ADVANCES IN KNOWLEDGE ORGANIZATION VOL. 15

INTERNATIONAL SOCIETY FOR KNOWLEDGE ORGANIZATION [ISKO]

Knowledge Organization for a Sustainable World: Challenges and Perspectives for Cultural, Scientific, and Technological Sharing in a Connected Society

Proceedings of the
Fourteenth International ISKO Conference
27-29 September 2016
Rio de Janeiro, Brazil

Organized by
International Society for Knowledge Organization (ISKO)
ISKO-Brazil
São Paulo State University

Edited by
José Augusto Chaves Guimarães
Suellen Oliveira Milani
Vera Dodebei

Ergon

Knowledge Organization for a Sustainable World: Challenges and Perspectives for Cultural, Scientific, and Technological Sharing in a Connected Society

Advances in Knowledge Organization, Vol. 15 (2016)

Knowledge Organization for a Sustainable World: Challenges and Perspectives for Cultural, Scientific, and Technological Sharing in a Connected Society

Proceedings
of the
Fourteenth International ISKO Conference
27-29 September 2016
Rio de Janeiro, Brazil

Organized by
International Society for Knowledge Organization (ISKO)
ISKO-Brazil
São Paulo State University

Edited by

José Augusto Chaves Guimarães
Suellen Oliveira Milani
Vera Dodebei

ERGON VERLAG

Editorial Support:

Daniel Martínez-Ávila
Isadora Victorino Evangelista

Predocumentation:
The volume contains: Introduction – Keynote Address – Epistemological Dimension of Knowledge Organization – Applied Dimension of Knowledge Organization – Social and Political Dimension of Knowledge Organization – List of Contributors and Author's Index

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at http://dnb.d-nb.de.

Cover Design: Jan von Hugo

www.ergon-verlag.de

ISBN 978-3-95650-221-7
ISSN 0938-5495

Table of Contents

Knowledge Organization for a Sustainable World:

Challenges and Perspectives for Cultural, Scientific, and Technological Sharing in a Connected Society

Proceedings of the 14th International Conference
International Society for Knowledge Organization

Introduction

Considering the current scientific and technological advances, we observe that the Knowledge Organization (KO) field, as a mediating element for produced and recorded knowledge to be socialized, accessible and appropriate and, consequently, generating new knowledge, is experiencing a moment of reflection and redirection, insofar new elements arise. Thus, on a theoretical level, it is necessary to systematize and consolidate knowledge (a set of knowledge) that can be verified in a given society at a given historical moment, with the objective of transmission, object of the study of philosophers related to the Theory of knowledge. As a result, it is important to rescue the knowledge recorded in documents, aiming its access and socialization, with the objective of retrieval, as shown by the studies of documentalists (Barité, 2001).

If, historically, the question came from a context of practical nature, aimed to solve problems related to information retrieval and access to documents in libraries and similar institutions, it was, however, especially after the second half of the twentieth century, with the concerns related to the scientific status of the area, that the theoretical approach on knowledge organization and representation gained more air, revealing its interdisciplinary nature. In this scenario, KO as a discipline arises, as García Marco recalls (1997, p. 8) in the 90s, at the crossroads of the so-called cognitive sciences, in the point among Psychology, Epistemology, Information and Communication Sciences, Semiotics, Linguistics, Mathematics, Logic and Computer Science.

In this sense, ISKO has been organizing, along more than two decades, biennial international conferences in order to promote a dialogical space among KO scholars and practitioners all over the world. For that, topics related to KO tools (Darmstadt, 1990), categories (Mysore, 2012), paradigms (Madras, 1992, Rome, 2010), structures (Lille, 1998), as well as the challenges that KO faces in questions related to quality management (Copenhagen, 1994), the tensions between dynamism and stability (Toronto, 2000) and the changes (Washington, 1996), the global information (London, 2004) and the global learning societies (Vienna, 2006), the integration of KO across boundaries (Granada, 2002) and the recognition of culture and identity (Montreál, 2008) and contexts (Mysore, 2012) as determinant factors in KO and the *razor´s edge* between historical patterns and future prospects in KO (Krakow, 2014).

This time, the ISKO International Conference takes place for the first time in the Southern Hemisphere as well as in Latin America and, therefore, it is built upon the

experience of the Brazilian Chapter of ISKO. During its seven years of existence, ISKO-Brazil was able to promote three national focused on the challenges and scientific perspectives for KO at the present time (Guimarães and Dodebei, 2012), the relationship between complexity and KO (Dodebei and Guimaraes, 2013) and KO in a context of cultural diversity (Guimarães and Dodebei, 2015).

Thus, the 14th ISKO International Conference ISKO (Rio de Janeiro, September 27-29, 2016), under the theme *Knowledge Organization for a sustainable world: challenges and perspectives for cultural, scientific, and technological sharing in a connected society,* comprises the epistemological dimension of KO (conceptual, historical, and/or methodological bases of KO as well as dialogs at the intersections of disciplines), the applied dimension of KO (KO models, formats, tools, products, and structures), and the social and political dimension of KO (education and professional practice in KO, ethics in KO, culture and identity in KO, and KO for a sustainable development). For this, the conference has the following objectives: a) to discuss the challenges and perspectives in KO in relation to cultural diversity, b) to verify the state of the art of international research in KO, c) to identify different perspectives of scientific dialog in the international environment of KO, d) to advance discussions on the current epistemological construction, interdisciplinary dialog, technological applications and the social dimension of KO, and e) to provide international visibility to KO research.

The conference is organized by the Brazilian Chapter of ISKO and the Graduate School of Information Science of São Paulo Satet University – UNESP and received the special support from ISKO International, Brazilian National Council of Technological and Scientific Development – CNPq, Brazilian Coordination for the Improvement of Higher Level -or Education- Personnel – CAPES, São Paulo Research Foundation – FAPESP, Getulio Vargas Foundation – FGV, Foundation for the Development of São Paulo State Foundtion – FUNDUNESP, VUNESP Foundation, and the Information Organization Research Group – GPFAPOI- UNESP.

The event includes the keynote "The influence of culture in knowledge organization", by María J. López-Huertas, and the roundtable "KO for a sustainable world: international perspectives", coordinated by D. Grant Campbell and with the participation of representatives of the Brazilian, British, Indian, French, North-American, Polish and German ISKO chapters. Of a total of one hundred and twenty original proposals submitted for evaluation by an international scientific committee composed of fifty two scholars from sixteen countries all over the word, seventy communications were selected for oral presentation at the event.

We do hope that this conference can contribute to the KO scientific environment worldwide, especially in this moment when the sustainability of the plane is considered a crucial question, derserving our deepest concerns.

José Augusto Chaves Guimarães
Marília, June 16, 2016

María J. López-Huertas (Keynote Address)

The Integration of Culture in Knowledge Organization Systems

Abstract

Culture has always been implied in the knowledge organization (KO) and the Knowledge Organization Systems (KOSs). It has been said that the latter are cultural artifacts because they are expressions of a given culture at a given time. It is evident that awareness of the importance that may have the cultural factor in this context has been slow but constant over time. This paper intends to reflect on the role of culture in KO and KOSs and to look at the main highlights and contributions in the field from the first reactions on this issue to the present day. That gives us the chance to observe how this topic evolves over time. Finally, some actual proposals for integrating culture in KOSs are given with special emphasis in two different models for approaching the problem: one for the integration of subcultures in a main common culture and the other for the integration different cultures in a general KOS.

Culture is a complex and ambiguous concept. It is considered that its complexity is due to its ambivalence that originates in the effort to reconcile the liberty and the human regulatory limits, the ambivalence between the creativity and the norms that rule humans in society (Cardoso Rodrigues 2015). It is an evolving concept which adds difficulty in conceptualizing it. Because of these difficulties, scholars have not reached an agreement about what we mean by culture. However, it is not our aim to go deep into the general concept of culture, and for that, the historicist conception of culture will be followed from here on. According to this approach, culture would be "the intellectual, artistic and moral aspects of a civilization or a country. So, we can talk about occidental culture, Hellenic culture and Brazilian culture" (Cardoso Rodrigues 2015). The concept of culture is transdisciplinary and refers to phenomena that make up the collective beliefs and activities of groups of people. Discussions of culture commonly refer to shared values, language, history, collective memory, social attitudes, preferences, practices, etc. (Beghtol 2002). When we talk about culture, the scope may vary. It can be very wide if we refer to the cultures of the world (Occidental, Oriental, etc.) or it can be much more restricted if we are talking about different cultures coexisting in a country, for instance.

The impact of culture in knowledge organization (KO) has been repeatedly recognized in the field of Library and Information Studies (LIS) and the need for research on solving the problems that this fact poses is a hot issue in KO. Some research has been carried out on this issue, but they do not usually address how to handle categories representing a given field of knowledge in a real setting in order to create a knowledge structure be able of harmonize them all with the aim of constructing a more communicative system.

This study will reflect on the meaning of culture in KOSs, starting with the first general classifications where culture meant some deviations in the selected subjects to be represented and in the way they were classified. Later, scholars refer to it as bias in the classification systems. Then will be a short tour through time that allows seeing

how this issue was gaining importance until a clear awareness is detected. An account of outstanding contributions on this subject, in our view, is given at the time that an increased interest for the study of the indigenous knowledge is detected. Finally, some of the proposals that favour cultural integration in global KOSs will be discussed.

1 Culture did not always mean a desired feature in KOSs

The inclusion of cultural view-points in KOSs is now considered as a desired action to be taken when designing and constructing those systems. This presence is demanded by many scholars and has created a body of literature behind it. But, had the cultural point of view the same effect in KOSs in all instances?

Culture, knowledge organization (KO) and knowledge organization systems (KOSs) are tightly bound together in a way or another. Since the first bibliographic classifications were published to the actual systems, culture is something inherent to the act of designing and making a conceptual structure as a tool for information recovery whether its presence is conscious or unconscious. It can also be said that the incorporation of cultural features in KOSs have not always meant something positive.

More than a century has passed since the first classifications appear, and it can be said that they were exponent of cultural inputs.As a clear example, let us take class 2 of the UDC, before this class was completely reshaped, for a quick test. The first 8 subdivisions are referred to Christianity and the rest of them were all put together under the 29 subclass with the general name of 'No Christian religions', not to enter in the categorization which is made to refer to some churches which are grouped under the name of sects. The aforementioned categorization and organization of the religious knowledge was influenced by the culture where the author of the classification was raised in. This fact gave place, some decades after, to a common concern that can be summarized in the following: classifications are an expression of the way in which its creators see the world or KOSs are cultural artifacts that have a powerful influence on individuals within a given culture (Beghtol 2001). First classifications were created in a time where the field was speculative and not sensitive enough about the problems that may cause to ignore the socio-cultural aspects of potential users in systems that were called universal. This is an example of a negative inclusion of culture. It really means exclusion because it reflects only the Occidental vision in a system that pretends to be of universal use. It is exactly the opposite to the current idea of incorporating the cultures in an integrated way.

Now, let us give a jump to the arrival of the Internet. It is said that there are no cultural borders since we can move around the world, but does it mean that cultural differences has been taken into account when knowledge has been organized on the Web? I would say that yes, we can move around but most of the times using a single model of culture that it is again the Occidental one. We are witnessing again that Internet is a cultural product, and that knowledge organization has been considered as a cultural form of new media (Andersen 2008). The reasons that make it happen are

different from that mentioned in the case of classifications, but still the final result is the exclusion of less favored cultures or the domination of a given culture on others. These are cases in which the cultural perspective negatively affects systems that have been created for universal use.

There has been a concern among scholars regarding this problem that has finally arrived at the revision of classifications to minimize these unwanted results.

2 Highlights in the integration of the cultural point of view in KOSs

International networks, international cooperation, projects and learning, global information systems of any kind have evidenced a reality already known in KO but never perceived to be as demanding as it is nowadays: cultural warrant. This concept needs reformulation according this the new circumstances. There is a need for representing and organizing cultural differences in an integrated way not only in KOS, although it is a major concern here, but also in other settings as could be the case of systems for e-learning. Global systems made it possible the coexistence of different cultures. In fact, the Web crosses and defies cultural and linguistic boundaries around the world and points to new uses and new users of information. Given this situation, the need for cross-cultural research have been detected by many scholars and the impact of these issues in information systems requires research in order to face the problems posed by new global information systems (Hunter & Beck 2000). Universal access bears a great relation with the capacity of systems to integrate cultures in theirs structures. As Treitler (1996) argued, without the integration of cultural differences in information systems, universal access cannot be guaranteed.

Some authors, while recognizing that cultural issues are often neglected in information systems, point out that "much research has focused on the effects these systems hold rather than viewing systems as tools to be designed given an understanding of socio-cultural context. Emerging research in community information systems and archives has highlighted possible interactions between system design and ethnographic research" (Srinivasan 2007, 723). There is a call for developing systems based on ethnographic knowledge and for concrete proposals regarding the design of such systems. Other studies urge reflection about the theoretical concept of multiculturalism as a "dangerous slogan and not sufficiently critical as to tackle the rights of diversity and singularity even within a given (but not real) mono-cultural society [...] Research on KO must be open to a new paradigm in which Critical Theory and hermeneutics go together" (García Gutiérrez 2002, 517).

The sensitivity to the role of culture in KOSs is old in our field to the point that has been considered a long term research question in KO (Gnoli 2008). Maybe one of the first manifestations of this interest was the biases detected in the first bibliographic classifications that last until now, as it is claimed in López-Huertas (2008). A token of that is the contribution of Rebecca Green (2015) who studied how indigenous people in the U.S. are represented in the Dewey Decimal Classification. She analyzes how they

are grouped or dispersed in the classification, how they are categorized, that is under which label they are represented, and the terminology used.She focused in marginalization through ghettoization, historicization, diasporization and missing topics.

It is also evident that we have witnessed a progressive concern about the importance of the cultural integration in KOSs that has been much stressed since the beginning of the 21st century coming to be seen as a sign of quality of the systems (López-Huertas 2008), although the last decade of the 20th century was also active regarding this matter. An expression of that concern can be found in two International ISKO Conferences, one in Granada in 2002 under the theme Integration of knowledge across boundaries (López-Huertas 2002) and the other in Montréal in 2008 devoted to culture and identity (Arsenalult & Tennis 2008). I would say that, as a result, a renewed interest in these issues emerged among researchers.

The increased concern of scholars towards the need for the inclusion of the cultural factor in KOSs, together with cross-cultural character of global information systems culminated with new theoretical formulations in the last decade, as it will be shown below.

2.1 The cultural warrant

One of the main contributions along the way to incorporate cultural points of view in KOSs was the formulation of the concept cultural warrant. This expression was used to draw attention to the need to take into account socio-cultural characteristics of users for which information systems were created in the belief that different cultures need different kinds of information.Some authors claim that culture plays an important role in the perception and recall of information, that different cultures may have different understandings of information (Kim 2013). The expression of cultural warrant was coined by Lee with the meaning of "the influence of socio-cultural factors in the semantic relationships of classification systems" (Beghtol 2001, p.104). It means that any kind of KOS can be appropriate and useful only if it is based on the values and assumptions of that same culture. Beghtol´s idea of cultural warrant includes the concept of user warrant that refers to the collaboration of potential users in the development of information systems. It is justified on the assumptions that users pertain to a certain culture and that they act as representatives of a given culture when they participate and use KOSs. She claimed that KOSs are maximally appropriate and useful for users in some culture only if they are based on the values, beliefs and assumptions of that culture. This quality will decrease when these conditions are not met.

Beghtol's 2002 article develops and deepens the concept of cultural warrant (Beghtol 2002). She expressed that, due to the increased globalization of information resources, there is a need to protect cultural and information diversity. In order to facilitate the incorporation of cultural viewpoints, she introduced the concept of

cultural hospitality, taken from the concept of hospitality as a required attribute of the notations of the classifications. She claims that the problems of globalization for KOSs can be approach by broadening the concept of hospitality in two ways: By concentrating on techniques for adding new concepts to KOSs and by adding not only new concepts but also the addition of different cultural warrants that in turn may include different user warrants. That is, "we need to make each knowledge representation and/or organization system, which by definition is based on some cultural warrant, 'permeable' to other cultural warrants and to the specific levels and layers of individual user choice within each culture" (Beghtol 2002, p. 518).

2.2 Integration of cultures in KOSs turns in an ethic issue

Cultural and social differences are important part of reality and they should occupy a prominent place in KO, especially when looking at global information systems either specialized or general. The importance of cultural issues to KO goes beyond its objective importance, it is a question closely related to professional ethics. It is also a question of being aware of what could be behind global systems in the sense that those systems might be using standardized views and KO models that are designed to fit certain visions of the world that reflect views and beliefs of dominant economies and cultures. There might be different good final reasons for addressing cultural topics in KO, but there is one that cannot be overlooked and this is the responsibility for us to watch over the information needs in non dominant cultural and economic regions or groups by representing them in global information systems. Users belonging to these areas have the right to access to information in an understandable way for them, and to be aware of it and to respond by creating the media to allow such a communication is an ethical question for KO researchers and professionals.

For these reasons, the idea of designing and constructing global KOSs that integrate the cultures to which they address goes beyond theoretical or methodological considerations directed to improve the retrieval systems. It has to do with responsibility of those who make these systems in order to get KOSs which are representative and fair for their users. So, we can understand that the concept of cultural, user warrant and cultural hospitality is tightly bound to that of ethics. It is openly expressed by Beghtol when she says that there is a need for providing "information globally, locally in any language, for any individual, culture, ethnic group or domain, at any location, at any time" (Beghtol 2002, p. 507). In fact she is concerned by the design, construction and maintenance of global information systems based on ethical principles and, in that scenario socio-cultural aspects play an important role. She arrives at the concept of ethical warrant for globalized KOSs that is based on three assumptions: "KOSs should be based on ethical principles, the ethical context(s) of cultural globalization should influence the design of ethically based KOSs and any discussion contains ethical preferences that may or may not be as explicit as is desirable" (Beghtol 2002, p.513). The absence of that causes biased representations that have been well documented in

the field, in special in cataloguing and classification practices and methodologies to recognize and avoid such biases are required. Lee (2015) reflects on the relationship between ethics, KO and culture arriving at the conclusion that the cultural issue raises ethical issues in KO

The ethical approach to KO is a current research question that is gaining attention among the specialists. I would say that the inclusion of a session and round tables on that matter in the International ISKO Conference held in Granada 2002 was an important point of departure. Since then, Ethics has been present in all ISKO Conferences to the point that it has been chosen as the topic in 2009, 2013 and 2015 Milwaukee Conference on Ethics on Knowledge Organization, organized by the Knowledge Organization Research Group and the Center for Information Policy and Research of the School of Information Studies at the University of Wisconsin-Milwaukee. Papers presented in this Conference has been recently published in the volume 42 (2015) of Knowledge Organization.

It can be said that the ethical approach to KO is considered today as a new core in the domain, a key component of KO that is actually supported by a new emergent cluster of authors (Smiraglia 2015).

3 A renewed interest in indigenous and local knowledge

A direct result of the vivid interest aroused by cultural and ethical issues is a renewed concern to get to know indigenous and local knowledge. Both actions are close related because you need to know the local culture in order for it to be represented and organized in information systems. There are a wide range of approaches to local cultures that goes from areas with oral traditions to regions with literary traditions. So, studies on indigenous knowledge at different levels and realms are emerging. There are contributions about how to manage indigenous knowledge (written or oral), how to organize it (Rao 2006, Kargbo 2005, Muswazi 2001, Espinhero de Oliveira 2002, Liew 2004 and Doyle 2006), how to carry out indexing activities using controlled languages in indigenous cultures (Monajami 2003) and how to construct controlled vocabularies for indigenous knowledge (Amaeshi 2001). Liew (2004) argues that the Maori language can be reconciled with worldwide use in digital libraries. Another attempt to have global systems accommodate the peculiarities of local environments is that described by Rolland & Monteiro (2002). The indigenous knowledge in India has been addressed by Rao (2006), although his emphasis lays on pointing out its importance for society in general and the need for documenting it. He defines it as a local and tacit knowledge that is unique of a given culture and claims that it is the basic component of any country knowledge system. He also identifies the characteristics of the indigenous knowledge and the types of it and states that is the key resource for social development and global issues. It is recognized as public knowledge. He suggests that this knowledge needs to be documented in order to make

it available and to facilitate due recognition to its holders. The knowledge organization structure of Taiwan's aboriginal cultures has also been object of study.

4 Some of the proposals that favour cultural integration in global KOSs

As it can be seen from the previous paragraphs, there is a deep concern regarding the need for cultural integration in global KOSs. This interest has given place to the formulation of theories where to support it, but unfortunately we do not have many examples that show how to do it. It is the aim of this section to give an account of the main proposals to our knowledge that would allow the construction of KOSs ethically founded and culturally representative.

4.1 Theories and methods favouring cultural integration

Cardoso (2015) suggests theories and methods that can meet the needs posed by cultural integrated KOSs. As a general frame, he states that the chosen theories should consider knowledge dynamism and be flexible so as to allow the continued hospitality of conceptual elements in the system. Cognitive based theories are also recognized to be of great help in doing this task. He mentioned five theories and methods that will help in accomplishing cultural integration: Cognitivism, Poly-representation, Domain Analysis, Faceted classification theory and Integrative levels theory.

The cognitive approach can collaborate in the improvement of communication between the system and the users' information needs. It will help in understanding the human cognitive mechanism in the process of information acquisition and its subsequent transformation into knowledge.

The principle of Poly-representation (Ingwersen, 1996) suggests that a representation of information according to the multiple users' need, problems and their states of knowledge. This statement deserves some comments. I would say that actually the application of the concept of poly-representation goes beyond the scope of the user, although in origin may have this sense due that Ingwersen belonged to this school of thought. However, it was soon taken to express an exhaustive representation of the relevance of each concept to be included in conceptual structure after applying discourse analysis to identify relevant textual elements for a given concept (López.Huertas 1997). This approach might be also useful for managing multicultural structures.

Domain analysis helps in the understanding of the domain through the eleven methods proposed by Hjørland (2002). This theory will help to understand and to identify the cultural structures in any culture, making possible the real understanding of them.

The Ranganathan facet theory and the Integrative levels theory will support aspects of the knowledge structuring related to practical questions, logic, norms and theory. The first one will help in the systematization of dynamics domains though its rules, canons, the continuous hospitality and its facet system. The second one will collaborate

as a model by contributing with its vision of systemic and interconnected elements that is how culture is conceptually understood by specialists. Cultural elements are interconnected and each element depends on state that others assume. Inside the system, there are rules so no element transgresses its limits to avoid the imbalance of the whole.

4.2 Changes in structural principles for KO

Some authors have studied the fundamental principles of classifications based on the Western logic to indicate what has to be changed in order to construct KOSs that meet cultural warrant. Hope Olson (2002) is an example of this trend. She addresses the cross-cultural issue and points out that the essential principles for KO in the Western world, such as mutual exclusivity, teleology and hierarchy, hinder multicultural inclusion in KOSs. She makes a deep reflection on the work Primitive classification by Durkheim and Mauss where logical classifications developed in the Western culture, inspired in Greek philosophers, and classifications made by primitive cultures that do not follow this pattern. Instead, the logic of primitive classifications is derived from social classification, arriving at the conclusion that primitive classifications do not meet the Western logics principles. Especially interesting is the reflection on the hierarchical principle because it is found to be an obstacle for cultural integration, claimed by Olson here and later developed by López-Huertas (2013) as it will be shown below. Olson finally claims that organizing knowledge based on different structural principles would favor cross-cultural understanding and enhance KO. She suggests other kind of structures to represent cultural knowledge and points that those structures should have contradictions, deviations and overlapping. To give an example of this approach, she uses her proposal to organize the knowledge of feminist culture which is "frequently a circle with variants including a spiral and a web" (Olson 2000, 8).

It seems to be quite clear that we have to move to conceptual structures not based on hierarchical logic in order to accommodate the cultural perspective in KOSs. That is to say, at least, that the designer of these systems needs an open mind in order to abandon traditional logics in favor of other solutions when needed.We can even say that it is a trend suggested elsewhere. An example of that can be seen in the content of the special issue of the journal Knowledge Organization, published under the title "Paradigms of knowledge organization: The tree, the net and beyond" (2013). We can find here different approaches explaining the move from hierarchies to other forms of KO in order to fit new needs.

4.3 Multicultural semantic warrant for global understanding

Cultural integration has much to do –I would say that it is the main issue- with the construction of conceptual structures where users from different cultures are familiar with the representation and the organization of concepts in those systems. The naming of categories takes the most important role here because one main goal would be to

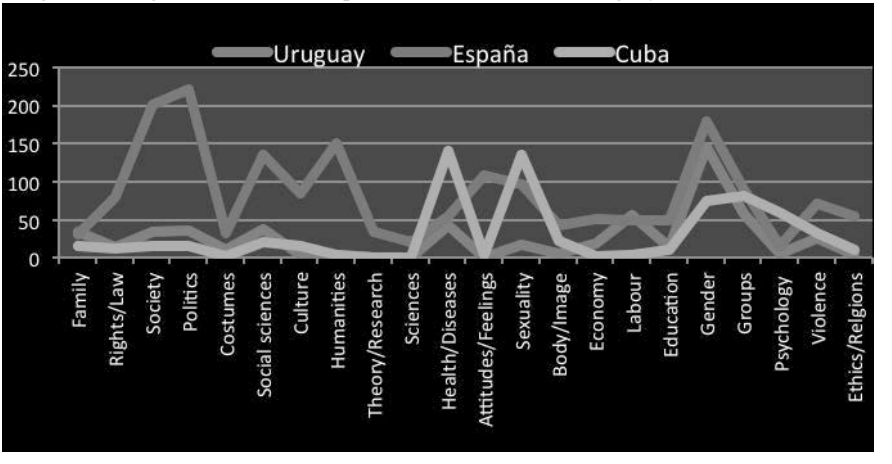
look for categories that are shared by the cultures for which the information system is made to reach. At the same time, the categories have the potential to organize knowledge, so are a key element in the construction of information systems.

If different cultures are to be represented in global KOSs, a deep knowledge of them is required for later integration. That is, knowledge generated inside them that reflects how a given topic is perceived and addressed to in a particular cultural area. One of the best ways to do it is by analyzing the content of the publications produced in that area.

The following is based on the results of several studies that show the representation of the same specialized, contextual knowledge in different cultures, the differences imposed by each culture and some suggestion for trans-cultural categorization. Two different situations are going to be discussed: the integration of subcultures, that use the same language, in a primary culture and the integration of several different cultures. From here, we can talk about two models for cultural integration: 1. Integration of subcultures that belong to a given cultural area, as it is the case of the Spanish, Cuban and Uruguayan cultures that belong to the Occidental and 2. Integration of different cultures as it is the case of the Occidental, the Hindu and Eastern Asian.

Model 1. Integration of subcultures that belong to the same primary culture. In this case, Gender Studies was the specialty chosen to carry out the research. In order to know how this topic is perceived and addressed to, specialized publications issued in Spain, Uruguay and Cuba were identified and later indexed. The extracted vocabulary was treated separately and later processed in order to get a primary structure for each culture. Then, a common broad structure could be built, as it can be seen in Figure 1:

Figure 1. Integrated structure of Spain, Cuba cultures and Uruguay for Gender Studies



This structure includes the main categories identified for the three regions and the weight that they have in those cultural areas. According to this, cultural integration in this model should have a conceptual structure that represents all of the identified classes, no matter that one of them is not representative of any particular culture, and has to go as deeper in the description as is required by any culture.

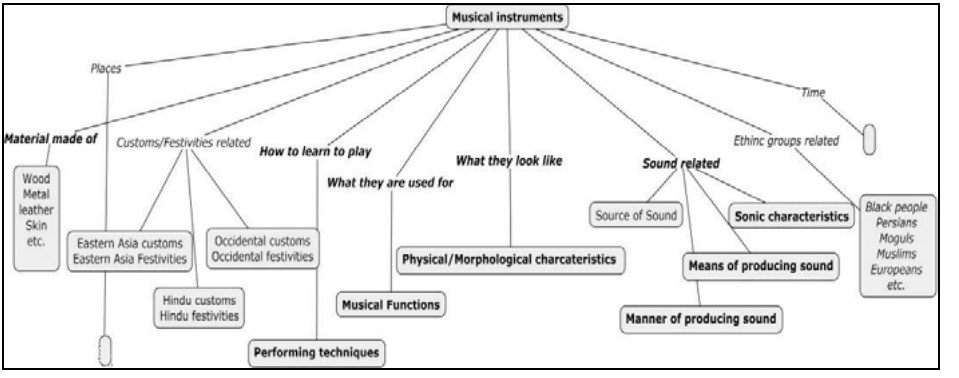
Model 2. It is referred to systems that have to integrate different cultures. The intention is to show that intercultural categories could be shared in a considerable extent, that the citation order that said categories receive in each culture is incompatible with a common structure and that the harmonization of the categories in a structure based on the meaning of categories could be an answer for building a structure able to communicate the cultures represented by those categories. Our goal here is give an idea about how to organize a common structure at the first steps of development because the categories at hand do not allow going any further.

Our example is concerned with Western, Hindu and Eastern Asia cultures. The topic chosen for the study is musical instruments. In this case, the point of departure is a number of categories identified for each region, based on Western dictionaries definitions (López-Huertas 2013) and on classifications of music representative of the Hindu and Eastern Asian cultures (Kartomi 1990). It is understood here that concepts are the units for knowledge representation and organization, understanding units related to semantic holism. Units are formed by characteristics according to which knowledge should be categorized and organized.) Many of the theories on concepts refer to characteristics defining the concept, called by Dalhberg knowledge elements, as essential elements for concepts definition. So, identifying these characteristics (knowledge elements) for a particular concept (knowledge unit) is a main goal for knowledge organization (Dalhberg 2011).

A comparison of categories is demanded in order to find out for them to be integrated in one scheme. If we pay attention to the literal translation of the categories, we find that the Hindu and Asian categories do not match with those found in the occidental culture. This was the case of Major and Minor limbs, Male and Female instruments, etc. However, a closer semantic analysis of categories allows discovering similarities between the three cultures. That is, if we look at the meaning of Female instruments, for instance, which means big instruments, we have the category Size which is in the Occidental scheme as well.So, if we apply this analysis to the example above, we found that 50% of the categories are shared by the three cultures, 10% are partially shared and 30% of the categories had no equivalence in either culture. This fact provides a basis for considering the integration. Regarding the citation order of categories found in the three schemes, the situation is quite different because there is no coincidence in the citing order of neither culture (López-Huertas 2013). This is due to the fact that the order of citation is much influenced by believes and customs which are responsible for assigning value or importance to the categories to be ordered in the

classification. As a result, the final structures have little in common. It is also in relation to the kind of music and the instruments related to it that identifies each culture in this case. For this reason, subordination needs to be avoided as much as possible. In principle, when subordination within a class takes place, it should be done by using categories shared by the cultures represented in the SOCs. For instance, material and physical characteristics subclasses, which are shared by the three cultures, could also be used for further subordination in other subclasses when needed. We believe that, by doing this, communication problems are expected to be reduced. Following this procedure the following first step classification emerges:

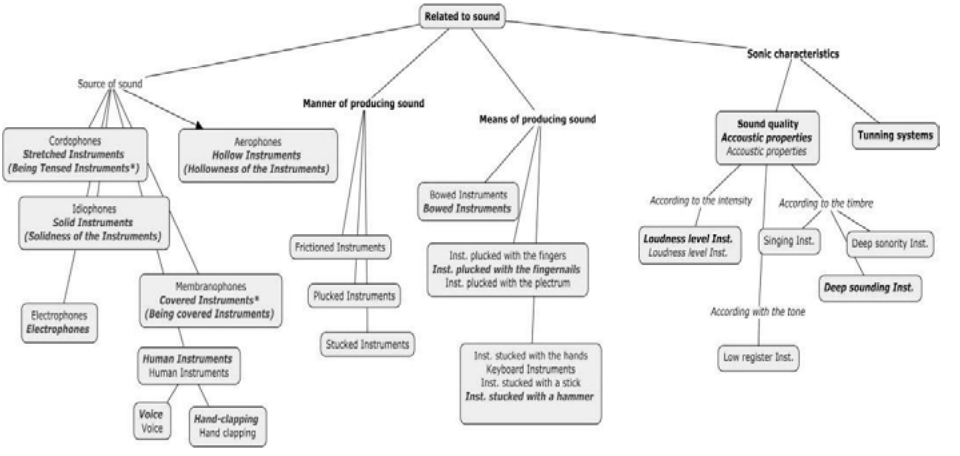
Figure 2. First level integrated structure for Musical Instruments



These tags have the function of organizing the knowledge and acting as information recovery and navigator keys. Place and time appear only in the Occidental case, so it is convenient to wait for more information to develop them. Categories in bold face are those completely shared. Categories from Eastern Asia are in italics- No subordination has been used for arranging categories under the sound related one.

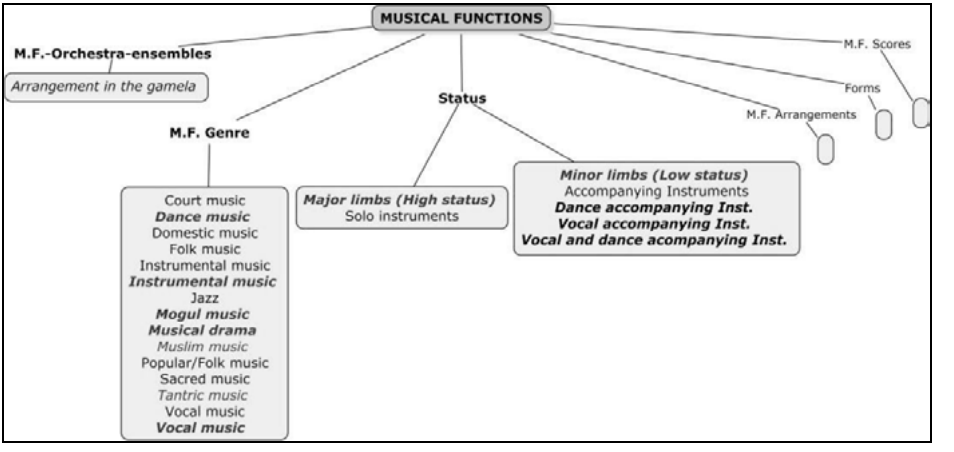
Three of them (the most representatives) were further developed: Sound related, Musical functionsand Physical/Morphological characteristics as it follows in the same order:

Figure 3. Development of Related to Sound Category



We can see in Figure 3 that there are four general categories subordinated to Sound. In red are the completely shared categories. When different wordings occur, these are given below the red one. In italics bold are the Hindu categories, in italics Eastern Asia categories and in black the Occidental categories. Stretches instruments* in Hindu do not only design cordophones, but also stretched-skin instruments. The case of covered instruments* is similar. It means that something covers an opening or hollow. It does not include certain membranophones such as the free kazoo type.

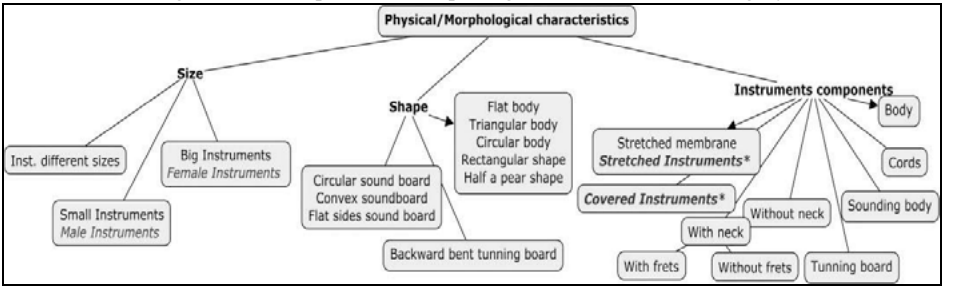
Figure 4. Development of Musical Functions category



By now, six subcategories formed the Musical Function category (Figure 4). Only three of them (M.F. Orchestra/ensembles, M.F. Genre and M.F. Status) could be developed because we only have information related to the Occidental culture for the rest of them.

In Figure 5, an example of the development of the Morphological characteristics is given. This category in Hindu and Easter Asia classifications is not much developed, although it is mentioned in the description of some instruments. The scheme below only includes categories in referred to in those aforementioned classifications. These are in italics (Eastern Asia), italics bold (Hindu) and black (Occident). The shared categories are in bold. Size is in bold although is formally mentioned in two cultures (Eastern Asia and Occident), but it is supposed to be found also in the Hindu culture when more data are at hand.

Figure 5. Development of Morphological Characteristics category



In a similar line of thought, Neelameghan and Iyer point that the global systems "have several impediments such as cultural bias, misinterpretation of concepts and non-existence or non-acceptance of ideas of one group by people of other cultures or faiths" (Neelameghan & Iyer 2002, p. 539). At the bottom of this claim, there is a need for representation elements that are common to all cultures for which the information system has been created, a need to identify categories shared by those cultures that mitigate the problem of communication across cultures. They study the site Mysticism in World Religion, where six religions were compared to show the possibility of finding these common elements. In this case, you will find categories common to all religions or a list of categories representing the six religions that allow for a search of information to the users of any of those religions (Neelameghan & Iyer 2002).

5 Conclusions

Awareness of the need for the integration of cultures in global systems is lately much demanded by KO scholars. It has been a slow but progressive concern that has given place to important contributions that have provided a point of departure to meet this demand.

There are contributions suggesting general theories to address cultural integration. However, there is a need for more research on real settings that face actual problems and that could offer solutions to those problems.

A deep knowledge of how the subject is represented (categorized) and organized in each of the cultures included in the structure is needed in order to find common categories to take into account to build an integrated structure.

Building the structure based on the meaning of categories no matter what they mean literally seems to be a good choice for cultural integration because it allows harmonizing the cultures involved. Potentially, it would increase communication between users of said cultures and the system.

It is expected that not shared categories are found. They also can be included in the KOS by representing them in the same way: according to the meaning of said categories

References

Amaeshi, Basil (2001). Subject indexing in the Nigerian National Corporation Library: design of an instrument. *Library Review,* 50(9): 457-60.

Andersen, Jack (2008). Knowledge organization as a cultural form. From knowledge organization to knowledge design. In Arsenault, Clément., & Tennis, Joseph. (Eds.) *Culture and identity in knowledge organization: Proceedings of the Tenth International ISKO Conference. Montréal, Canada, 5-8 August, 2008*. Advances in knowledge organization, 11. Würzburg: Ergon. Pp. 269-74.

Arsenault, Clément & Tennis, Joseph. (Eds.) (2008). *Culture and identity in knowledge organization: Proceedings of the Tenth International ISKO Conference*. Montréal, Canada, 5-8 August, 2008. Advances in knowledge organization, 11. Würzburg: Ergon.

Beghtol, Clare (2001). Relationships in classificatory structure and meaning. In Bean, Carol and Green, Rebecca (Eds.). *Relationships in the Organization of Knowledge.* Dordrecht: Kluwer Academic Publishers. Pp. 99-113.

Beghtol, Clare (2002). A proposed ethical warrant for global knowledge representation and organization systems. *Journal of Documentation*, 58(5): 507-32.

Cardoso Rodrigues & Anderson Luiz (2015). A cultura e a organizaçao do conhecimento: desafíos teórico-metodológicos. *Información, Cultura y Sociedad*, (35).[http://www.scielo.org.ar/scielo.php?script=sci_arttext&pid=S1851-17402015000100003]

Dahlberg, Ingetraut (2011). How to improve ISKO's standing: Ten desiderata for knowledge organization. *Knowledge Organization*, 38(1): 68-74.

Doyle, Ann (2006). Naming and reclaiming indigenous knowledges in public institutions: Intersections oflandcapes and experiences. In Budin, Gerhard, Swertz, Christian and

Mitgutsch, Konstantin (Eds.). *Knowledge Organization for a Global Learning Society: Proceedings of the Ninth International ISKO Conference*. Vienna July 2006. Würzburg: Ergon. Pp. 435-42.

Espinhero de Oliveira, María Odaisa (2002). Kowledge representation from Amazonian Narratives. In: López-Huertas, María (Ed.). *Challeges for Knowledge Representation and Organization for the 21st. Century. Integration of knowledge across boundaries*: *Proceedings of the 7th International ISKO Conference*. Würzburg: Ergon. Pp. 546-51.

García Gutiérrez, Antonio (2002). Knowledge organization from a "Culture of the Border": Towards a transcultural ethics of mediation. In: López-Huertas, María (Ed.). *Challeges for Knowledge Representation and Organization for the 21st. Century. Integration of knowledge across boundaries: Proceedings of the 7th International ISKO Conference*. Würzburg: Ergon. Pp. 516-22.

Gnoli, Claudio (2008). Ten long-term research questions in knowledge organization. *Knowledge Organization,* 35(2/3): 137-49.

Green, Rebecca (2015). Indigenous people in the U.S., sovereign nations, and the DDC. *Knowledge Organization*, 42(4): 211-21.

Hjørland, Birger (2002). Domain analisys in information science:eleven approaches – traditional as well as innovative. *Journal of Documentation*, 58(4): 422-62.

Hunter, M. Gordon & Beck, John (2000). Using repertory grids to conduct cross-cultural information system research. *Information Systems Research*, 11(1): 93-101.

Ingwersen, Peter (1996). Cognitive perspectives of information retrieval interaction elements of a cognitive IR theory. *Journal of Documentation*, 52: 3-50.

Kargbo, John Abdul (2005). Managing indigenous knowledge: what role for public librarians in Sierra Leone? *International Information and Library Review*, 37 (3): 199-207.

Kartomi, M. (1990). *On Concepts and classifications of musical instruments. Chicago: University of Chicago Press.*

Kim, Ji-Hyun. Information and culture: Cultural differences in the perception and recall of information. *Library & Information Science Research*, 35(3): 241-50.

Lee, Wan-Chen (2015). Culture and classification: an introduction to thinking about ethical issues of adopting global classification standards to local environments. *Knowledge Organization*, 42(5): 302-7.

Liew, Chern Li (2004). Cross-cultural design and usability of a digital library supporting access to Maori cultural heritage resources: an examination of knowledge organization issues. In: McIlwaine, Ia (Ed.). *Knowledge Organization and the Global Information Society: Proceedings of the 8th International ISKO Conference*. Würzburg: Ergon Verlag. Pp. 127-32.

López-Huertas María J. (2013). Transcultural categorization in contextualized domains. *Information Research*, 18(3). [http://www.informationr.net/ir/18-3/colis/paperC16.html#.V0QqaCuLWOE]

López-Huertas, María J. (1997). Thesaurus structure design. A conceptual approach for improved interaction. *Journal of Documentation*, 53(2): 139-77.

López-Huertas, María J. (2008). Some current research questions in the field of knowledge organization. *Knowledge Organization*, 35(2/3): 113-36.

López-Huertas, María J. (Ed.). (2002). *Challenges in knowledge representation and organization for the 21st century: Integration of knowledge across boundaries: Proceedings of the*

Seventh International ISKO Conference. Granada, Spain, 10-13 July, 2002). Advances in knowledge organization, 8. Würzburg: Ergon.

Mazzochi, Fulvio & Fedeli, Gian C. (2013). Paradigms of knowledge and its organization: The Tree, the Net and beyond. *Knowledge Organization* 40(6):366-74.

Muswazi, Paiki (2001). Indigenous knowledge management in Swaziland: perspectives. *Information Development*, 17(4): 250-55.

Neelameghan, A. & Iyer, Hemalata (2002). Some patterns of information presentation, organization and indexing for communication across cultures. In López-Huertas, María (Ed.). *Challeges for Knowledge Representation and Organization for the 21st. Century. Integration of knowledge across boundaries*: *Proceedings of the 7th International ISKO Conference*. Würzburg: Ergon. Pp. 539-45.

Olson, Hope (2000). Reading "primitive classification" and misreading cultures: The metaphysics of social and logical classifications. In: Beghtol, Clare and Williamson, Nancy (Eds.). *Dynamism and stability in Knowledge* Organization: Proceedings of the 6th International ISKO Conference. Würzburg: Ergon. Pp. 3-9.

Rao, Siriginidi Subba (2006). Indigenous knowledge organization: An Indian scenario. *International Journal of Information Management*, 26(3): 224-33.

Smiraglia, Richard (2015). Ethics in knowledge organization: Two Conferences point to a new core in the Domain.*Encontros Bibli: Revista eletrônica de Biblioteconomia e Ciência da Informaçao*, 20(1).

Srinivasan, Ramesh (2007). Ethnomethodological architectures: Information systems driven by cultural and community visions. *Journal of the American Society for Information Science and Technology*, 58(5): 723-33.

Treitler, Inga (1996). Culture and the Problem of Universal Access to Electronic Information Systems. *Social Science Computer Review*, 14(1): 62-4.

Mariângela Fujita and Lena Vania Ribeiro Pinheiro

Epistemology as a Philosophical Basis for Knowledge Organization Conceptions

Abstract

This paper presents an analysis of Popper's ideas on knowledge organization systems with an emphasis on the foundations of critical-rationalist epistemology, particularly objective knowledge. Knowledge Organization as a field of study is concerned with the processes and knowledge organization systems aimed to develop more qualitative proposals to other fields of knowledge. Knowledge organization systems are representations of structures whose contents, organized in controlled vocabularies of terms, represent concepts. The function of concept organization and representation is the most important characteristic of these systems that relates them to Popper's objective knowledge theory. The knowledge organization system and Popper's objective knowledge, shown in his Table of Ideas, and the relationship between concepts and theory formulation are analyzed. The results demonstrate that epistemological aspects can be applied to knowledge organization systems. It can be concluded that analyses of Popper's objective knowledge and Epistemology in general provide further development of theoretical issues in knowledge organization.

Introduction: some concerns and epistemological research

Knowledge Organization (KO) as a human activity is linked to cognition in social, professional, and intellectual actions and is part of the daily life of every person. In the same way, knowledge areas, such as Chemistry, Physics, Biology, etc., have a continuous development and, for this reason, they must be systematized and organized by creating specific terminologies and using taxonomies to meet their needs.

These ideas encourage thinking about some theoretical approaches between KO and Epistemology. The first one comes from the Theory of Knowledge, the most remote origin of KO since the Ancient Times (Hjørland, 1994, Barité, 2001), although its institution as a field of knowledge only occurred when the International Society for Knowledge Organization (ISKO) was founded. Another epistemic approach is the conceptual one, pointed out by Dahlberg (2006), who considers KO as a new science formed by a huge set of concepts and that it complies with the anthropological and propositional concepts of Alwin Diemer's science (1970 and 1975). More than theoretical approximations, the question would be: how will Epistemology be capable of clarifying the foundations of the epistemic construction of KO and strengthen such theoretical studies?

Epistemology was taken into consideration by Japiassu (1977, p. 9, 25) after analyzing studies by authors, such as Blanché, Carnap and Lecourt. According to him, Epistemology would be the "genesis, development, structuring and articulation processes of scientific knowledge" or the "critical study of the principles, hypotheses and results of several sciences". Japiassu views science discourse as a strategic theory and science historicity as essential to an epistemological critique. The key concept is knowledge derived from scientific knowledge, which was initially linked to knowledge-state and, afterwards, to knowledge-process, understood as becoming

(*devenir*). According to Japiassu (1977, p. 27), the task of Epistemology would be to know this becoming (*devenir*) by analyzing "all stages of its restructuring", which would always result in "a temporary knowledge that is never finished or definitive".

Among the various lines of thought in Epistemology, Japiassu (1977) focuseson Piaget's genetic epistemology, Foucault's archaeological epistemology, Bachelard's historical epistemology and Popper's critical rationalism, which will be analyzed here because of its close relationship with the theme of the present research.

Epistemology has been studied in Information Science from both the historical and scientific perspectives, especially referring to interdisciplinarity. The theoretical foundations of Epistemology gave support to the first interdisciplinary studies in the area, particularly the ones by Japiassu, Ivani Fazenda and Olga Pombo, a Portuguese theoretician, among many other Brazilian and foreign authors.

Although Hjørland (2003) considers it difficult to outline the theoretical and scientific progress of KO, because different lines of thought seem to coexist, it is important to investigate the epistemology of its conceptions.

The theoretical-conceptual nature of this research is related to Epistemology and KO in the organization, structuring and systematization of knowledge for the constitution of knowledge organization systems: classification systems, thesauri, taxonomies, among others. Its objective is to analyze Popper's ideas on knowledge organization systems with an emphasis on the foundations of critical-rationalist epistemology, particularly objective knowledge, and the relationship between concepts and formulation of theories.

Knowledge Organization and Representation: current issues about knowledge organization systems

The development of KO as a theoretical field is linked to the foundation of theInternational Society for Knowledge Organization (ISKO) by Dahlberg, who was its first president. Before Dahlberg and the foundation of ISKO, the literature recognizes the decisive contribution of theoreticians, such as Ranganathan, Otlet and La Fontaine, Austin, Farradane, Kaiser, Coates and the Classification Research Group(CRG) to the theoretical conception and historical evolution of KO. The processes, tools, methods and knowledge organization systems, particularly the Universal Decimal Classification and the Dewey Decimal Classification, are well known and used throughout the world

In this respect, it is important to know Hjørland's conception of KO as an area of study that is specifically related to Information Science, Libraries and Archives and comprises activities such as "[...] document description, indexing andclassificationperformed in libraries, bibliographical databases, archives and other kinds of 'memory institutions' by librarians, archivists, information specialists, subject specialists"(Hjørland, 2008, p. 86). In a broader sense, i.e.,related to other disciplines, KOis about the way knowledge is organized at universities and other institutions, structures of disciplines, etc.

Hjørland believes thatthe relationship between the broader and narrower senses of KO can contribute to understanding how, in a narrower sense, it will be developed by means of systems and processes, that is, knowledge of the way other disciplinary fields, such as Chemistry or Biology, handle the organization of knowledge in their domains. Understanding this relationship is important and decisive to study the processes and systems of KO.

According to Hjørland (2008, p. 86), “Knowledge Organizationas a field of study is concerned with the nature and quality of such knowledge organizing processes (KOP) as well as the knowledge organizing systems (KOS) used to organize documents, document representations and concepts”. Therefore, KO is a field dedicated to the study of processes and systems of knowledge organization aiming to develop more qualitative proposals to other areas of knowledge.

In the context of KO, Zeng and Chan (2004, p. 377) consider “knowledge organization systems (KOS) is a general term referring to the tools that present the organized interpretation of knowledge structures”. According to the authors, nowadays several information retrieval systems coexist in a digital environment, such as bibliographic databases, online catalogs, digital libraries, institutional repositories, web-based subject directories, among others. Each information retrieval system has been structured over the years with languages with different vocabularies organized in various logical-hierarchical structures. Thus, there are many typologies of KOS and, as pointed out by Zeng and Chan (2004, p. 377-8), “they can be grouped under three general categories according to their structure and complexity, relationships between terms and historical function”, as follows:

1. Term lists, which include glossaries, authority files, dictionaries and geographical dictionaries;

2. Classifications and categorization schemes: libraries classification schemes, taxonomies and categorization schemes, and

3. Relational vocabularies: subject headings lists, thesauri, semantic networks and ontologies.

These knowledge organization systems are representations of knowledge organization structures at macro level whose contents, organized in controlled vocabularies of terms, represent concepts. This function of organization and concept representation is the most important characteristic of these systems that relates to Popper's theory of objective knowledge.

Popper's ideas on objective knowledge

The approach to Popper’s knowledge organization system and his notions of objective knowledge (1975, p.106) require considering his theory of the three worlds: the first world would be “the world of physical objects or of physical states”; the secondly “the world of states of consciousness, or of mental states,” and the third one, which is analyzed in this communication, refers to “the world of objective contents of

thought, especially of scientific and poetic thoughts and of works of art".

In the essay "Epistemology without a knowing subject" (Popper (1975, p. 106-152), the author discusses some aspects of his third world and relates it to Hegel's ideas and Plato's in particular, in order to identify points of convergence and divergence, as well as theoretical closeness or distance at different levels. Therefore, it is fundamental to draw a relationship between Popper's thoughts to Plato's theory of Forms and Ideas, as well as to Hegel's Objective Spirit, even though Popper himself said "my theory differs radically, in some decisive respects, from Plato's and Hegel's" (Popper, 1975, p. 106).

Popper (1975, p. 107) explains that the expression "third world" was adopted by convenience and provocative effect on those he calls "belief philosophers", among them Descartes and Kant. His problem was "to find better and bolder theories", because oft he importance of critique and not of belief. Although he recognizes that both theoretical systems, and especially problems and problem-solving situations, are part of the third world, he emphasizes the major relevance of critical arguments, state of discussion or state of critical argument, and adds: "and, of course, the contents ofjournals, books and libraries".

According to Popper (1975, p. 108), Epistemology is a "theory of scientific knowledge" and there are two different senses of knowledge or thought: in a subjective sense, it consists "of a state of mind or of consciousness or a disposition to react"; and "knowledge or thought in an objective sense, consisting of problems, theories and arguments as such". On comparing traditional epistemology with the one that he advocated, Popper (1975, p. 108) emphasized the relevance of "the study of scientific problems and problem situations, of scientific conjectures […], of scientific discussions" and he concluded that "the third world of a largely autonomous objective knowledge is decisively important for epistemology".

Henceforth, Popper poses questions, produces arguments and counterarguments, until he comes to a point that directly serves the purpose of a part of this paper: human language as a byproduct of the objective third world. "The world of language, of conjectures, theories and arguments - in brief, the universe of objective knowledge – is one of the most important of these man-created, yet at the same time, largely autonomous universes." (Popper, 1975, p.118) In Popper' theory, the idea of autonomy is central, although the third world is a human byproduct, one that creates its own autonomy domain. Autonomy, in turn, relates to another concept created by Popper (1975, p.120), the feedback, as follows: "The autonomy of the third world and the feedback of the third world upon the second one and even the first are important facts about the growth of knowledge." Moreover, "The most important of human creations, with the most important feedback effects upon ourselves and especially upon our brains, are the higher functions of human language: more especially, the descriptive function and the argumentative function" (Popper,1975, p. 119).

These two functions are considered higher and the most important ones in human

languages: the first one is “regulative truth” or “a description which fits the facts”. According to Popper (1975, p. 120), “further regulative or evaluative ideas are content, truth content, and verisimilitude.”

Plato believed that the third world, the world of Forms or Ideas, would generate ultimate explanations, that is, explanations by essences, expressed by hypostatized words. The objects of the third world were conceived by Plato as non-material things and, for this reason, “became concepts of things, or essences or natures of things, rather than theories or arguments or problems.” Popper states that, consequently, “from Plato until today, most philosophers have either been nominalists or else what I have called essentialists. They are more interested in the (essential) meaning of words than in the truth and falsity of theories” (Popper 1975, p.123). He presents the problem in the form of a table (see Table 1). According to him, “the left side of this table is unimportant, as compared to the right side: what should interest us are theories; truth; argument.”

Table 1: Ideas and the relationship between concepts and formulation of theories (Popper, 1975, 124).

IDEAS	
that is	
DESIGNATIONS or TERMS or CONCEPTS	STATEMENTS or PROPOSITIONS or THEORIES
may be formulated in	
WORDS	ASSERTIONS
which may be	
MEANINGFUL	TRUE
and their	
MEANING	TRUTH
may be reduced, by way of	
DEFINITIONS	DERIVATIONS
to that of	
UNDEFINED CONCEPTS	PRIMITIVE PROPOSITIONS
the attempt to establish (rather than reduce) by these means their	
MEANING	TRUTH
leads to an infinite regress	

This is one of the most relevant issues in KO, since either the left side in Table1, which is directly related to KO, or the right side, which represents theories and theoretical systems, truth or the truth of assertions, are the object of KO. According to Popper, “concepts are partly means of formulating theories, partly means of summing up theories. In any case their significance is mainly instrumental; and they may always be replaced by other concepts” (Popper 1975, p.123-124).

Changes in concepts (left side) also occur in knowledge organization systems because their tools represent concepts of knowledge fields (right side), which are transitional and liable to changes within time.

Discussion on knowledge organization systems and Popper's ideas on objective knowledge

When we consider Popper's thought in relation to ideas about objective knowledge, as discussed in the previous section, the relationship with knowledge organization systems, especially classification schemes and thesauri, is revealed.

In the literature about Information Science and KO, scholars have carried out a body of research on Epistemology and the theoretical foundations of KO: Zins (2003), Zeng e Chan (2003), Tennis (2008), Hjørland (2013) from abroad, and Miranda (2002) in Brazil, among others, deserve mention. The latter studies the need to establish a relationship between Information Science and objective knowledge. In his article, Tennis (2008) approaches Epistemology, theory and methods in KO with a view on classification, metatheory and research framework. Hjørland (2013) deals with knowledge and KO theories as well. Considerations in both fields are relevant in general and to this study in particular, but they are not analyzed here because they are out of the scope of this paper.

Zins (2004, p. 49), in an article about the epistemological perspective of KO,follows a research line similar to the one in this work, especially the relationship between the foundations of KO and the development of classification schemes and knowledge maps. Initially, the author points out six main stages that make up the philosophical argumentation and distinguishes objective knowledge, equivalent to an object or a thing, from subjective knowledge, meaning knowledge of a subject or of an "individual knower" and he admits the correlation between the two of them.

The fifth stage proposed by Zins (2004) closely relates to the issues discussed in this paper. The author claims that classification schemes of objective knowledge, such as the Library of Congress Classification Scheme (LCCC), influence our cognitive map and are subject to empirical scientific verification. In the sixth stage, Zins (2004) states that epistemological analysis contributes to distinguish between two kinds of structures: "conceptual cognitive pre-experiential structures and external recorded or documented structures". Consequently, there are two major structuring approaches: "rationalistic (i,e., phenomenological or conceptually based) and empirical structuring methods" (2004, p. 50).

Zins (2004) points out that the concept of "knowledge" is used in various meanings and contexts. He asserts that "knowledge as a state of mind is a product of a synthesis" (Zins 2004, p.50-51).He also gives some examples based on a concept map that, "after being recorded or documented, becomes an object or thing". As such "it becomes part of the objective, or rather universal, knowledge"(Zins 2004, p.53).

As mentioned earlier, Dahlberg (2006) states that KO is a new science encompassing a huge set of concepts, theories, methods, practices, etc. Its connection with Epistemology reinforces the idea that scientific knowledge goes from knowledge-state to knowledge-process, that is, the becoming (*devenir*), which is the task of

Epistemology. And how is it done? Through an analysis of its restructuring stages, which would have a provisional not a definitive result, as Popper presents in his Table of Ideas, and would lead to an "infinite regress". Epistemological aspects are applied to knowledge organization systems tools by interpreting the structures of KO (Zeng and Chan, 2004, p.377), as it occurs, for instance, in classification schemes, glossaries, thesauri and semantic networks, and ontologies.

The specialized literature in the field of KO recognizes that concepts are fundamental and regulate their tools, such as classification schemes and thesauri. Therefore, as they represent concepts of scientific knowledge in various specific areas, they are transitional and liable to changes. It can be concluded that research based on the application of Popper's objective knowledge (Popper 1975) and Epistemology in general can contribute to further development of theoretical issues.

References

Campos, Maria Luiza (1995) Linguagens documentárias: núcleo básico de conhecimento para seu estudo. *Revista da Escola de Biblioteconomia da UFMG*, 24(1): 52-62,

Dahlberg, Ingetraut (2006). Knowledge organization: a new science? *Knowledge Organization*, 33(1): 11–19.

Hjørland, Birger. (2008). What is Knowledge Organization (KO)? *Knowledge Organization*, 35(2/3): 86-101.

Hjørland, Birger. (2013). Theories of knowledge organization: theories of knowledge. In *Meeting of the German ISKO, 13th. Keynote Potsdam, 19th to 20th March 2013*. [https://scholar.google.com.br/scholar?q=theories+of+knowledge+organization+%E2%80%94+theories+of+knowledge&hl=pt-BR&as_sdt=0&as_vis=1&oi=scholart&sa=X&ved=0ahUKEwiL34uj6dzMAhWCG5AKHVuWCWcQgQMIGjAA].

Japiassu, Hilton (1977) *Introdução ao pensamento epistemológico*. 2.ed. Rio de Janeiro. F. Alves.

Miranda, António. (2002). A Ciência da Informação e o conhecimento objetivo: um relacionamento necessário.In Aquino,Mirian de Albuquerque. *O campo da ciência da informação: gênese, conexões e especificidades.*João Pessoa: Editora Universitária/ UFPB. Pp.9-24.

Popper, Karl R. (1975) *Objective knowledge: an evolutionary approach*. Oxford: Clarendon Press.

Platão (1999). *Fédon*. São Paulo: Nova Cultural.

Tennis, Joseph T. (2008). Epistemology, theory, and methodology in knowledge organization: toward a classification, metatheory, and research framework. *Knowledge Organization*, 35(2/3): 102-12.

Zeng, Marcia Lei & Chan, Lois Mai (2004) Trends and issues in establishing interoperability among knowledge organization system. *Journal of the American Society for Information Science and Technology*, 55(5): 377-95.

Zins, Chaim (2004) Knowledge organization: an epistemological perspective. *KnowledgeOrganization*, 31(1): 49-54.

Paula Carina de Araújo and José Augusto Chaves Guimarães

Epistemology of Knowledge Organization: A Study of Epistemic Communities

Abstract

The article analyzes the theoretical relationship between the authors that research about Epistemology of Knowledge Organization domain through author bibliographic coupling from the scientific production indexed in the databases Web of Science (WoS) and Scopus. We identified a corpus of 66 articles on Epistemology of Knowledge Organization and we established a cut in the corpus and selected only the scientific production of the authors represented with two or more articles, a total of 22 articles. The implicit relationship between Hjørland and Gnoli is the strongest in the bibliographic coupling because they share 21 theoretical references. We found similarity between Hjørland and Smiraglia because they have 18 related theoretical references. It is also noteworthy that Hjørland is the most representative author in the author bibliographic coupling both for having the largest number of articles in the corpus, and for having the strongest implicit relationships with other authors and for being widely cited in the analyzed studies. There are two important considerations about this research that is the new knowledge we present: first, there is the need to consolidate and systematize the terminology in the domain to describe the studies on Epistemology of Knowledge Organization because there is a restricted circle of researchers in KO using the same terminology to describe their research on this subject. Second, there is a lack of thematic representation in key fields of the articles (title, abstract and keywords), which generates an informational gap. For further studies, we suggest developing co-citation analysis and content analysis to deepen these research findings.

Introduction

"Epistemology is *how we know*. In KO we make implicit epistemic statements about knowledge of concepts, acts (such as representations), entities, and systems". The author understands that "in so doing, we create knowledge, and our epistemic stance dictates what kind of knowledge that is. [...] There is the added burden of embodying your epistemic stance in your method and in your writing, which leads to a number of misunderstandings in scholarly communication". (Tennis 2008, 103).

Hjørland explains that in KO context, "epistemology is the philosophical study of knowledge, and epistemologies are theories or approaches to knowledge". And, the "epistemology can be seen as the generalization and interpretation of collected scientific experience". (2002, 439).

In this context, the main objective of this research is to analyze the theoretical relationship between the authors that research about Epistemology of Knowledge Organization domain through author bibliographic coupling from the scientific production indexed in the databases Web of Science (WoS) and Scopus.

The proposal was to use one kind of citation analysis, the author bibliographic coupling, to analyze part of the domain. We agree with Castanha and Gracio that "bibliometric approach provides a valuable understanding both to the information design and to the theoretical understanding of the social process that permeates the information, including historical processes". (2014, 173).

This choice can be justified by the comprehension that epistemological studies

allows the self-knowledge, self build and interdisciplinarity of the domain like was proposed by Rendon Rojas (2008, 1). We believe that "considering Knowledge Organization as a domain in a continuous process of theoretical-methodological consolidation, it becomes important to identify its epistemological configuration and "epistemic communities" in order to measure its impact on society and academy." (Guimarães, Martínez-Ávila and Alves 2015, 1).

"Epistemic communities work through connectivity, perhaps not so much by connecting people, but by connecting objects and subjects, people and places, production and distribution, individuals and collectives, histories and futures, the virtual and the concrete". (Meyer and Molyneux-Hodgson 2010, 5).

Therefore, we understand that in a domain, in addition to discursive communities, we need to consider the epistemic communities, as the connection between authors, scientific production and theory through implicit relationships are evident, for example, as proposed in this study, to provide a network of bibliographic coupling on the following pages.

Methodology

A qualitative and exploratory research was conducted through the collection of scientific articles available on Web of Science (25 articles) and Scopus (41 articles) databases. We used the following search strategy: [(epistemolog* OR "theory of knowledge") AND ("knowledge organization" OR "information organization")] and the fields title, keywords and abstract were considered. We applied the filter "scientific articles" to retrieve only this kind of document in both databases. The bibliographic manager Zotero was the tool used to collect, store and organize the articles. The author bibliographic coupling was the technique used to present and to analyze the results, as it is considered one method of citation analysis. (Marshkova 1981).

We decided to use the author bibliographic coupling in this study to have an initial view of the Epistemology domain of KO. We are in line with Zhao and Strotmann (2008) that through bibliographic coupling, it is possible to recognize the authors that influence a domain.

Epistemology of Knowledge Organization

"It is the interactions of the ontological, epistemological, and sociological priorities that define a domain's work as productive activity and thus reveals its critical role both in the evolution of knowledge and in the comprehension of knowledge as a scientific entity". (Smiraglia 2015, 7). Smiraglia explains the importance of the relation between epistemology, ontology and methodology to determine a domain. He believes that "just as domain analysis for knowledge organization has incorporated many theoretical perspectives, so has it been demonstrated to be a multimethod paradigm". (2015, 15).

We consider Epistemology of Knowledge Organization as a domain in this research and the analysis of the relation between the authors and theories they are based is the

first step to recognize the foundations of this domain. As López-Huertas (2015, 578) we agree that “the identification of paradigmatic structures in disciplines might not be an easy task, but disciplines have the advantage of having a well defined and delimited discourse, have a historical background, a tradition, that helps in keeping the trace and the evolution of their paradigms and theories”.

The epistemological approach is perceived in many of Hjørland's studies. The author claimed that research is always based on specific epistemological ideals. In his thinking, “epistemology is, however, the best general background that is possible to teach people within information science. It is the best general preparation we can provide for people in order to study any domain”. (2013, 169).

Following this same line, Tennis explains that “epistemology is an important part of the KO armature because it reflects our assumptions about language, the primary material of Knowledge Organization Systems (KOS). Dousa’s research (2014, 152) demonstrates this assumption. He analyzed Julius Otto Kaiser’s method of Systematic Indexing (SI) and Brian Vickery’s method of facet analysis (FA) for document classification and it was possible to identify the epistemological and methodological eclecticism in the construction of Knowledge Organization Systems (KOSs) based on Hjørland’s typology of epistemological position (2003).

Therefore, we understand “even a casual glance at the literature shows that epistemic, theoretical, and methodological concerns constitute the driving force behind argument and findings in much of the conceptual work of KO”. (Tennis, 2008, p. 102). Following this thought, we propose the analysis of the Epistemology domain of KO. This analysis is proposed to be developed using bibliometric studies.

“Bibliometrics is a strong approach because it shows many detailed and real connections between individual documents” (Hjørland 2002, 433). Smiraglia (2015, 11) argue that “author productivity is frequently a bibliometric measure that can help identify both research fronts and invisible colleges”. We agree with Hjørland (2002) that the combination of bibliometric analyses and epistemological studies represents a support to domain analysis and it allows to identify the epistemic community of Epistemology of KO.

“An epistemic community is a network of professionals with recognized expertise and competence in a particular domain and an authoritative claim to policy-relevant knowledge within that domain or issue-area”. Therefore, “what bonds members of an epistemic community is their shared belief or faith in the verity and the applicability of particular forms of knowledge or specific truths”. (Haas 1992, 3).

This is the initial study that will support the development of a wide research that proposes to identify epistemology positions and their relation to the methods used. The findings allow to describe the domain and its epistemic community entirely. In this research, we identified some of the members of the epistemic communities and their theoretical influences as described in the following section.

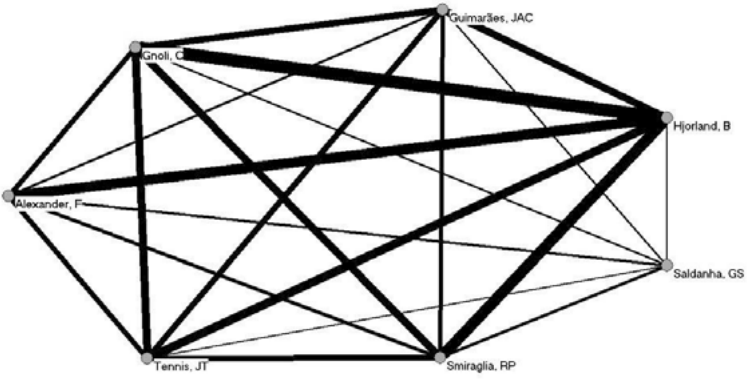
Epistemic Communities in the Epistemology Domain of Knowledge Organization

The corpus of this research was formed by 66 scientific articles indexed in WoS and Scopus. The most representative journal in the corpus of this research is Knowledge Organization represented by 23 articles, and secondly the Journal of Documentation with 7 articles. The results point out that the analyzed scientific production is mostly written in single authorship, with a total of 46 articles. Multiple authorship was identified in 20 articles.

The authors with 2 or more papers in the corpus were selected, and we only considered the first author in the papers written in co-authorship because in Information Science the first author is responsible for the research. Therefore, we developed author bibliographic coupling for 7 main authors with 2 or more, which led to the analysis of 22 articles, forming the new corpus of this research.

Birger Hjørland was the most representative author in the corpus, considered for the author bibliographic coupling. He is recognized by his research on epistemology and socio-cognitive approach in KO. We verified the co-occurrence of authors in the scientific production of researchers identified in the corpus of this study through their articles about epistemology of KO indexed in WoS and Scopus, the main multidisciplinary and international databases, as we can see in Figure 1.

Figure1: Author bibliographic coupling network between 7 authors with 2 or more items in the search corpus.



We realized that the strongest relationship is between Hjørland, B and Gnoli, C. In the scientific production analyzed in this research, the authors cited papers from fundamental authors to the domain, such as: Bliss, HE; Dahlberg, I; Olson, H; Ranganathan, SR; among others. The authors used 21 common authors in the articles analyzed in this research. Therefore, we can state that there is an implicit theoretical relationship between the two researchers, considering that there is great similarity between the theoretical references used by both.

There is also strong proximity between Hjørland and Smiraglia. The authors share 18 theoretical references in the articles analyzed. They also cited seminal authors in the KO domain (Bliss, Dahlberg, Frohman, Olson, Ranganathan, etc.), but also used the other papers from authors recognized by their studies in epistemology of KO, such as Hjørland, Mai and Tennis. Therefore, the theoretical proximity between research authors from the results of bibliographic coupling is confirmed.

In this context, we recall Kessler (1965) when he stated that the set of references used by authors in their article highlights the intellectual environment in which they work and if two articles have similar bibliographies, there is an implicit relationship between them.

Alexander and Hjørland's papers also have a strong implicit relationship, as they share 17 theoretical references in their research. We note Hjørland in stronger relationships of author bibliographic coupling network. His research is also highlighted among the citations of coupled authors. This is because Hjørland is one of the leading authors in the Epistemology of KO domain, which can be seen in most of his publications. On the other hand, Saldanha was the author that had the lowest implicit relationship network with the other authors, especially Hjørland and Tennis with whom he shared only two theoretical references, despite his research theme is closely related to these authors' thematic.

Conclusion

The results of this research show that author bibliographic coupling is an important tool for visualizing a domain, as it enables to recognize the actors involved, and characterizes its epistemic community. It is noteworthy the need to use metric studies to analyze a domain and better recognize it.

The most notably author throughout the analysis was Hjørland, B, as he is the most productive author in the corpus (8 papers), representing a significant scientific production on the theme Epistemology of KO. In addition, he is the author with the strongest implicit relationships in the bibliographic coupling network, and he is also one of the authors with more cited papers by the researchers that make up the network. The three findings are interconnected, because, since he is the most productive author on the subject, and one of the precursors on the epistemological approach in KO, naturally, he is also an important theoretical reference for the domain.

There are two important considerations highlighted in this research and which can be understood as the new presented knowledge: First, there is a restricted circle of researchers in KO using the same terminology to describe the Epistemology studies of KO. This observation leads us to conclude that even in an area turned to KO, the need to consolidate and systematize the terminology is evident.

Second, when we use title, abstract and keywords as search fields to locate the articles on Epistemology of KO, we assume that these fields accurately present the thematic representation of research. However, given that many authors who develop

epistemological studies in KO were not included in the corpus of this research, we can state that there is a lack of thematic representation in key fields of their articles, which generates an informational gap. Also, possibly, these authors do not take the Epistemology of Knowledge Organization as a theme for their research.

The results also allow us to infer that the epistemic community recognized in this research is formed by researchers with a consolidated work in KO, and in most cases, often publish on the theme Epistemology in KO in single authorship. It is also possible to identify an important theoretical group for KO.

It is suggested that reference bibliographic coupling could be made to deepen the description of theoretical proximity among the authors as an opportunities to deepen the analysis of this domain. Other metric studies can also be used to recognize this domain.

References

Castanha, Renata Cristina & Grácio, Maria Cláudia Cabrini (2014). Bibliometrics contribution to the metatheoretical and domain analysis studies. *Knowledge Organization*, 41(2): 171-174.

Dousa, Thomas M. & Ibekwe-SanJuan, Fidelia (2014). Epistemological and methodological ecleticism in the construction of Knowledge Organization Systems (KOSs): the case of analytico-synthetic KOSs. *Advances in Knowledge Organization*, 14: 152-159.

Guimarães, José Augusto Chaves, Martínez-Ávila, Daniel& Alves, Bruno Henrique (2010). Epistemic communities in knowledge organization: an analysis of research trends in the Knowledge Organization Journal.Paper presented at *the meeting of the International Society for Knowledge Organization – Chapter United Kingdon, London, UK, July 13-14, 2015*.

Haas, Peter M. (1992). Introduction: epistemic communities and international policy coordination. *International Organization*, 46(1) Winter: 1-35.

Hjørland, Birger (2002). Domain analysis in information science: eleven approaches - traditional as well as innovative. *Journal of Documentation*, 58(4) January: 422-462.

Hjørland, Birger (2003).Fundamentals of knowledge organization. *Knowledge Organization*, 30: 87-111.

Hjørland,Birger (2013). Theories of knowledge organization: theories of knowledge. *Knowledge Organization*, 40(3): 169-181.

Kessler, Meyer Mike (1965). Comparison of the results of bibliographic coupling and analytic subject indexing. *American Documentation*, 16(3): 223–233.

López-Huertas, María J. (2015). Domain Analysis for Interdisciplinary Knowledge Domains. *Knowledge Organization*, 42(8): 570-580.

Marshakova, Irina V. (1981). Citation networks in information science. *Scientometrics*, 31(1): 13-16.

Meyer, Morgan& Molyneux-Hodgson, Susan (2010). Introduction: the dynamics of epistemic communities.*Sociological Research Online*, 15(2): 1-7.

Rendón Rojas, Miguel Ángel (2008). La ciencia de la información en el contexto de las ciencias sociales y humanas: ontología, epistemología, metodología e interdisciplina. *Datagramazero: revista de ciência da informação*, 9(4).

Smiraglia, Richard P. (2015). D*omain analysis for knowledge organization: tools for ontology extraction*. Chandos Information professional Series. Waltham: Elsevier Chandos Pub.

Tennis, Joseph T. (2008). Epistemology, theory, and methodology in Knowledge Organization: toward a classification, metatheory, and research framework. *Knowledge Organization*, 35(2/3): 102-112.

Zhao, Dangzhi & Strotmann, Andreas (2008). Evolution of research activities and intellectual influences in Information Science 1996–2005: introducing author bibliographic-coupling analysis. *Journal of the American Society for Information Science and Technology* 59(13): 2070-2086.

Laura Ridenour and Richard P. Smiraglia

How Interdisciplinary is *Knowledge Organization*? An Epistemological View of Knowledge Organization as a Domain

Abstract

The epistemic influences that shape research in knowledge organization come from many disciplines, but this interdisciplinarity is rarely made explicit in research in the domain. In order to gain understanding about the multiple influences on our domain, and to discover one aspect of interdisciplinarity in knowledge organization we have attempted to measure the interdisciplinarity of the journal Knowledge Organization. For this first attempt at empirical observation of interdisciplinarity we have chosen to study KO between 2011 and 2015.

Background

Although knowledge organization as a practice is (arguably) eternal, the science of knowledge organization (KO) is sometimes reckoned to the origins of the International Society for Knowledge Organization (ISKO), which was founded by Ingetraut Dahlberg in 1989 (Dahlberg 2006, 11). Another identifiable starting point for the science of knowledge organization is the birth date of the domain's only scientific journal. In 1974 the journal began publishing with the title *International Classification* (12). Beginning with number 1 in 1993, the journal's title was changed to the current *Knowledge Organization*. The rationale for this change was described by Dalhberg (1993, 1):

> The new title, denoting a superordinate concept to "classification," clearly indicates that we do not wish to confine ourselves to the problems falling under the "classification" concept, but rather are interested—as in fact, we always have been, although many a one did not notice it—in all questions of knowledge organization such as they are now alluded to in the subtitle of our journal: hence in Conceptology, Classification (including Thesaurus Problems), Indexing, and Knowledge Representation (including the relevant Linguistic Problems and Terminology.

For twenty-three years the journal has continued as the sole purveyor of the science generated by ISKO members. Although other journals contain core KO literature as well, *Knowledge Organizatio*n is the only journal sponsored by ISKO and solely devoted to research in KO. But, research in KO, as Dahlberg noted in 1993, is intended to be inter- and multi-disciplinary and as such, bounded not by the normal boundaries of disciplinary theoretical paradigms, but rather by the core interest in the ordering of concepts, the organization of knowledge, which is endemic to all disciplines. Considering the eclectic reach of our field, an interesting research question unfolds: how interdisciplinary is Knowledge Organization?

Several domain analytical studies of the literature of KO have produced some basic facts about the epistemological ordering of the domain. For instance, domain analyses of KO rely on the contents of the journal, added to the proceedings of the biennial international conferences of ISKO, the biennial conferences of ISKO's regional

chapters, and the papers presented at each annual classification workshop conducted by the Special Interest Group for Classification Research (SIG-CR) of the organization now known as the Association for Information Science and Technology (ASIST). In 2012 Smiraglia's keynote presentation to the Mysore international ISKO conference contained results of a meta-analysis of domain analyses of KO. In that paper we learned, for example, that (6):

> KO as a domain has robust and continuous formal publication venues that help to maintain domain coherence. In KO, theoretical poles are both conceptual and methodological. The domain is scientific, but also has deep roots in humanistic methods and modes of thought.... Thus we see consistently marked dimensions within the domain theoretical versus applied on one continuum, humanistic versus scientific on another.

For example, analyses of recent international ISKO conferences has led to the rough hypothesis (Smiraglia 2014, 345) that "there is an observable dichotomy in KO in which roughly equal numbers of research papers are epistemologically either empirical science or humanistic narrative. The former tend to have few recent citations, and the latter tend to have many older citations." In the 2014 Krakow conference, works cited had a mean age range between 4.7 and 41.3 years. Works cited in conference papers were split roughly between 49% research papers and 35% monographic sources (348).

The most recent meta-analysis (Smiraglia 2015, 603) demonstrates that the majority of papers in the domain appear in conference proceedings, most papers are either informetric or terminological, most works cite recent scholarship, but a large proportion of humanistic methodologies keep the mean age of works cited at around 11.45 years (605). There is no clear influence from any one region, although most papers originate in North America, Denmark, or Brazil. Typically (610) there is an even division in the research front between papers reporting empirical results and those relying on humanistic approaches; this is evident in the split between journal and monographic sources cited by authors in the domain.

Measures of interdisciplinarity

Interdisciplinary research has become increasingly an increasingly popular topic, especially in the sciences. Reasons for the increasing interest in interdisciplinary research vary, but an increasing amount of research funding for interdisciplinary groups and problems has been made available since 2003 (Porter and Rafols 2009). In order to understand how interdisciplinary something is, one must first create a measure of interdisciplinarity.The simplest measure of how interdisciplinary a document is can be calculated based on the number of categories cited by the document (Moed 2015).Much research examining the interdisciplinary draw of an area of research has been conducted using Thomson Reuters Essential Science Indicator categories to determine the spread of citations (Moed 2015). In this paper, we apply the same principles to Scopus data and Scopus categories.

Units of measurement in scholarly communication and science indicators illuminate different aspects of structures of knowledge. These units include authors, articles,

journals, institutions, and identified areas of research. In this analysis, we engage in "navel gazing" to better understand the spread of intellectual and epistemologicalinfluences of KO as evidenced through the spread of the journal's citations of broad science categories found in Scopus.

Scopus contains classifications for journals and conference proceedings from 1980 forward, which are available in a master file from Scopus. Up to five All Science Journal Classifications (ASJC) codes to each type of publication venue (Scopus 2015). A total of 334 four-digit classification codes are contained in ASJC. Classifications are a strict taxonomy, with a 2-digit identifier indicating the top-level classification to which each child classification code is assigned. We assume that the more interdisciplinary a publication venue is, the more Scopus categories will be assigned to the journal from widespread top-level categories.

Methodology

Similarly to how Larivière et al. (2012) examined the interdisciplinarity of information science journals, we examine the interdisciplinarity of *Knowledge Organization* over time. Our dataset consists of five years of the journal *Knowledge Organization*, from 2011-2015, as indexed by Scopus. A bibliographic list of citations and publications was gathered from Scopus. Cited references were gathered by year of the publication to make the analysis easier, that is to say, all articles published in 2011 were separated out, and the references from those records were gathered in Scopus.

Journals and conference proceedings were matched to their classification codes by joining the file to the codes in SQL, and processed in R to gather the counts of both categorical and journal citations per year. Though all categories were examined, only top-level classification was included in this analysis due to word limit constraints.

Results

The top journals cited are shown in Table 1. Unsurprisingly, the top citations were to LIS journals including *Journal of the Association for Information Science and* Technology and its forerunners (*JASIST*), the *Journal of Documentation*, and *Knowledge Organization*.

Table 1. Source Titles with Nine or More References from 2011-2015.

Source Title	2015	2014	2013	2012	2011	Total
Journal of the American Society for Information Science and Technology etc.	110	105	102		148	465
Journal of Documentation	32	50	48	22	38	190
Knowledge Organization	33	56	28	24	47	188
Scientometrics	76		40		44	160
Information Processing and Management	50	40				90

Journal of Information Science		18	20			38
Advances in Knowledge Organization				27		27
Conference on Human Factors in Computing Systems - Proceedings					24	24
Annual Review of Information Science and Technology			20			20
Axiomathes			16			16
Library Trends					12	12
Cataloging and Classification Quarterly				9		9
Library Journal		8				8
Total	**301**	**277**	**274**	**82**	**313**	**1247**

Twenty-three categorical codes appeared in the dataset. These terms represent the visible interdisciplinarity of the domain broadly and are shown in Table 2.

Table 2. AJSC Category Codes in the Dataset.

Agricultural and Biological Sciences	General
Arts and Humanities	Health Professions
Biochemistry, Genetics, and Molecular Biology	Immunology and Microbiology
Business, Management, and Accounting	Materials Science
Chemical Engineering	Mathematics
Chemistry	Medicine
Computer Science	Neuroscience
Decision Sciences	Nursing
Earth and Planetary Sciences	Physics and Astronomy
Economics, Econometrics, and Finance	Psychology
Engineering	Social Sciences
Environmental Science	

The top categorical citations were to Computer Science (436), Social Sciences (378), and Arts and Humanities (123). These counts are shown in Figure 1. A co-occurrence map was produced by recording each co-occurrence of two codes in a top tier journal (those shown in Table 1 above). A multi-dimensionally scaled plot was produced using IBM-SPSS™, and that plot is shown in Figure 2. Essentially the plot shows two regions. One, with the dark boundary, contains both mathematics and the arts and humanities. It seems likely that this region indicates disciplinary space that is perceived as most distant from knowledge organization itself. The other region contains two sub-regions, both indicated with dashed boundaries; one comprises

engineering and computer science, and the other related region contains decision sciences and social sciences. Obviously this represent the technical dimension of interdisciplinarity, and all sectors of this region are places where aspects of knowledge organization may intersect if not interact with the named disciplines.

Figure 1: Top Category Citations by Year.

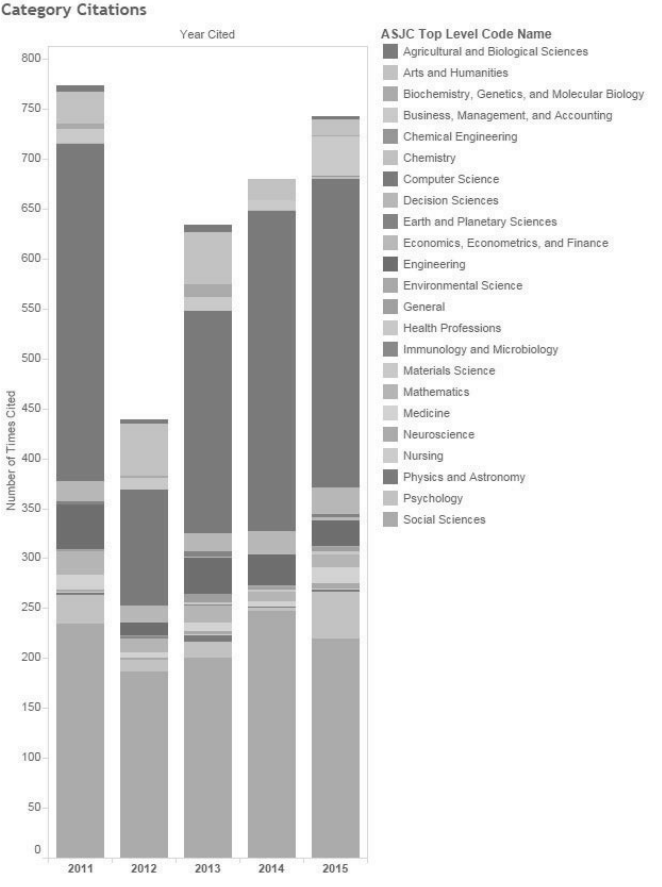
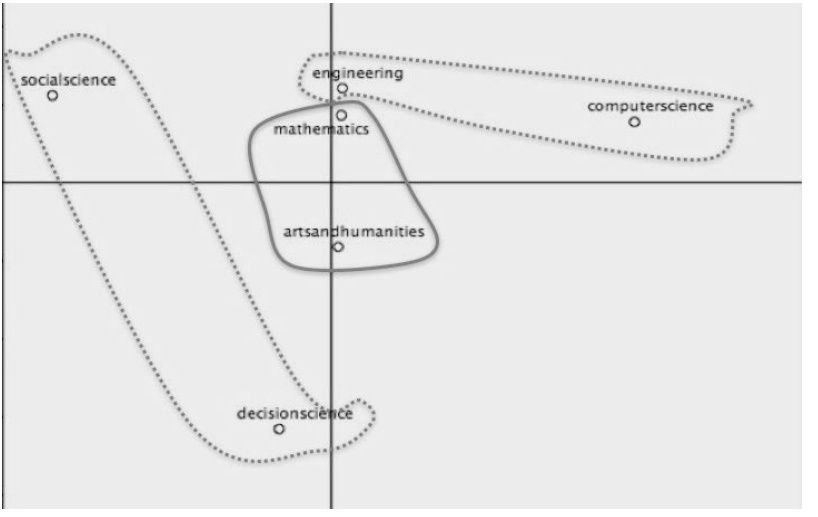


Figure 2. Category Co-occurrence Plot (stress = 0.6624 R-squared = 0.59217).



Another opportunity for visualization comes from the fact that the most-cited journals co-occur within the AJSC categorical regions. These co-occurrences were plotted and a multi-dimensionally scaled map was produced using IBM-SPSS™ (Figure 3).

Figure 3. Journal co-occurrence plot (stress = 0.18024 R-squared = 0.96020).

This plot is less representative; there are twenty-four journals in the matrix but only major nodes are visualized here. There are two regions shown. The region bounded by the solid line contains classical information science, ranging from *JASIST* and *IPM* to *Webology, Applied Ontology, Online Information Review* and *Information Retrieval and Canadian Journal of Information and Library Science* (not shown). The region is tightly bounded and fairly distant from all of the rest. The remaining journals range from *First Monday* to *Journal of Documentation, Information Systems, Journal of Digital Information, Cognitive Psychology, ARIST* and *International Journal of Human Computer Studies* and nine others not shown. This demonstrates the breadth of journal productivity that is considered relevant to knowledge organization research, and also points to interdisciplinarity, especially as viewed in Figure 2.

Conclusions

KO the domain relies on a broad disciplinary spread of journals as the basis on which research to create new knowledge is undertaken. Like its sibling (or cousin) information science, KO borrows from almost all disciplines because its primary concern, the conceptual order of knowledge, is meta-disciplinary. As such, it is not a surprise to discover interdisciplinarity in *Knowledge Organization* the journal. The breadth of topical categories is testimony to the wealth of interdisciplinary thought that undergirds the domain of KO.

However, there is an interesting fly in this ointment, and that is the close-knit clustering of journals from neighboring (or sibling or cousin) information science, which clearly contribute much to this picture of interdisciplinarity. The question arises, then, how interdisciplinary is this interdisciplinarity? It is clear that the most influential journals are those from information science and computer science. The rest of the interdisciplinary range emerges from social sciences, decision sciences, mathematics and even arts and humanities, all serving as test-beds for conceptual theories that arise in the core of KO.

This study was limited by the availability of data within Scopus; only five years of *Knowledge Organization* are indexed in Scopus, but *Knowledge Organization* as a journal has been available as *KO* or International Classification since 1974. To better understand the evolution of disciplinary influence in the journal, a larger span of data for a longitudinal study will be required.

References

Dahlberg, Ingetraut (2006). Knowledge Organization: A New Science? *Knowledge Organization*, 33: 11-19.

Dahlberg, Ingetraut (1993). Editorial: Why ‘Knowledge Organization’? The Reasons for IC’s Change of Name. *Knowledge Organization*, 20: 1.

Moed, Henk F. (2015). The Future of Research Evaluation Rests With an Intelligent Combination of Advanced Metrics and Transparent Peer Review. In *Scholarly Metrics Under the Microscope: From Citation Analysis to Academic Auditing*, edited by Blaise Cronin and

Cassidy R Sugimoto, 693–712. ASIS&T Monograph Series. Medford, NJ: Information Today, Inc.

Larivière, Vincent, Sugimoto, Cassidy R. & Cronin, Blaise (2012). A bibliometric chronicling of library and information science's first hundred years. *Journal of the American Society for Information Science and Technology*, 63: 997-1016.

Porter, Alan, and Rafols, Ismael (2009). "Is science becoming more interdisciplinary? Measuring and mapping six research fields over time." *Scientometrics* 81: 719-745.

Scopus (2015). “All Science Journal Classification Codes.” Elsevier.

Smiraglia, Richard P. (2012). “Universes, Dimensions, Domains, Intensions and Extensions: Knowledge Organization for the 21st Century.” In *Categories, Contexts and Relations in Knowledge Organization: Proceedings of the Twelfth International ISKO Conference, 6-9, August 2012, Mysore, India*, ed. A. Neelameghan and K.S. Raghavan. Advances in Knowledge Organization v. 13. Würzburg: Ergon-Verlag, pp. 1-7.

Smiraglia, Richard P. (2014). “ISKO 13’s Bookshelf: Knowledge Organization, the Science, Thrives—An Editorial.” *Knowledge Organization* 41: 343-56.

Smiraglia, Richard P. (2015). “Domain Analysis of Domain Analysis for Knowledge Organization: Observations on an Emergent Methodological Cluster.” *Knowledge Organization* 42:602-11.

Jay H. Bernstein

Anthropology and Knowledge Organization: Affinities and Prospects for Engagement

Abstract

Anthropology, like knowledge organization (KO), studies the structuring of knowledge through classification and categorization, but it provides a distinctive perspective by analyzing knowledge as a cultural phenomenon. Unlike KO, which examines knowledge in documents and collections, anthropology studies knowledge embedded in lived activities and practices, paying attention to socially conditioned concepts of what counts as knowledge and how it is legitimized. The study of concepts is one of several topics that would benefit from an approach combining anthropological and KO perspectives.

Knowledge organization (KO) has traditionally focused on bibliographical records and texts and, as such, has been practiced mainly within library and information science (LIS). However, a broader context of KO would include the organization of knowledge on all levels and in all domains. Birger Hjørland (2003, 2008), a leader in the field, argues for a maximally broad definition of KO that includes both the intellectual or cognitive organization and the social organization of knowledge, and he suggests that to evolve as a science, KO must engage with these broader issues. Although he discusses philosophical schools of thought and psychological theories in presenting an overview of central issues in and approaches to KO, he does not mention anthropology, even though the latter discipline covers several topics within KO's broad context as identified by him, such as symbolic and conceptual systems. The low occurrence of interchange and mutual awareness between anthropology and KO are unfortunate because they would appear to share much common ground.

Anthropology is salient among the academic disciplines that study knowledge. Not only is it associated with a distinctive approach to knowledge, but it would not be an exaggeration to say that it is largely focused on knowledge and its structuring. To the extent that it includes a description and analysis of knowledge as a human phenomenon, the anthropological approach contributes to the broader context of KO mentioned by Hjørland, bringing together theoretical and scholarly approaches to the organization of knowledge in all senses of the word, including the approaches to knowledge and its organization across the entire spectrum of domains and disciplines.

Of course, the anthropological perspective and orientation on knowledge is quite different from that of KO, as are the intent and purpose of the two disciplines. The divergence in their approaches to knowledge can be explained in part by their genealogies. Anthropology coalesced following the age of discovery and exploration, and its closest affinities are with sociology, geography, history, psychology, and biology (Vermeuluen, 2015). Approaching the problem of knowledge as a social and behavioral science, anthropology seeks to describe, document, and analyze how humans relate to knowledge, including how they organize it, but does not prescribe or

set about to create or improve the organization of knowledge. By contrast, KO originated as an applied science growing out of the development of bibliographic records in managing libraries and a practical focus on the classification of books and documents. KO scholarship, both in the narrow and the broader sense, serves as an intellectual foundation for the practical goal of doing the work of organizing knowledge in all media through the development of knowledge organizing systems. KO in this sense is a design science (Simon, 1969) aimed at improving and perfecting systems of access to knowledge and an informing science aimed at serving the needs of a clientele (Eli Cohen 2009).

Anthropology and the Concept of Culture

Anthropologists approach knowledge as a cultural object or product. Culture is a kind of phenomenon unto itself comprising traditions, customary practices, aesthetics as found in clothing, architecture, dance, music, and verbal arts, legends, folklore, stories, and all kinds of minutiae. Although the notion of culture as it relates to human existence existed before anthropologists took it up, it had been applied only to higher civilization (as in classical music, architecture, or poetry) until 1871, when the British anthropologist Edward Burnett Tylor wrote a two-volume book, *Primitive Culture*, that effectively seized the culture concept as the primary conceptual tool for anthropology. In his opening sentences Tylor set forth the first anthropological definition of culture as "that complex whole which includes knowledge, belief, art, morals, law, custom, and any other capabilities and habits acquired by man as a member of society." Although Tylor's own concept of culture was not a modern anthropological one, as many have supposed (Stocking, 1968; Singer, 1968; cf. Kroeber & Kluckhohn, 1952/1963), the wording of his definition gave a warrant for later generations of anthropologists to connect culture to knowledge and cognition in a distinctively anthropological way. The word "capabilities" in Tylor's definition connoted the "ability of people to acquire and produce knowledge, beliefs, etc." (Blount, 2011, 13), and this pointed the way for anthropologists to think of culture in terms of knowledge and belief. By the 1960s, many anthropologists were occupied by issues of semantics, classification, and knowledge more generally. Exactly a hundred years after the publication of Tylor's seminal text, a leading anthropologist, Ward H. Goodenough, defined culture as "what is learned," and "the things one needs to know in order to meet the standards of others" (Goodenough 1971, 19). Viewing culture in this way, Goodenough (1971, 20) asserted that "a valid description of a culture as something learned is one that predicts whether or not any particular action will be accepted by those who know the culture as conforming to their standards of conduct." Anthropologists who accepted Goodenough's model sought to elicit frames for vocabulary words to represent cognitive models that fit into folk taxonomies of various domains (Tyler, 1969).

While Goodenough's mentalistic approach to culture was not universally accepted and was only one of several theories of culture that have been debated (Keesing, 1981),

anthropologists have generally viewed culture as intimately related to knowledge and have written extensively in their ethnographies about the uses of knowledge, especially in traditional, non-literate societies. The ethnographic method of participant observation was first developed by anthropologists and was designed for research in small-scale, village-based, and ideally self-contained societies, though it has increasing become an important qualitative methodology throughout the social sciences in all kinds of social environments (Erickson, 2011). Anthropology's traditional focus on such societies speaks to this discipline's origins and development in connection with exploration, museums, and colonial administration, but it also indicates the application of anthropological research to documenting and developing conceptual approaches for analyzing the development and uses of knowledge in societies where traditional knowledge is not recorded in standard accessible codes but only transmitted orally. Although formerly widespread notions about primitive, pristine, and even traditional culture underlying ethnography and anthropology have been severely questioned, the expectation for ethnography to elucidate local knowledge persists (Turnbull, 2008).

The Anthropology of Knowledge

The term "anthropology of knowledge" was first articulated in 1973 when Mary Douglas compiled an anthology of selections on what she called "the anthropology of everyday knowledge" organized around themes including tacit conventions, the logical basis of constructed reality, orientations in time and space, the limits of knowledge, and provinces of meaning. Douglas's version of the anthropology of knowledge focused on assumptions embodied in behavior and her selections included at least as much philosophy from phenomenological traditions as it did anthropological contributions. Douglas (1973) followed in the tradition of Emile Durkheim and Marcel Mauss (1903/1963) in analyzing the social forces underlying culturally constructed classification systems in traditional non-literate societies. The classifications she and her followers were interested in tended to be culturally salient, emotionally provocative, and even morally loaded binary oppositions such as those between pure and impure, sacred and profane, right and left, and male and female (Ellen & Reason, 1979). The first review article on the emergent subject to apply the term "anthropology of knowledge" was produced by Malcolm Crick in 1982. Such a term manifestly invokes an earlier term, "sociology of knowledge," popularized much earlier by the German sociologist Karl Mannheim (1936).

The anthropological focus on culture helps explain the perceived need to create an anthropology of knowledge when the sociology of knowledge already existed. One can trace its origins an arrangement brokered in the late 1950s by Alfred Louis Kroeber and Talcott Parsons (1958), considered the deans of mid-20th century American anthropology and sociology (cf. Kuper, 1999), that encouraged anthropology and sociology to develop in divergent directions by directing their attention on different aspects of the same phenomena. According to this plan, anthropologists focus on

culture (including knowledge and its products) while sociologists focus on social forces and structures (interaction between individuals and collectivities). Therefore, the anthropology of knowledge studies symbolic and semantic systems, especially in traditional, premodern societies, while the sociology of knowledge, besides concentrating on modern industrial and postindustrial society, focuses on social pressures shaping opinion, such as mass media, and social institutions including government, corporations, and academia itself (especially in the sciences).

Beginning in the 1970s, anthropologists wrote specifically and extensively about knowledge. A 1975 study of the Gnau people who inhabit 23 small villages in Papua New Guinea called *Knowledge of Illness in a Sepik Society*, (Lewis, 1975) examines the processes of recognizing illness, concepts and categories of its causation, and cases of diagnosis and explanation of illness to present an indigenous concept of knowledge as it pertains to illness. About the same time, the eminent anthropologist Fredrik Barth (1975) studied concepts of knowledge in the context of male ritual initiation in age grades among the Baktaman, another small-scale New Guinea society. Such studies have focused on social aspects of secrecy, concealment, and mystification (see also Crook, 2007). Yet another important ethnography of the time was *Knowledge and Passion* by Michelle Z. Rosaldo (1980), which examined semantics, discourse, and social interactions relating vernacular concepts translated as knowledge and passion to notions of age, gender, and concepts and associations concerning the heart and human development in a village-based society in the Philippines. Over the years, the corpus of literature on the anthropology of knowledge has grown. As one practitioner has stated, the anthropologist of knowledge studies "not only who knows what but who claims to know what, how such claims are evaluated, legitimated, and accepted, and their consequences for social relations, especially for power, morality, and what Douglas has called social accountability" (Lambek 1993, 10).

Anthropology's contribution to the study of knowledge is in the analysis of the construction or constitution of knowledge domains, the process whereby knowledge is legitimized, the classification, categorization, and underlying connotations and symbolism of knowledge within a cultural context, its valuation, its storage in external memory systems, and its transmission or deployment in discourse, including political considerations. It also covers the encoding and decoding of knowledge through symbolism of various kinds and the social uses of memory.

These topics seem to articulate clearly with KO's concerns, and to the extent that KO ought to strive to bring together all approaches to knowledge, as argued by Hjørland (2003, 2008), they ought to be included in the broader set of subjects studied within the KO curriculum. Yet anthropological approaches are rarely mentioned in the KO literature. A recent essay by Wen-Chang Lee (2015) cites anthropological statements on culture as well as providing a review of its occurrence in the KO literature, as in Clare Beghtol's (1986) suggestion that culture provides a possible basis

for literary warrant. Lee suggests that the concept of culture could be used in KO to study the ways people interact with classification schemes, as in an office setting or while working with a library catalog or database, or perhaps another activity in cyberspace. Mentioning ethical implications of culture in classification, she refers to Bowker and Star's (1999) book *Sorting Things Out*, which can be said to straddle anthropological and KO approaches. This text remains the most significant attempt to make an anthropological or at least ethnographic approach to classification directly relevant to KO researchers by specifically studying classification in documentation by examining the cultural basis and uses of the International Classification of Diseases and the Nursing Interventions Classification.

Toward a Convergence of Interests and Agendas

Addressing her fellow anthropologists, Emma Cohen (2010, S193) has written that they are "uniquely positioned" to address challenges concerning "the emergence, spread, persistence, and transformation of knowledge," but stresses that

> If we truly aspire to understand 'how we come to know', to espouse theories of knowledge acquisition, storage, retrieval, and communication processes, and to account for the importance of bodily and mental states in learning and performance, we simply cannot afford to ignore the vast and increasingly sophisticated scholarship on such issues in neighboring disciplines. (2010, S194)

While it is unlikely that Cohen had her eyes on KO for an intended partnership with anthropology, a wide opening for just such a project had just been created when Hjørland (2009) himself produced a brilliant article on concept theory, identifying it as an important area for development in KO. The article covers interdisciplinary research on concepts, the function of concepts, conceptual stability and change, and theories of knowledge corresponding to various theories of concepts without once mentioning anthropology or even culture. Besides the major theories of knowledge within Western philosophical traditions to which Hjørland refers, one could extend one's analysis to include alternative indigenous knowledge and epistemologies and ontologies, including those studied by anthropologists using emic analysis and indigenous ethnography (Bernstein, 2010, Smith, 2012).

Approaches to the study of concepts have been developed by anthropologists and are relevant to KO. These include methods for the discovery of complex cognitive structures developed by Roy D'Andrade (2005), in which schemas revealing shared cultural models are elicited and verified through interviews in naturalistic settings. D'Andrade's approach shows how concepts are organized by condensing complex concepts that require long explanations or narratives into short terms that are interrelated in a shared conceptual system. This approach also seeks to explain the reasoning underlying propositions within concepts and their organization in conceptual domains (see also Blount, 2011). Other anthropologists who have grappled with the study of conceptual spaces and their elicitation and who connect ethnographic approaches and anthropological theories to studies in other disciplines are Maurice

Bloch (2012) and David Kronenfeld (1996). Not surprisingly, both completely omit any mention of contributions from KO, which, as usual, is shunted far away from main scholarly debates about knowledge in the humanities and the social sciences.

KO in Larger Debates about Knowledge

As KO grows as a scholarly subject, it seems natural for practitioners to aspire for it to expand into a larger intellectual space, and one can only agree that KO has a wider relevance beyond technical reading for fellow specialists. Hjørland's insistence on a broader context for KO that includes psychology, linguistics, philosophy, and the sociology of knowledge appears to allude to such a prospect. With the continuing evolution of academic disciplines and knowledge creation in and outside of universities and traditional institutions, KO can play a central and coordinating role in a transdisciplinary project on the uses and interpretation of knowledge (Bernstein, 2014). As this study has demonstrated, however, the full suite of approaches to knowledge for this improved, broader KO to draw on is incomplete without anthropology and its distinctive approach to knowledge and classification that examines them as cultural phenomena.

Yet anthropologists who study knowledge have also been amiss in ignoring KO's own unique contributions. Even as some anthropologists study knowledge organization in their own fashion, they are unaware that an entirely separate field called knowledge organization exists. While a few anthropologists have examined humans' interactions with machines (Suchman, 1987) and cyberspace more generally (Hakken, 2003) none to date have engaged with theories from KO such as facet classification, information retrieval, bibliometrics, or domain analysis. Given that documents, libraries, archives, and other memory institutions are part of culture and academia, this blind spot deserves to be called out.

For a connection to be made between the two fields one should not expect the outreach to come from one direction only. Since ethnographic approaches are already being used in LIS, one can hope for KO scholars to engage more meaningfully with anthropological studies of knowledge than is presently the case and for the valuable work done on KO to engage the interests of anthropologists studying knowledge.

References

Barth, Fredrik. (1975). *Ritual and Knowledge among the Baktaman of New Guinea*. Oslo: Universitetsforlaget and New Haven, CT: Yale University Press.

Beghtol, Clare (1986). Semantic validity: Concepts of warrant in bibliographic classification systems. *Library Resources & Technical Services* 30(2/3): 109-125.

Bernstein, Jay H. (2010). Folk concepts. In H. James Birx (Ed.), *21st Century Anthropology: A Reference Handbook*. Thousand Oaks, CA: Sage. Pp. 848-855.

Bernstein, Jay H. (2014). Disciplinarity and transdisciplinarity in the study of knowledge. *Informing Science*, 17: 241-273.

Bloch, Maurice (2012). *Anthropology and the Cognitive Challenge*. Cambridge: Cambridge University Press.

Blount, B.G. (2011). *A history of cognitive anthropology*. In David B. Kronenfeld, Giovanni Bernardo, Victor de Munck, & Michael D. Fischer (Eds.), *A Companion to Cognitive Anthropology*. Chichester, UK: Wiley-Blackwell. Pp. 11-29

Bowker, Geoffrey C. & Star, Susan Leigh (1999). *Sorting Things Out: Classification and its Consequences*. Cambridge, MA: MIT Press.

Cohen, Eli B. (2009). A philosophy of informing science. *Informing Science*, 12: 1-15.

Cohen, Emma. (2010). Anthropology of knowledge. *Journal of the Royal Anthropological Institute*, 16: S193-S202.

Crick, Malcolm R. (1982). The anthropology of knowledge. *Annual Review of Anthropology* 11: 287-313.

Crook, Tony (2007). *Anthropological Knowledge, Secrecy, and Bolivip, Papua New Guinea: Exchanging Skin*. Oxford: Oxford University Press/British Academy.

D'Andrade, Roy. (2005). Some methods for studying cognitive cultural structures. In Naomi Quinn (Ed.), *Finding Culture in Talk: A Collection of Methods*. New York: Palgrave Macmillan. Pp. 83-104

Douglas, Mary (Ed.). (1973). *Rules and Meanings: The Anthropology of Everyday Knowledge*. Harmondsworth, UK: Penguin.

Durkheim, Emile, & Mauss, Marcel. (1963). *Primitive Classification* (Rodney Needham, Trans.) Chicago: University of Chicago Press. (Original work published 1903).

Ellen, Roy, & Reason, David (Eds.) (1979). *Classifications in their Social Context*. London: Academic Press.

Erickson, Frederick (2011). A history of qualitative inquiry in social and educational research. In Norman K. Denzin. & Yvonna S. Lincoln (Eds.), *The Sage Handbook of Qualitative Research*. Thousand Oaks, CA: Sage. Pp. 43-59.

Goodenough, Ward H. (1971). *Culture, Language, and Society*. Addison-Wesley Module in Anthropology, no. 7. Reading, MA: Addison-Wesley.

Hakken, David. (2003). *The Knowledge Landscapes of Cyberspace*. New York: Routledge.

Hjørland, Birger (2003). Fundamentals of knowledge organization. *Knowledge Organization* 30(2): 87-111.

Hjørland, Birger (2008). What is knowledge organization (KO)? *Knowledge Organization*, 35(2/3): 86-101.

Hjørland, Birger (2009). Concept theory. *Journal of the American Society for Information Science and Technology*, 60(8): 1519-1538.

Keesing, R. M. (1981). Theories of culture. In Ronald W. Casson.*Language, Culture, and Cognition: Anthropological Perspectives*. New York: Macmillan. Pp. 42-66.

Kroeber, A. L., & Kluckhohn, Clyde. (1963). *Culture: A Critical Review of Concepts and Definitions*. New York: Vintage Books. (Original work published 1952).

Kroeber, A. L., & Parsons, Talcott. (1958). The concept of culture and of social system. *American Sociological Review*, 23: 582-583.

Kronenfeld, David B. (1996). *Plastic Glasses and Church Fathers: Semantic Extension from the Ethnoscience Tradition*. New York: Oxford University Press.

Kuper, Adam. (1999). *Culture: The Anthropologists' Account*. Cambridge, MA: Harvard University Press.

Lambek, Michael (1993). *Knowledge and Practice in Mayotte: Local Discourses of Islam, Sorcery, and Spirit Possession*. Toronto: University of Toronto Press.

Lee, Wan-Chen. (2015). Culture and classification: An introduction to ethical issues of adopting global classification standards to local environments. *Knowledge Organizati*on, 42(5): 302-307.

Lewis, Gilbert (1975). *Knowledge of Illness in a Sepik Society: A Study of the Gnau, New Guinea*. London: Athlone Press.

Mannheim, Karl (1936). *Ideology and Utopia: An Introduction to the Sociology of Knowledge* (Louis Wirth & Edward Shils, Trans.). New York: Harcourt, Brace.

Rosaldo, Michelle Z. (1980). K*nowledge and Passion: Ilongot Notions of Self and Social Life*. Cambridge: Cambridge University Press.

Simon, Herbert A. (1969). *The Sciences of the Artificial*. Cambridge, MA: MIT Press.

Singer, Milton. (1968). The concept of culture. In David Sills (Ed.), *International Encyclopedia of the Social Sciences* (vol. 3). New York: Macmillan & The Free Press. Pp. 527-543.

Smith, Linda Tuhiwai (2012). D*ecolonizing Methodologies: Research and Indigenous Peoples*, 2nd ed. London: Zed Books.

Stocking, George W., Jr. (1968). *Race, Culture, and Evolution: Essays in the History of Anthropology*. New York: The Free Press.

Suchman, Lucy A. (1987). *Plans and Situated Actions: The Problem of Human-Machine Communication*. Cambridge: Cambridge University Press.

Turnbull, David. (2008). Knowledge systems: Local knowledge. In Helaine Selin (Ed.), *Encyclopedia of the History of Science, Technology, and Medicine in Non-Western Cultures*, 2nd ed. Berlin, Germany: Springer. Pp. 1198-2003

Tyler, Stephen A. (Ed.). (1969). *Cognitive Anthropology: Readings*. New York: Holt, Rinehart & Winston.

Tylor, Edward B. (1871). *Primitive Culture: Researches into the Development of Mythology, Philosophy, Religion, Art, and Custom*. London: John Murray.

Vermeulen, Han F. (2015). *Before Boas: The Genesis of Ethnography and Ethnology in the German Enlightenment*. Lincoln: University of Nebraska Press.

Andre Vieira de Freitas Araujo, Fiammetta Sabba and Giulia Crippa

Semantic Order in the 16th Century: An Introductory Discussion of Conrad Gesner's *Pandectae*

Abstract

This work consists of a brief discussion regarding the semantic aspects contained in *Pandectarum sive partitionum universalium...* or *Pandectae* (1548) of Conrad Gesner (1516-1556). It situates the contributions of Gesner in the field of Knowledge Organization (KO) from a bibliographical and especially historical-documentary perspective. The *Pandectae* propose an innovative approach to the semantic treatment of documents, classifying them with a model which is an expansion from the medieval one while providing with orientation about the preparation of indexes. However, the effectiveness of the *Pandectae* is not necessarily in their semantic organization, but in their conceptual framework. Both *Bibliotheca Universalis* and in particular the *Pandectae* feature as fundamental historical documents for the understanding of the basis of knowledge organization. They are doubtlessly the earliest works to discuss bibliographical techniques (Wellisch, 1981) while materializing classificatory sensitivity - which is theorized, formulated and applied with logical *finezza* and acute sense of categorical multidimensionality (Serrai & Sabba, 2005).

Introduction

The historical dimension of Knowledge Organization (KO) finds fertile grounds for its development in the ancient bibliographical practices, because earlier forms of registration, organization and mediation of knowledge can shed light on questions, problems, technologies and products that encompass the field of KO today.

Historically there are numerous examples of projects related to knowledge organization, such as: *Pinakes* of Callimachus in Antiquity and *Institutiones Divinarum Litterarum* of Cassiodorus in the Middle Ages. The latter expressed the real "bibliographical gesture" that confirms the existence of bibliographical experiences and consequently the existence of the organization of knowledge prior to Modernity (Crippa, 2015).

In Early Modern Europe, the concept of Bibliography undergoes a breakthrough moment during which there is an expansion of the figure of the humanists. According to Araujo (2015, 127):

> interested in the ancient texts and the ways in which they should be organized, many [humanists] have dedicated themselves not only to the classification of beings, but also of knowledge. In the modern bibliographers there was a dissection posture of the elements of natural sciences which was transferred to the dissection of knowledge, constituting the idea of an anatomy of knowledge.

In this context the contributions of Conrad Gesner (1516-1565) - considered the father of Bibliography - are inserted.

The aim of this paper is to present a brief discussion of the semantic aspects contained in *Pandectarum sive partitionum universalium Conradi Gesneri Tigurini, medici & philosophiae professoris, libri XXI : Ad lectores. Secundus hic Bibliothecae nostrae tomus est, totius philosophiae & omnium bonarum artium atque studiorum locos communes & ordines universales simul & particulares complectens...* or *Pandectae* (plural form), of Gesner.

From a bibliographical and historical-documentary perspective, this research intends to set the contributions of Gesner in the field of KO.

The Bibliographical Project of Conrad Gesner

Conrad Gesner was a Swiss scholar, scientist and bibliographer. He was a typical Renaissance "polymath", with the ability to articulate and discuss numerous areas of knowledge. He published books on multiple topics such as linguistics, medicine, theology, botany, zoology, paleontology and mineralogy. His scientific and scholarly activity,

> [...] can be built around two distinct phases: 1) the phase of studies that included classical studies, application in Medicine, the profession, and interest in Botany [...] 2) the mature phase, that is, the scientific production: a) bibliographical works b) linguistic-philological works c) medical works, physical and natural sciences (Sabba, 2012, 30).

The vocation of Gesner for the preparation of bibliographical works, such as the *Catalogus Plantarum* (1542), resulted in the most ambitious bibliographical project of Modern Europe: *Bibliotheca Universalis*.

The work was published in four parts between 1545-1555: 1) *Bibliotheca Universalis, sive Catalogus omnium scriptorum locupletissimus, in tribus linguis, Latina, Graeca, et Hebraica...* (1545); 2) *Pandectarum sive partitionum universalium...* (1548); 3) *Partitiones theologicae* (1549) and 4) *Appendix bibliothecae* (1555).

Bibliotheca Universalis, sive Catalogus... is an alphabetical name catalog featuring 5031 authors of works composed in the three Sacred Languages, namely Latin, Greek and Hebrew. The catalog is organized alphabetically by first names of authors, but is accompanied by a separate list organized by surnames (Sabba, 2012). It also presents a summary and extracts of the documents listed.

One of the major motivations for the development of *Bibliotheca* is the fact that Gesner felt great concern when the Turks burned and looted, in 1527, the Library of Matías Corvino, during the attack in the city of Buda. This event would have exerted influence on Gesner who decided to do later work in order to safeguard the testimony of thought, in the case of the disappearance of books (Malclès, 1960).

In this sense, the *memorial* aspect of *Bibliotheca* is to be noted as Gesner attains the materialization of a broad and representative bibliographical universe in a knowledge map. His interest is also in bibliographical mediations between *Bibliotheca* and a community of scholars concerned with the quality of their sources of study.

According to Blair (2010), the justification for the universalist ambition of Gesner is that in his considerations all authors are worthy to be remembered or cataloged. Gesner lists them up, leaving selection and judgement to others.

Pandectae

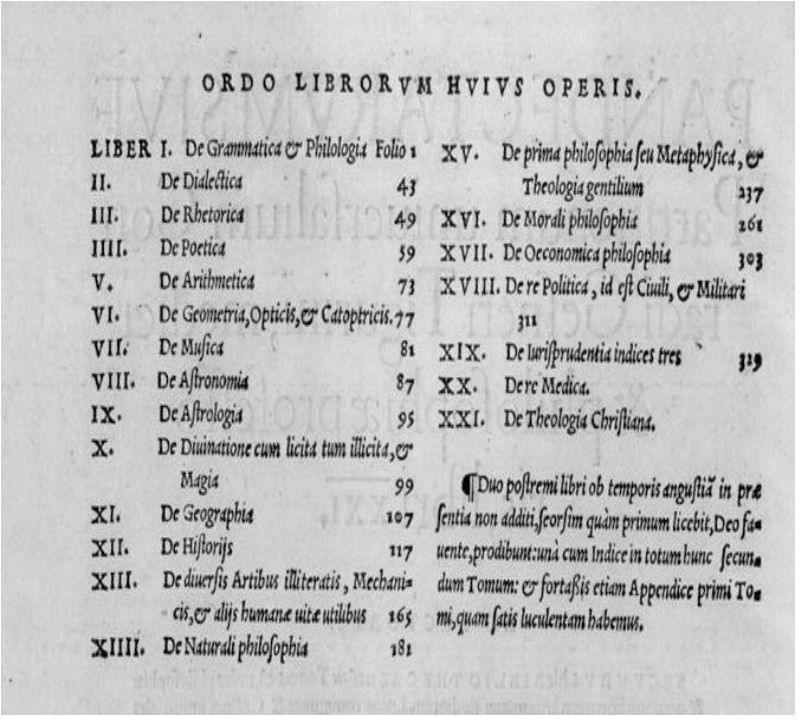
The *Pandectae* contemplate the classified display of the documentary material, that is, the exposition of knowledge extracted from the book of the first part.

Initially, the scheme was designed in XXI Partitions: *Pandectae* comprise XIX

Partitions, so that *Partitiones theologicae* comprises the last Partition, published separately in 1549 and named Partition XXI. The XX Partition designed to present the semantic catalog of medical works was not completed (Serrai, 1990).

The back of the title page of the *Pandectae* presents the complete scheme of general classes (Fig. 1), actually resulting in a systematic repertory.

Figure 1. Complete scheme of the general classes in the *Pandectae* (Gesner, 1548, http://www.e-rara.ch/zuz/content/pageview/67861)



ORDO LIBRORVM HVIVS OPERIS.

¶ Duo postremi libri ob temporis angustiã in præsentia non additi, seorsim quàm primum licebit, Deo faciente, prodibunt: unà cum Indice in totum hunc secundum Tomum: & fortasis etiam Appendice primi Tomi, quam satis luculentam habemus.

The classification system in the *Pandectae* stems from the seven liberal arts spanning to categories of complementary subjects and of interest to scholars of the Renaissance.

The *Pandectae* are based on the scheme of Philosophy, thought of as comprising all arts and sciences. Sciences were divided into *Preparantes* (preparatory sciences) or *Substantiales* (substantial sciences); the first split into *Necessariae* and *Ornantes*. *Necessariae* into *Sermocinales* and *Mathematicae*. *Sermocinales* comprehend: 1) *Gramatica*, 2) *Dialetica*, 3) *Rhetorica*, 4) *Poetica*. *Mathematicae* comprehend: 5) *Arithmetica*, 6) *Geometria*, 7) *Musica*, 8) *Astronomia*, 9) *Astrologia*. *Ornantes*: 10) *Historiarum cognitio*, 11) *Geographia*, 12) *Divinationis et magiae cognitio*, 13) Va*ria cognitio de artibus illiteratis, Mechanicis, et aliss humanae vitae utilibus*.*Substantiales*are: 14) *Physica*, 15) *Methaphysica et Theologia gentilium*, 16)

Ethica, 17) *Oeconomica*, 18) *Politica*, 19) *Jurisprudentia*, 20) *Medicina*, 21) *Theologia Christiana* (Serrai, 1977).

From this framework, Gesner elaborates the *Pandectae* with the following structure: 1) Grammar (and Philology), 2) Dialectic, 3) Rhetoric (representing the *trivium*), 4) Poetics, 5) Arithmetic, 6) Geometry, 7) Music, 8) Astronomy (the last four classes representing the *quadrivium*). These are followed by the sciences included in the medieval university curriculum: 9) Astrology, 10) Divination and Magic, 11) Geography, 12) History, 13) Mechanical Art, 14) Natural Philosophy, 15) Metaphysics, 16) Moral Philosophy, 17) Economic Philosophy, 18) Politics and finally, 19) Law, 20) Medicine as well as 21) Theology.

The reason for this division is the fact that these partitions reproduced the scopes of subjects according to Philosophy. Gesner believed Philosophy to be the constituent element of knowledge.

Each *Pandectae* class represents a book corresponding to a partition. Each book, in turn, is organized as follows: book title (the title of the respective partition), the dedicatory (whom the book is dedicated to) and the overall exposition of the subclasses subordinated to the class that names the book or partition. Finally, the titles are listed according to all subclasses.

According to Serrai & Sabba (2005, 56-57):

> The Partitions, or main classes, are represented in Titles; the Titles are subdivided into Parts - sequentially. The Parts contain the *Loci*, followed by a brief reference to the authors and the works which deal with the themes identified or placed in those *Loci*. References are accompanied with information about the book and chapter of the work. The book itself is referred to implicitly because it is mentioned only under the author's name. The *Loci* are usually presented in a thematic sequence, which means, they cover the scope of subjects connected to a particular Part. Otherwise, the *Loci* appear in alphabetical order – such as in lists which include philosophers, animals, plants, gods, oracles, etc.

In short: the books listed in *Bibliotheca Universalis, sive Catalogus*... are arranged in alphabetical order by authors. In the *Pandectae*, the listed books are ordered according to the *loci communes* and are gathered by their subjects (Serrai, 1990).

According to Malclès (1960), Gesner's classification system is unique as it expands the seven liberal arts of the Middle Ages.

Therefore, the gesnerian scheme reflects his conception of the division of knowledge directly linked to the classical thought of the Renaissance man.

Index in Gesner

The *Pandectae* distribute and hierarchize disciplines/subjects in a system of 21 classes displaying the contents of documents represented as semantic places or *Loci*.

According to Serrai (2007), the *Loci* and the classification scheme are the logical elements on which the *Pandectae* structure is based. The bibliographic research is structured and based on those very elements.

The Loci are concepts or categories that express the themes and the core elements

which are considered significant and representative of a document as part of one of the areas of interest and study of a specific culture. The *Loci* are thus able to express information and the intellectual content of that document. In bibliographical terms, the *Loci* - or the semantic indexes of a document - are also known as subjects or objects (Serrai, 2007).

Loci communes and *particulares*, which make up the gesnerian index, are extracted from each part of the editions (comments, prefaces, etc.), but Gesner also uses *loci* books as Maximus Planudes, Agathia etc.

According to Gesner (1548), indexing a book is a known process. The indexing process, in the gesnerian view, is clearly exposed by Considine (2015, 490-491):

> You should write out its key points on one or more sheets of paper, in any order, single-sided, marking the words by which multiword items are to be ordered. Then you should cut up the sheets of paper so that each unit of information is on a separate slip. The slips which are imagined here will be so narrow that if each is sorted as soon as it has been cut away from the page, it will still cling to the blade of the scissors used to cut it (a larger slip will of course immediately fall away from the scissors under its own weight). Sorting the slips into piles on a tabletop, or into containers, may be done in one or more passes. […] After sorting, they may be copied out in their final order, or, preferably, they may be glued down onto sheets of paper. This should be done with a water-soluble paste so that the slips can be detached from their backing sheets if they have to be rearranged. As an alternative to pasting, sheets of heavy paper can be prepared to hold two columns of slips by passing threads through them so that slips can be held in place by the threads at each side of each column. Such sheets can be bound up into volumes of a hundred folios—though no more than that, since each folio will be heavy to start with and made heavier by the slips fastened to it.

In the process of indexing, Gesner was innovative: he was the first to recommend the use of slips to create an alphabetical index. Each alphabetically ordered item would be copied on a single-sided sheet of paper and cut out into slips (Blair, 2010).

The indexes developed in *Bibliotheca Universalis* and the *Pandectae* are sophisticated if compared to other works by Gesner's contemporary scholars.

An example of this is the main *Pandectae* index that was printed as the last part of the Book XXI (Fig. 2): it takes 77 columns of 26 folios and contains around 4000 entries alphabetically arranged (Wellisch, 1981).

Figure 2. *Index communis in libros XX*
(Gesner, 1548, http://www.e-rara.ch/zuz/content/pageview/678610)

INDEX COMMVNIS IN
LIBROS XX. PANDECTARVM CONRADI
GESNERI. a.b.c.d. CVIVSQVE FOLII PRIMAM, SECUNDAM, TERTIAM, AUT QUARTAM COLUMNAM SIGNIFICANT. T. LITERA LIBRUM DE THEOLOGIA.

aa

In the *Index communis in libros XX*, there are both numbers and letters next to the alphabetically arranged entries.

These locators/codifications are organized by page number and letters. Numbers refer to a certain page. Letters refer to a column on a certain page (a=first column, b=second column, c=third column, d=fourth column). The letter "T" refers to the

Theological Book, i.e., Book XXI.

Gesner is paramount in the development of indexes, as they appear in other works of his, such as in those related to Botany, Pharmacology, Linguistics etc.

Final Considerations

The *Pandectae* propose an innovative approach to the semantic treatment of documents, classifying them with a model which is an expansion from the medieval one while providing with orientation about the preparation of indexes. However, the effectiveness of the *Pandectae* is not necessarily in their semantic organization, but in their conceptual framework.

For Serrai & Sabba (2005), the *Pandectae* offer the advantage of making a universal structure of loci communes to mirror the totality of science and art, something that no one had tried before.

Gesner's desire, registered in the preface to his work, was that others could follow on from him. That attests the historical process underlying the different forms of record, organization and mediation of knowledge. Such desire may have come to design/define the practices in the areas of information and knowledge at different historical moments.

The echoes of the gesnerian bibliographical project is materialized on two fronts: 1) Bibliography takes its nature as a discipline/subject from Gesner and 2) other universalists have developed projects of registration, organization and mediation of documents and information over the centuries.

In this context, Paul Otlet is not to be forgotten. In his search for an environment and for tools of international collaboration, he created the International Institute of Bibliography proposing a Universal Book which comprised individual cards instead of slips. Even the card catalog has its origin in Gesner's experiments with paper slips (Wright, 2014).

Gesner´s *oeuvre* has proved to be an irrefutable inspiration along the centuries and might have certainly influenced Otlet as well: "Conrad Gesner had created his great bibliography by cobbling together material from a wide range of existing sources. Otlet e La Fontaine followed his example, building their Universal Bibliography largely by drawing on previously published material" (Wright, 2014, 71).

Both *Bibliotheca Universalis* and in particular the *Pandectae* feature as fundamental historical documents for the understanding of the basis of knowledge organization. They are doubtlessly the earliest works to discuss bibliographical techniques (Wellisch, 1981) while materializing classificatory sensitivity - which is theorized, formulated and applied with great logical *finezza* and acute sense of categorical multidimensionality (Serrai & Sabba, 2005). Herein lies one of the many keys to understand the contributions of the so-called father of Bibliography to KO in contemporary times.

References

Araujo, Andre Vieira de Freitas (2015). Pioneirismo bibliográfico em um polímeta do Séc. XVI: Conrad Gesner. *Informação & Informação*, 20(2) May/Aug.: 118-42.

Blair, Anne M. (2010). *Too much to know: managing scholarly information before the modern age*. New Haven, CT: Yale University Press.

Considine, John (2015). Cutting and Pasting Slips: Early Modern Compilation and Information Management. *Journal of Medieval and Early Modern Studies*, 45(3) September: -

Crippa, Giulia (2015). Cassiodoro e as Institutiones Divinarum Litterarum como fonte histórica para a discussão sobre práticas bibliográficas e organização do conhecimento. *Informação & Informação*, 20(2): 86-117.

Gesner, Conrad (1548). *Pandectarum sive partitionum universalium Conradi Gesneri Tigurini, medici & philosophiae professoris, libri XXI : Ad lectores. Secundus hic Bibliothecae nostrae tomus est, totius philosophiae & omnium bonarum artium atque studiorum locos communes & ordines universales simul & particulares complectens [...]*. Tiguri: excudebat Christophorus Froschouerus. [http://www.e-rara.ch/zuz/content/titleinfo/624958]

Malclès, Louise Nöelle (1960). *La bibliografía*. Buenos Aires: EUDEBA.

Sabba, Fiammetta (2012). *La 'Bibliotheca Universalis' di Conrad Gesner: monumento della cultura europea*. Roma: Bulzoni Editore.

Serrai, Alfredo (1977). *Le classificazioni: idee e materiali per una teoria e per una storia*. Firenze: Leo S. Olschki Editore.

Serrai, Alfredo (1990). *Conrad Gesner*. Edit by Maria Cochetti. Roma: Bulzoni Editore.

Serrai, Alfredo (2007). *I Pandectae di Conrad Gesner*. Bibliotheca. 1: 11-37.

Serrai, Alfredo, & Sabba, Fiammetta (2005). *Profilo di Storia della Bibliografia*. Milano: Edizioni Sylvestre Bonnard.

Wellisch, Hans (1981). How to make an index - 16th century style: Conrad Gesner on index and catalogs. *International Classification*, 8(1): 10-15.

Wright, Alex (2014). *Cataloging the world: Paul Otlet and the birth of the information age*. Oxford: Oxford University Press.

Rodrigo de Sales

Knowledge Organization in the Brazilian Scientific Community and Its Epistemological Intersection with Information Science

Abstract

In order to investigate the epistemological intersection between Knowledge Organization (KO) and Information Science (IS), we chose to explore two typical aspects of epistemological studies: KO's nature (as an essential condition of KO) and its interdisciplinary interface with IS. To do so, as our main target, we investigated how the Brazilian scientific community has conceived knowledge organization and its relation with Information Science in the 21st century. The specific goals of the investigation were: a) to indentify different perspectives to define KO's nature and its relation with IS in the international scenario; b) to understand, based on ISKO-Brasil, how Brazilian researchers have defined KO's nature; and c) to analyze how ISKO-Brasil researchers have related KO and IS. The methodology used was the technique of Content Analysis of the ISKO-Brasil Proceedings. The results showed that most Brazilian studies in this context, regarding KO's nature, consider it an autonomous research area. Regarding its relation with IS most of them understand that they are distinct areas, but in constant dialogue.

Introduction

After the second half of the 20th century, Knowledge Organization (KO) research in the Brazilian scientific community began to be developed in the context of Information Science (IS). It was in the Brazilian Institute of Information Science and Technology (IBICT) that the studies on KO first found room to be discussed and developed and later gained momentum after the creation of the first Graduate Course in Information Science in Brazil in 1972. However, it was within the context of the National Association for Research and Graduate Studies in Information Science (ANCIB), founded in 1989, more specifically in the *GT2 – Knowledge Organization and Representation Work Group*, that Brazilian researchers managed to leverage and strengthen the development of KO in the country. The ANCIB work groups represent major specialized themes in the area of Information Science, which indicates that KO appeared in the Brazilian scenario as a theme within IS, as research of that nature began to be discussed and applied within IS.

Thus we can infer that in the Brazilian context knowledge organization first found epistemological and institutional conformation within Information Science, which shows an indispensable relation between KO and IS. In this scenario, the relation between KO and IS is especially evident in the "subject approach to information" developed over the 20th century (Foskett 1973), markedly in the classification, cataloguing and indexation approaches.

In 2007, researchers linked to ANCIB's GT2 approved the statute that officially defined the creation of ISKO's Brazilian Chapter (International Society for Knowledge Organization). With the foundation of ISKO-Brasil, the research on KO reached a new field for dialogue and development, intensifying not only its studies in Brazil but also its concatenation with KO's international scenario.

Sales (2015 a, b) contextualizes ISKO's international scenario and shows that in the first decade of the 21st century, knowledge organization was predominately defined as an "activity" of "operational" nature (Garcia, Oliveira, Luz, 2000; Green, 2002; Gárcia Gutiérrez, 2002), whose objects of investigation were mainly the concepts and conceptual structures(Kent, 2000; Green, 2002; Ohly, 2008, Smiraglia, 2010) instrumentally formalized in the systems of knowledge organization, such as the classification systems, thesauri and ontologies (Albrechtsen, 1990; Kent, 2000; Green, 2002; Zherebchevsky, 2010; Souza; Tudhope and Almeida, 2010). In this sense, we can observe a conception of KO which is very similar to the understanding found in the Brazilian context of ANCIB, i.e., connected to the (practical and intellectual) activities referring to Information Science and Library Science, corroborating Foskett's thematic approach of information (1973).

However, regarding ISKO, two central authors – Dahlberg (since the 1990s) and Hjørland (since 200) – have imprinted other perspectives as they defined KO's nature. Dahlberg (1993, 1995, 2006 and 2014) believes that KO consists of an autonomous study field, independent from Information Science, characterized as a subfield of Science of Science. Although Hjørland (2008) also sees KO as an autonomous study field, he clearly states that it maintains a strong relation with IS and Library Science, especially regarding KO's *narrow meaning*.

In this context, it is possible to define at least three different perspectives for the conceptions of KO's nature and its relation with IS: *Perspective* 1: based on Dahlberg's ideas, knowledge organizations is an autonomous study field, independent from Information Science; *Perspective* 2: according to Hjørland's conception, knowledge organization is considered an autonomous study field which is constant dialogue with Information Science, mainly concerning cognitive knowledge organization; *Perspective* 3: knowledge organization is traditionally set as an integral part of Information Science and does not seem to claim independence as an autonomous study field, but only contributes to Information Science's central and mediating space. This viewpoint is supported by the tradition that sets KO as a subfield of IS, as it occurs in ANCIB.

For this piece of research the three perspectives were used as a basis to guide the investigation that sought to understand how ISKO-Brasil researchers have related KO and IS.

It is precisely this epistemological clarification that this research aims to explore – KO's nature and its relation with IS within the Brazilian scientific field. If at the end of the 20th century there seemed to be a consensus (for Brazilian research) about the fact that KO was a theme or subfield of IS, could the emergence of ISKO-Brasil have made KO research horizons brim over IS's limits to the point of outlining a new area? Did a new perspective emerge in the Brazilian scientific community to reflect on KO? These are only a few questions that served as inspiration for this research.

Methodology

We stress that ISKO's Brazilian chapter is already one of the biggest in the international scenario in number of researchers as well as in amount of ongoing studies. Considering knowledge organization carried out and theorized by Brazilian researchers, what is under investigation here is the understanding of how the Brazilian scientific community approaches the relation between KO and IS. In this way, in order to bring to light some results that may contribute to more precise understanding on this regard, as our main goal, we sought *to investigate how the Brazilian scientific community has conceived knowledge organization and its relation with Information Science in the 21st century*.

In this sense, some specific goals were set: a) to indentify different perspectives to define KO's nature and its relation with IS in the international scenario; b) to understand, based on ISKO-Brasil, how Brazilian researchers have defined KO's nature; c) to analyze how ISKO-Brasil researchers have related KO and IS.

The scope to form the corpus of analysis was defined by all the texts published by Brazilian researchers in the Proceedings of ISKO-Brasil (2011, 2013 and 2015). Out of 138 publications, only 46 showed explicit definitions regarding KO's nature or its relation with IS. Therefore, 46 texts comprised the corpus of analysis.

To analyze and interpret the texts, we applied the technique of Content Analysis defined by Bardin (2003). Once the analysis categories were defined as *The Nature of KO* and *The Relation between KO and IS*, some inference variables were set to allow deeper investigation of the ideas proposed by the researchers. Thus, the inference variables for category 1 (*The Nature of KO*) were: a) research area, b) action/activity, c) object of research and d) Science.

Regarding category 2 (*The Relation between KO and IS*), the variables were inspired by the perspectives identified in the international literature, as follows: Perspective 1 (variable a): *KO as an autonomous study field (research area) without any relation with IS*; Perspective 2 (variable b): *KO as an autonomous study field (research area) but related to IS*; Perspective 3 (variable c): *KO as a theme within IS, i.e., KO as a subfield of IS*.

The analysis of the publications was done in the following order: 1) reading the texts that comprised the corpus of analysis; 2) surveying the information related to the categories and the inference variables in each text; 3) grouping the authors of the texts based on the categories and inference variable. In the same way, by relating the authors to the categories and their respective inference variables, it was possible to identify how the authors have defined KO's nature and its relation with IS. Table 1 shows the synthesis of the analysis:

Table 1: Analysis Synthesis

Analysis categories	Inference variables	Authors
Nature of KO	Research area	Guimarães & Dodebei (2011); Alves, Gracio & Oliveira (2011); Abdalla & Kobashi (2011); Mota & Silva (2011); Dodebei (2011); Miranda et al. (2011); Café (2011); Lima & Maculan (2011); Oliveira & Alves (2013); Guimarães (2013); Guimarães (2015); Weiss & Bräscher (2015); Bufrem; Arboit; Freitas (2015); Guimarães et al. (2015); Sales (2015); Ramalho (2015); Alves & Moraes (2015); Nakano, Padua, Jorente (2015); Bernardino et al. (2015); Alves; Oliveira; Grácio (2015); Lima et al. (2015); Suenaga & Cervantes (2015); Araújo & Guimarães (2015); Farias, Almeida,Martínez-Ávila (2015); Lima (2015); Orrico (2015); Souza (2015); Bufrem (2015); Dodebei (2015).
	Action/Activity	Oliveira, Santos & Oliveira (2011); Baptista (2013); Varela & Barbosa (2013); Tognoli & Barros (2015); Fonseca & Rodriguez (2015); Santos (2015) Fujita (2015).
	Object of Research	Lara (2011); Bufrem, Silveira & Nascimento (2013); Baptista (2013); Ravazi & Moreira (2015)
	Science	Andrade et al. (2011)
Relation between KO and IS	Perspective 1	Andrade et al. (2011)
	Perspective 2	Guimarães & Dodebei (2011); Dodebei (2011); Abdalla & Kobashi (2011); Lima &Maculan (2011); Oliveira & Alves (2013); Guimarães (2013) Fujita (2013); Weiss & Bräscher (2015); Guimarães (2015); Sales (2015); Guimarães et al. (2015);

	Castanha & Grácio (2015); Jorente (2015); Suenaga & Cervantes (2015); Araújo & Guimarães (2015); Lima (2015).
Perspective 3	Lara (2011); Café (2011); Bräscher (2011 e 2013); Bufrem, Silveira & Nascimento (2013); Ramalho (2015); Evangelista, Moreira & Moraes (2015); Ravazi & Moreira (2015); Lima et al. (2015); Moreira & Moraes (2015).

Results

What Regarding the nature of KO, the results obtained after the analysis of the publications of ISKO-Brasil (2011, 2013 and 2015) show that most of the texts by Brazilian researchers (63%) define KO as a research area, while 15.2% define KO as an action/activity, 8.7% see it as an object of research and only 2.2% define it as Science. Some 11% of the texts analyzed made no reference to KO's nature

Regarding the relation between KO and IS (category 2), the main observation point of this study, we observed that most of the texts analyzed (34.8%) converge with perspective 2, as they understand that KO is an autonomous research area but which maintains a strong connection with IS. It is closely followed by the 16.6% who believe KO is in fact a specialized theme or a subfield of IS, included in perspective 3. Only 2.2% of the texts converge with perspective 1, which states that KO is an autonomous area without any relation with IS, or, as Dahlberg (2006) prefers, that KO is a new Science. Some 46% of the texts analyzed made no reference to the existing relation between KO and IS.

Figure 1 shows the results obtained regarding Category 1 (Nature of KO) based on this category's inference variables:

Figure 1: Nature of KO

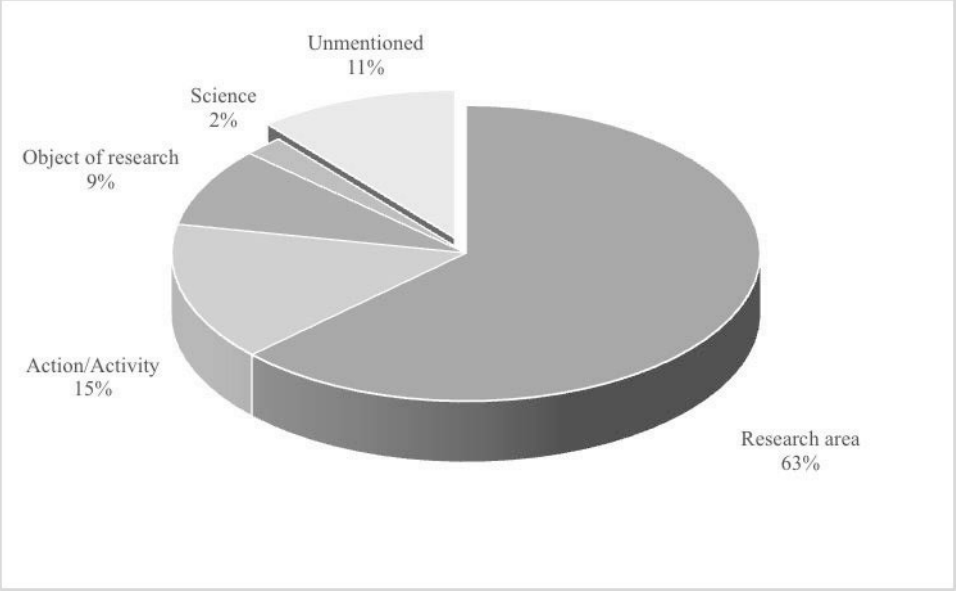
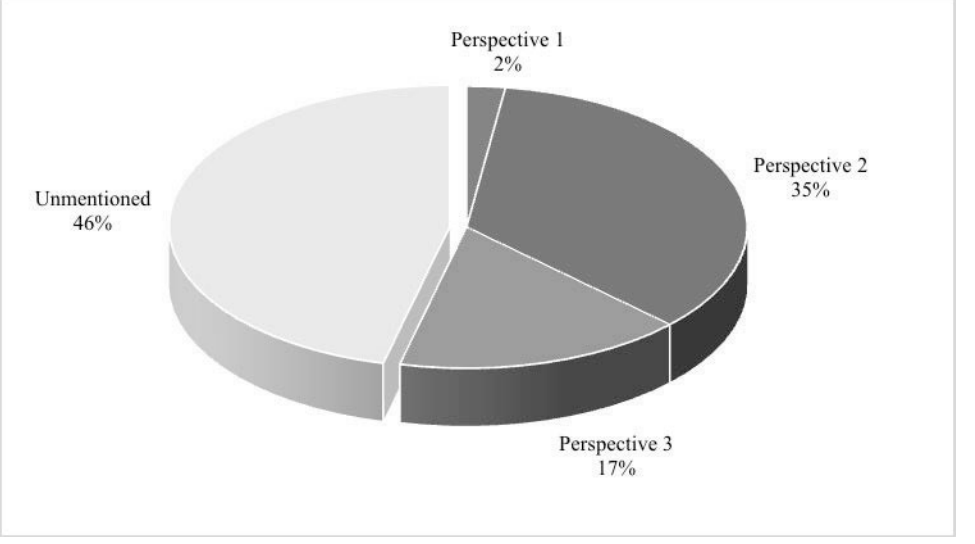


Figure 2 shows the results of the analysis carried out in Category 1 (Relation between KO and IS) based on this category's inference variables:

Figure 2: Relation between KO and IS



Thus the emerging scenario in ISKO-Brasil in the 21st century is that most Brazilian researchers consider that KO has become an autonomous research area, but it is in

constant dialogue with Information Science, imprinting its own epistemological features and appearing as a new emerging area.processes.

Conclusion

Although the research on knowledge organization in Brazil has developed epistemologically and gained institutional strength within Information Science (mainly within the scope of ANCIB), we can observe that a new perspective is emerging in ISKO-Brasil at the begging of the 21st century. KO, which throughout the 20th century was seen as a subfield of IS, seems to have brimmed over the latter's scientific limits to start producing its own space and to be considered predominantly an autonomous research area.

However, KO's gaining space in the Brazilian scientific reality does not represent a rupture of the connection between KO and IS. In fact, this research shows that most ISKO-Brasil researchers believe that the two areas are linked and in constant dialogue, although they recognize certain autonomy of KO as a research field.

We believe that research aiming to investigate KO's nature and disciplinary interfaces can contribute substantially to understand knowledge organization's epistemology itself.

References

Albrechtsen, Hanne.(1990). Software concepts: knowledge organization and the human interface. In: *Tools for knowledge organization and the human interface: Proceedings of the 1st International ISKO Conference*. Frankfurt: Indeks. Pp. 48.

Bardin, Laurence. (2003). *L'analyse du contenu*. 7ème. Paris: PUF.

Dahlberg , Ingetraut. (1995). Current trends in Knowledge organization. In Garcia Marco F. J. (Org.). *Organización del conocimiento em sistemas de información y documentación*. Zaragoza: Universidad de Zaragoza. Pp. 7-25.

Dahlberg, Ingetraut. (2006). Knowledge organization: a new science? *Knowledge Organization*, 33(1): 11-19.

Dahlberg , Ingetraut. (1993). Knowledge organization: its scope and possibilities. *Knowledge Organization*, 20(4): 211-222.

Dahlberg , Iingetraut. (2014). What is knowledge organization. *Knowledge Organization*, 40(1): 85-91.

Foskett, A.C. (1973). *A abordagem temática da informação*. Tradução de Antônio Agenor Briquet de Lemos. São Paulo: Polígono; Brasília: Ed.UnB.

Garcia, S. M. M., & Oliveira; Luz, G. M. S. (2000). Knowledge organization for query elaboration and support for technical response by the internet. In *Dynamism and stability in knowledge organization: Proceedings of the Sixth International ISKO Conference*Würzburg: Ergon. Pp. 189.

Gárcia Gutiérrez, Antonio L. (2002). Knowledge organization from a culture of the border: towards a trascultural ethics of mediation. In *Challenges in knowledge representation and organization for the 21st century: integration of knowledge across boundaries: Proceedings of the Seventh International ISKO Conference*. Würzburg: Ergon.

Green, Rebecca. (2002). Conceptual universals in knowledge organization and representation. In *Challenges in knowledge representation and organization for the 21st century: Integration of knowledge across boundaries: Proceedings of the Seventh International ISKO Conference*. Würzburg: Ergon. Pp.15.

Hjørland, Birger. (2003). Fundamentals of knowledge organization. *Knowledge Organization*, 30(2): 87-111.

Hjørland, Birger. (2008). What is knowledge organization (KO)? *Knowledge Organization*, 35(3/2): 86-111.

ISKO-Brasil: International Society for Knowledge Organization (2016). [http://isko-brasil.org.br]

Ohly, H. Peter. (2008). Knowledge organization pro retrospective In*Culture and identity in knowledge organization: Proceedings of the Tenth International ISKO Conference*. Würzburg: Ergon. Pp. 210.

Kent, Robert. E. (2002). The information flow foundation for conceptual knowledge organization. In *Dynamism and stability in knowledge organization: Proceedings of the Sixth International ISKO Conference*. Würzburg: Ergon. 2002, Pp. 111.

Sales, Rodrigo. (2015a). A Relação entre Organização do Conhecimento e Ciência da Informação na Comunidade Científica Brasileira: uma investigação no âmbito da ISKO-Brasil. In José Augusto Chaves Guimarães, & Vera Dodebei. (Org.). *Organização do Conhecimento e Diversidade Cultural*. 1ed.Marília, SP: ISKO-Brasil; FUNDEPE, v. 1. Pp. 73-84.

Sales, Rodrigo. (2015b). O diálogo entre a Organização do Conhecimento e a Ciência da Informação na comunidade científica da ISKO-Brasil. In X*VI Encontro Nacional de Pesquisa em Ciência da Informação (XVI ENANCIB), 2015, João Pessoa, PB. Informação, Memória e Patrimônio: do documento às rede*s. João Pessoa, PB: Universidade Federal da Paraíba, v. XVI. Pp. 1-21.

Smiraglia, Richard P. (2010). Perception, knowledge organization and noetic affective social tagging. In *Paradigms and conceptual systems in knowledge organization*. Würzburg: Ergon. Pp. 64.

Souza, RenatoRocha, Tudhope, Douglas& Almeida, Maurício Barcellos (2010). The KOS spectra: a tentative faceted typology of knowledge organization systems. In *Paradigms and conceptual systems in knowledge organization: Proceedings of the Eleventh International ISKO Conference*. Würzburg: Ergon. Pp. 122.

Zherebchevsky, Sergey (2010). Formalism in knowledge organization. In *Paradigms and conceptual systems in knowledge organization: Proceedings of the Eleventh International ISKO Conference*. Würzburg: Ergon. Pp. 98.

Jiri Pika

Knowledge Organization in Sciences – As a Classificatory Performance and Classification Design Model for Humanities

Abstract

The paper provides an overview of natural science classification scheme development with major control of classification criteria presented in the Linnaean taxonomy. Based on natural laws, the Linnaean taxonomy has been accepted worldwide. Unlike the indexing of the natural sciences items that follows the logic and systematics of natural laws – a real challenge still exists in classification of documents originating from human intellectual activity.Items, produced as a human output are a particular phenomenon and as such, follow no common rules. This lack of evident natural law as a basis for a common classification can be substituted by practices of facet classifications and Information Coding Classification (ICC) [1] that advances to the field of classifying literature. Their common feature is to analyse the information content with a set of categorical questions and to express the answers in exact terms, concepts and notations. The ensuing categorizations are certainly both concise and unequivocal: essentially Linnaean, or better!

Introduction

Among the numerous examples of knowledge organization in sciences, one case is particularly interesting, mainly from the documentary point of view (Umstätter 2009). Ever since the Swedish naturalist Carl Linnaeuswasknighted to become Carl von Linné, in recognition of his classificatory work in 1761, we have learned that the natural arrangement of objects of intellectual and physical environment leads to knowledge. Linnaeus thus made a significant contribution to the development of documentary sciences, without being adequately appreciated in this area. How revolutionary his idea was, can be seen by the fact that his work "Systema Naturae" (1735) was listed on the "Index Librorum Prohibitorum" by the pope (Jahn 2000). His influence in the 18th century was so great that J.W.v. Goethe on 7. November 1816 wrote to his friend Carl Friedrich Zelter [2]:

> This day I have reread Linnaeus and I am shocked by this extraordinary man. I have learned so much from him, but not Botany. With the exception of Shakespeare and Spinoza, I know no one among the no longer living who has influenced me more strongly.

Even his opponent, the director of the royal gardens in Paris, Georges-Louis Leclerc, Comte de Buffon had to accept the systematics of Linnaeus on royal behest in 1774.

The systematic arrangement of living creatures by Linnaeus came as a result of the increasing travel activities of naturalists and their plant and animal descriptions. To name a few: Andrea Cesalpino described in his book "De Plantis" (1583) more than 1500 plants and Gaspard Bauhin in his "Pinax Theatri Botanici" (1623) described 6000 plant species. Joseph Pitton de Tournefort (Eléments de botanique ou method pour connaître les plantes) characterized nearly 7000 in 1694 and John Ray described over 18 000 by 1704 in "Historia plantarum" (1686-1704).

To organize this vast amount of information, trying to cope with various classifications was at that time extremely important, especially for the purpose of

medicine and agriculture (Hansen, 1902). In particular, the use of different names for the same plant has led to dangerous misunderstandings.

> The rules, which Linné gives in his Philosophia Botanica for choosing the name, are masterful. He points out the absurdity of most of the old names and calls the botanists to choose their Nomina vera with the words: idiotae imposuere nomina absurda.

Linnaeus in his “Philosophia Botanica” (1751) characterized other botanists as “Fructistae“, “Corollistae“, “Calycistae“ and several other classes of botanists [3] (Hansen 1902), depending on which part of the plant his botanist colleagues (Linnaeus1751, Rádl 1905) used to design their classifications.

> Whereas other botanists are classed as Fructistae, Corollistae, Calycistae, under the Sexualists [4] stands a solitary, proud “ego”, which is correct, since he is the sole inventor of the “sexual system”, but it bears a strong aftertaste of the most sovereign self-confidence (Hansen 1902).

Linnaeus regarded himself [5] as “Sexualist” because he based his system on the classification of plant sex organs. Linnaeus made it clear that sexuality is a ubiquitous phenomenon of nature. This, at that time truly brilliant discovery, can be found in his thesis (1730) – an account of plant sexual reproduction: [6] “Praeludia Sponsaliorum Plantarum“ (=On the prelude to the wedding of plants). He relied on knowledge of Rudolph J. Camerarius (1665-1721), professor of medicine and director of the botanical garden in Tübingen, who had demonstrated by his publication (De sexu plantarum epistola 1694) that plants have sexuality. Nevertheless it was Nehemiah Grew (1641-1712), who had actually discovered this fact, but wasn’t able to prove it.

Linnaeus considered this phenomenon highly anthropomorphic. When he repeatedly talks about the bridal bed, or in connection with the "Polyandria", he asserts that in a flower with 20 stamens and one stylus, there are '20 males or more in the same bed with the female', a state of affairs enjoyed by the poppy (Papaver) and the linden (Tilia).

He opens his dissertation:

> “In spring, when the bright sun…The actual petals of a flower contribute nothing to generation, serving only as the bridal bed which the great Creator has so gloriously prepared, adorned with such precious bed-curtains, and perfumed with so many sweet scents, in order that the bride-groom and bride may therein celebrate their nuptials with the greater solemnity”) [7] Blunt (1971).

The introduction of sexuality as classification criterion led to the theory of evolution expressed later by Darwin, and found its basic fundament exactly in this classification. So it is understandable that the "Systema Naturae" was banned by the Pope and placed on papal Indexes of Prohibited Books (The Index Librorum Prohibitorum).Linnaeus pointed out that science is established primarily by its classification system, which arranges the knowledge relations within the specific system. Today we would say: integrated into semiotic networks (Umstätter 2009).

Although Linnaeus initially regarded his system as artificial - today it could be called constructivistic - it soon became evident that it was a natural one. His system depicted the natural evolution of living nature, because it applied the sexual kinship of

species as a classificatory criterion. Thus he transformed his system from a pure constructivism into an evolution model (Umstätter 2009).

> Linnaeus did not suppose that his classification of the plant kingdom in the book was natural, reflecting the logic of God's creation. His sexual system, where species with the same number of stamens were treated in the same group, was convenient, but in his view artificial. Linnaeus believed in God's creation, and that there were no deeper relationships to be expressed. He is frequently quoted to have said: "Deus creavit, Linnaeus disposuit" ("God created, Linnaeus organized") [8]

Linnaean taxonomy

In 1727 Linnaeus became aware of a newspaper article, which reported on a public lecture by Sébastien Vaillant, the member of the Academy of Sciences and director of the royal garden in Paris, on the sexuality of plants. In it was the indication that the pollen of the plants have the same function as sperm. The sexualistic system of Linnaeus (Rádl 1905) became accepted despite the resistance of many botanists, because it was clear, consistent and provocative (Umstätter 2009). Thus, it could not have been ignored by the world of experts (Mayr 1982). Since then the newly discovered creatures could be classified and recognized again by a standardized procedure.

Another important achievement of Linnaeus is the establishment of still-in-use standardized botanical nomenclature [9] (Paterlini 2007). In the "Genera Plantarum" (1737) he has determined the rules according to which the genera of plants should be named

> The name of a plant should be two-fold: a genus name equals to the human family name and a name of a species, as the name in daily life (nomina trivialia). The diagnosis depends on the associations in kinship circle of the respective species…. (Jahn 1985).

Equally important were the terminology introduced by Linnaeus and his instructions about how to describe the plant species. He introduced and defined about 1000 botanical terms in "Fundamenta Botanica" by 1736. Crucial for the classification was the clear distinction of significant and insignificant characteristics. As insignificant Linnaeus recognized characteristics, such as color, odor and size, because it was obvious to him that these could vary easily even within one species. In contrast, the sexual system was largely a type- or species-consistent categorization.

During his life Linnaeus realized ever more clearly that the species that he initially thought to be immutable can hybridize. Moreover, he observed some adaptation of plants to their environment and towards the end of his life he considered the origin of new species by hybridization to be quite feasible (Mallet 2007).

Cladistics, Knowledge Organization and Phylogenetic Classification

The question of what can be used in a classification as a division-criterion for categorization has proved crucial in Linnaeus' work. The key idea in the cladistics is to let the classes branch according to their relationship. Whereas the development of

many library classifications for routine indexing of publications requires no cladistic considerations,these are essential in the organization of knowledge,because knowledge develops epigenetically (Umstätter 2009).

In the search for alternative approaches to disciplinary classification, Gnoli (2006) reviews and evaluates classification schemes, bibliographic classifications and facet analysis [10] and proposes a model called phylogenetic classification. It integrates both evolutionary order and similarity as its main criteria: "phylogenetic method seems to have some potential to give a significant contribution to the development of more satisfying and generally valid classification schemes".

Status quo in contemporary classification systems

Contrary to the Linnaean Period, which brought consensus and a worldwide-accepted system for organizing plants and animals, nowadays the world of complex information and documents seems to be still running in Pre-Linnaean Period, as can be seen:

- The publishing rate is increasing – analogous to the rapidly increasing rate of plant and animal descriptions in the Linnaean Period.
- Analogous "origin of new species by hybridization" is arriving in libraries: hybrid documents with all types of media formats, imposing a need for document and library-rules re-arrangement.
- The flood of new scientific disciplines, concepts and terminologies keeps rising.
- No agreement regarding one common classification system exists. On the contrary:
- The increasing publishing rate is accompanied by an increasing number of thesauri and classifications. Dahlberg (1982) quantified their amount, which reached 2261 in 1982. Today's amount is possibly a multiple of that.
- Most of the classification systems are basically "mark and park" type (Slavic 2000), helping to create signatures, but providing hardly any document content description.
- A real challenge exists in the classification of the documents from human intellectual activity. Human output (Dahlberg 2014) is a particular phenomenon and as such, follows no common rules: "inanimate objects (Mayr 1982) should be classified by principles different from those used in biology, because they lack any evolutionary history" (Mayr & Bock 2002).
- Classification systems, thesauri or the knowledge organization possess different levels of constructivism. Most of the classifications are constructivistic in their systematic arrangements, as they deal with non-natural science phenomena i.e. human output, and they use arbitrary methods to accommodate their information about the documents (Dahlberg 2014, 2015).

- Indexing consensus across various cultures - Since not all cultures worldwide understand one item equally and consistently (Tillett 2015), it follows that the indexing of human output may differ or might be biased.

Yet, the two following schemes help to categorize the human output almost analytically.

- Information Coding Classification: to answer the challenge of indexing literature and various kinds of information became a goal for the ICC, a classification system covering almost all existing 6500 knowledge domains. “Its conceptualization goes beyond the range of the well known library classification systems, such as DCC, UDC, and LCC by extending into knowledge systems that so far have not afforded to *classify literature.* ICC actually presents a flexible universal ordering system for both literature and other kinds of information, set out as knowledge fields” [11]. It has nine ontical levels, grouped under three captions [12]: 1. Prolegomena, 2. Life Sciences and 3. Human Output.
- Facet-analysis extracts information treated in the document and expresses the document content in the catalogue. This objectively uniform assessment consists of sequential query, extracting a set of facts such as subject, place, time and form - thus summarizing the document content. Examples are UDC 13, PMEST 14 or CRG 15 as main schemes for facet arrangement of concepts. Similarly Soergel (2009) suggested scrutinizing the document text with the set of categorical questions and expressing the responses in a sequence of exact terms, concepts and notations.

Significance of the Linnaean taxonomy for Classification Design

Although Linnaean taxonomy has been challenged by contemporary genomics and DNA sequencing technology, its value as initial spark for the worldwide accepted classification remains unequalled. Its integrative impact serves as an example of best practice to establish one Unified Classification System for sharing uniform metadata among libraries and any other kind of collections. The use of above mentioned schemes like PMEST, SVOPT 16 or of UDC syntactic rules can perform this kind of classification competently. Their common feature is to analyse the information content with a set of established questions and to convey the answers in precise arrangement of concepts and notations. In case ofUDC this is expressed with hierarchically expressive notations that are friendly to navigate and use. Their complex notations can be deconstructed accurately into simple UDC concepts. Today's aim is to deposit a quest for such a classification system by consensus - conceivably carried out by the next generation of KO experts.

Endorsement - quest for a synthetic classification system by consensus

Everyone should benefit from commonly accepted, comprehensive classification rules! The approach: extracted metadata, arranged in fixed categories, shared in a catalogue. The objective: clearly structured search and finding with high precision / high recall. All the ensuing categorizations are concise and unequivocal - essentially Linnaean, or even better!

Examples of classification dynamics

New concept and hierarchy - Quaternary, the youngest geological period we live in, has been used since 1759 as a concept for the period younger than Tertiary. As a name, it is contained in titles of hundreds of books and it is present in thousands of articles and names of numerous quaternary research groups, working groups, commissions and conferences almost in all countries. A similar concept is Tertiary, used since 1750 to define the second youngest geological period and it is used in a similar amount of titles, documents and organizations. Tertiary has been divided into the Paleogene and Neogene periods. Currently, based on scientific evidence, both these concepts together with Quaternary were joined into one term called Cenozoic ("628" Cenozoic 17).Use of Tertiary is discontinued, and Quaternary became a subdivision of Cenozoic. This fact needs a clear referencing for the older, current and future users, as both geological periods are archives of climate changes in the past and therefore all the written records hold important information for calculating climate models.

Since the scientific language development, reflected by up-to-date classifications, will never cease to advance, this results in a steady enrichment of the well-maintained classifications. These relations can be visualized by “additional referencing” as discussed by Gnoli (’Commerce, see Rhetoric’, 2015) for relationships other than hierarchical, i.e. in cross-discipline relationships, by using: ‘see also’ as in DDC, LCSH 18 and in UDC that points to related classes in other hierarchies. NEBIS 19 system applies related terms (RT) for pointing into poly-hierarchies (Pika & Pika 2015). For that reason the newly added concepts, interlinked in due way, must enable any query for every particular term in its conceptual environment. Alas: the designers of library management systems should be aware of that too.

Adjustment of redundant concepts - Monitoring of science terminology development versus classification schemes in the past 25-30 years 20 reveals that vast amount of information has been faultlessly classified and contributes to enhance the search yield (Pika 2010). Only sporadically the metadata were incorrect or inappropriate due to re-labelling or biased indexing.

Labelling, re-labelling - One of the redundant sources of vocabulary enrichment is the red tape: some new expressions for current, still valid terms, originate from the fact that in many countries the grant organisations would not fund scientific projects bearing titles like “climate change” or “plate tectonics” for several years consecutively, as they seem out-dated. Hence a new scientific label “skin tectonics” instead of plate

tectonics, has much better chances to obtain further funding. As a consequence, coining of the new concept of “skin tectonics”, which expresses the same phenomenon as plate tectonics, achieved the funding. This particular success enriches the vocabulary, however the information entropy increases. Though these artefacts are scanty, they must be disambiguated and placed in correct context.

Bias in point of view leads to different categorization - A “glacier movement”, seen and indexed by a physicist, would be frequently expressed as a “mechanics of continua” combined with “ice”. For a glaciologist or an earth scientist a “glacier movement” is a class descriptor on its own, manifested by its advancing and retreating due to climatic change. The truth is: both classifications are useful, though it is a costly arrangement. A user from natural sciences department would surely opt for the second description search, whereas a physicist would search under the first formulation.

Conclusion

It is both challenging and rewarding to understand the way our users perceive. Our task is to adjust to their level of communication, nothing more, nothing less. The cooperatively attentive cognition yields the appropriate meaning for different terms (Umstätter 2009):

“From the classificatory point of view, we basically have to realize that the human-limited mind can comprehend this world only by its partial generalization. It starts with children who, at an early age commence arranging the world into "quack quack" and "woof woof". Strictly speaking, it is the Aves and the Mammalia, but a lot of parents believe that these are ducks and dogs. With increasing knowledge, our thesaurus grows truly universal! A genuinely interesting challenge”!

Acknowledgement

At the ISKO 2014 Conference I met a number of interesting people and some of them, from the Brazilian ISKO Chapter, sparked my curiosity in their country. In October 2015 during the Lisbon’s UDC Seminar I found the culture of Portuguese language and people very appealing and this prompted the idea of going to ISKO 2016 Conference in Rio. At the same time the work of Ingetraut Dahlberg and Walther Umstätter continues to be equally inspiring and encouraging.

Notes

[1] Information Coding Classification https://de.wikipedia.org/wiki/Information_Coding_Classification
[2] Goethe, J.W.v. 1816:27/7539 http://www.zeno.org/Literatur/M/Goethe,+Johann+Wolfgang/Briefe/1816.
[3] Philosophia Botanica http://www.scientificlatin.org/philbot/pbbibl.html
[4] http://www.scientificlatin.org/philbot/pb31.html
[5] Philosophia Botanica http://www.scientificlatin.org/philbot/pb31.html
[6] Praeludia Sponsaliorum Plantarum, in quibus Physiologia earum explicatur, Sexus demonstratur, modus Generationis detergitur, nec non summa Plantarum cum Animalibus

analogia concluditur. Dec.1729. Original manuscript 1730 preserved in Uppsala University Library, printed in 1908 In: *Skrifter af Carlvon Linné. Utgifna af Kungl. Svenska Vetenskapsakademien*. Band 4, Nr. 1, 1908, S. 1-26

[7] Blunt (1971), p. 244 and p. 34

[8] https://en.wikipedia.org/wiki/Systema_Naturae - cite_note-NG-10

[9] http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1973966/

[10] Facet classification (FC)

[11] Information Coding Classification https://de.wikipedia.org/wiki/Information_Coding_Classification

[12] Literally from ICC: 1. Unbelebtes, 2. Belebtes 3. Produziertes Sein = 1. Inanimate, 2. Animate and 3. Human Production Or the formal version: 1. Prolegomena, 2. Life Sciences and 3. Human Output.

[13] Universal Decimal Classification - UDC codes can describe any type of document or object to any preferred level of detail https://en.wikipedia.org/wiki/Universal_Decimal_Classification

[14] In Colon classification (CC), facets describe "Personality" (subject), Matter, Energy Space and Time: PMEST. It is a system of library classification developed by S. R. Ranganathan.

[15] CRG categories: an expansion of Colon classification to 13 categories: describes any desired level of detail - see Vickery‘s (1960) list for classifying scientific domains: Substance, Organ, Constituent, Structure, Shape, Property, Object of action (patient, raw material), Action, Operation,Process, Agent, Space, Time.

[16] SVO, SVOPT, SVOMPT - Subject S, Verb V, Object O, Place P, Manner M, Time T. - English Grammar Word Order. https://en.wikipedia.org/wiki/Subject/verb/object

[17] "628" Cenozoic (65.5 MYBP -now) in http://www.udcsummary.info/php/index.php

[18] DDC, LCSH: Dewey Decimal Classification, Library of Congress Subject Headings

[19] NEBIS library network, ETH-Bibliothek, Zurich.

[20] KVK - Karlsruhe Virtual Catalog. https://www.google.ch/#q=kvk

References

Blunt, Wilfrid & Stearn, William T. (1971). *The compleat naturalist: a life of Linnaeus.* London: Collins

Blunt, Wilfrid (1994). *The Art of Botanical Illustration.* Dover Publications.

Dahlberg, Ingetraut (Hrsg.) (1982). *International Classification and Indexing Bibliography (ICIB 1): Classification systems and thesauri 1950-1982*. INDEKS Verlag, Frankfurt 1982

Dahlberg, Ingetraut (2014). *Wissensorganisation – Entwicklung, Aufgabe, Anwendung, Zukunft*. Würzburg: Ergon Verlag

Dahlberg, Ingetraut (2015). Warum Universalklassifikation? Lecture at *ESZ-Kolloquium 24.10.2015, Darmstadt*. (Personal communication from Ingetraut Dahlberg: Bad König)

Gnoli, Claudio (2006). Phylogenetic Classification. *Knowledge Organization* 33(3):

Gnoli, Claudio, De Santis, Rodrigo, & Pusterla, Laura (2015). Commerce, see also Rhetoric: cross-discipline relationships as authority data for enhanced retrieval. In *The International UDC Seminar entitled "Classification & Authority Control: Expanding Resource Discovery" in National Library of Portugal in Lisbon, on 29-30 October 2015*.

Goethe, Johann Wolfgang V. (1816). *Briefe*: 27/7539. [http://www.zeno.org/Literatur/M/Goethe,+Johann+Wolfgang/Briefe/1816]

Hansen, Karl Adolf (1902). *Die Entwicklung der Botanik seit Linné.* Gießen.

Hjørland, Birger (2012*).* Facet analysis: The logical approach to knowledge organization. *Information Processing & Management*, 49 (2), March 2013: 545–557

Jahn, Ilse, Löther, Rolf &Senglaub, Konrad. (1985). *Geschichte der Biologie, Theorien, Methoden, Institutionen, Kurzbiographien.* Ed.2, 864 Seiten, VEB Fischer, Jena

Jahn, Ilse (2000). *Geschichte der Biologie*: Heidelberg: Spektrum, 2000 Spektrum Akademischer Verlag; 3. Aufl. (8 Aug 2000).

Linnaeus,Carl (1730). *Praeludia sponsaliorum plantarum* [https://de.wikipedia.org/wiki/Praeludia_Sponsaliorum_Plantarum]

Linnaeus,Carl (1735). *Systema Naturae. Lugduni Batavorum [Leiden, the Netherlands]* [https://en.wikipedia.org/wiki/Systema_Naturae#cite_note-NG-10]

Linnaeus,Carl (1751). *Philosophia Botanica,* P. 12, ed. 1, Stockholm & Amsterdam

Mak, Christian (2011). *"Kategorisierung des Datenbestandes der EuropeanaLocal-Österreich anhand der ICC"* (Bericht des Instituts "Ang. Inf. Forschungsgesellschaft mbh" (AIT) (Graz)

Mallet, James (2007). Hybrid speciation. *Nature*, 446(15).

Mayr, Ernst (1982). *The growth of biological thought: diversity, evolution, and inheritance.* Cambridge (Mass.), London: Belknap Press.

Mayr, Ernst, & Bock, W. J. (2002). *Classifications and other ordering systems,J. Zool. Syst. Evol. Research*, 40: 169–194

Paterlini, Marta (2007). *There shall be order. The legacy of Linnaeus in the age of molecular biology.*European Molecular Biology Organization Report 2007 Sep; 8(9): 814–816. doi: 10.1038/sj.embor.7401061

Pika, Jiri (2010): *Erschließungssysteme in der Schweiz und in der ETH-Bibliothek.* KIT Karlsruhe, 23.7. 2010 : (urn:nbn:de:bsz:ch1-qucosa-64942). 34. Jahrestagung der GfKl

Pika, Jiri, & Pika-Biolzi, Milena (2015). Multilingual subject access and classification-based browsing through authority control: the experience of the ETH-Bibliothek, Zürich. In:*The International UDC Seminar entitled "Classification & Authority Control: Expanding Resource Discovery" in National Library of Portugal in Lisbon, on 29-30 October 2015.*

Rádl, Emanuel (1905). *Geschichte der biologischen Theorien seit dem Ende des 17. Jahrhunderts*. Leipzig, 1905, 2 Bde, 1905-09

Slavic, Aida (2016) *"mark and park" tool (DDC and LCC) A Definition of Thesauri and Classification as Indexing Tools* [http://dublincore.org/documents/thesauri-definition]

Soergel, Dagobert (2009). Illuminating Chaos. Using Semantics to Harness the Web. Presentation at *UDC Seminar, The Hague, 29-30 October 200*9

Tillett, Barbara B. (2015). Complementarity of perspectives for resource descriptions. In *The International UDC Seminar entitled "Classification & Authority Control: Expanding Resource Discovery" in National Library of Portugal in Lisbon, on 29-30 October 2015.*

Umstätter, Walther (2009). *Zwischen Informationsflut und Wissenswachstum,* Berlin: Simon Verlag für Bibliothekswissen. 340 Seiten

Umstätter, Walther (2012). *Email from Inetbib: Re: [InetBib] Leser oder Surfer? Die Zukunft der NYPL* - 6 June 2012.

Vickery, Brian Campbell (1966). *Faceted classification schemes.* New Brunswick, NJ: Graduate School of Library Science at Rutgers University (Rutgers series on systems for the intellectual organization of information, edited by S. Artandi, V. 5).

Joseph T. Tennis

Structure of Classification Theory: On Foundational and the Higher Layers of Classification Theory

Abstract

Provides an interpretation on the structure of classification theory literature.

Introduction

Classification theory has enjoyed an at least 200-year history if we observe its beginnings with treatises that compared different rationales for orders of classes and particular kinds of subdivisions (e.g., Horne,1814). We have benefitted from the intellectual work of many bright thinkers who took serious the task of intellectualizing the methods and concerns we have in building the best classification scheme. Many arguments have been advanced in the literature as to what constitutes the best classification scheme and how we go about building it. What we are faced with at this moment in the field is the need to clearly and concisely articulate what we know, and where we disagree to state what arguments are being advanced in what contexts. From there we need argue where we need to further explore the nature, methods, and core concepts of classification theory.

If we follow Jens-Erik Mai, we understand that there are different kinds of knowledge organization problems – based on size. Further, he argues that size determines whether the knowledge organization problems are tractable. He identifies three sizes of knowledge organization problem, or KOP. Large KOPs are those that are large in scale, both in collection size, number of concepts in the scheme, and number of user types. Medium KOPs focus their attention on collections, concepts, and users in a particular domain. These seem more tractable. Small KOPs are those that focus on individual collections, with their idiosyncratic concept formations, and for a single user, (Mai,2010). This is a helpful characterization of the problem space of knowledge organization. However, there are other ways to organize our thinking about the literature. For example, we can look at the objects of discussion in the classification theory literature, for example design methods or scheme interoperability, and from there, divide the different lines of research. We can see, from inspecting the literature along these lines, at least four distinct kinds of topics.

Four Orders of Classification Theory

The first is that literature which examines the methods and practices of designing classification schemes. This is a large literature, and forms a core component of our canon (e.g. Berwick Sayers,1955 Ranganathan,1967; Vickery,1969; Hunter,2002). Innovations in this literature unfolded in the twentieth century with the advancement of faceted classification and analytico-synthetic classification by Ranganathan and the

Classification Research Group (CRG). This literature we will call *first order* classification theory literature. It is first, because, we need to have classification schemes in order to study how best to improve on them.

The second kind of literature is that which concerns itself with what to do with classification schemes once they are built. The maintenance of schemes as concepts change, the concern with how schemes might interoperate or crosswalk with other schemes, and the application of schemes built for one context are deployed to another. These are *second order* concerns in classification theory. For example, the work on subject ontogeny and scheme versioning (Tennis, 2012) is concerned with how to preserve the functionality of classification schemes over time, while allowing scheme designers to keep with literary warrant. Likewise, interoperability research attempts the seemingly impossible. It looks into ways that schemes built for a particular purpose can be deployed in service to another context (Dahlberg,1996a, 1996b, 1996c; Panzer and Zeng, 2009).

A *third order* is the study of schemes as part of a whole population. They represent those studies that examine a population of schemes and compare them – both the theory of the scheme and the scheme as it is deployed in a context. While this may seem a newer kind of work, Dewey did this in his comparisons of classification schemes at Columbia (Dewey Archives). Contemporary examples of this work are Kipp (2011), and Tennis (2014). These studies want to draw out both the differences and similarities that obtain between the universe of classification schemes (and their allied constructions in the realm of indexing languages).

Finally, the remaining literature can be called *foundational* classification theory. This literature is concerned with defining the concepts and terms we use to discuss classification at all the other orders. Questions that can be posed at this level concern the domains (Tennis, 2003) classes (Broadfield, 1946), relationships between hierarchy and socio-political structures like patriarchy (e.g., Olson,2007), among other definitional concerns. The authors of the publications listed above must deal with foundational and definitional issues in order to do their work. This is true of all publications on all the orders – that some foundational work is usually found in them. This work is variable. Some of it is consciously concerned with laying foundations, like Svenonius (1999), Bowker and Star (2001), and Smiraglia (2003). Others look to not only establish our foundations of classification theory, but also look to destabilize us so that we are thinking critically about our assumptions (Furner,2009, Mai, 2002).

The Use of Orders in Advancing Classification Theory

What do we gain from understanding classification theory to be divided into these components? It can be argued that we know quite a bit about first order classification theory. We have knowledge of the methods of classification scheme construction, and we understand the arguments advanced for designing and building schemes in particular ways. In the polemical sense, one could argue that at this point we do not

need to preoccupy ourselves with the design of schemes. What is more, we might consider the best use of our time as researchers is not in repeating what we know about classification scheme design, but rather focusing on unsolved problems.

The only exception that might surface in this First Order is any work that attempts to generalize the ethics that play out in the design and implementation of contemporary classification schemes. While we have a growing literature on the ethics of classification and the larger field of knowledge organization, we seem to be focused on case studies, rather than abstracting from these cases some basic principles. Exceptions do exist (Mai, 2013).

With regard to the foundational literature, it seems to be a ripe time to interrogate the agreements and the contested ground we have established in this kind of literature. What is the relationship between the structures we have historically cared about and others? What is the relationship between language and our work? We do have particular arguments in this regard, and with the rise of big data corpora we may return to our assumptions about speech acts and language games (cf., Blair, 1990). The remaining two kinds of literature seem ripe for systematization and further exploration at the same time.

Conclusion

In this paper we have provided an argument that the structure of the classification theory literature has at least four layers. We have labeled them orders. We have further advanced the idea that we can use these orders to understand past, present, and future research fronts. If we follow Tennis in his understanding of theory in knowledge organization, "[t]heory is a set of propositions used to explain some phenomena, a narrative," (Tennis, 2008 102), then we must ask ourselves, given the structure of our past narratives, what will the future stories be that we tell about classification schemes?

References

Berwick Sayers, W.C. (1955). *A Manual of Classification for Librarians and Bibliographers*. 3rd ed. London: Grafton.

Blair, David (1990). *Language and representation in information retrieva*l. Amsterdam:Elsevier.

Bowker Geoffrey & Star, Susan Leigh (2001). *Sorting Things Out: Classification and Its Consequences*. MIT.

Broadfield, A. (1946). *Philosophy of Classification*. London: Grafton.

Dahlberg, Ingetraut. (1996a). Library catalogs in the internet: switching for future subject access. *Advances in knowledge organization*, 5: 155-64.

Dahlberg, Ingetraut. (1996b). The compatibility guidelines – a re-evaluation. In *Compatibility and Integration of Order Systems (Research Seminary Proceedings of the TIP/ISKO Meeting, Warsaw, 13-15 September, 1995*. Warsaw Wydawnictwo SBP.

Dahlberg, Ingetraut. (1996c). Compatibility and Integration of Order Systems 1960-1995: an annotated bibliography. In *Compatibility and Integration of Order Systems (Research Seminar Proceedings of the TIP/ISKO Meeting, Warsaw, 13-15 September, 1995*. Warsaw Wydawnictwo SBP.

Dewey Archives. Columbia University.

Furner, Jonathan. (2009). Interrogating Identity: A philosophical approach to an enduring issue in knowledge organization. *Knowledge Organization* 36(1): 3-16.

Horne, Thomas Hartwell (1814). *An introduction to the study of bibliography: to which is prefixed A memoir on the public libraries of the antients*. London: Caldwell and Strand.

Hunter, Eric J. (2002). *Classification made simple*. 2nd ed. Aldershot: Ashgate.

Kipp, Maragaret E. (2011). User, Author and Professional Indexing in Context: An Exploration of Tagging Practices on CiteULike. *Canadian Journal of Library and Information Science* 35(1): 17-48.

Mai, Jens-Erik. (2002). Is classification theory possible? Rethinking classification research. In *Challenges in knowledge representation and organization for the 21st century: Integration of knowledge across boundaries: Proceedings of the Seventh International ISKO Conference 10-13 July 2002 Granada*. Würzburg, Ergon. Pp. 427-78.

Mai, Jens-Erik. (2010). Classification in a social world: bias and trust. *Journal of Documentation* 66(5): 627-42.

Mai, Jens-Erik. (2013) Ethics, Values and Morality in Contemporary Library Classifications. *Knowledge Organization*, 40(3): 242-53

Olson, Hope (2007). How we construct subjects: A feminist analysis. *Library Trends*, 56(2): 509-41.

Panzer Michael, & Zeng, Marcia Lei. (2009). Modeling Classification Systems in SKOS: Some Challenges and Best-Practice Recommendations. In *Proceedings of the International Conference on Dublin Core Metadata and Metadata Applications*.

Ranganathan, S. R. (1967). *Prolegomena to Library Classification* 3rd Edition. Madras.

Smiraglia, Richard P. (2003). The history of 'the work' in the modern catalog. *Cataloging and Classification Quarterly* 35(3/4): 553-67.

Svenonius, Elaine. (1999). *The Intellectual Foundation of Information Organization*. MIT.

Tennis, Joseph T. (2012). The strange case of eugenics: a subject's ontogeny in a long-lived classification scheme and the question of collocative integrity. *Journal of the American Society for Information Science and Technology*, 63(7): 1350-59.

Tennis, Joseph T. (2003). Two axes of domain analysis. *Knowledge Organization*, 30(3/4):191-5.

Tennis, Joseph T. (2008). Epistemology, Theory, and Methodology in Knowledge Organization: Toward a Classification, Metatheory, and Research Framework. *Knowledge Organization*, 35(2/3): 102-12.

Tennis, Joseph T. (2014). Description and Différance: Archives, Libraries, and Museum Descriptive Traditions and their Educational Communities and Cultures. In Urban, R. J., Coleman, L-E., Marty, P. F. *Libraries, Archives, and Museums: Connecting Educational Communities and Cultures. In Proceedings of the ASIST Annual Meeting*. Seattle, Washington.

Vickery, Brian C. (1960). *Faceted classification: A guide to construction and use of special schemes*. Aslib.

Akhigbe Bernard Ijesunor, Aderibigbe Stephen Ojo, Kayode Anthonia Aderonke, Afolabi Babajide Samuel, Adagunodo Emmanuel Rotimi

Towards Better Knowledge Organization Systems: Exploring the *Uc*-Paradigm of Evaluation

Abstract

On realizing the significance of KOS, several efforts have been made (some ongoing) at revising the conventional Knowledge Organization (KO) tools. Some are aimed at designing new tools, while others based on Web enabled collaboration have resulted in folksonomies, and so on. Motivated by the understanding that the forgoing signifies the possibility of not only revising KO System (KOS), but improving them; this paper contributes a Measurement Model (MM) of Actionable Attributes (AA) as personas. The Web Search Engine - a manifestation of KOS - served as an example case, upon which the characteristics of Knowledge itself was explored using the User-centred (Uc) paradigm of evaluation. The Theory of Information Processing (ToIP) was drawn on and the Web analytic technique provided the empirical guide and the praxis to translate the supporting theory - ToIP - to action. The presented AA were modelled based on the experiences of users as they bother on Uc interactive software design using the exploratory factor analytic technique. The AA (as it were) would contribute to the re-engineering of end-users' needs towards better KOS. This is because the AA as personas evidently encapsulated "behavioural data" that were gathered from the empirical analysis of actual users. In future, there will be need to develop a detailed narrative of each persona and possibly select more realistic and corroborating scenarios (and photos) that will be easily translated into believable characters.

Introduction

The modern world has become almost (if not) a connected society. Cultural, scientific and technological experiences (in the form of *knowledge*) can be shared anytime-anywhere without much difficulty. Information and Communication Technology (ICT), the Internet and the ubiquity of the World Wide Web (3W) are responsible for this (Ahuja,, 2015). *Knowledge* is a justified belief. It is subjective, dynamic, not self-contained, socially constructed, and affective in characteristics (Loebbecke, 2012). It exists in an individual's mind, and can be influenced by cultural, scientific and technological experiences (Alavi and Leidner 2001). This influence is derivable from *knowledge* artifacts, which are physically stored information in documents, records, or videos (Davenport and Prusak 1998).

Interestingly, existing Knowledge Organization Systems (KOS) - thesauri, taxonomies, folksonomies and others are composed of registers (Shiri,2014), which are important elements of search engines (Garud and Kumaraswamy,2005). They enhance the functionalities of search systems (Horrocks,2008; Shiri,2014), and despite these attributes, KOS are still purely systemic in fashion. This systematism limits KOSs and the characteristics of *knowledge* is undermined. No doubt the algorithms that implement the KOS are useful for retrieval (Castellano et al.,2004). However, they are self-contained unlike *knowledge*, and as such inhibits KOS from seeking additional information. The assessment of how *knowledge* is organized such that its

characteristics are not tempered with has also continued to receive less attention. This paper debates the fact that the real end-users of KOS are useful sources of *knowledge*, which can be elicited. The information can be used to inform how search and retrieval systems are designed to make them better. The question is; why will KOS not be able to modify its own behaviour in response to changes in its operating environment? They should be open-adaptive (Oreizy et al.,1999), and able to support and adjust to new users' behaviours continually. There are sufficient end-user inputs in its operating environment in the form of queries, users' interactive experiences (and so on) that it can learn from (or adapt to). In software engineering (even way back - Oreizy *et al*, 1999) it has been possible to engineer algorithms that are not self-contained. So, why will this not be possible for KOS? There may be need to re-engineer existing KOS algorithms considering the foregoing. However, application developers need to answer several questions when developing a self-adaptive software system (Oreizy *et al*., 1999; Krupitzer *et al.*,2015). One of such questions (which this paper seeks to answer) is; how would KOS Developers (KDs) understand and find the need(s) of users' that are informed from patterns of their behaviour? This implies that for KOS and its manifestation - the Web Search Engines (WeSEs)) (Baker et al., 2013) to reflect the characteristics of Knowledge, and also be truly usable, the real end-users must be examined.

The foregoing represents the main motivation for this paper. Its focus is on assessing KOSs - the WeSE as a case. Its users particularly on the 3W; their direct interaction with WeSEs' interfaces; their associative relationships with KOS; and how they achieve deep search formed the context of this paper. The rest of the paper is structured with section 2.0 dedicated to literature review; 3.0 methodology; 4.0 result, and 5.0 conclusion.

Literature

The work of Kumar (2013) highlighted the fact that users are still not able to communicate clearly with search systems. The query language and domain terminology are difficult to understand, such that the context of user's query are incomprehensible. Nevertheless, there are efforts to create KOS based on Human Need(s) that are understood by machines (Ahuja, 2015). Thus, the KO community seem set to do things differently and develop KOS with flexibility and subjectivity in approach (Matthes, 2012). However, there is dearth of user-centred Actionable Attributes (AA) that can guide *adaptation decision* for the design of self-adaptive KOS. It is interesting to note that users' behavioural traces are ubiquitous on the Web, and have not been harnessed. In terms of conceptual relationships and logical theories, existing semantics of domains could be exploited to make KOSs more self-adaptive.

In Binzabiah and Wade (2012), it was affirmed using the report of Sharif (2009) that KOSs are strict and formal, and are characterized by several features of formality, solidarity and immutability. The work of Grey *et al.* (2012) also supported the claim by

Binzabiah and Wade (2012), and Sharif (2009) that vocabulary controlled tools like the KOSs are strict and static. Based on the foregoing, we believe the time is ripe to motivate a new perspective on the need for non-static (non-self-contained) KOS with open-adaptive capability as a better way to represent *Knowledge*.

There is dearth of appropriate theoretics to distinctly conceptualize *knowledge* and emphasize its role using Knowledge Portals (KP) as a unifying networking and repository KOS. Loebbecke (2012) identified three challenges with which a set of hypotheses about the successful deployment of KPs were postulated. For theorizing purpose, the outcome variable of knowledge reuse was explored and validated. The researcher explored their applicability through a review of 42 empirical KP-related studies (Loebbecke, 2012). However, like in the other work the provision of *Uc* AAs as guide for *adaptation decision* towards the design of self-adaptive KOS was not communicated. Motivated by this postulation, this paper offers information on how to possible leverage the potentials of *Uc* to contribute useful requirement(s) based on users' context and prior experience of system use.

Methodology

This section contains three subsections - the subsection on theoretical background, methodological framework, and the method used in this paper.

Theoretical Background

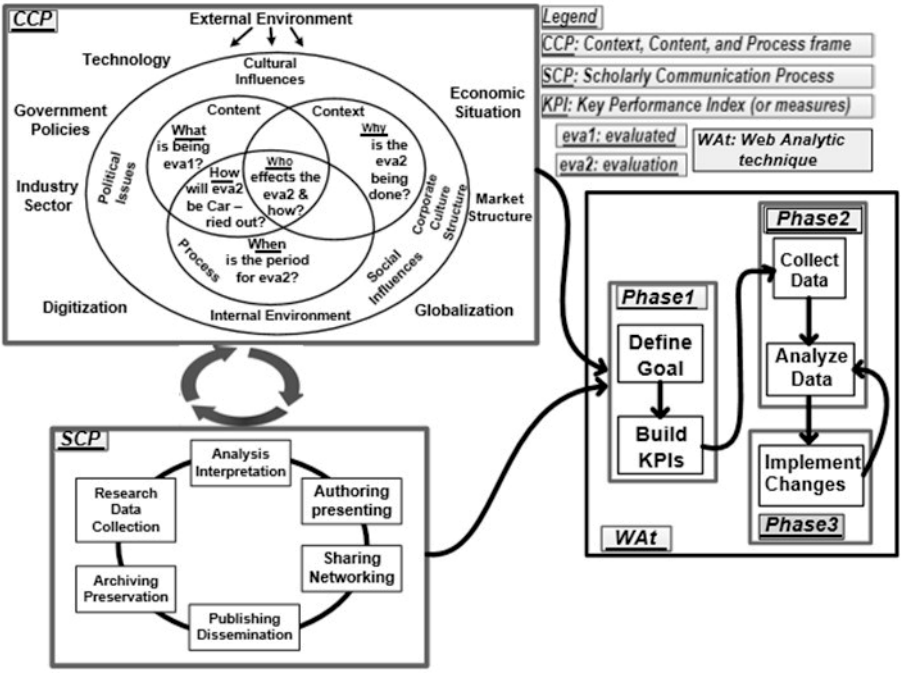
In Hjørland (2015), KOSs were referred to as non-static theoretical representations. Since, KOSs are dynamic in properties, and respond to empirical evidence (Hjørland, 2015); the concept of Web Analytics (WA) (Fagan, 2014) and the theory of Information Processing (TIP) (see Gao *et al*., 2012; Ortiz-Cordova and Jansen, 2012) were drew on. These theoretical underpinnings were necessary to offer new perspective(s) with clear implications for the practical use of empirical evidence towards better KOS (Hjørland, 2015). The foregoing theoretics are the core of the *Uc* paradigm. This paradigm draws from the concept of user-centricity with its roots in user-centred design. Its practice represents a general philosophy that seeks to bring users into the design process (Miaskiewicz and Kozar, 2011). Satisfying and fulfilling users' need(s) are the central concern of the Uc. This can be difficult to attain; hence the introduction of personas in this paper to provide alternative method to represent and communicate users' needs (Miaskiewicz and Kozar, 2011). The challenge of directly involving users in large design processes can be tasking. Time, cost, and logistics are other constraints (Marshall *et al.*, 2014). These were managed using personas to provide approximations to intended end user requirement(s).

The Methodological Framework

The characteristics of knowledge is evidently *Uc* (Loebbecke, 2012). The *Uc*-Paradigm of evaluation - a persona-oriented was therefore adopted in this paper. To use

it; the framework in Figure 1, was leveraged. As a parsimonious frame, the CCP and SCP frames by Irani (2008) and Rieger (2010a) provided the context and viability to assess users based on a wide variety of evaluative situation(s). The CCP highlights human (information) needs as being from social and natural environments, which are further influenced by several situations and factors of the society (see Figure 1) (Irani, 2008; Ahuja, 2015). The SCP frame on the other hand raises a range of processes such as create, represent, organize, analyze, and communicate "scholarly or user" content as how users' of ICT perceive their information needs (Rieger, 2010). In synergy with the WAt, they provided an extensible context to Identify, Conceptualize and Operationalize (ICO) intended measures broadly and qualitatively. The ICO is a qualitative and quantitative research methodology (Hussein, 2015). The ICO activities were carried out in Phase 1, and the WAt in Phase 2 guided the quantitative aspect. The circular arrows in between the CCP and SCP means both context were thoroughly considered using the participant-observatory approach in Endres et al. (2008). This is because the example KOS - as a case is pervasive, and requires broader (CCP and SCP contexts) approaches to perform the ICO activities to elicit evaluative data. Additionally, for pervasive systems like WeSEs the ICO activities were guided by the what, why, who, how and when factors of evaluation. By implication, the processes highlighted in Figure 1 is supported by the TIP, which emphasizes the processes users go through to make decisions (Gao et al., 2012). Thus, we assumed that for a praxis-oriented evaluation exercise that is Uc, the users understood the nuances, influences and perceptions of the elements involved in the evaluation in question. The influence of the CCP and the SCP helped in this perception of the context as conceptualizable, though complex. The postulations of the CCP and SCP as underpinned by the TIP made the sense-making doable. Details of what happenned in Phase 2 and 3 are presented in section 3.3.

Figure 1: The praxis-oriented framework based on the Web analytic technique



Methods

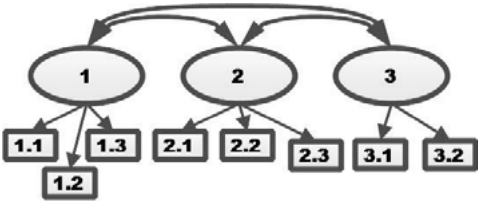
In Phase 1 goals were defined, and used to build KPIs using the ICO activities as earlier stated. Following similar effort in Akhigbe *et al*. (2014; 2015), the ICO yielded decision variables with which data were elicited as required in Phase 2. The data collected using the questionnaire survey technique (Nielsen and Hansen, 2014) are *Uc* data (or ordinal data); they represented the cognitive thoughts of 250 inferred sample size frame. To explore the *Uc*-Paradigm of evaluation the Factor Analytic (FA) technique was relied on to analyse and model the *Uc* data as personas. A representative model that is also a Measurement Model (MM) since it is re-usable was formulated using result from exploratory FA that captured users' focus, their individual goals and what they wanted to achieve. The MM is made up of AA (see Figure 2), which are ready to be used to inform relevant changes (see Phase 3). When this said changes are made, if there is need for adjustment, this can be made as the arrow suggests.

Result

Three measures (or factors) were adopted for translation into personas, which are user requirements following Idoughi *et al*. (2012). The profile of the users in the sample frame are implicated in the personas (see Table 1) as practice in LeRouge *et al*.

(2013). We present 3 personas (see Table 1.) following the standard in Goodwin (2010) and Courage and Baxter (2005). Identity, general profile, knowledge and experience, goals, and expectations are the personal componenets that guided the persona creation in Table 1 (Idoughi *et al*., 2012; Maliki *et al*., 2015). The analytics from the MM in Figure 2 resulted in the personas from three parsimonious decision variables - 1.1, 1.2, and 1.3; 2.1, 2.2, and 2.3; and 3.1 and 3.2 that explained factors 1-*Interactivity*, 2-S*ystem efficiency*, and 3-*Self-adaptivity* respectively. The plausibility of the factors is shown in the degree of reliability of the factors (1, 2 and 3 - see Figure 1), which score of decision variables were ≥ 0.5 threshold point.

Figure 2: The measurement model



Conclusion

Knowledge is not self-contained, but existing KOS are; and as earlier argued, this inhibits it from seeking additional information. The benefits of the Uc paradigm of evaluation have not been harnessed in the KO research community. Less attention has been given to it. Interestingly, the users of WeSEs - the example case are so many and data (their opinion) from them cannot be ignored. Actionable Attributes (AA) can be formulated from them, and used by KDs to inform better and usable KOS. This paper therefore provided empirical evidences to support this argument. To answer the question of how KOS developers will understand and find the need(s) of users' as informed from patterns of their search behaviour; a measurement model was presented (see Figure 2). This paper also contributed AA, with the belief that the KO research community should be alerted of the possibility of developing KOSs that are open-adaptive. This was motivated by the belief that KOS, should be able to (i) learn from users' behaviours that are modelled from users' need(s) based on users' experiences, and (ii) adapt by adjusting itself to user-oriented attributes towards the provision of better search experiences.

Table 1: Sample personas as actionable attributes based on the measurement model

Personas	Personas
(1): Interactivity **(1.1)** Users' in interactivity category, wants To be able to select from choices; **(1.2)** exert less effort to locate information relevant to their IN; **(1.3)** need system that is responsive and use a system without feeling a pressure. ***(2): System efficiency*** **(2.1)** Users want to use a system that empower them to have control; **(2.2)** easily use the system even with little or no technological experience; **(2.3)** keep track of system activities, and be assisted by the system with just-in-time support, even if it means adapting to varying changes that is orchestrated by the user previous search behaviour.	***(3): Self-adaptivity*** **(3.1)** *Personalization factor:* They wanted a system that is tailored to their personalization wrt IN; a system that is able to adjust their behaviour in response to the perception of end-user/environment and of the system itself; **(3.2)** *Institutionalization factors:* Users need Search system that will reason along the context of their IN by characterizing their (entity) situation as possibly as possible. *Time* (proactive & reactive), *Technique* (parameter, structure & context), and *Reason* (change in context, change caused by the users, & so on) are some useful clue for IN characterization.
Wrt: *(with respect to)*	***IN:*** *(Information Need)*	***Personas:*** *between Ages ≥ 16 to > 56*
Personas' Identity/General Profile: *undergraduate students, University Lecturers, Researchers*		
Personas' Knowledge/Experience: *They use the Internet on a daily, weekly and monthly basis*		

The foregoing was achieved using the *Uc* paradigm of evaluation based on the exploratory factor analytic technique. The actionable attributes that resulted from the exercise sufficed as feature(s) of users' requirements for KOS. The level of validity of the measurement model that was presented showed this (see Figure 2), since the models' decision variable (items) were all above the standard threshold point. The AA as personas and *Uc* attributes will help KOS's designers stay on track with what work and does not work for users. The perspective of users of KOS can be easily toggled on and off to provide useful Mental Framework (MeF) for KO experts who will rely on them to quickly readjust their perspective(s) to that of the real user. These personas can also be adopted as representation of real end users, when users are not available or unknown. Even the process through which the MM was formulated can be reused at low cost, since it will not be conceived afresh to study and create users' attributes.

Overall, our research contribution seeks to motivate a new perspective(s) on the need for nuance metric(s). From the metric(s) as AA; models, goals, rules/policies that can influence the *adaptation decision* to assist in the design of KOS with open-adaptive qualities can be delivered. We hope to concentrate on re-engineering the AA to achieve usable target systems for real users. The absence of real narrative for each persona and realistic scenarios cum photos and quotes from the primary data is a drawback.

Acknowledgment: Many people supported this work, and they are all acknowledged. Finally, thanks to Obafemi Awolowo University, Ile-Ife, Nigeria for their support through TETfund.

References

Ahuja, Jayasree (2015). KO Systems Based on Human Needs Approach - To Bring Harmony between Homogeneity and Heterogeneity of Future Information Environment. *Advances in Classification Research Online*, 25(1): 1-15.

Akhigbe, Bernard Ijesunor, Afolabi, Babajide Samuel, & Adagunodo, Emmanuel Rotimi. (2014). A Baseline Model for Relating Users' Requirements of Web Search Engines. *Advances in Knowledge Organization*, 14. Pp. 374 - 81.

Akhigbe, Bernard I., Afolabi, B. Samuel, & Adagunodo, Emmanuel R. (2015). Towards a Conceptual Praxis-oriented Evaluative Framework: A Web Analytics Approach. In *Proceedings of the 10th Conference of ISKO-France-Knowledge Organization Systems and Digital Humanities, Strasbourg, France November 5th– 6th, 2015*. Pp. 1-19.

Alavi, Maryam, & Leidner, Dorothy E. (2001). Knowledge Management and Knowledge Management Systems: Conceptual Foundations and Research Issues. *MIS Quarterly*, 25(1): 107-36.

Al-Maliki, Malik, Ncube, C., & Ali, R. (2015). Adaptive Software-based Feedback Acquisition: A Personas-based Design. In *Proceedings of the 9th IEEE International Conference on Research Challenges in Information Science, May 13-15, Athens, Greece* Pp. 100-11.

Baker, Thomas, Bechhofer, Sean, Isaac, A., Miles, A., Schreiber, Guus, &Summers, Ed (2013). Key Choices in the Design of Simple Knowledge Organization System (SKOS). *Web Semantics: Science, Services and Agents on the WWW*, 1-15.

Binzabiah, Reyad, & Wade, Steve (2012) Proposed Method to Build an Ontology Based on Folksonomy. In *Proceedings of the International Conference on Information Society. i-Society*, IEEE, London, UK. Pp. 441-6.

Blomkvist, Stefan. (2002). The User as a Personality. Using Personas as a Tool for Design. In P*roceedings of the Workshop "Theoretical Perspectives in Human-Computer Interaction" at IPLab, KTH, Septem*ber 3, 2002.

Castellano, C., Cecconi, F., Loreto, V., Parisi, D., & Radicchi, F. (2004). Self-contained Algorithms to Detect Communities in Networks. *European Physical Journal B*, 38: 311-9.

Courage, C., & Baxter, K.(2005). *Understanding Your Users: A Practical Guide to User Requirements Methods, Tools, and Techniques*. San Francisco: Elsevier.

Davenport, Thomas H, & Prusak, Laurence (1998). *Working Knowledge: How Organizations Manage What They Know*. Boston: Harvard Business School Press.

Fagan, J. C. (2014). The Suitability of Web Analytics Key Performance Indicators in the Academic Library Environment. *Journal of Academic Librarianship* 40: 25-34.

Garud, Raghu, & Kumaraswamy, A. (2005). Vicious and Virtuous Circles in the Management of Knowledge: The Case of InfoSys Technologies. *MIS Quarterly*, 29 (1): 9-33.

Goodwin, Kim (2010). *Designing for the Digital Age: How to Create Human-centred Products and Services*. Indianapolis, USA: Wiley Publishing.

Grey, April, Hurko, Christine R. (2012). So You Think You're an Expert: Keyword Searching vs. Controlled Subject Headings, Codex.*The Journal of the Louisiana Chapter of the ACRL*, 1(4): 16-26.

Hjørland, Birger. (2015). Theories are Knowledge Organizing Systems (KOS). *Knowledge Organization*, 42(2): 113-28.

Horrocks, Ian (2008). Ontologies and the Semantic Web. *Communications of the ACM*, 51(12): 58-67.

Idoughi, Djilali, Seffah, Ahmed, & Kolski, Christophe. (2012). Adding User Experience into the Interactive Service Design Loop: A Persona-based Approach. *Behaviour and Information Technology*, 31(3): 287-303.

Krupitzer, Christian, Roth, Felix M., VanSyckel, Sebastian, Schiele, Gregor, & Becker, Christian. (2015). A Survey on Engineering Approaches for Self-adaptive Systems. *Pervasive and Mobile Computing*, 17: 184-206.

Kumar, A. (2013). A Comparative Analysis of Taxonomy, Thesaurus and Ontology. *International Journal of Applied Services Marketing Perspectives*, 2(1): 251

Larsson, Magnus. (2013). Theoretical highlight 10: Users, Tasks and Requirements. [http://www.it.uu.se/edu/course/homepage/hcinet/ht13/lectures/Lecture10/part2]

LeRouge, C., Ma, J. Sneha, S.& Tolle, K. (2013). User profiles and personas in the design and development of consumer health technologies. *International Journal of Medical Informatics*, 82: e251-e268.

Loebbecke, Claudia, & Crowston, Kevin. (2012). *Knowledge Portals: Components, Functionalities, and Deployment Challenges.* [http://crowston.syr.edu/sites/crowston.syr.edu/files/KP%20to%20distribute.pdf]

Matthes, Florian, Neubert, Christian, & Steinhoff, Alexander. (2012). Structuring Folksonomies with Implicit Tag Relations. In *Proceedings of the 23rd ACM conference on Hypertext and social media*. Pp. 315-6.

Nielsen, Lene, & Hansen, Kira Storgaard (2014). Personas is Applicable - A Study on the Use of Personas in Denmark. In *the proceedings of CHI, Toronto, ON, Canada*. [https://pure.itu.dk/ws/files/74977691/06_01_2014_chi_2014_personas_are_applic_revised_final.pdf]

Oreizy, Peyman, Gorlick, Michael M., Taylor, Richard N., Heimbigner, Dennis, Johnson, Gregory, Medvidovic, Nenad, Quilici, Alex, Rosenblum, David S., &Wolf, Alexander L. (1999). An Architecture-based Approach to Self-adaptive Software. *IEEE Intelligent Systems*, 3: 54-62.

Sharif, A. (2009). Combining Ontology and Folksonomy: An Integrated Approach to Knowledge Representation. In *IFLA 2009 Satellite Meeting in Florence, Italy, 1-13*.

Shiri, Ali. (2014). Linked Data Meets Big Data: A Knowledge Organization Systems Perspective. *Advances in Classification Research Online*, 24(1): 16–20.

Richard P. Smiraglia

Extending Classification Interaction: Portuguese Data Case Studies

Abstract

"Classification interaction" describes an emergent research stream in which an attempt is made to use the network structure embedded in classification strings to navigate pathways through large bibliographic datasets by discovering predictable correlations of co-occurrence among data elements. Three datasets from Portuguese libraries were used for further analysis. The datasets came from BND Livre (the National Library of Portugal's Digital Library), BNP Catalogo (The National Library of Portugal catalog), and BNP PORBASE (the Portuguese national bibliographic utility). Results are consistent with earlier results demonstrating how elements of a KOS can be statistically associated with certain bibliographic characteristics of the documents they describe. Results have extended our comprehension of classification interaction and also demonstrated the value of continued research in more and more diverse contexts.

Classification interaction

"Classification interaction" is the term used to describe an emergent research stream in which an attempt is made to use the network structure embedded in classification strings to navigate pathways through large bibliographic datasets by discovering predictable correlations of co-occurrence among data elements. The prospect that it might be possible to use classification in this manner emerged from a series of papers about elementary structures of knowledge (Smiraglia and van den Heuvel 2013; Smiraglia, van den Heuvel and Dousa 2011), in which thought experiments matched bibliographic elements (such as editions or series membership) with bibliographic phenomena (such as membership in complex faceted subject domains). In fact, the elementary theory suggests such interactions might be predicted by association with specific patterns of complex classification, such as those represented by faceted classification strings. The Universal Decimal Classification, which employs complex synthesis and a faceted auxiliary structure represents an intriguing host system for seeking empirical evidence of classification interaction.

Indeed, further empirical pointers arose from research by the Dutch Knowledge Space Lab (KSL) team1 who used empirical ontogenetic analysis of the UDC as a control set alongside its analysis of the emergent network of knowledge in Wikipedia. In a series of papers this team demonstrated the extent and complexity of the UDC, demarcated its evolution over the past century and demonstrated empirically its underlying network structure (Akdag Salah et al. 2012, Smiraglia et al. 2013). These analyses were based on UDC datasets from the OCLC WorldCat and the library of the Katholische Universiteit Leuven.

Three papers on classification interaction (Smiraglia 2013, 2014a, 2014b) have reported empirical data to date demonstrating some statistically significant results based on datasets of bibliographic records classified with the Universal Decimal

Classification. In these papers MARC-coded bibliographic characteristics were correlated with each other and with elements of deconstructed UDC strings. Statistically significant correlations of both types occurred, demonstrating the potential for creating a means of data-mining the underlying network structure. That is, it was demonstrated that in works post-1970 the presence of an ISBN (International Standard Book Number) is strongly correlated with membership in a series, and furthermore that these two bibliographic characteristics are strongly correlated with UDC main classes in the sciences. Attempts to correlate specific semantic values using place names, publisher names and terms from topical subject indexing were less successful, partially because the datasets were not well-indexed or were too small to generate statistically-significant correlations. For example, few publisher names or place names appeared more than five times in the entire dataset from the WorldCat. Still, some semantic clustering techniques were demonstrated in association with elements of UDC strings in the records. This abstract extends that research with new and different bibliographic datasets provided by the Portuguese National Library.*

Portuguese datasets

In 2015 the KSL team received three datasets from Portuguese libraries for further analysis of UDC population. The Portuguese datasets were thought to be useful for extending the research, because unlike many libraries contributing to the OCLC WorldCat, Portuguese libraries used UDC as a form of subject indexing rather than for shelving. This means, for example, that records frequently contain several UDC strings, each matching a different aspect of the document's subject content. The team received files created using Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH), which yielded XML files containing local records in UNIMARC format. MARC field coding was used to extract specific data. The datasets came from BND Livre (the National Library of Portugal's Digital Library), BNP Catalogo (The National Library of Portugal catalog), and BNP PORBASE (the Portuguese national bibliographic utility). Thus the datasets are themselves diverse, and together they add quite a lot of richness to the analytical capability based on the UDC.

From BNP PORBASE we received 1.1 million records, from the BNP Catalogo we received approx. 880,000 records and from the BND Livre we received 21,000 records. For the present research, random samples of 400 records were extracted from each dataset. The sampling technique (described in Smiraglia 2013 and 2014b) randomly selected records from each dataset with a sample size calculated to yield results at 95% confidence ±5%. In the prior trials, generalizability was indicated by the matching sample and population date of publication and UDC main class distributions. In the present study bibliographic elements were extracted from the records and arrayed in spreadsheets prior to processing with the IBM-SPSS Statistical package. Most data-cleaning consisted of regularizing dates of publication, which had been transcribed or recorded according to bibliographic practice rendering them unusable if (e.g., "ca.

1564" or "imp. 1702" or "18-?"). It is recognized that some bibliographic subtlety is lost in the regularization; this is considered a reasonable limitation of the methodology.

BND Livre

The basic descriptions of the three sample datasets show how clearly these represent new and more diverse data than the WorldCat or Leuven datasets. The BND Livre (National Library of Portugal's Digital Library) sample had 387 functional bibliographic records. Few were very fulsome in bibliographic detail. Only one had an ISBN, three had an edition statement, and seventeen had series statements. Eight languages were represented; fourteen records had no language indicated. Almost half of the records report Portuguese, another third are in French or Latin. Almost half have no place of publication recorded, Lisbon occupies 21.6%, Paris 9.6%. No publisher's name occurs more than twice in this dataset. Interestingly, most of the topical descriptors were musical terms.

Dates of publication range from 1500 to 2004. Most of the works date from the 18th and 19th centuries. 40% were published in 1700, 4% in 1850, and the rest spread evenly. The UDC main classes show the collection is predominantly populated by history and geography, arts entertainment and sport, and social science and politics. A quarter of the records occur in class 9, another quarter in class 7 and 20% in class 3. No records occur in class 1. There are 1,050 UDC strings assigned to the 387 records; the mean number of strings per record is 2.7, the range is from 1 to 7. 80 topical subject headings were assigned to 28 records, for a mean number of headings per record of 2.8.Obviously the majority of records had no topical indexing assigned, but the rate of 2.7 to 2.8 indexing strings (if we compare subject heading assignment to UDC string assignment) is consistent across the dataset.

BNP Catalogo

The Catalog of the National Library of Portugal had 400 usable records in the sample dataset. These were more bibliographically dense than BND Livre. 157 of the records or about 40% had ISBNs. 268 or a bit more than half had series statements, but none occur more than twice. Only 3 had geographical subject headings, 62 had topical subject headings, and 11 had other subject terms. Eleven languages are reported in the records, but 80.8%, the vast majority, were in Portuguese, with 6.3% in French and 5.3% in English. Places of publication are diverse with 110 different place names reported, but the majority, 39.5%, are from Lisbon, 13% from Porto, and 5% from Coimbra. Publisher names are diverse as they were in the WorldCat with 2.3% from Porto Editora, 1.3% from Asa, and most of the rest occurring 1-3 times. Most of the records are dated post-1997. The population of the UDC in the BNP shows all main classes are represented. A little more than a quarter are in social sciences, another large cluster are in literature and language, followed by applied sciences and arts. 168 topical subject headings are assigned to 62 records; the mean is 2.7 headings per record. 891 UDC strings are assigned to 400 records; the mean is 2.2 strings per record. The

longest UDC string in the three datasets occurs in this dataset: 316.344-055.2(469.512)"1567/1776"(091).

BNP PORBASE

The PORBASE bibliographic network sample also contained 400 usable records. Like the BND Livre, there is less formal bibliographic data in this set. Only 78 records had ISBNs, 35 had edition statements, 200 or half contain place names and publisher statements, 61 have series statements, 4 have geographic subject terms and 28 have topical subject headings. Nine languages are recorded, of which 31.8% are Portuguese, 13% are English and 7.2% are French. Among 63 place names 13.3% are Lisbon, 5% Porto and 3.8% Paris; Coimbra, London and New York contribute approximately 2% each; Sao Paolo, Braga, Madrid and Vila Real approximately 1% each. None of the rest occur more than twice. Publisher names are similar to those in the BNP Catalogo, with Asa occurring 4 times (1%) and all the rest only once or twice. Only three series statements occurred more than once. Dates of publication span the range from 1575 to 2014.This distribution of dates is roughly the same as that in the BNP, with the majority of the works occurring between the mid-twentieth century and the present. UDC Main class population shows the social sciences, literature and applied sciences predominate. This time religion is in the middle of the distribution and natural sciences at the bottom. In this dataset 711 UDC strings were assigned to 400 records; the mean is 1.7 strings per record. 68 topical subject headings were assigned to 28 records; the mean is 2.4 headings per record.

Results

Classification interaction describes the extent to which elements of a KOS interact with characteristics of the documents they describe. To begin, the main UDC class associated with each sample item was cross-tabulated with the presence of ISBN, edition statement and series statement using IBM-SPSS™. Results for all three datasets are are shown in Table 1.

Table 1. Main UDC classes by bibliographic characteristics in all three datasets

UDC Class	ISBN	Edition	Series
0			
1			
2	.003/.005 BNP	.007/.019 BNP	
3		.027/.033 PORB	.006/.013 BND
5			
6		.009/.008 BNP	.003/.003 BNP
7	.003/.002 BNP .027/.026PORB	.025/.027 BNP	
8	.000/.000 BNP .002/.004PORB	.000/.000 BNP 0.011 BND .000/.000 PORB	.000/.000 BNP .007/.011
9			

The presence of consistent associations is what is most remarkable. In each case the Chi-squared significance level is given followed by the level from Fisher's exact test if it also was calculated. Class 8 is the most clearly associated; this makes sense given that class 8 contains literature, but more interesting is to consider that we are looking at three different datasets ranging from a national library to a digital library in all three cases there is a clear association between the presence of these elements and the class number 8. One might expect all books to have ISBNs, but in fact most in these datasets do not, but most literature has ISBN and an edition statement and a series statement. Classes 2,3,6 and 7 also show some associations; and classes 0,1,5 and 9 do not.

Another approach to classification interaction is to analyze the relationship between the presence of classification elements and semantic indexing in the bibliographic records. The three datasets analyzed here varied in their use of topical indexing terms, but there were no formal subject headings per se in any of the datasets. Instead topical terms appeared in a small proportion of bibliographic records in each dataset. Because the use of multiple UDC strings in each record provides the fullest topical coverage in all three datasets, it seems fair to assume that topical terms are added only when necessary to express concepts such as forms or genres not easily expressed with the UDC. In BND Livre 28 of 387 bibliographic records contained topical terms; In BNP Catalogo there were topical terms assigned to 62 of 400 bibliographic records; and, in PORBASE 28 of 400 bibliographic records carried topical terms. To analyze the interaction of topical terms assigned with UDC main classes a cross-tabulation was conducted using IBM-SPSS™. Results for all three datasets are are shown in Table 2.

Table 2. Main UDC Classes by Presence of Topical Index Terms

Dataset/UDC Class	0	1	2	3	5	6	7	8	9
BND Livre topic	.034//058						.014/.020		
BNP Catalogo				.006/.013			.000/.000	.003/.002	
PORBASE	.000/.000			.05/.035		.05/.035	.000/.000		

Most of the associations occurred in the PORBASE dataset; associations in classes 3 and 6 were borderline Chi-squared associations, shown by Fisher's exact test (shaded areas) to be one-directional, thus in these cases the presence of the topical terms is associated with the main class but not the other way around. Class 7, which contains music, has the most and some of the strongest associations. Literature in the national library catalog also is strongly associated with the addition of topical index terms. All of this strengthens the supposition reported above that most of these terms supply form and genre indications not represented in typical UDC strings. To further analyze the semantic content of the topical index terms, frequency distributions for each dataset were prepared. Terms used more than once in each dataset are shown in Table 3.

Table 3. Most frequently occurring topical index terms

BND Livre		Estudos de caso	3	Lisboa (Portugal)	2
Gravuras	15	Língua portuguesa	3	Qualidade de vida	2
Instrumentos musicais de cordas	11	Matemática	3	**PORBASE**	
Música manuscrita	5	Música …	4	Música para piano	3
Música impressa	3	Química	3	Portugal	3
Crimes--Sentenças--Portugal--1320-1864	2	Actas	3	Teses	3
Documentos administrativos--Portugal--1320-1864	2	Dicionários	3	Baixo Alentejo (Portugal)	2
Música para canto e piano	2	1999-2002	3	Catálogos	2
BNP Catalogo		Alunos portugueses do ensino básico	2	Ensaios	2
Portugal	8	Ciências	2	Escolas do ensino básico primeiro ciclo	2
Teses	7	Desportos	2	Exposições	2
Libretos	5	Física	2	Gestão	2
Ensaios	4	Formação profissional	2	Teoria	2
Música para piano	4	História	2	Trabalhadores	2
Bailados	4	Importância da adaptação psicossocial	2		

It is clear from this view of the topical terms that most are form or genre terms, and many are musical or dramatic terms, although a fair population of historical and legal indicators also is present. In an earlier analysis it was noted that individual terms were too infrequently present to provide statistically-significant correlations, but given that we can see from the data in Table 2 that certain domains are typically those where topical index terms are added, it might be possible to identify semantic clusters that could be used as pointers in those domains. Table 4 shows the clusters that appear in these three datasets.

Table 4. Semantic clusters among topical index terms

BND Livre	**BNP Catalogo**	**PORBASE**
Crimes--Sentenças--Portugal--…	1910-…	Influências …
Documentos administrativos--Portugal--…	Alunos …	Língua …
Gravuras	Desenvolvimento …	Música …
Instrumentos musicais de …	Estudo …	Séc. …
Missas--…	Influências …	Teoria …
Músic[a]/[os] ...	Importância …	
Música [impressa]/[manuscrita]	Músic[a]/[os] ...	

	Mulheres …	
	Organizaç* ...	
	Representações …	
	Séc. …	

These clusters show clearly the associations in the main classes from table 2 that are statistically-significant. Musical, dramatic, legal and political form and genre terms make up most of the semantic clusters that could be used as nodes for extending understanding of the UDC strings.

Finally, the complexity of the UDC strings was indicated by tallying the occurrence of auxiliary indicators "/," ""","" "()," "n=" and ":." "+" and "*" did not appear; "n=" was used only in the PORBASE dataset. Because there were one-to-many relationships between the UDC strings and the bibliographic records no correlations were drawn for this paper. Table 5 shows the comparative tallies.

Table 5. Auxiliary UDC usage

	BND Livre		**BNP Catalogo**		**BND Livre**	
	Maximum	**Mean**	**Maximum**	**Mean**	**Maximum**	**Mean**
/	36	4.7778	12	4.7778	11	3
()	276	83.1111	212	70.3333	107	34.1111
"	104	13.8889	58	9.1111	28	4.4444
n=	0	0	0	0	3	0.6667
:	3	0.4444	5	0.7778	10	2.2222

Conclusion

These results are consistent with earlier results (Smiraglia 2013, 2014a, 2014b). Taken together we can say there is a clear indication that elements of the KOS can be statistically associated with certain bibliographic characteristics of the documents they describe. More research using more granularity in the KOS should not only uncover more areas of association but also the specific "footprint" of these predictable associations. This is the main point of classification interaction: to generate heuristics for datamining using KOS strings.

Another clear conclusion comes from the value of the diversity of these datasets. They were diverse in several ways, including context (national library, national union catalog, digital library), and contents (much music and non-textual digital content), and in their advanced use of UDC for complex indexing. Results have extended our comprehension of classification interaction and also demonstrated the value of continued research in more and more diverse contexts.

Acknowledgment: The datasets discussed here were provided to the KSL team by Maria Inês Cordeiro, the director of the National Library of Portugal.

References

Akdag Salah, Gao, Cneg, Suchecki, Kryzstof, Scharnhorst, Andrea, & Smiraglia, Richard P.(2012). The Evolution of Classification Systems: Ontogeny of the UDC. In A. Neelameghan, & K.S. Raghavan eds. *Categories, contexts, and relations in knowledge organization: Proceedings of the Twelfth International ISKO Conference, 6-9 August 2012, Mysore.* Würzburg: Ergon. Pp. 51-7.

Smiraglia, Richard P. (2013). Big Classification: Using the Empirical Power of Classification Interaction. *Advances in classification research online* 24(1).

Smiraglia, Richard P. (2014a). Classification Interaction Demonstrated Empirically. In Wiesław Babik, ed. *Knowledge Organization in the 21st Century: Between Historical Patterns and Future Prospects: Proceedings of the Thirteenth International ISKO Conference 19-22 May 2014, Kraków*. Würzburg: Ergon. Pp. 176-83.

Smiraglia, Richard P. (2014b). Extending the Visualization of Classification Interaction with Semantic Associations. In *Advances in Classification Research Online*.

Smiraglia, Richard P., & Heuvel, Charles van den. (2013). Classifications and Concepts: Towards an Elementary Theory of Knowledge Interaction. *Journal of documentation* 69: 360-83.

Smiraglia, Richard P., Scharnhorst, Andrea, Salah, Almila A., & Gao, Chen (2013). UDC in action. In Slavic, Aida, Almila A. Slah, & Sylvie D. eds., *Classification and Visualization: Interfaces to Knowledge, Proceedings of the International UDC Seminar, 24-25 October 2013, The Hague*. Würzburg: Ergon. Pp. 259-72.

Smiraglia, Richard P., Heuvel, Charles van den,&Dousa, Thomas M. (2011). Interactions Between Elementary Structures in Universes of Knowledge. In Slavic, Aida & Civallero, Edgardo eds., *Classification & Ontology: Formal Approaches and Access to Knowledge: Proceedings of the International UDC Seminar 19-20 Sept. 2011, The Hague*. Würzburg: Ergon, 2011. Pp. 25-40.

Francisco-Javier García-Marco

The Interaction between the Systematic and Alphabetical Approaches to Knowledge Organization and Its Subjacent Mechanisms: a Long-term Primary Wave?

Abstract

A hypothetic basic wave affecting the history of knowledge organization (KO) is discussed from a cognitive and evolutionary perspective, exploring its logic and implications. This wave is provoked by the fluctuation of the knowledge organization effort throughout a knowledge space that has a natural tendency to growth, but is subjected to cognitive and technological limits. During more than two thousand years, the KO effort has fluctuated between two technologies invented by men for recovering knowledge records in retrieval spaces, e.g., logic-based classification and alphabetical ordering. Both have been used to engage the tendency of knowledge to grow continuously, stressing the organization of retrieval spaces and, therefore, their predictability. The proposed denomination of this wave is 'KO semiotic wave', because the KO effort fluctuates between both semiotic planes (signified and signifier). A preliminary list of its crests, troughs and transition points is provided for further discussion, which comes from two basic areas where an intense transdisciplinary KO effort has traditionally existed: encyclopaedias and document repositories.

Aims

The search for cycles is a constant in such different disciplines as physics, climatology, ecosystems, physiology, history, sociology, economy and many others.

The interest in cycles is easily understandable. Cycles are about recurrence, and therefore their detection allows for programmable, effective action. Once a cycle has been found and described, it is possible to make predictions from selected indicators, and to produce consequences by creating the appropriate conditions. So, mastering the basic cycles of a domain is key for any activity or science, practical (focused on applications) or theoretical (oriented towards explanation and prediction).

The existence, recurrence, characteristics and typology of cycles are usually the object of important controversies and debates. They also constitute the background for many important steps forward towards the construction of the scientific theories that define these disciplines. This kind of regularities can be detected either structurally (in space or form) or evolutionally (in time). Structural cycles are usually represented by closed circumferences (or circuits, if very complex). On the other hand, historic or evolutionary cycles are typically denoted by open "circumferences", unfolded along the line of time, e.g., waves. In waves, the values of some variables recur, but the context change, according to the flow of time.In this paper, we will try to show that one of such waves might exist in the field of knowledge organization (KO), at least for historical times 1, we explore its logic, and we discuss its implications for the research agenda of the discipline.

Usually, waves are the result of a source of energy impacting a medium that is flexible enough to let it cross through without provoking structural changes, but absorbing part of it.In particular, the potential KO wave that we are exploring would be

provoked by the way in which knowledge organization efforts fluctuates between two technologies invented by men for recovering knowledge records in retrieval spaces, which have a tendency to growth, disorganization and retrieval entropy. They have been used for thousands of years to engage the tendency of knowledge to grow continuously, stressing the organization of retrieval spaces and, therefore, their predictability. These two technologies are logic-based classification and alphabetical ordering.

Methods

Two main research methods have been used in this paper. On the one hand, a semantic analysis of the basic concepts in discussion was carried out with the aim of conceptually modelling the wave. On the other, a historical survey was pursued, looking for different potential examples of the wave's crests, troughs and optima.

The semantic analysis of the main concepts in question was carried out in five steps. Firstly, the main approaches to knowledge organization—conceptual and alphabetical—were analysed as technologies, that is, as inventions to improve a practical need: the recall of accumulated knowledge. Secondly, the distinction between natural and artificial, logic-based classifications was set up. Thirdly, the application of the alphabetical order to information storage and retrieval was reframed as a new KO paradigm, invented during the Hellenistic period. Fourthly, both systems were systematically compared and classified according to a set of relevant criteria, grounded on the theory of signs (particularly, on the distinction between signifier and signified, Saussure, 1916). Finally, the subjacent reasoning (cognitive) processes were identified: from signals to concepts (induction) and from concepts to signals (deduction).

Concerning the historical approach, a review of relevant historical events related to knowledge organization was carried out trying to find evidence to support the model, e. g., the great systematic compilations of the Antiquity and the Middle Ages, the move from classified encyclopaedias to alphabetical ones, Cutter's interconnection between the alphabetic catalogue and his bibliographic classification, the evolution of Encyclopaedia Britannica, the thesaurus synthesis, the directories v. search engines in the starting age of the Internet, and the evolution of Goggle and other competing modern search engines

The findings have been organized in four parts: an evolutionary perspective of KO, mainly regarding the realm of cognition and particularly memory (sections 3, 4, 5); the basic modeling of the wave (section 6); and the case studies (section 7).

Theoretical perspective: an evolutionary perspective of KO

This paper adopts a cultural, evolutionary approach (see Bates 2005; García & Esteban, 1993; García, 2010) towards the history of knowledge organization, more frequent in the field of knowledge management. In principle, knowledge allows human beings to better adapt to their environments, and even transform them or move to more

suitable ones. So, there is a clear incentive to codify, share and accumulate knowledge.

But knowledge is only useful if it can be memorized, e.g., stored and retrieved. A first stage in knowledge organization was the invention of new ways to relate concepts in more complex ways of representation: the signalization of their environment (transforming it in a kind of proto-document) and narrations. But, as traditional societies rest on inner memories to retain knowledge, there is a limit to their capacity.

The next stage came with the invention of external memories: documents. Documents are the result of a long process of miniaturizing 'semantized' environments; that is, of using the keys of physical spaces as keys to abstract concepts.

Documents successfully extend the limits of the human memory, but, after they accumulate up to a point, such accumulation also surpasses the recall capacity of the brain. If the retrieving space is small enough, the searcher can completely memorize it. But when it grows up to the point of surpassing the capacity of the memory, alternative approaches are needed 2. Artificial, non-natural systems of ordering had to be invented to store and retrieve the keys that lead to the units of knowledge recorded in the documents. So arises the third stage in knowledge organization in the modern sense of the world, bibliographic organization, setting on new waves of changes and adaptations.

So, historically, knowledge organization as a practice and later as a discipline is the result of the accumulative knowledge gained from the series of efforts carried on to enhance human memory in social environments through the invention of new technologies—both soft and hard—. In this way, knowledge organization can be viewed as the conscious and evolutionary effort of recording and storing knowledge in a way that it can be efficiently and effectively retrieved and reused, dealing with the entropy that results from the increase in the size and availability of knowledge that such KO effort produces: Initially, KO technologies allow for the growth of a society's cultural heritage; but after reaching the maximum complexity they are able to manage, they begin to fail and become challenged, and the need for new, more effective tools arise. So, each time that a more effective and efficient technology to achieve this purpose is found, accumulated and reusable knowledge can grow till the limits of the technology are reached, and the cycle is reproduced.

Systematic and alphabetical organization as invented 'technologies'

Conceptual creation, interrelation and specification are the natural ways in which humans develop the domain representations of their environment, and classification constitutes the backbone of the resultant concept maps (García-Marco & Esteban-Navarro, 1993), which allow not only for recognition but also for categorical inference. Scholars and philosophers, notably in the fields of logic and mathematics, improved natural classification, inventing the theory of concept, definitions and taxonomies, allowing for improved classification and more complex and rich knowledge storages 3.

But when conceptual domains grow in extension or complexity, and especially when

such an expansion is fast and produces cracks in the basics communal conceptual agreements, classification gets more complicate and increasingly idiosyncratic. In this way, it becomes a problem instead of a solution, and its efficacy is put in doubt until a new consensus is found.

Fortunately, around the Hellenistic period an innovative use of the method for learning phonetic writing systems (alphabets) was applied to the organized storage and retrieval of knowledge records: e.g., the alphabetic ordering, first documented in the Alexandria's Library (Daly, 1967). Thus, a complete new paradigm of knowledge organization was born, because alphabetic ordering does not require the mastering of the structure of concepts to retrieve information, but only knowing the term that express it. In addition to that, alphabetic ordering can be easily computed. Whereas concepts must be properly understood in order to be organized and searched, alphabet ordering can be delegated and eventually automated. Thus, ordering and retrieving become easier and cheaper, but there are also disadvantages: no map exists that communicates the extension, subdivisions and relations of the conceptual domain, and the possibility of categorial inference is lost.

The logic behind the complementarity of the conceptual and alphabetical approaches to KO

In human memory, information is recalled by two processes of opposite directions. One of them goes from the sensible information (the signal) to its meaning, and is eminently inductive. For example, if we see dark clouds, we assume there is a strong probability of rain. The other process goes from the concepts towards the signals, which is deductive. For example, if we are very bored, we think about possible things to do, and then we look for the entities that can be useful to fulfil these activities, for example a friend to talk with, a guitar to play music, a ball to play, etc. The first process is triggered by the form of the sign; the second is governed by its meaning, the conceptual plane.

From the perspective of retrieval, the most important aspect of these ordered spaces of recorded knowledge is their predictability (Fugmann, 1993). Humans have invented several technologies to make external cognitive spaces retrievable that are based in both processes, meaning that they support a set of operations that can be launched to fulfil a search in a way that is simple and efficient enough. In both systems—classificatory and alphabetic—there is a simple principle allowing a trained searcher to predict where an item should be in the ordered space.

Table 1. Pros and cons of the conceptual and alphabetical approaches to knowledge organization

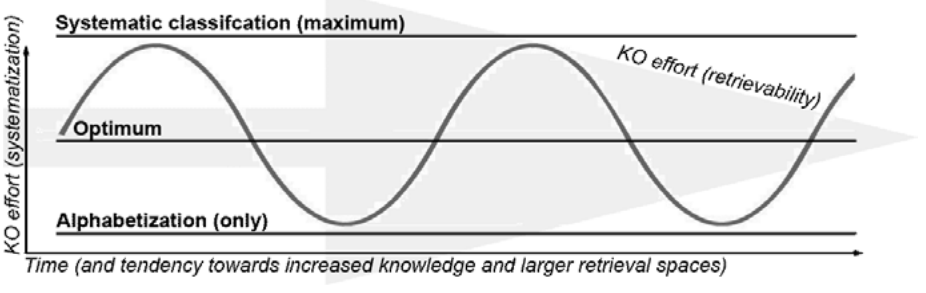
Dimension	*Plane*	*Dimensions*	*Ordering*	*Advantages*	*Disadvantages*
Conceptual plane	Signified	Multi-dimensional	Classification (inclusivity)	Map of knowledge, richness, multiple perspectives and alternative orders	Difficult, potentially idiosyncratic, complex computation
Form of the sign	Signifier	One-dimensional	Alphabetic	Conventional, easily computable, delegable	No explicit semantics, no alternative paths

Basic model of the 'KO semiotic wave'

So, both systems have advantages and disadvantages (Table I) and are mostly complementary. This is the reason why they have been interacting through history, although not linearly, but fluctuating between them, and, so, generating the cycles we are exploring. In its basic form, the cycle consists of five components:

- The energy source is the human work devoted to knowledge organization, which results in the commission of time and other human resources, but also in the development of technologies to improve such human efforts.
- The medium is the potential retrieval space that results from the interaction between the actual accumulation of knowledge (with its extension and granularity) and the potential demands posed over it.
- The upper line represents complete systematization of knowledge.
- The bottom line denotes simple alphabetization of the basic recall keys.
- The middle line represents the typical, functional, well-organized information systems, where knowledge is not completely systematized, but a good and practical balance is reached using both classification and alphabetization.

Figure 1. The model



Inside both limits, knowledge organization efforts fluctuate according to these three basic dynamics:

- When the classification tools become good enough, interaction and diversification become easier and the retrieval space can gain in extension and granularity. However, as a result, it finally becomes too dense for the previous

systematization tools and its effectiveness collapses, because they are too rigid or out-dated to deal with the new corpus of knowledge or promote its growing.

- When systematic schemes are unable to manage a system of knowledge that has become too complex or is growing too quickly, alphabetic indexes can provide a basic but effective access. However, when alphabetical tools grow very much, they finally reach a point in which the interconnection between knowledge units gets too scattered, and the system becomes entropic.
- The fluctuations have become more rapid as history has advanced and technological change has accelerated.

Case studies

In our historical survey, several preliminary case studies have been selected (García-Marco, 2016), that correspond to different momenta of the basic wave (Table 2).

Table 2. Prospective momenta of the KO waves

Age	*Knowledge device*	*Momentum*
Late classic Greek	Aristotle's synthesis	Crest
Hellenistic period	From encyclopaedic synthesis to alphabetic ordering	Trough
Late Antiquity	Saint Isidore's Etymologiae	Crest
High Middle Ages	Thomas Aquinas's Summa Theologiae	Crest
(Central) Modern period	Diderot's French Encyclopaedia, Enc.Britannica, etc.	Trough
1870s	Cutter's KOS integration	Optimum
1945	Vannevar Bush' hypertext asystematic approach	Trough
1953	Taube's uniterms	Trough
1970s	Thesauri	Optimum
1980s	Faceted KOS, etc.	Cresting
1984	Encyclopædia Britannica's Propaedia	Crest
1994	Yahoo	Crest
1996-1999	Google engine	Trough
2004-2012…	Google synonyms, Google Graph	Optimum
¿?	Advanced semantic web	Crest

From systematic synthesis to alphabetical encyclopaedias, and vice versa

After the application of alphabet to knowledge retrieval, alphabetic ordering was increasingly used to deal with the growth of knowledge stored in documents. But also, at the end of such periods, new systematic synthesis were developed, trying to pack 'all' the relevant knowledge of its age in an organized way. This is the case of the Aristotelian synthesis, Saint Isidore's Etymologiae and Saint Thomas Aquinas's Summa. On the contrary, the extraordinary growth of sciences and techniques during the Early Modern Age was dealt with the first alphabetically organized reference tools, epitomized by Diderot's French Encyclopaedia (alphabetic), as it happened with the application of the alphabetical ordering during the Hellenistic period.

Cutter's KOS integration

A highly successful attempt to interconnect the conceptual and alphabetic methods of access was achieved in the last third of the 19th century by Charles Ammi Cutter by linking bibliographic classifications with the alphabetical dictionary. However, the enumerative nature of the classifications of the time and the librarians' reliance on manual procedures made this attempt fail as the scientific and information explosion accelerated.

Encyclopædia Britannica

Encyclopædia Britannica (Glasgow, 2002; Auchter, 1999, Encyclopaedia… 2014) exemplifies the tensions between knowledge creation and organization extremely well (García-Marco, 2016): from its first edition of two volumes in 1768-71, it grew very quickly in size to the 21 volumes, and from the first index volume in 1830–1842 to the 28 volumes of the 11th edition in 1910-1911.The physical size of the encyclopaedia reached a limit —an infrastructural one, related to the costs of distribution and storage—. So, the expansion of knowledge during the 20th century had to be addressed by reducing the wording of entries and considerably increasing them, up to 700.00 in the printed version of 2007. This huge alphabetically ordered repository of knowledge required a classificatory device to facilitate the navigation by the users, and also to produce a more compact, less redundant content. Thus, the Propaedia outline of knowledge was introduced in the 15th edition of 1974-1984, after the great knowledge explosion of the world and cold wars. Finally, as a result of the digital information explosion, another way of compiling and organizing knowledge for reference was needed, both more compact and usable and, thanks to the hypertext model (Bush, 1945), Wikipedia took up the torch from the Britannica and other traditional encyclopaedias.

From descriptors to thesauri and faceted KOS

The big growth of knowledge during the World Wars period lead to great dissatisfaction with the existing knowledge organization tools. As a result, two alternative paradigms emerged: the alphabetical approach adopted by scientific and technical experts in documentation using the new ordering machines based on the Boolean logic, and the analytic approach to classification developed by some innovative librarians (Bliss, Ranganathan).

Both the alphabetic and faceted approaches were explored to their extremes very soon. In the alphabetic front, Taube (1953) proposed a system of indexing based on postcoordinable simple terms, which was highly criticised outside very highly specialized environments (Moers, 2003:818; Aitchinson & Clarke, 2004:7). In the analytic-synthetic side, hierarchical and associative relations were considerable expanded in some systems, as in PRECIS (Austin & Butcher, 1969). Both extremes ultimately failed, and a new consensus was reached around thesauri and faceted

classifications, which also interacted to a great extent. Dahlberg (1991, 1992, 1995, 2011) shows how dissatisfaction with simpler thesauri led to a renew interest in classification, giving birth to the field of knowledge organization.

The early KO efforts in the Internet and its evolution

In a way that resembles an accelerated recapitulation of the previous century of knowledge organization efforts, knowledge organization in the Internet started with hierarchical structures of concepts (FTP services, bulletin boards, Gopher services, Yahoo categories…), moved towards keyword access (Archie, Veronica, Jughead), increasingly after the web explosion with the aid of probabilistic models (web search engines), confronted better relevance and contextual input (Google), and is now struggling with a renewed need for carefully conceptual access and processing, so it can become a Semantic Web (Berners-Lee, Hendler & Lassila, 2001).

Google and its competing search engines: towards semantic searchers

Google and other search engines have become the gate to Internet, and thus to a fast-increasing part of the recorded knowledge. But they are used in the simplest possible way: users search only by those keywords that immediately come to their minds; do not try to find synonyms or related terms, to use commands or advanced features, to expand or refine their searches or examine metadata; and do not read but those results than are in the first page (some references in Garcia, 2016). This situation has posed a great challenge to search engines, which have had to take the semantic load over them.

So, after relaying on automatic indexing, vector and probabilistic models, and citation, big search engines have begun to incorporate an explicit semantic dimension to their search model in response to the ever-increasing size of the web and the related growing users' demand for greater precision, mainly by exploiting big knowledge repositories that have been formalized in "knowledge graphs". In particular, Google incorporated synonyms 'rings' to its latent semantic indexing model in February 2004 (Mooz, 2015), and in May 2012 began to implement of its own "Knowledge Graph" that codifies people, places, things, and the relations among them, remarkably presented as retrieving "things, not strings" (Singhal, 2012).

Conclusions and future lines of research

An evolutionary approach towards KO allows us to recognize knowledge as an adaptive advantage, whose proper storage and retrieval improves human life. This contributes to explain why human beings have so systematically devoted efforts to invent hard and soft technologies that can improve KO, allowing for increased knowledge accumulation up to their inherent limitations.

For thousands of years, inventing new concepts, setting environmental keys and organizing concepts in multi-level structures and narrations were the ways in which knowledge was 'naturally' codified and systematized, so it was retrievable. Around the

5th century B.C., classification was improved with the aid of logic, leading to the invention of concept theory, definitions and taxonomies. They helped knowledge to flourish during the Hellenistic period, making the previous Aristotelian synthesis unable to organize all the available knowledge. To deal with this overflowing knowledge environment, a new KO model was developed counting on a new 'soft' technology: alphabetic ordering. Since then, the wave has unfolded once and again till reaching our times: different efforts to systematize existing knowledge have been developed, only to be overcome by a new knowledge explosion that could at least be partially indexed so that it major part was retrievable. We have prospectively called this wave the 'KO semiotic wave", because the upper and bottom lines of the fluctuation correspond to the planes of the sign: signified and signifier, respectively.

The signifier is the easily manipulable, computable part of the sign, because it deals with its sensible and perceptual layers. However, as many theoreticians in the field have stated (Dahlberg, 1995), truly organization of knowledge occurs only in the layer of the signified, that is, concepts codified in a language (Saussure, 1916)—.On the other hand, the signified layer can be computed making explicit its relations to other signs, expressed by their signifiers, and this is just the approach that, in different ways, classifications, subject headings, thesauri, ontologies or latent semantic models take. So, both planes are important for an effective knowledge organization effort, though the importance of each pole in each historical moment is different, depending on the relative efficacy of the available KO technologies.

If the KO research community would consider such a historical cycle in KO to be an interesting hypothesis, much work would remain to be done. The survey of key events should be expanded, with the aim of producing a prospectively complete time-line. In this way, the cycles could be drawn with more precision, and more conclusions could be obtained. At this stage, the model has only a descriptive value, and no computation or numerical calculi can be drawn from it. Tentatively, it can be affirmed that the waves have shortened as a result of technological revolutions, showing accelerated fluctuations (Table II). Likewise, the survey should be expanded to non-textual records of knowledge, such as images, maps or music records. This is very important from a theoretical perspective, because it should help a better generalization of the dynamics behind and between signified- and signifier-based KO efforts and systems.

Also in the theoretical plane, at least two other fronts deserve attention. Being KOS invented technologies, theories of technological innovation developed in the fields of Archaeology and Engineering should be taken into account. Secondly, the model should be linked to other very important retrieval paradigms, particularly those regarding physical organization, temporal organization (records), the social division of cognitive work, intra- and inter-textual reference (including hypertext), and, more recently, the application of mathematics to the detection of latent semantics.

Acknowledgments: This research was funded by the project CSO2015-65448-R (MINECO/FEDER).

Notes

[1] In theory, if knowledge organization is defined as the effort to organize knowledge in a way that is retrievable, this effort is previous to historical times, when knowledge began to be recorded in documents, e.g., external memories.

[2] Another problem that is not being considered here is the reproduction of the knowledge system. New-comers and the new generations must be introduced to its use, which also poses a limit to its social efficacy.

[3] Classification and alphabetization are not the only KO 'soft' technologies that should be considered. There are others tools of great importance for organizing and recovering knowledge. One of them is organizing information by date, as in stories (in a single discourse) or records (several messages). Another is reference, relating two pieces of information that are not contiguous.

References

Aitchison, Jean & Clarke, Stella Dextre (2004). The thesaurus: a historical viewpoint, with a look to the future. In Roe, S. K., A. R. Thomas. *The Thesaurus: Review, Renaissance, and Revision*. NY: Haworth Press, 5-21.

Auchter, Dorothy (1999). The evolution of the Encyclopaedia Britannica: from the Macropaedia to Britannica Online. *Reference Services Review* 27(3): 291-299.

Austin, Derek W., &Butcher, Peter. (1969). *PRECIS: a rotated subject index system*. London, British National Bibliography.

Bates, Marcia J. (2005). Information and knowledge: an evolutionary framework for information science. *Information Research* 10(4): paper 239. [http://InformationR.net/ir/10-4/paper239.html]

Berners-Lee, Tim, Hendler, James & Lassila, ora. (2001). The semantic web. *Scientific American* 284(5): 76-88.

Bush, Vannevar (1992). As we may think. *Atlantic Monthly* 176 (July 1945): 101-8.

Dahlberg, Ingetraut (1991). Knowledge organization, thesauri, and terminology. *International Classification* 18(3): 133.

Dahlberg, Ingetraut (1992). Knowledge organization and terminology: philosophical and linguistic bases. *International Classification* 19(2): 65-71.

Dahlberg, Ingetraut (1995). Current trends in knowledge organization. *Organización del conocimiento en sistemas de información y documentación* 2: 7-26. [http://www.iskoiberico.org/wp-content/uploads/2014/07/007-026_Dahlberg.pdf].

Dahlberg, Ingetraut (2011). How to Improve ISKO's Standing: Ten Desiderata for Knowledge Organization. *Knowledge Organization* 38(1): 68-74.

Daly, Lloyd. W. (1967). *Contributions to a history of alphabetization in Antiquity and the Middle Ages*. Bruxelles: Latomus.

"*Encyclopaedia Britannica". World Heritage Encyclopedia*, 2014. [http://community.worldheritage.org/articles/Encyclopædia_Britannica].

Fugmann, Robert (1993). *Subject analysis and indexing: theoretical foundation and practical advice*. Frankfurt am Main: Indeks Verlag.

García-Marco, Francisco-Javier (2011). La pirámide de la información revisitada enriqueciendo el modelo desde la ciencia cognitiva". *El profesional de la información* 20(1).

García-Marco, Francisco-Javier (2016). The evolution of thesauri and the history of knowledge organization: between the sword of mapping knowledge and the wall of keeping it simple. *Brazilian Journal of Information Studies: Research Trends* 10(1): 1-11 [http://www.bjis.unesp.br/revistas/index.php/bjis/article/view/5786]

García Marco, Francisco Javier & Esteban Navarro, Miguel Angel (1993). On some Contributions of the Cognitive Sciences and Epistemology to a Theory of Classification. *Knowledge Organization* 20(3): 126-32.

Glasgow, Eric (2002). Scotland and the Encyclopaedia Britannica. *Library Review* 51(5): 263–267. [http://dx.doi.org/10.1108/00242530210428764]

Google (2012). Introducing the Knowledge Graph: things, not strings. *Google Official Blog*, May 16, 2012. [https://googleblog.blogspot.com.es/2012/05/introducing-knowledge-graph-things-not.html]

Mooers, Calvin N. (2003). Descriptors. In *Encyclopedia of Library and Information Science,* Marcel Dekker, Pp. 813-821. DOI: 10.1081/E-ELIS 120008981.

Mooz (2015). Google Algorithm Change History. *Learn Seo*. [https://moz.com/google-algorithm-change]

Saussure, Ferdinand de (1916). *Cours de linguistique générale*. Paris: Payot.

Singhal, Amit (2012). Introducing the Knowledge Graph: things, not strings. *Google Official Blog*, May 16, 2012. [https://googleblog.blogspot.com.es/2012/05/introducing-knowledge-graph-things-not.html]

Taube, Mortimer & associates (1953). *Studies in coordinate indexing*. Washington: Documentation Inc.

Inkyung Choi and Hur-Li Lee

A Keyword Analysis of User Studies in Knowledge Organization: The Emerging Framework

Abstract

The purpose of the paper is to examine how scholars in knowledge organization (KO) studied users of knowledge organization systems by analyzing the keywords used in the literature. In the study, two data sources in English over a ten-year period, from 2005 to 2014, were used: three KO-focused journals and KO dissertations from Canada and the United States. Through a quantitative text analysis, the study identified keywords from the titles and abstracts of the selected works. The authors then performed a qualitative content analysis on the derived keywords to formulate a framework of user studies. As intended, the emerging framework will contribute to an improved understanding of users and help identify gaps in KO user studies.

Introduction

Practitioners and theorists in knowledge organization (KO) have long maintained that information systems should give priority to users' needs. As one of the leading cataloging experts in the late 19th century put it, "The convenience of the public is always to be set before the ease of the cataloger" (Cutter, 1904, p. 6). To meet users' needs, it is sensible and necessary to base system design on insight gained through user studies. Accordingly, user studies have indeed been a significant part of KO research. The purpose of this paper is to examine how KO scholars researched users by analyzing the keywords used in the literature over a ten-year period, from 2005 to 2014. A major goal of the authors is to place the central concepts derived from the published literature into a framework for better understanding of KO user studies. The paper is intended to contribute to theory building in user studies, not only in KO but also in information science.

Literature Review

As in many other disciplines that have developed a user-centered approach, information science has conducted many user studies that reflect users' perspectives in recent years (see, for example, an overview by Nahl, 2003). The interest in user studies particularly has grown as improved technology allows general users to have direct access to information, which was previously restricted to experts only, making it a desired goal of information systems to be user-friendly and thus increasing the need for user studies. Bawden (2006) notes that a history of the study of users' information needs is not short. Yet, such studies often lack a clear foundation in methods and conceptual frameworks (Wilson, 1981 & 1997). Similar criticism has also appeared in KO (Hjørland, 2013). Within the domain of KO, user studies by and large examine users' interactions with one or more KO systems (KOSs), sometimes by themselves but more often within an information retrieval system such as a library catalog.

Catalog use

In KO, a big portion of the early user studies focus on use of the library catalog. Most of the works on catalog users before 2000 reach the same conclusion: users prefer a small learning curve and usually do not understand library and information science (LIS) systems (Larson, 1991; Lewis 1987). The major research method applied in those studies is transaction log analysis in a variety of forms since catalogs became automated (Cochrane and Markey, 1983; Seymour, 1991). While such studies of catalog use continued for a long time, Wilson (2014) more recently proposes to adopt an ethnographic approach to catalog user studies that will provide more in-depth understandings of users' knowledge, experience, and expectations.

Catalog use studies cover a wide range of issues that are both KO-related (e.g., use of the subject index in searching) and not KO-related (e.g., interaction with interface features and search strategies). Generally speaking, the tendency in such studies is to treat the catalog as an information retrieval tool, giving limited consideration to KO principles and KOS development.

Other information retrieval systems

A user-oriented approach has also been adopted for studies of other types of information retrieval systems. For example, Shiri and Revie (2006) investigate the query expansion behavior of end-users in interacting with a thesaurus-enhanced search system on the World Wide Web. Koch, Golub, and Ardö (2006) explore users' browsing behavior in a Web service that applied the *Dewey Decimal Classification* to organizing its information resources. Studies like these mainly focus on users' searching or browsing behavior in a service supported by a KOS.

KOSs

Comparatively speaking, user studies involving KO concepts and KOSs as the primary research objects appear to be a minority. One of the more notable projects has produced a series of works concerning users' input in the development of the Book House System for classification of fiction in Denmark (e.g., Pejtersen, 1980). Shiri and Revie (2005) carry out a user evaluation of a pilot terminology service to investigate users' thought processes, perceptions, and attitudes to inform development of a full service. Recently, Smith (2011) reviews research that targets a specific user group—i.e., health information consumers and patients—in studying their terminology use in searching.

Critiques

Hjørland (2013) examines the theoretical basis of user studies in KO, pointing out a lack of theory, questioning their usefulness for the development of the core principles of KO, and challenging the assumption of the desirability of user-based system design. Despite his reservations about the usefulness of user studies, he mentions a few

successful examples that reflect users' points of view in the design of KOSs that actually work, specifically detailing the case of the Book House System cited above. His major point, however, is that user studies conducted for the Book House System did not have a significant impact on the System's design or improvement as claimed by the system developer Annelise Mark Pejtersen. Citing the findings of a dissertation on the System (Eriksson, 2010), he suggests that Pejtersen's literary studies background contributed more to the quality of the System.

Another renowned KO scholar Francis Miksa holds a different view toward KO user studies who considers "user" one of the essential concepts in KO (2009). In recounting the conceptualization of information users and use in relation to KO over time in the West, Miksa declares the existing conceptions of users and information use to be mostly mysterious and under-defined. He also questions whether the concepts of users and use in past studies are still valid: 1) because of the Internet and World Wide Web, the purpose of information use is no longer to simply acquire the information objects; and, 2) information use in reality is not isolated as sequential occasions of searching for and using information objects as premised in most of the current models of information use. Miksa favors more contextual interpretations of the concepts of information users and use for future research.

Overall, user study has been a significant approach to research in KO. This approach and its resulting body of research, however, have received insufficient attention. Miksa's critical essay (2009) presents a challenge to KO. To move the field forward, it will be of great importance to pay due attention to users. The first step, we propose, is to better understand how users are conceptualized in the KO literature. Such understanding will in turn serve as an indispensable basis for theory development.

Methodology

The study intends to establish a framework of KO user studies through an analysis of the keywords used in KO literature, to understand how scholars have researched users and use of KOSs. Previously, Choi (2015) conducted a similar study to examine user studies published in the proceedings of four international conferences of the International Society for Knowledge Organization between 2006 and 2012. The current study turns to two other major sources in English: selected academic journals and doctoral dissertations published between 2005 and 2014.

Data sources

The study limits itself to the following data sources for analysis:

1) Academic journals: 10 years (2005-2014) of published research articles from three major journals in English that focus on KO research streams: *Knowledge Organization*, *Cataloging & Classification Quarterly*, and *Library Resources & Technical Services*. Non-research publications such as editorials, book reviews, journal updates, news, reports of events, interviews, and etc., were excluded.

2) Doctoral dissertations: 10 years (2005-2014) of doctoral dissertations in KO from 30 ischools in Canada and the United States, each of which contains "information science" as an assigned subject in the *ProQuest Dissertations & Theses Global*. Among them, 40 were determined to be KO-focused and thus included for the study. Their actual publication years range from 2008 to 2014.

Selection of user studies

From the collected research publications, we further selected user studies for analysis based on the two criteria established in Choi (2015). The first criterion is "perspective of the end user" and the other "instrument used."

a. End user perspective: a study examining the perspectives of end-users, as opposed to information professionals.
b. Instrument: a study collecting data directly from human participants by employing instruments such as survey, interview, focus-group, transnational logs, and experimentation. But the data does not necessarily reflect the perspectives of end-users. Information professionals and information institutions can also be data sources for evaluating or understanding KO systems.

All the works that met at least one of these two criteria form a general group (General) for analysis. On the other hand, a core set (Core) of user studies emerged from the intersection of the two criteria that concern both users' perspectives and use real data from end-users. In all, there are a total of 1,023 collected works, 41 of which (4.0%) are in the Core set of user studies (13 from KO, 18 from CCQ, 3 from LRTS, and 7 doctoral dissertations) and 117 (11.5%) in the General group (25 from KO, 53 from CCQ, 23 from LRTS, and 16 doctoral dissertations). The year 2011 shows the highest number of total user studies in publication (n=23) and the year 2009 has the highest ratio of the Core set to the General set (5:7).

Data analysis

This study consists of two stages of data analysis: a quantitative text analysis and a qualitative content analysis; the former for identifying keywords and the latter for constructing a conceptual framework to better understand KO user studies. The previous study by Choi (2015) identified the following four clusters of keywords by performing a quantitative co-word analysis on the selected ISKO conference papers on user studies in KO:

- Cluster 1: access, analysis, information, user, system and systems, and knowledge
- Cluster 2: approach, subject, results, search, users
- Cluster 3: classification, library, and online
- Cluster 4: retrieval, searching, thesauri, and thesaurus

In the first stage of the current study, the same text analysis was conducted. The tool used in this analysis was WordStat 6.0. It was the preferable tool because of its

dictionary feature that provides better control of terms. For example, it allows for multi-word terms such as "knowledge-organization" to be included. It also makes it possible to group similar words (e.g., words with the same stem). For example, the dictionary function can be customized to identify the words "search," "searches," and "searching" as one group and a single count of frequency can then be generated for all three words combined. For the text analysis, the titles and abstracts of the 117 studies in the General data set were included.

The second stage of data analysis was a qualitative content analysis. Through the use of important keywords identified in the text analysis, it attempted at placing users and the elements of KOSs into a useful framework for examining and understanding KO user studies. The researchers performed this analysis on the 41 articles in the Core set because these articles address end-user perspectives with user-engaged research design. To facilitate the analysis, the computer program Nvivo 11 was used because it enables the researchers to analyze textual data easily by generating nodes based on the manually coded data and visualizing the coding process and results.

Findings: Frequency and proximity analysis

All the keywords that showed a frequency count above 10 across the 117 studies were selected, with certain stopwords removed. Words with the same stem were merged; for example, the keyword cataloging includes catalog, cataloger, catalogers, catalogers', cataloging, and catalogs. This procedure produced a list of 68 keywords.

The top 30 keywords are: (1) KOS-related – library (227), cataloging (178), metadata (120), systems (73), tags (70), subject (58), terms (57), classification (54), academic (52), collections (50), and digital (48); (2) user-related – use (177), users (136), organizing (69), searching (68), process (54),and needs (50); (3) method-related terms – study (137), research (100), results (79), data (76), survey (70), examining (49), participate (47). * Frequency in parentheses.

Both use and users are on top as expected. Among the top keywords appearing more than 100 times, cataloging and metadata are essential KO concepts. Other KO concepts such as tags, organizing, subject, and classification are not far behind.

Proximity with the targeted terms, user and users, was calculated for the keywords on the frequency list (by Jaccard distance). The 10 keywords that were the most related to user/users are: searching (0.552), systems (0.432), tags (0.358), task (0.348), access (0.312), knowledge (0.291), information (0.289), terms (0.289), data (0.269), and cataloging (0.268). * Proximity in parentheses.

Many essential KO concepts such as tags appear on top. One noticeable difference between the two lists is that more method-related terms such as task, participants, interview, and approach come up on top 30 keywords on the second list, demonstrating their closer relationship to user/users. Also, the second list shows knowledge as the sixth most related keywords to user/users, while the previous list for word frequencies has knowledge at twenty ninth.

Generation and refinement of codes

The researchers first created a provisional start list of codes that contained four initial codes: environment, KOS, user, and user interaction with KOS. We then added more codes indicating specific examples or refined concepts

Applying all the codes, one of the researchers acted as the coder and followed the three-step procedure borrowed from Mayring's qualitative content analysis (Flick, 2014, p. 430) to code individual keywords from the above-mentioned quantitative text analysis in an attempt to establish an internal structure for these key concepts. First, we consolidated the list of keywords by deleting several terms that are synonyms or near-synonyms (i.e., summarizing content analysis). Next, the coder identified the contexts of use of individual terms in the data (i.e., explicative content analysis). In view of the identified contexts, new codes were then generated to overarch the terms sharing similar contexts. At last, all the revised codes were clustered under 6 main codes (i.e., structuring content analysis). Among the 6 main codes, three—Context of KOS, KOS type, KOS functionality—are associated with KOSs, and two others—User and User interaction with KOS—are related to users. The last one—User study implementation—concerns research itself. Table 1 shows all codes including the main codes and sub-codes in the finalized structure.

Table 1. Code report

Code	Description	Sub code	Notes
Contexts of KOS	Where KOS is used - e.g. library or web - but not limited to spaces. This code also includes situational use of KOS - e.g. multilingual or collaborative research	*Digital library*	
		Library	Excludes digital library
		Multilingual	
		Online	
		Scientific Communication	KOS in research setting, especially with the focus on sharing research data or related information
KOS type	KOSs for description, categorization, and vocabulary control. Retrieval tools supported by KOSs	*Categorization - Classification*	
		Categorization - Taxonomies	
		Controlled Vocabulary - Ontologies	
		Controlled Vocabulary - subject headings	
		Controlled Vocabulary - thesaurus	
		Controlled Vocabulary - unspecified	Indicate controlled vocabulary explicitly but not fully specified in the abstract
		Description - Metadata	
		Retrieval - Catalogue	
		Uncontrolled	

		Vocabulary - Tagging	
		Uncontrolled Vocabulary - unspecified	Indicate uncontrolled vocabulary explicitly but not fully specified in the abstract
KOS functionality	Whatever the KOSs are, their functionalities or purposes are not exclusive but inclusive. So they are multi-functional - retrieving, describing, and categorizing. And the research context or research goal indicates what functionality of KOS is examined in the use of the KOS. For example, tags are considered a retrieval tool or organizing tool	*Categorizing*	
		Describing	
		Retrieving	
USER	User as the subjects to be studied or as study participants	*User group - disciplinary or institutional*	
		User group - use of system or materials	User groups in the study are engaged in use of a certain system or materials, with no focus on specific backgrounds of the user group
USER interaction with KOS	How users interact with the KOS. External interaction is observable user activities in using KOS Internal interaction is users' mental processes which influence users' activities in using KOS	*External interaction - Browse*	
		External interaction - Organize	
		External interaction - Search	
		External interaction - general	This sub code does not specify the behavior or activities listed above but looks into some practices or activities of users in use of KOSs
		Internal interaction - needs	
		Internal interaction - perception	Users' understanding, attitudes, or conceptualization of the systems and system uses– e.g. mental models of web users, user comprehension of thesauri, etc.
User study implementation	Concerning research itself	*Design*	
		Evaluation	

In the process, we adjusted some of the initial codes, all informed by the data. Two of the significant refinements are described below.

KOS type and KOS functionality: There are many types of KOSs as subjects in user studies. At first they were classified into four groups: categorization, description, controlled vocabulary, and uncontrolled vocabulary. The library catalog, though not a KOS in the strict sense, was placed here as the fifth group: retrieval tool supported by major KOSs. Taking a closer look at the data, we recognized the need for a new category of KOS functionality, for each type of KOS serves multiple functions that have been examined separately by various studies. For example, researchers might examine how users use a KOS for its categorization function even though it is intended to be a descriptive type of KOS.

User interaction, External and Internal: Studying users' interactions with KOSs is not limited to users' external activities. Perceptions and comprehension of users, for example, are not explicitly manifested in user activities but influence what users do in their use of KOSs. Considering many LIS user studies have focused on searching and retrieval (i.e., external activities), studies of users' internal interactions with KOSs are of great importance. We thus clarified the main code of User interaction to explicitly note both the external as well as the internal interactions between the users and KOSs.

Conclusion

As shown above, over 11% of the research in the three English KO journals and doctoral dissertations during the past decade were user studies, attesting to a vibrant research area. The study analyzed the keywords derived from the titles and abstracts of the selected works to formulate a framework for a systematic understanding of the approaches to researching users and their interactions with KO systems. The emerging framework is also useful for identifying the gaps and weaknesses in this body of literature. In these regards, the study moved the field a step forward to a better conceptualization (or, as suggested by Miksa, a reconceptualization) of users in KO, which will no doubt lead to improved system design and service.

In future research, the developed coding scheme and framework of KO user studies will be tested by 1) exploration of correlations among main codes and 2) application of the coding scheme as indexing keywords for subsequent publications of user studies.

References

Bawden, David (2006). Users, User Studies and Human Information Behaviour: A Three-Decade Perspective on Tom Wilson's "On User Studies and Information Needs". *Journal of Documentation* 62(6): 671-9.

Choi, Inkyung (2015). Is User Studies User-oriented? Domain analytic approach to User Studies in Information Organization. *iConference 2015 Proceedings*.

Cochrane, Pauline A., & Karen Markey (1983). Catalog Use Studies--Since the Introduction of Online Interactive Catalogs: Impact on Design for Subject Access. *Library and Information Science Research*, 5(4): 337-63.

Cutter, Charles A. (1904). *Rules for a Printed Dictionary Catalog*. 4th ed. Washington, D.C.: U.S. Government Printing Office.

Eriksson, Rune (2010). *Klassifikation og indeksering af skønlitteratur – et teoretisk og historisk perspektiv*. Ph.D. dissertation. Copenhagen: Royal School of Library and Information Science.

Flick, Uwe (2014). *An introduction to qualitative research*. Los Angeles: SAGE.

Hjørland, Birger (2013). User-based and Cognitive Approaches to Knowledge Organization: A Theoretical Analysis of the Research Literature. *Knowledge Organization* 40(1): 11-27.

Koch, Traugott, Golub, Koraljka, & Ardö, Anders (2006). Users browsing behaviour in a DDC-based web service: a log analysis. *Cataloging & classification quarterly* 42(3-4): 163-86.

Larson, Ray R. (1991). The decline of subject searching: Long-term trends and patterns of index use in an online catalog. *JASIS* 42(3): 197-215.

Lewis, David W. (1987). Research on the use of online catalogs and its implications for library practice. *Journal of academic librarianship* 13(3): 152-7.

Miksa, Francis (2009). Information organization and the mysterious information user. *Libraries & the cultural record* 44(3): 343-70.

Miles, Matthew. B., Huberman, A. Michael, & Saldana, Johnny (2013). *Qualitative data analysis: A methods sourcebook*. SAGE Publications, Incorporated.

Nahl, Diane (2003). The user-centered revolution: Complexity in information behavior. In *Encyclopedia of Library and Information Science Online Second Edition*, edited by Miriam Drake. Florida: CRC Press. Pp 3028-42.

Pejtersen, Annelise Mark (1980). Design of a classification scheme for fiction based on an analysis of actual user-librarian communication, and use of the scheme for control of librarians search strategies. In *Theory and application of information research*, edited by Ole Harbo & Leif Kajberg. London: Mansell.

Seymour, Sharon (1991). Online Public Access Catalog User Studies: A Review of Research Methodologies, March 1986-November 1989. *Library and Information Science Research* 13(2): 89-102.

Shiri, Ali, & Crawford Revie (2006). Query expansion behavior within a thesaurus-enhanced search environment: A user-centered evaluation. *Journal of the American Society for Information Science and Technology* 57(4): 462-78.

Shiri, Ali, & Crawford Revie (2005). Usability and user perceptions of a thesaurus-enhanced search interface. *Journal of Documentation* 61(5): 640-56.

Smith, Catherine A. (2011) Consumer language, patient language, and thesauri: a review of the literature. *Journal of the Medical Library Association: JMLA,* 99(2): 135-44.

Wilson, Thomas Daniel (1997). Information Behaviour: An Interdisciplinary Perspective. *Information Processing & Management* 33(4): 551-72.

Wilson, Tom D. (1981). On User Studies and Information Needs. *Journal of documentation* 37(1): 3-15.

Wilson, Victoria (2014). Catalog Users "In the Wild": The Potential of an Ethnographic Approach to Studies of Library Catalogs and Their Users. *Cataloging & Classification Quarterly* 53(2): 190-213.

Ann M. Graf

Describing an Outsider Art Movement from Within: The AAT and Graffiti Art

Abstract

Knowledge organization is the study of the order, whether natural or imposed, of knowledge. As researchers in this field of science have increasingly acknowledged the importance of different epistemologies, or ways of knowing, that merit not only acceptance but investigation, I have chosen to examine how a particular artistic community describes their processes and products via historical discourse found in graffiti zines from the mid-1980s to 1990s in comparison with the overarching art community discourse as evidenced by a popular controlled vocabulary. The focus of this research project is to examine the sufficiency of vocabulary contained within the Getty Art and Architecture Thesaurus (AAT) for use in representing concepts from the graffiti art movement.

Introduction

Graffiti art has been studied from a number of perspectives, citing the movement into criminal justice, sociology, history, and art (Ferrell, 1993; Lachmann, 1988; Forster et al., 2012; Masilamani, 2008). The works themselves have value to researchers from all of these areas of study and as a global art movement and ever expanding online archive of artistic output, it behooves library and information science to be aware of the processes in place to collect, organize, and access these uncoordinated collections. As precursor to a larger examination by the author, this domain analytic study will advance discussions on the epistemological construction of an art community and the organic knowledge organization revealed in the social construction of terminology by its members as evidenced in the selected graffiti zines.

Major historical art movements are represented with descriptive terminology available in large structured vocabularies such as the Getty Research Institute's Art and Architecture Thesaurus (AAT). Participants in newer, smaller, or outsider art movements may not find terminology in such readily available vocabularies to represent concepts, contexts, and methods commonly used by their community of practice.

Sometimes referred to as graffiti art, the movement began in the late 1960s in Philadelphia and New York with the proliferation of graffiti tagger names in concentrated areas of these cities and quickly spread and evolved into larger, more detailed and artistic renderings on city walls, subways, and the ubiquitous train cars (Austin, 2001; Castleman, 1982). As the movement spread across the country, several graffiti magazines, or zines, began to appear. These were originally photocopied sheets of photos of graffiti art compiled by graffiti artists themselves, sometimes with artist or crew names and general locations of works contained therein. Some of the popular zines became more like modern magazines, printing in color and accepting subscriptions from around the country and even overseas. As the Internet developed,

most of these zines moved online and there is now a burgeoning number of websites, blogs, and social media feeds such as those on Flickr and Instagram that feature works by the graffiti art community. Due to the ephemeral nature of graffiti, photography is the most currently reliable, albeit not perfect, means to preserve a record of these works (Wacławek, 2011).

The AAT

The AAT is a structured vocabulary "that can be used to improve access to information about art, architecture, and material culture" (About the AAT, 2015). The resource began in the late 1970s as art libraries and art journals were looking for ways to index and describe their collections in the face of new computer cataloging technologies. The AAT has been developed by submissions from numerous sources and thus is a collaborative project, evolving and continuously expanding.

> Its scope includes terminology needed to catalog and retrieve information about the visual arts and architecture; it is constructed using national and international standards for thesaurus construction; it was initially a hierarchy inspired by the tree structures of MeSH (Medical Subject Headings Thesaurus); it is based on terminology that is current, warranted for use by authoritative literary sources, and validated by use in the scholarly art and architectural history community; and it is compiled and edited in response to the needs of the user community. … The AAT is a hierarchical database; its trees branch from a root called Top of the AAT hierarchies (Subject_ID: 300000000). There may be multiple broader contexts, making AAT polyhierarchical (About the AAT 2015).

This study follows in the tradition of a postmodern conception of knowledge organization (Mai, 1999) and is built upon domain analytic methods as introduced by Hjørland and Albrechtsen (1995) and Hjørland (2002), and expanded upon by Smiraglia (2015) and Smiraglia and Lee (2012), among others. While studies combining KO and graffiti art are rare, Gottlieb (2008) used a modified Delphi questionnaire method with 11 graffiti experts to develop a classification of graffiti art styles for use by image catalogers. Her research resulted in 14 style categories, 14 facets for each style, and additional foci for each facet. Ørom (2003) examined knowledge organization systems in the domain of art studies and suggested that newer art historical paradigms, which often cross traditional domains, might be well served by the polyhierarchical structure of the AAT (2003). Due to this structure, the AAT can be expanded for new paradigms, methods, and styles more easily than monohierarchical systems. With this optimism in mind, and because of its popularity and ease of use, the AAT was chosen for this comparative examination.

Methods

I examine terminology relating to the graffiti art movement through textual analysis of a series of three graffiti zines, International Graffiti Times, Can Control, and Flashbacks. These three zines were chosen from a number that were available to me through the generosity of Dr. Joe Austin at the University of Wisconsin – Milwaukee. After discussing my research proposal with him, he offered me, from within his large

graffiti zine collection, a sampling of titles that are best known, respected, and long-lived. From within this group of about a dozen titles, I chose three of which he has the most complete runs. These zines contain numerous photographs of graffiti art, sometimes with additional information about the individual photos, as well as essays, reviews, and interviews with graffiti artists. Text from the zines was transcribed into a Microsoft Word document so that it could be more easily manipulated. This transcription process resulted in a document 124,443 words in length.

Graffiti art descriptive terminology related to processes, products, and style was extracted manually and entered into an Excel spreadsheet. From there terms were normalized to account for varied spellings, misspellings, casual variants, and pluralization. Terminological preference was given to noun forms over verbs and adjectives. For example, the term piece was most often used as a noun, sometimes in gerund form (piecing), but it was also used as a verb (to piece) or as a past participle (pieced). Examples of normalization can be seen in Table 1, the first line representing the normalized term used for each group of variant terms.

Table 1. Examples of term normalization

end to end	*top to bottom*	*throw-up*
E to E	T to B	throw up
E-E	t-b	throw ups
end 2 end	top 2 bottom	throwups
E2E	T2B	thro-ups
ends to end	top-to-bottoms	throw-ups
E-to-E	T-to-B	

Once a typology of the most commonly appearing terms from the zines was developed, comparison was made between the terms and the available terminologies that might be used to represent them from within the AAT.

Results

After normalization a list of terms resulted that could be sorted according to frequency. This can be seen in Table 2, below. The most commonly occurring term, unsurprisingly, is *graffiti*. This term appeared four times more often than the second most commonly occurring term, *piece*. Words that occurred at least ten times or more were kept for comparison with terms from the AAT.

Table 2: Most commonly occurring terms and their frequencies

graffiti	741	piecing	25
piece	185	wildstyle	23
bomb	79	burner	17
throw-up	41	graffiti art	16

whole car	38	end to end	15
aerosol	36	blackbook	14
character	35	insides	14
spray paint	35	subway art	13
mural	30	aerosol art	11
top to bottom	29	production	11

To better understand how well search results for these terms in the AAT correspond to the actual meanings of the zine terms, summary definitions for the zine terms are provided in Table 3.

Table 3: Graffiti zine terms defined

Graffiti: "Typically refers to words, figures, and images that have been written, drawn and/or painted on, and/or etched into or on surfaces where the owner of the property has NOT given permission" (Ross, 2016, 476).
Piece: "(short for 'masterpieces') Large, colorful, elaborate, detailed, and stylistically intricate rendering of letters and images. Pieces require a greater amount of time and expertise to create than 'throw-ups' and 'tags'. (Usually deserving of more respect from other graffiti artists/writers)" (Ross, 2016, 477).
Bombing: "The prolific writing of one's tag [chosen name]. Bombing usually involves saturating a given area with a large number of one's 'tags' and/or 'throw-ups'. Often regarded as an important avenue for achieving recognition among other graffiti writers" (Ross, 2016, 475).
Throw-up: "(also known as throwies) … Produced with spray paint, throwies spell out a graffiti writer's name in bubble-style letters. These letters are usually produced and filled in quickly with a single color, and then outlined with a second color of paint. Throwies may also be done with a single can of paint, in which case the graffiti writer will produce a quick series of letters. In the more recent history of graffiti, throwies have increasingly come to be recognized as a distinct and valuable part of a graffiti writer's repertoire, often leading to the production of multi-colored throw-ups. Unlike masterpieces, throw-ups allow graffiti writers to cover more surface area relatively quickly" (Ross, 2016, 478).
Whole car: A large piece that covers an entire train car. This references the size of the piece and is related to 'end to ends', 'top to bottoms', and 'window-downs'. (Snyder, 2009)
Aerosol: Aerosol can refer to spray paint (see below) or it can be used instead of the word graffiti, as in an aerosol artist or an aerosol artwork. While a writer or artist may refer to a piece as aerosol art instead of graffiti art, graffiti art remains illegal, while aerosol art could be carried out legally on a canvas or other legal surface.
Character: "A term used to describe pictorial elements of graffiti works, especially renditions of creatures or personas. Characters are often used in conjunction with elaborate pieces of a graffiti writer's name/tag, and often incorporate gestures that draw the viewer's attention to the name" (Ross, 2016, 475).
Spray paint: Paint in a can that is applied using internal pressure and aerosol spray caps of varying sizes to change how it behaves when leaving the can.
Mural: "Large paintings on walls, sides of buildings etc. where the artist/s have been given express permission by the owner, and/or has been commissioned to do the piece (e.g. the work of Diego Rivera). Often depicting historical and/or religious events, themes, individuals, etc." (Ross, 2016, 477).
Top to bottom: A top to bottom (T-B, T2B, T-to-B) is a piece that covers a train car from top to bottom. (Snyder, 2009)
Piecing: Piecing refers to the making of pieces, or "masterpieces." See the definition for *piece* above.
Wildstyle: "Energetic pieces of graffiti with interlocking, highly stylized and often cryptic lettering" (Ross, 2016, 479).
Burner: "A graffiti piece that is regarded as high quality. To 'burn' is to outdo the work of others" (Ross,

2016, 475).
Graffiti art: "Graffiti art is a face-to-face, social practice with clear aesthetic intentions and unlike traditional graffiti, the semantic content of graffiti art is secondary to its visual aspirations. The identity of the individual (name and/or signature) is a crucial component of both, but graffiti art developed and is practiced collectively within skilled, locally organized subcultures" (Austin, 2010, 35).
End to end: An end to end (E-E, E2E, E-to-E) is a piece covering a train car from one end to the other. (Snyder, 2009)
Black book: "Writers carry sketchbooks that they call blackbooks which they use to practice outlines and to get autographs from other writers" (Snyder, 2016, 211n3).
Insides: The insides of subway trains. Graffiti artists can paint insides or outsides. There are many different ways to describe outsides, but insides are not commonly places to bomb or to piece, but rather to simply tag, which is to quickly write one's stylized name, usually in black marker. Graffiti writers speak of doing *insides* or *outsides* as a type of work.
Subway art: Another way of referring to graffiti art that was typically practiced on the subway cars in New York City in the late 1960s to 1980s.
Aerosol art: Graffiti is sometimes referred to as *aerosol art*, but this term was not common in the zines until issue 10 of *IGT*, when the editor explicitly stated disdain for the term *graffiti* and began to use *aerosol art* instead almost exclusively in all issues going forward (IGT 10, 1988). The term was used in *IGT* as aerosol art, aerosol archives, and aerosol artists. Interestingly, the title *IGT* formerly stood for *International Graffiti Times*, but by issue 8 the G in the acronym appears to have changed from graffiti to "Get Hip".
Production: These are larger and more involved pieces that involve several artists (often from the same crew) to work together. They are done on legal walls, where permission has been granted or the work commissioned. They require a larger amount of time, supplies, and people, all of which are prohibitive without permission. (Snyder, 2009)

Using these definitions, meaningful comparisons can be made between them and matching or related terms from the AAT. Results from this comparison are presented in Table 4. The original zine term is given first, then the term match from the AAT, as well as the name of the facet and the hierarchy under which the AAT term is found. Results that are shaded are those that provided a match or related term, but that were either not sufficient to describe the zine term or were completely different in meaning. NR indicates that although a result was found in the AAT, it was not related to the zine term meaning. A dash indicates that no result was found for the zine term in the AAT.

Table 4: Results of zine term search in the AAT

Zine terms	AAT	facet	hierarchy name
graffiti	graffiti	objects	visual and verbal communication
	graffiti artists	agents	people
	subway graffiti	object	visual and verbal communication
piece	NR		
bomb/bombing	NR		
throw-up	throw up [book binding action]	activities	processes and techniques
whole car	railroad cars (subdivides into freight cars, passenger cars)	objects	furnishings and equipment
Aerosol	aerosol	materials	materials

character	NR		
spray paint	--		
mural	mural painting (image-making)	activities	processes and techniques
	mural paintings (visual works)		
top to bottom	--		
piecing	piecing [quilting]	activities	processes and techniques
wildstyle	--		
burner	--		
graffiti art	--		
end to end	--		
black book	black books (graffiti)	objects	visual and verbal communication
insides	interior	physical attributes	attributes and properties
subway art	subway cars	objects	furnishings and equipment
aerosol art	--		
production	working drawings	objects	visual and verbal communication

This comparison resulted in three term matches: *graffiti*, *mural*, and *black book*. While *graffiti artists* were found in the AAT, graffiti itself was not referred to as *graffiti art*. *Subway art* is not found in the AAT, but *subway graffiti*is, as a sub-category of *graffiti*. The term *aerosol* is found in the zines and in the AAT, though the meaning is different. In the AAT *aerosol* is a material only, not a way to describe graffiti or graffiti art as it is in the zines.

It is really no surprise to find *mural* in the AAT as this is a recognized art term in common usage, associated with specific artists such as Diego Riviera, for example. What is notable is the inclusion in the AAT of the graffiti artist's *black book*, something that is basically a sketchbook but specifically so named by graffiti artists for their purposes.

Limitations

Graffiti zines were chosen for use as data in this research because of their place in the early history of the graffiti art movement. As a written record by and for graffiti artists, they document the language used by those creating graffiti art *themselves*, such language not being accessible by looking at the artworks or photographs of the artworks. While the knowledge contained in the zines is rich, it must be noted that it is from a specific era of the movement's development. The zines used in this study were from the 1980s and 1990s, but the art movement has continued to grow and develop. Further study is needed to examine the evolution of terminology to the present.

Knowing how graffiti artists talk about and describe their artworks and artistic practices is important to inform how systems for the organization of art and cultural objects make representations of the artworks available for further study and appreciation, yet it is not the only input upon which such systems can, and probably should, be created. This is a highly contested issue, that of who gets to decide how a

movement is described – those within it, those studying it, those in power, or a combination of all of these and more. This study has examined only one small part of this puzzle in demonstrating the lack of inclusion of one relatively recent artistic subculture's vocabulary within the AAT.

Summary

The results of this study show that, other than a few very general concepts, the terminology of the graffiti art community is not well represented by what is available in the AAT. The AAT is well suited to represent mainstream and traditional historical art movements, but does not include most terminology that is used by the graffiti art community to describe their own work. Reasons for this gap in AAT terminology may be due to the relatively young age of the graffiti art movement, to the well documented rejection of mainstream institutional art ideology by graffiti art community members, or a combination of both. Without further study it is not easy to say what users of graffiti art collections need from them, but by looking at what facets of information are offered by the zines, a picture of community needs develops based on these facets.

There is a need to be aware of newer movements in not only art, but other various aspects of a constantly evolving society. Especially considering how quickly artistic, political, and social movements now spread across virtual space and thus the world, continuing research of knowledge specialists must consider not only these issues and how to meet the demands of varied users to access knowledge of all types, but also to develop ways to change and/or expand subject vocabularies in faster and more flexible ways to meet these needs. The graffiti art movement is an important example of a subdomain of art that can be shown to warrant further research into the epistemological dimensions of KO.

References

Austin, Joe (2001). *Taking the train: How graffiti art became an urban crisis in New York City*. New York: Columbia University Press.

Austin, Joe (2010). More to see than a canvas in a white cube: For an art in the streets. *City*, 14(1/2): 32-47.

Can Control (1990-2000). North Hollywood, CA: Ghetto Art Productions.

Castleman, Craig (1982). *Getting up: Subway graffiti in New York*. Cambridge: MIT Press.

Ferrell, Jeff (1993). *Crimes of style: Urban graffiti and the politics of criminality*. New York: Garland.

Flashbacks (1991-1996). New York: John Edwards.

Forster, Alan M., Vettese-Forster, Samantha & Borland, John (2012). Evaluating the cultural significance of historic graffiti. *Structural Survey* 30(1): 43-64.

Gottlieb, Lisa (2008). *Graffiti art styles: A classification system and theoretical analysis*. Jefferson, NC: McFarland & Company, Inc.

Hjørland, Birger, & Albrechtsen, Hanne (1995). Toward a new horizon in information science: domain-analysis. *Journal of the American Society for Information Science* 46(6): 422-62.

Hjørland, Birger (2002). Domain analysis in information science: Eleven approaches-traditional as well as innovative. *Journal of documentation* 58(4): 422-62.

The J. Paul Getty Trust. (2015). *About the AAT*. [http://www.getty.edu/research/tools/vocabularies/aat/about.html]

The J. Paul Getty Trust (2015). *The Art & Architecture Thesaurus Online*. [http://www.getty.edu/research/tools/vocabularies/aat/index.html]

Lachmann, Richard (1988). Graffiti as career and ideology. *American Journal of Sociology*, 94(2): 229-50.

Mai, Jens-Erik (1999). A postmodern theory of knowledge organization. In *Knowledge: Creation, Organization and Use. Proceedings of the ASIS Annual Meeting*, 36: 547-56.

Masilamani, Rachel (2008). Documenting illegal art: Collaborative software, online environments and New York City's 1970s and 1980s graffiti art movement. *Art Documentation* 27(2): 4-14.

Ørom, Anders (2003). Knowledge organization in the domain of art studies: History, transition and conceptual changes. *Knowledge Organization*, 30(3/4): 128-43.

Ross, Jeffrey Ian (2016). *Routledge handbook of graffiti and street art.* New York: Routledge.

Schmidlapp, David (1984-1993). *International Graffiti Times*. New York: IGTimes.

Smiraglia, Richard P. (2015). *Domain analysis for knowledge organization: Tools for ontology extraction*. Waldham, MA: Chandos.

Smiraglia, Richard P., & Lee, Hur-Li (eds.) (2012). *Cultural frames of knowledge*. Würzburg: Ergon Verlag.

Snyder, Gregory J. (2009). *Graffiti lives: Beyond the tag in New York's urban underground*. New York: New York University Press.

Wacławek, Anna (2011). *Graffiti and street art*. London: Thames & Hudson.

Michael Kleineberg

Integral Methodological Pluralism: An Organizing Principle for Method Classification

Abstract

In indexing theory, a pragmatic turn has taken place, emphasizing the context of meaning production. This demands multi-perspectival knowledge organization systems in order to cope with the challenge of epistemic pluralism. This paper is concerned with the methodological dimension of human knowledge, including epistemic activities such as applied methods and techniques that are grounded in broader methodologies or foundational paradigms. An expressive cataloging or indexing of methods requires a systematic organization beyond a merely inductively derived listing of common research procedures and practices. Therefore, integral methodological pluralism (IMP) based on integral theory and deduced from fundamental formal-pragmatic distinctions will be proposed as an organizing principle for a classification of methods.

Introduction

In semiotic terms, bibliographic records or resource descriptions are traditionally limited to either descriptive cataloging or indexing based on syntactics (information on author, title, publisher etc. as mere characters regardless of their meaning) or subject cataloging or indexing based on semantics (information on the aboutness or meaning of a document), whereas the field of pragmatics (information about the context of meaning production) has been largely neglected (Kleineberg 2013). In indexing theory, however, a pragmatic turn has taken place, paying more attention to the analysis of underlying epistemic frameworks and activities (Weinberg 1988; Frohmann 1990; Tibbo 1994; Bies 1995; Hjørland 1997; Jacob 2000; Andersen & Christensen 2001; Szostak 2004; Mai 2005; Biagetti 2006; ISKO Italy 2007; Szostak, Gnoli & López-Huertas 2016). Now it is widely accepted that under the condition of epistemic pluralism, the desideratum of a "multi-perspective knowledge organization" (Kaipainen & Hautamäki 2011, 509) needs to be addressed. As argued elsewhere, this task would require a formal indexing of context including at least both viewpoint indexing (theory) as well as method indexing (praxis) based on adequate organizing principles (Kleineberg 2013).

This contribution is concerned with the method aspect (for a twin paper dealing with the viewpoint aspect see Kleineberg 2014) and advocates that an inductively derived listing of common procedures and techniques should be succeeded by a systematic organization based on their interrelations in order to provide a more expressive indexing of methods. Therefore, integral methodological pluralism will be introduced as a basic schema of primordial and irreducible perspectives or methodologies, something that Brier (2000, 438) would call a "meta-frame for qualitatively different types of knowledge."

Classifying Methods

The need to include the methodological dimension in indexing theory in order to cope with the challenge of "method plurality" (Dervin 2003, 125) while resisting any kind of "method hegemony" (Esbjörn-Hargens & Zimmerman 2009, 7) or a relativistic "anything goes" (Feyerabend 1975, 35) is often articulated (Hutchins 1975; Tibbo 1994; Szostak 2004; Tennis 2008; Taylor & Joudrey 2009; Gnoli 2012). Frequently it is emphasized that interdisciplinary research might benefit from a comprehensive classification and indexing of methods or techniques since researchers need to know, for example, what kinds of methods are already applied to a particular object of interest (and what kinds are not), to what extent methods can be imported from or exported to other fields of study, or in which way they can be combined in mixed or multiple methods research (Szostak, Gnoli & López-Huertas 2016).

In indexing practice, context information about applied methods or techniques is, if provided, relegated to the fringe, hidden in metadata fields like annotation or footnote, and lacking any documentary language. In the case that methodological issues are made explicit in metadata or contributed as additional keywords by the authors, they are usually freely chosen index terms without controlled vocabulary.

Therefore, a conceptual clarification of the close relation between method, methodology, or paradigm is required. As noted by Dervin (2003), the notion of methodology is often reduced to method, although it refers rather to the theoretical analysis of methods. Likewise, Hjørland (2000) argues for a clear-cut distinction of methods that refers exclusively to techniques versus methodologies that are concerned with problems of epistemology or the philosophy of science. Furthermore, Cibangu (2010) emphasizes that both methods as specific research strategies including procedures like data collection and data analysis as well as methodologies as sets of such methods are grounded in foundational paradigms.

In the field of knowledge organization (KO), an initial approach to the methodological dimension is made by Langridge's(1989) general distinction between topic and form of knowledge, that is, between the objects of study and the ways in which these objects are perceived. For example, zoology is considered to be the science (form) of animals (topic), or ethics the philosophy (form) of morals (topic). Langridge identifies at least twelve qualitatively different forms of knowledge and argues that, in contrast to disciplines, such fundamental forms are few in number, stable in time and mutually exclusive, even though they could be divided further into overlapping specializations. In a similar way, Szostak(2004) argues that research practices and techniques can be reduced to about a dozen scholarly methods that are common in different disciplines and often labeled by the same terms.Further typologies of research methods, without claiming comprehensiveness, are provided by Cibangu (2010) or Chu (2015) for the field of library and information science (LIS) (see Table 1).

Table 1. Examples of inductively derived listings of forms of knowledge or research methods

Langridge (1989)	Szostak (2004)	Cibangu (2010)	Chu (2015)
Prolegomena	Experiment	Focus group	Bibliometrics
Philosophy	Surveys	Ethnography	Content analysis
Natural science	Interviews	Grounded theory	Delphi study
Technology	Mathematical models	Interviews	Ethnography, field study
Human science	Statistical analysis	Discourse analysis	Experiment
Social practice	Ethnographic,	Content analysis	Focus groups
History	observational analysis	Survey research	Historical method
Moral knowledge	Experience, intuition	Historical research	Interview
Religion	Textual analysis	Case studies	Observation
Art	Classification	Naturalistic research	Questionnaire
Criticism	Mapmaking	Cultural studies	Research diary, journal
Personal experience	Hermeneutics, semiotics	Ethnomethodology	Theoretical approach
	Physical traces		Think aloud protocol
	Case studies		Transaction log analysis
			Webometrics

Integral Methodological Pluralism

It is important to note that such inductively derived listings, that is, term lists without any conceptual relations, can only be a first step towards an expressive knowledge organization system which is able to indicate in which ways all these different methods and techniques are interrelated. The second step from an unbounded methodological pluralism to a specification of its internal relations is emphasized by Wilber's (2002, 10) integral methodological pluralism:

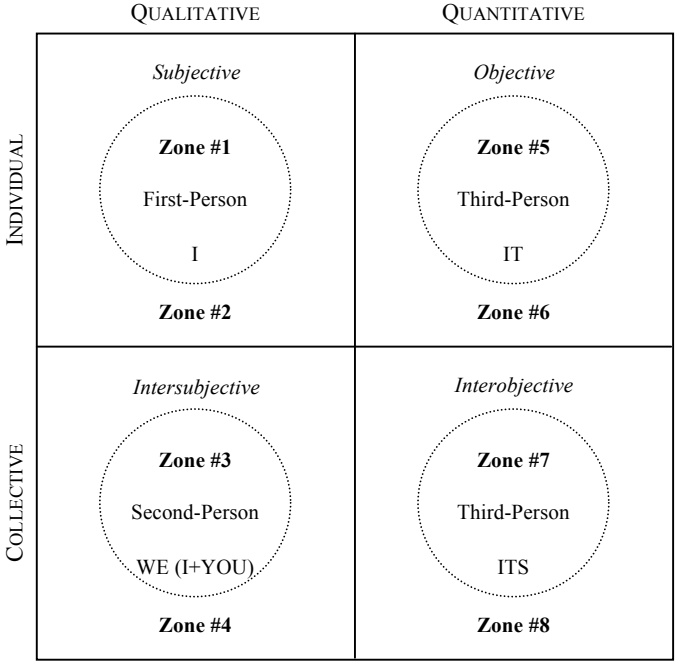
> "Integral," in that the pluralism is not a mere eclecticism or grab bag of unrelated paradigms, but a meta-paradigm that weaves together its many threads into an integral tapestry, a unity-in-diversity that slights neither the unity nor the diversity. "Methodological," in that this is a real paradigm or set of actual practices and behavioral injunctions to bring forth an integral territory, not merely a new holistic theory or maps without any territory. And "pluralism" in that there is no one overriding or privileged injunction (other than to be radically all-inclusive).

Based on integral theory, this intended synthetic approach applies three heuristic principles, namely, non-exclusion (one methodology cannot be used by itself to exclude other legitimate methodologies), unfoldment (some methodologies are more encompassing or more inclusive than others), and enactment (phenomena are brought forth or co-constructed by injunctions, paradigms, or social practices). While the principle of non-exclusion is widely accepted in KO discourse, typical modernist approaches tend to reject the principle of enactment since phenomena are considered to be independent from the observer (ISKO Italy 2007), whereas typical postmodernist approaches tend to reject the principle of unfoldment since different methodologies or paradigms are per se seen as incommensurable (Hjørland 2000; Jacob 2000).

This contribution aims to demonstrate, however, that both theoretical camps might benefit from a comprehensive classification of methods that relies on some

fundamental formal-pragmatic distinctions as analyzed by Habermas (1998) and adopted, for example, by Wilber's IMP as well as by some KO theorists (Brier 2000; Gracioso 2012; Ma 2012). In contradistinction to empirical pragmatics (e.g., sociolinguistics), Habermas's formal pragmatics seeks to analyze general patterns of communicative action that apply to all languages and all contexts of knowledge or information exchange such as typologies of perspectives (e.g., third-person, second-person, first-person), world relations (e.g., objective, social, subjective), or validity claims (e.g., truth, rightness, truthfulness).

Figure 1. Integral methodological pluralism (based on Wilber 2006)



As shown in Figure 1, these formal-pragmatic features are, to some extent, reflected by important methodological distinctions such as quantitative methodologies vs. qualitative methodologies (Hjørland 2000; Dervin 2003; Ma 2012; Chu 2015), methodological individualism vs. methodological collectivism (Hjørland 1997, 2000; Ritzer 2001), or inside vs. outside approaches as described, for example, by Ma (2012, 1864) with regard to the study of human culture:

> To attain an insider, participant view and an intersubjective understanding of the culture, the researcher must take a position (including the first-, second-, and third-person positions) with the cultural group, for observing as an "outsider" or a "neutral researcher" (i.e., maintaining a third-person position without taking a first- or second-person position) implies a subject-object relationship with the cultural group.

Based on these three formal-pragmatic distinctions, IMP deduces eight fundamental

zones that can be considered as major "methodological families" (Esbjörn-Hargens 2006, 84), each including sub-classes of specific methods or techniques (see Table 2).

Table 2. IMP as an example of a deductively derived basic schema for method classification

Main class				Representative methodology
Qualitative	Individual	Inside	#1	Phenomenological analysis (e.g., Husserl)
		Outside	#2	Cognitive analysis (e.g., Piaget)
	Collective	Inside	#3	Hermeneutic analysis (e.g., Gadamer)
		Outside	#4	Structuralist analysis (e.g., Lévi-Strauss)
Quantitative	Individual	Inside	#5	Autopoietic analysis (e.g., Maturana & Varela)
		Outside	#6	Behavioral analysis (e.g., Skinner)
	Collective	Inside	#7	Autopoietic social systems analysis (e.g., Luhmann)
		Outside	#8	Social systems analysis (e.g., Bertalanffy)

For example, zone #1 can be represented by classical phenomenology using Edmund Husserl's procedure of phenomenological reduction or epoché (direct approach: researcher's own consciousness), even though not every phenomenological approach is strictly limited to this particular zone (Martin 2008; Küpers 2009). By contrast, zone #2 can be represented by cognitive psychology using Jean Piaget's clinical method, that is, psychometric tests combined with open-ended interviews (indirect approach: consciousness of others) (for a detailed overview of the IMP zones see Wilber 2002, 2006; Esbjörn-Hargens 2006; Esbjörn-Hargens & Zimmerman 2009; Kleineberg 2013).

The decisive point is that each methodological zone requires its own narrative or "description system" (Brier 2000, 435) and cannot be reduced to other zones because different kinds of practices are concerned with different kinds of phenomena. Nevertheless, IMP emphasizes the complementary character of these zones and offers a coherent framework or "methodological relationism" (Ritzer 2001, 126; see also Juckes & Barresi 1993) for both identifying hidden reductionism as well as applying multiple methods research beyond a mere "methodological eclecticism" (Dousa & Ibekwe-SanJuan 2014, 152; see also Olson 1995; Hjørland 2000).

Method Analysis and Indexing

A short sketch of an IMP-based method analysis and indexing will be presented, focusing on the example of LIS research investigating users of information systems such as libraries or online environments. As noted by Hjørland (2000, 515), user

studies cover a broad range of divergent methodological approaches such as behaviorist, cognitivistic, hermeneutic, sociological, or domain analytic. According to IMP, a convenient way to identify a methodological zone, even if the applied methods or foundational paradigms are not made explicit, is to ask three simple questions:

a) Is it a qualitative (inter-/subjective) or quantitative (inter-/objective) approach?

b) Is the focus on an individual (element) or a collective (system)?

c) Is it an inside (direct) or an outside (indirect) view?

In LIS research, one might expect that individual users are studied either quantitatively, for example, by using zone #6 methods (e.g., information behavior analysis, log file analysis, eye-tracking observation, questionnaire, survey), or qualitatively by using zone #2 methods (e.g., cognitive analysis, interview) or zone #1 methods (e.g., thinking aloud protocol, journal writing). Since individual users are always already culturally and socially embedded in knowledge domains or user groups, they cannot be adequately understood without using also methodological collectivism. This too might be done either quantitatively by using, for example, zone #8 methods (e.g., network analysis, informetrics, social systems analysis), or qualitatively by using zone #3 methods (e.g., hermeneutic analysis, focus group, participant observation), or zone #4 methods (e.g., discourse analysis, domain analysis, detached observation).

Furthermore, the IMP zones offer a disambiguation tool for approaches that cover different methods under the same label. For example, the umbrella term "domain analysis" refers to both quantitative methods (e.g., citation analysis: zone #8) as well as qualitative methods (e.g., discourse analysis: zone #4). Likewise, information behavior research applies both quantitative methodological individualism (e.g., behaviorist analysis: zone #6) as well as qualitative methodological collectivism with regard to "values and norms of cultural and social groups" (Ma 2012, 1865) associated with zones #3 and #4. In the same way, the multiple methods approach cognitive work analysis might be simultaneously concerned with a sociocultural dimension (e.g., zones #3, #4), an environmental or organizational dimension (e.g., zones #7, #8), and an individual dimension (e.g., zones #1, #2, #5, #6). In this respect, an IMP analysis is also able to identify methodological reductionism or even new research directions.

As often noted, the methodological dimension is deeply intertwined with the ontological and epistemological dimensions of human knowledge (Olson 1995; Gnoli 2012; Ma 2012; Kleineberg 2013). Therefore, Wilber's (2002, 24) so-called "integral indexing" is an attempt to include and interrelate all three of them within a non-relativistic integral perspectivism that is formalized by an elaborated IMP notation system (Esbjörn-Hargens & Zimmerman 2009; Fuhs 2010).

Conclusion

In addition to merely inductively derived listings of research practices, integral methodological pluralism, deduced from fundamental formal-pragmatic distinctions,

offers a comprehensive basic schema for a classification of methods that is able to show the complementary character of different approaches, to identify methodological reductionism, and to guide interdisciplinary or multiple methods research. This contribution emphasizes that expressive context indexing beyond term lists calls for new organizing principles. In more general terms, one might conclude that in the field of knowledge organization the pragmatic turn requires a formal-pragmatic twist.

References

Andersen, Jack & Christensen, Frank S. (2001). Wittgenstein and indexing theory. In Albrechtsen, Hanne & Mai, Jens-Erik, eds., *Advances in classification research*, Vol. 10. Medford: Information Today. Pp. 1-21.

Biagetti, Maria T. (2006). Indexing and scientific research needs. In Budin, Gerhard et. al., eds., K*nowledge organization for a global learning society*. Würzburg: Ergon. Pp. 241-6.

Bies, Werner (1995). Pragmatische Inhaltserschließung: Grundlagen, Probleme und Perspektiven. In Meder, Norbert; Jaenecke, Peter & Schmitz-Esser, Winfried, eds., *Konstruktion und Retrieval von Wissen*. Frankfurt am Main: Indeks. Pp. 134-42.

Brier, Søren (2000). Trans-scientific frameworks of knowing: Complementary views of the different types of human knowledge. *Systems Research and Behavioral Science*, 17: 433-58.

Chu, Heting (2015). Research methods in library and information science: A content analysis. *Library & Information Science Research*, 37(1): 36-41.

Cibangu, Sylvain K. (2010). Paradigms, methodologies, and methods. *Library & Information Science Research*, 32: 177-8.

Dervin, Brenda (2003). Given a context by any other name: Methodological tools for taming the unruly beast. In Dervin, Brenda et al., eds., *Sense-making methodology reader: Selected writings of Brenda Dervin*. Creskill: Hampton Press. Pp. 111-32.

Dousa, Thomas M.,& Ibekwe-SanJuan, Fidelia (2014). Epistemological and methodological eclecticism in the construction of knowledge organization systems: The case of analytico-synthetic KOSs. In Babik, Wiesław, ed., *Knowledge organization in the 21th century: Between historical patterns and future prospects*. Würzburg: Ergon. Pp. 152-9.

Esbjörn-Hargens, Sean (2006). Integral research: A multi-method approach to investigating phenomena. *Constructivism in the Human Sciences* 11: 79-107.

Esbjörn-Hargens, Sean,& Zimmerman, Michael (2009). *Integral ecology: Uniting multiple perspectives on the natural world*. Boston: Integral Books.

Feyerabend, Paul (1975). *Against method: Outline of an anarchistic theory of knowledge*. London: New Left Books.

Frohmann, Bernd (1990). Rules of indexing: A critique of mentalism in information retrieval theory. *Journal of Documentation*, 46(2): 81-101.

Fuhs, Clint (2010). An integral map of perspective-taking. In Esbjörn-Hargens, Sean, ed., *Integral theory in action*. New York: SUNY Press, 273-302.

Gnoli, Claudio (2012). Metadata about what?: Distinguishing between ontic, epistemic and documental dimensions in knowledge organization. *Knowledge Organization*, 39(4): 268-75.

Gracioso, Luciana S. (2012). Language philosophy in the context of knowledge organization in the interactive virtual platform. *Journal of Systemics, Cybernetics & Informatics*,10(6): 64-7.

Habermas, Jürgen (1998). *On the pragmatics of communication*. Cambridge: MIT Press.

Hjørland, Birger (1997). *Information seeking and subject representation: An activity-theoretical approach to information science*. Westport: Greenwood.

Hjørland, Birger (2000). Library and information science: Practice, theory, and philosophical basis. *Information Processing & Management*, 36: 501-31.

Hutchins, William J. (1975). *Languages of indexing and classification: A linguistic study of structures and functions*. Stevenage: Peregrinus.

ISKO Italy (2007). León manifesto. *Knowledge Organization*, 34(1): 6-8.

Jacob, Elin K. (2000). The legacy of pragmatism: Implications for knowledge organization in a pluralistic universe. In Beghtol, Clare; Howarth, Lynne C. & Williamson, Nancy J., eds., *Dynamism and stability in knowledge organization*. Würzburg: Ergon. Pp. 16-22.

Juckes, Tim J.,& Barresi, John (1993). The subjective-objective dimension in the individual-society connection. *Journal for the Theory of Social Behaviour*, 23(2): 197-216.

Kaipainen, Mauri,& Hautamäki, Antti (2011). Epistemic pluralism and multi-perspective knowledge organization: Explorative conceptualization of topical content domains. *Knowledge Organization*, 38(6): 503-14.

Kleineberg, Michael (2013). The blind men and the elephant: Towards an organization of epistemic contexts. *Knowledge Organization*, 40(5): 340-62.

Kleineberg, Michael (2014). Integrative levels of knowing: An organizing principle for the epistemological dimension. In Babik, Wiesław, ed., *Knowledge organization in the 21th century: Between historical patterns and future prospects*. Würzburg: Ergon. Pp. 80-7.

Küpers, Wendelin M. (2009). The status and relevance of phenomenology for integral research: Or why phenomenology is more and different than an “upper left” or “zone #1” affair. *Integral Review* 5(1): 51-95.

Langridge, Derek W. (1989). *Subject analysis: Principles and procedures*. London: Bowker-Saur.

Ma, Lai (2012). Some philosophical considerations in using mixed methods in library and information science research. *Journal of the American Society for Information Science and Technology*, 63(9): 1859-67.

Mai, Jens-Erik (2005). Analysis in indexing: Document and domain-centered approaches. *Information Processing & Management*, 41(3): 599-611.

Martin, Jeffery A. (2008). Integral research as a practical mixed-methods framework: Clarifying the role of integral methodological pluralism. *Journal of Integral Theory and Practice*, 3(2): 155-64.

Olson, Hope A. (1995). Quantitative “versus” qualitative research?: The wrong question. In Olson, Hope A. & Ward, Dennis B., eds., *Connectedness: Information, systems, people, organizations*. Edmonton: University of Alberta. Pp. 40-9.

Ritzer, George (2001). *Explorations in social theory: From metatheorizing to rationalization*. London: Sage.

Szostak, Rick (2004). *Classifying science: Phenomena, data, theory, method, practice*. Dordrecht: Springer.

Szostak, Rick, Gnoli, Claudio, & López-Huertas, María (2016). *Interdisciplinary knowledge organization*. Cham: Springer.

Taylor, Arlene G., & Joudrey, Daniel N. (2009). *The organization of information*. Westport: Libraries Unlimited.

Tennis, Joseph T. (2008). Epistemology, theory, and methodology in knowledge organization:

Toward a classification, metatheory, and research framework. *Knowledge Organization*, 35(2/3): 102-12.

Tibbo, Helen R. (1994). Indexing for the humanities. *Journal of the American Society for Information Science*, 45(8): 607-19.

Weinberg, Bella H. (1988). Why indexing fails the researcher. *The Indexer*, 16(1): 3-6.

Wilber, Ken (2002). *Excerpt B. The many ways we touch: Three principles helpful for any integral approach*. [http://www.kenwilber.com/writings/read_pdf/84]

Wilber, Ken (2006). *Integral spirituality. A startling new role for religion in the modern and postmodern world*. Boston: Shambhala.

John M. Budd and Daniel Martínez-Ávila

Epistemic Warrant for Categorizational Activities in Knowledge Organization

Abstract

The classification of works is, of course, a complicated matter. Many systems exist, and many ideas have been propounded over the years. The present proposal suggests that assessing the warrant of a work is a means to categorize content. To be more specific, both literary and epistemic warrant can be used to develop classification mechanisms and controlled vocabularies as new subject headings are proposed and established using literary warrant when a cataloger is cataloging an item and is not satisfied with the available subject headings. In this vein, epistemic warrant is proposed as a means of enrich literary warrant in the construction of knowledge organization systems.

Introduction

The concept of "warrant," in knowledge organization, can be understood as "the rational justification for the introduction of a term or concept into a controlled vocabulary […] Warrant provides the limits a classificationist sets on source of concepts and terminology, and as a result on the inclusion or exclusion of concepts and terminology" (Tennis 2005, p.86). The "ANSI/NISO Z39.19-2005 Guidelines for the construction, format, and management of monolingual controlled vocabularies" prescribes "using warrant to select terms" for the construction of classifications and other knowledge organization systems, stating that "The process of selecting terms for inclusion in controlled vocabularies involves consulting various sources of words and phrases as well as criteria based on: -the natural language used to describe content objects (literary warrant), -the language of users (user warrant), and -the needs and priorities of the organization (organizational warrant)" (p.16). The previous "ANSI/NISO Z39.19-1993 Guidelines for the construction, format and management of monolingual thesauri" also considered literary warrant and user warrant but omitted the organizational warrant.

Literary warrant

The concept of Literary Warrant was introduced in 1911 by E. Wyndham Hulme, and as Mario Barité (2009, p.13) pointed out "Since then, it has evolved slowly but steadily to become one of the basic and unquestionable foundations of knowledge organization for information retrieval." Hulme (1911 cited by Chan et al. 1985, p.48, and Barité et al. 2010, p.124) explained literary warrant as: "meaning that the basis for classification is to be found in the actual published literature rather than abstract philosophical ideas or concepts in the universe of knowledge or the order of nature and system of the sciences." In 1995, Claire Beghtol discussed the possibility of applying a type of domain analysis of fiction studies considering two concepts, "literary warrant" (Hulme, 1911-1912) and "consensus" (Bliss,1939), and for the purpose of the project

she characterized literary warrant “as the topics around which a literature has become established” (p.31).

Although today the concept of literary warrant is common among classificationists, Beghtol (1986, p.112) also pointed out that, historically, some authors have argued uses of the term that differ from Hulme's original conception. For instance, some authors do not distinguish book classifications from scientific or philosophical classifications of knowledge as Hulme does. Indeed, the British Classification Research Group (CRG) played a great role in the shaping of literary warrant while omitting, borrowing, and transforming the concept during the 20th century. As Beghtol (1986,p.113) reported, the CRG narrowed Hulme's original idea from “literary” to what might be called' “terminological” warrant. In this form, the system would not be based on the subjects of books but on the terminology of a subject field (i.e., the terms that the authors in the literature of the subject use). The context of this move can be pinned in the efforts of the CRG on facet analysis and search for terms in the foci and subdivisions. This might be perhaps related, in context, to Hjørland’s observation (2007) of Mooers’(1972) arguable criticism of the principle of literary warrant: “Mooers does not directly say that the ideas are not to be found in the literature, but rather that the specific expressions found there should not be used.” Today, the definition of literary warrant by the ANSI/NISO Z39.19-2005 (more concerned with “words and phrases” than ideas) seems to be in the CRG “terminological” line too. In a prescriptive way, the ANSI/NISO Z39.19-2005 defines literary warrant as the “Justification for the representation of a concept in an indexing language or for the selection of a preferred term because of its frequent occurrence in the literature” (p.6, 162).

Epistemic Warrant

What follows is a proposal that states epistemic warrant as a means that can assist information seekers in locating content that closely matches the desired purpose of searches. This can be helpful for both assigning categories to a work and incorporate those categories into the knowledge organization system. It is known that new subject headings are proposed and established using literary warrant when a cataloger is cataloging an item and is not satisfied with the available system, such as the LCSH (Strader 2012, p.238). Categorization and classification of information carries the explicit implication that terms assigned to works are indicative of the content of the works. Means by which the contents of works may be classified are guided by the purpose of helping information seekers to locate content which matches the ideational implicature of the works in question. The effort to classify content is a direct application of the purpose of location. Knowledge organization is (and should be) concerned with accomplishing the purpose of classification, and it is suggested here that epistemic and literary warrant can be employed in categorizational activities.

Epistemic warrant has a substantial pedigree in a couple of ways, including argumentation and the theory of knowledge. Stephen Toulmin (1958) has developed a

well-defined schema for assigning warrant within argumentation. In argumentation, data proceeds towards claims, with reason(s) being the linking element. Warrant—logical reasons for believing and accepting premises—is necessary if a claim is to be articulated. Working backwards, warrant is necessary for the evaluation of a claim. Anyone stating an argument, according to Toulmin, must integrate warrant into the formulation of all components. Without reasoned warrant, a claim stands upon shaky ground, and it is doubtful whether it can (or should) be accepted.

One of the most prominent proponents of epistemic warrant is Alvin Plantinga. He (1993a) sums up his conceptualization succinctly; warrant is "that which distinguishes knowledge from mere true belief" (p. 3). He (1993b) elaborates:

> a belief *B* has warrant for you if an only if (1) the cognitive faculties involved in the production of *B* are functioning properly (and this is to include the relevant defeater systems as well as those systems, if any, that provide *propositional* inputs to the system in question); (2) your cognitive environment is sufficiently similar to the one for which your cognitive faculties are designed; (3) the triple of the design plan governing the production of the belief in question involves, as purpose or function, the production of true beliefs (and the same goes for elements of the design plan governing the production of input beliefs to the system in question); and (4) the design plan is a good one: that is, there is a high statistical or objective probability that a belief produced in accordance with the relevant segment of the design plan in that sort of environment is true (p. 194).

One example that can be used to illustrate epistemic warrant is Plantinga's own work, *Warrant and Proper Function*. The Library of Congress Subject Headings assigned to the work are: "Knowledge, Theory of;" "Belief and Doubt;" "Cognition." The subject headings are not inaccurate, but it is questionable whether all are necessary and sufficient. Since the work is primarily about epistemic warrant, a subject heading such as "Epistemology, Warrant" would be useful to anyone seeking information about the topic. Also, other subject headings that would be useful to connecting an information seeker to the work's content could include "Probability, Epistemic," "Induction (Logic)," "Naturalism," and "Testimony." Further, Plantinga draws heavily from the thought of Thomas Reid, so "Reid, Thomas" would be helpful. The headings are warranted, according to the proposal articulated above, by the substantive, manifest, and apparent inclusion of the topics within the work. The LCSH operates under the principle of literary warrant. However, these subject headings are not applied to this example, actually some of them do not even exist in the system (exceptions are "Induction (Logic)," "Naturalism," and "Testimony (Theory of knowledge)"). The consideration of epistemic warrant in the LCSH would help to identify and populate the system with alternatives that can be helpful for some communities and contexts.

Conclusion

Epistemic warrant can be used to determine the content of a work in order to either categorize it or propose categories in the literary warrant process. Epistemic warrant can be considered thus as a new kind of warrant for the classification of works, and for the warrant of knowledge organization systems of any kind, since the institutions in

charge of the systems, such as the Library of Congress, usually consider terms that are determined by classificationists and indexers at the time of the analysis of the work.

References

Barité, Mario. (2009). Garantía literaria y normas para construcción de vocabularios controlados: aspectos epistemológicos y metodológicos. *Scire*, 15 (2): 13-24.

Barité, Mario, Fernández-Molina, Juan Carlo, Guimarães, José Augusto Chaves, & Moraes, João Batista Ernesto de. (2010). Garantia literária: elementos para uma revisão crítica após um século. *TransInformação*, 22(2): 123-38.

Beghtol, Clare (1986). Semantic Validity: Concepts of Warrant in Bibliographic Classification Systems. *Library Resources and Technical Services*, 30(2), 109-125.

Beghtol, Clare (1995). Domain Analysis, Literary Warrant, and Consensus: The Case of Fiction Studies. *Journal of the American Society for Information Science*, 46(1): 30-44.

Bliss, Henry Eevelyn (1939). *The Organization of Knowledge in Libraries and the Subject-Approach to Books*, 2d ed. rev. New York: Wilson.

Chan, Lois Mai, Richmond, Phyllis A., & Svenonius, Elaine. (Ed.). (1985). *Theory of subject analysi*s. Littleton, Co.: Libraries Unlimited.

Hjørland, Birger. (2006). Literary warrant (and other kinds of warrant). In *Lifeboat for Knowledge Organization*. [http://www.iva.dk/bh/lifeboat_ko/CONCEPTS/literary_warrant.htm]

Hulme, E. Wyndham. (1911-1912). Principles of book classification. *Library Association Record*, 13, (1911: Oct. 14, 354-358; Nov. 15, 389-394; Dec. 15, 444-449) and 14, (1912: Jan. 15, 39-46; Mar. 15, 174-181).

Mooers, Calvin N. (1972). Descriptors. In: *Encyclopedia of Library and Information Science*. ed. by Allen Kent & Harold Lancour. 7. New York: Marcel Dekker. Pp. 3145.

National Information Standards Organisation (1994). *ANSI/NISO Z39.19-1993 Guidelines for the construction, format and management of monolingual thesauri*. Bethesda, MD: NISO Press.

National Information Standards Organisation (2005). *ANSI/NISO Z39.19-2005 Guidelines for the construction, format and management of monolingual controlled vocabularies*. Bethesda, MD: NISO Press.

Plantinga, Alvin (1993a). *Warrant: The current debate*. Oxford: Oxford University Press.

Plantinga, Alvin (1993b). *Warrant and proper function*. Oxford: Oxford University Press.

Strader, C. Rockelle (2012). Citation Analysis: Do Age and Types of Materials Cited Correlate with Availability of Appropriate Library of Congress Subject Headings? *Library Resources & Technical Services*, 56(4), 238-53

Tennis, Joseph T. (2005). Experientialist Epistemology and Classification Theory. *Knowledge Organization*, 32(2) 79-92

Toulmin, Stephen. (1958). *The uses of argumen*t. Cambridge: Cambridge University Press.

Toulmin, Stephen. (2001). *The return to reason*. Cambridge, MA: Harvard University Press.

Mario Barité

Literary Warrant Revisited: Theoretical and Methodological Approach

Abstract

Hulme introduced the literary warrant (LW) notion in 1911. He advocated that the terms of a classification system or any other knowledge organization system (KOS) had to come from literature rather than theoretical or philosophical criteria, scientific considerations or classifications. LW was considered a "pivotal" or "focal" concept but also a marginal topic, relegated a kind of conceptual purgatory, at least until 1984. Later, Hulme's heritage received several recognitions, but always in a discrete way. One significant LW problem is that its original meaning has been expanded, restricted, and also misunderstood. Hulme committed the original sin of presenting the concept without explicit detailed explanations. We compile, analyze and compare 49 LW definitions, formulated along 105 years, to assess similarities and differences between them, and identify the variety of meanings the 'LW' term has nowadays. As a result, we find that LW is seen from, at least, five different perspectives: theoretical principle, methodological tool, body of literature about one topic, state-of-art of KOS evaluation tool, and prediction tool of research. The five LW perspectives can be summarized into three: theoretical, methodological and applicative viewpoints. They are not necessarily exclusive since they interact and mutually influence each other. The comprehension of Hulme's notion oriented their applicability in all information contexts (prints, audiovisuals and digital environments).

Literary warrant: a crucial concept

Literature about warrants followed a sinuous way in Knowledge Organization (KO), since Edward Wyndham Hulme introduced the literary warrant (LW) notion in his work Principles of Book Classification, published serially between 1911-1912.

First, a few words about the man. Hulme was, for many years, the librarian of the British Patent Office (today named Intellectual Property Office), and a prominent member of the still existing Newcomen Society devoted to the history of Engineering and Technology (see http://www.newcomen.com). In these roles, specially with the help of his easy access to patents, he published many books and articles about different topics like an early history of the English patent system, the invention of English flint glass, the statistical history of the iron trade of England and Wales between 1717 and 1750, Gallic fortification in Caesar's time or a history of the grated hearth, the chimney, and, the air-furnace. Hulme is also considered a pioneer history of bibliometrics studies, because he was the first author to introduce the term 'statistical bibliography' in 1922 (substituted for 'bibliometrics' in 1969 by Pritchard), when he presented a statistical analysis of science history (Hood & Wilson, 2001). In words of Hjørland, Hulme used 'Statistical bibliography' "to describe the process of illuminating the history of science and technology by counting documents" (Hjørland, 2005). In the development of their research work on patents and technology histories, Hulme probably deposited the idea that relevant topics of documents could be counted and weighted, and considered as a quantitative base to select appropriate terminology for

classification systems in libraries.

In this way, Hulme wrote: "...a class heading is warranted only when a literature in book form has been shown to exist, and the test of the validity of a heading is the degree of accuracy with which it describes the area of subject-matter common to the class" (Hulme, 1911, 447). He added that "definition, therefore, may be described as the plotting of areas preexisting in literature. To this literary warrant a quantitative value can be assigned so soon as the bibliography of a subject has been definitely compiled (Hulme, 1911, 447). In these few words, he established a basic notion that requires further analysis of its implications, potential and projections.

A good definition to better understand the LW concept belongs to Beghtol, who says that "LW may be generally characterized as the topics around which a literature has become established" (Beghtol, 1995).

An original conception of LW is supported, then, by a "solid and tangible foundation: the content of books" (Rodríguez, 1984, 19). Hulme advocated that the terms of a classification system -or, for extension, into any other knowledge organization system (KOS): thesauri, taxonomies, lists- had to come from literature rather than theoretical or philosophical criteria, scientific considerations or classifications (Foskett, 1996, Yee, 2001). Thus, the subjects of documents act as a catalyst of the processes through which the conceptual structures destined to classification and indexing of information resources are created, thinking in users' requirements and retrieval of documents by topics.

Hulme's contribution could be considered revolutionary at least for three reasons: 1. He dared the traditional justification of classification terms, based on the authority of philosophers or scientific organization thinkers (like Bacon or Leibniz), or on scientific consensus (as happens with the Cutter Expansive Classification), or on the authority of the same XIX century classificationists (Brunet and their classification for Parisian booksellers, Brown and others). 2. He shifted the axis from the authority of classificationists and specialists -always contaminated by subjectivity or social mentalities- to the authority of knowledge, as it is registered and socialized. 3. Hulme also proposed a quantitative approach to the management of documentation in libraries and in other contexts of information.

Nevertheless LW always was a marginal term made invisible, pushed into a kind of conceptual purgatory (Howarth & Jansen, 2014). The LW topic was only sporadically and superficially dealt with for seventy years (Farradane, 1961; Olding, 1968; Lancaster, 1977; Fraser 1978). The fact that Hulme`s initiatory book, after the 1911-12 publication, only had two further editions (Hulme, 1950, 1980 facsimile) talk about their hiding impact, in the manner of way as cinema remakes movies wich come back every 30/40 years.

The only relevant repercussions -but carrying with discretion- were the consideration of LW as a basis for the development of the Library of Congress

Classification (LCC), the Library of Congress Subject Headings (LCSH) and the Dewey Decimal Classification (DDC), as well as the inclusion of LW as a helpful sequence principle by Ranganathan in his Prolegomena (Ranganathan, 1957). But always under the critical magnifying glass of all those who considered that an elementary method like counting cannot be seriously considered as a procedure of terminology selection. That's why Rodriguez wrote: “LW is one of the most fundamental principles of subject analysis [but] the term is rarely encountered today, and the name of Hulme is virtually forgotten” (Rodriguez, 1984, 17), and claimed for their rediscovery.

In response to this alarm warning, Hulme's heritage received several recognitions. The responsibles for a compilation of LIS canonical texts transcribed the pages in which Hulme formulated the LW principle, with the certainty that his contribution accounted for three selection criteria of fundamental texts of/in the following fields: theoretical emphasis, significance and impact, as well as perspicuity (Chan, Richmond & Svenonius; 1985, p. xiv, 48ss.). Afterwards, LW applicability was extended from classification systems to thesauri (Lancaster, 1986). In the same year, Beghtol proposed for the first time a generic definition of 'warrant', and studied in-depth four types of semantic warrants: literary, scientific/philosophical, educational and cultural warrants (Beghtol, 1986). With this work, warrants studies were open to additional contributions (Cochrane, 1993; Beghtol, 1995; Hjørland, 2005; Dabney, 2007; Barité, 2011; Nunns, Peace & Witten, 2015), and its suitability for electronic resources and Web environment has been explored (Vizine-Goetz & Beall, 2004; Campbell, 2008). The LW principle was also incorporated into prestigious standards (NISO, 2010), and began to be considered as a “foundational”, “pivotal” or “focal” concept (Beghtol, 1995; Singh, 2001; Huvila, 2006), but always in the twilight of a disciplinary corner.

Probably a certain LW methodological insufficiency caused the emergence of other autonomous warrant forms, like user (Lancaster, 1977; Rosso & Haas, 2010), cultural (Lee, 1976), organizational (NISO, 2010), phenomenological (Ward, 2000), market warrant (Martínez-Ávila, 2012), and others ones. In any case, no author could write the death certificate of the LW; on the contrary, they enriched the original concept.

One significant LW problem is that its original meaning has been expanded, restricted, and also misunderstood. Hulme committed the original sin of presenting the concept without explicit detailed explanations. Ranganathan used LW with a slight variation, like a tool to arrange the focus of a facet in a decreasing sequence, considering the quantity of documents published on every focus (Ranganathan, 1957). For those responsible of the DDC, LW is a measure of what is enough: they require twenty works published on a topic to incorporate a new number into the system (Beall, 2003). For others, LW could be used as a device for the evaluation of scientific and technological production (Dahlberg, 1993, Barité, 2011).

This study intends to contribute to demonstrate the usefulness of LW tools in many

ways and forms, from their different approaches over time in LIS literature. Here we compile, analyze and compare an important number of LW definitions, formulated along 105 years, to assess similarities and differences between them, and finally identify the variety of meanings and applications that LW has nowadays. We intend to make a contribution to the comprehension of Hulme's notion and its applicability in all the contexts of information (prints, audiovisuals and digital environments).

Methodology

The inductive methodology used to identify different meanings of the LW principle is based in the KO literature, and was organized as follows:

- identification of original definitions (not merely transcriptions of the term), and useful elements for an ideal definition of LW, in LIS works published between 1911 and 2016;
- breakdown of each definition into its essential elements, to answer the following questions: what is it? Which are their elements? What role do they play? In which processes are they involved? What information do they supply?;
- creation of a chart divided by questions and the respective answers;
- comparative analysis of definitions;
- determination of the various meanings of LW;
- presentation of results;
- conclusions.

The corpus was made up of 44 works containing 49 LW definitions or descriptive and critical concept developments. The corpus was integrated with 16 articles, 4 books, 2 PHD theses, 8 dictionaries (which records 13 different term senses), 11 communications to congresses, 1 standard and 2 classification systems preliminaries.

Results

We find, analyzing the definitions found, that LW is seen from at least five different perspectives:

i) Theoretical principle. LW is mentioned as “criterion” (Clason, 1973), “concept” (Olson, 2002) and “principle” (Yee, 2001). It is therefore seen as an objective -and consequently, external- expression (Howarth & Jansen, 2014), like a systematic and consistent approach to the knowledge organization oriented to information retrieval. As a theoretical formulation it can be applied to all knowledge areas and it enhances the value of knowledge recorded in documents as a common pattern of scientific and technological understanding.

ii) Methodological tool. First, many authors agree to consider LW to justify the selection and hierarchization of terms (and related terms) to be included (or to be excluded because of their low justification) in any KOS.

Second, LW is considered relevant in processes of KOS creation, evaluation and

revision (specially in operations of quality evaluation of terminology). Its potential has been proven by their regular application by those responsible for LCC, DDC,

LCSH, Unesco Thesauri and others (Beall, 2003; Vizine-Goetz, & Beall, 2004; Green & Panzer, 2014). This is recognized, for example, by Scott who wrote that "the modern history of DDC is generally dated from 1958, with the publication of a refocused Edition 16 (...). Changes were kept to a mininum, reflecting only those most urgently needed to accommodate existing knowledge and literary warrant" (Scott, 1998, 2). We must also remember, to sizing the LW impact, that "the LCC is based entirely on the Library of Congress collection" (Hallows, 2015, 88), and Library of Congress is the major library in the world.

Third, LW could justify and arrange terms in mapping fields of knowledge, to order topics or to select the first focus in a facet (Rajaram, 2015), and to decide the inclusion/exclusion of dictionary and glossary terms (Cabré, 1993).

iii) Body of literature about one topic. In this sense, LW is expressed in the attribution of a quantitative value, like a material dimension of documentation. It is very useful to indicate the real documentation warrant of online thesauri and Web taxonomies. The CDD and LCC editorial rules only consider an expansion or a new number when there is enough documentation on a topic, represented by a concrete number of works devoted to a subject (Beall, 2003; Wood, 2010). In this body it is possible to distinguish the relative weight of the various types of works: canonical texts, manuals, dictionaries and other reference works, theses, technical or descriptive monographs, critical and legal documents, articles in specialized journals or regular proceedings. In other words, in this way it is possible to consider the relative importance of every type of document in the general production of a discipline. This could be useful, for example, to perform comparative studies about the internal integration of specialized documentation in different disciplines.

iv) State-of-art of KOS evaluation tool. LW allows to compare the situation of knowledge fields structures versus the situation of KOS conceptual structures, their quality and currentness. If the KOS was constructed according with the state of knowledge and enables to reasonably classify and index all types of specialized documents, LW could contribute to visualizing areas more or less explored by research, and zones of obsolescence, through the measurement of scientific, technical and critical production under every topic. This demands a large compilation of the bibliography in a subject field, duly classified or indexed throughout a lengthy period by one KOS. Dahlberg (1995) used LW (without mention of the term) to analyze current trends in KO based upon the bibliographic references published in the KO literature bulletins, incorporated as supplements of the Knowledge Organization Journal, and classified by the Classification System for Knowledge Organization Literature, in the period 1991-1993. Barité (2011) extended the study over the 1994-2009 period. Both studies applied the systematifier as a cut of scientific production tool in three axis (theory,

praxis, environment), and identified more productive and less studied subareas in the KO domain. In this way, LW contributes to weight the quality of conceptual structures of KOS to reflect the reality of a field, and the adequacy, timeliness or obsolescence of a KOS.

v) Prediction tool of research. Beghtol wondered twenty years ago "is statistical analysis of existing indexing bibliographic records predictive of trends in different subject domains?" (Beghtol, 1995, 4). If LW can establish the state-of-art of domains, it is also able to identify gaps as well as areas with an increasing production, suitable for research purposes. For prediction, it is necessary to have a collection of disciplinary documents, for periods of ten or more years, classified by the same updated classification system. Diachronic studies covering in this way enough scientific or specialized academic production, show predictive trends.

The five LW perspectives can be summarized into three: theoretical, methodological and applicative. They are not necessarily exclusive since they interact and mutually influence each other. Thus, it could be agreed that LW is a theoretical principle which supports a method; or that it is at the same time a principle, a methodology and a product. In its nature, LW has an essentially multiple sided value, taking into account different approaches and utilities in the scientific discourse representation.

LW can be useful whether it is seen as a conceptual orientation, or an organized set of analytical tools (Huvila, 2006), and it could also be considered as an application or merely a material dimension (body of works). Then, in this nature, LW has an essential polyhedral value, that is to say, the capability to exhibit different faces, facets and vertices, those who could be integrated into a common figure, with its own identity, considering approaches, purposes and utilities in the scientific discourse representation.

Conclusions

Hulme installed LW in the notional system of KO, and incorporated, surreptitiously as well as firmly, a notion of warrant as a rationale element of subject headings validity in KOS.

The LW principle was revolutionary, at least for three reasons: 1. It rejected the idea, firmly established, that classification systems should necessarily be constructed like a mirror of the scientific and philosophical order of knowledge. 2. It shifted the axis from the authority of classificationists and specialists -inevitably contaminated by subjectivity or social mentalities- to the authority of knowledge, as it is registered and socialized. 3. It proposed a quantitative approach to the management of documentation in libraries and in other contexts of information.

LW is still a marginal concept in spite of being the basis of revision procedures of LCC or DDC systems, but in the last thirty years its relevance and projection has risen its value. At present, LW shows its full applicability to digital environments, through the number of works or references seen at any time in databases and web taxonomies.

Although LW has undergone a diversification of meanings and applications, this variety can be considered as a manifestation of the Promethean nature which enables it to split into close significations, all of them functional to the purposes of KO. Also due to its Promethean nature, LW was considered from five different perspectives in KO literature: theoretical principle, methodological tool, body of literature about one topic, state-of-art of KOS evaluation tool, and prediction tool of research. They can, however, essentially be reduced to three approaches: theoretical, methodological and applicative viewpoints. The three perspectives, at the same time of being autonomous, are complementaries and allows to establish a documented map of knowledge.

Today the comprehension of Hulme's notion directs its applicability into all information contexts (print, audiovisual and digital environments). Since Internet allows to show in numbers, at any time, the body of works or references in an information search, or in databases, the existence of LW usefulness is assured, because it can be used as a quantitative measure in any research about information management and access.

By the way, it is relevant to mention that the real KOS authority can be supported - beyond the particular opinions of their designers, and beyond the established science and technology formal structures- in the accumulated document production of the human race, and in the relationships that authors knit between topics of reality or imagination. One thing cannot be confronted: if a number classification or a descriptor does not have documents to be assigned, they neither play any role nor do they offer any usefulness to users.

Over a century after this first enunciation, LW continues to be a useful tool, and has proven an incredible capacity of adaptation to the profound changes and the new challenges in the world of information.

In a discipline like LIS, where Hjørland identified four serious problems in research tradition, “lack of explicit empirical methods, lack of methodological updating, lack of comparing its own approach with other approaches, lack of formal recognition within LIS” Hjørland, 2002, p. 451), LW shines as an exotic and still unexplored star, with so much potential to bring out new theoretical and methodological approaches, in a digital and refined contexts of information systems.

Maybe it is time to bring Hulme out of the unstable gloom in which it has been plunged all these years, and to place him in a privileged position into the Knowledge Organization pantheon of great personalities.

References

Barité, Mario (2011). *La garantía literaria como herramienta de revisión de sistemas de organización del conocimiento: modelo y aplicación*. Tesis doctoral. Granada: Universidad de Granada. [digibug.ugr.es/bitstream/10481/17583/1/19711864.pdf]

Beall, Julianne (2003). Approaches to expansions: case studies from the German and Vietnamese translations. In *World Library and Information Congress: 69th IFLA General Conference*

and Council 1-9 August 2003, Berlin. [http://www.ifla.org/IV/ifla69/papers/123e-Beall.pdf]

Beghtol, Clare (1986). Semantic validity: concepts of warrant in bibliographic classification systems. *Library Resources & Technical Services*, 30(2): 109-23.

Beghtol, Clare (1995). Domain analysis, literary warrant, and consensus: the case offiction studies. *Journal of the American Society for Information Science*, 46(1): 30-44.

Cabré, Maria Teresa (1993). *La terminología: teoría, metodología, aplicaciones*. Barcelona: Antártida.

Campbell, D. Grant (2008). Derrida, logocentrism, and the concept of warrant on the Semantic Web. In *Proceedings of the Tenth International ISKO Conference*. Pp. 222-8.

Chan, Lois Mai, Richmond, Phyllis A., & Svenonius, Elaine (eds.) (1985). *Theory of subject analysis: a sourcebook.* Littleton, Colorado: Libraries Unlimited.

Clason, W.E. (1973). *Elsevier's dictionary of Library Science, Information and Documentation*. Amsterdam: Elsevier.

Cochrane, Pauline (1995). Warrant for concepts in classification schemes. *Advances in Classification Research*, 4. Medford: Information today. Pp. 35-46.

Dabney, Daniel (2006). The universe of thinkable thoughts: literary warrant and West's key number system. *Law Library Journal*, 99(2): 229-47.

Dahlberg, Ingetraut (1993). Current trends in knowledge organization. In *Organización de conocimiento en sistemas de información y documentación: Actas del I Encuentro de ISKO-España, Madrid, 4-5 noviembre 1993*. Ed. por Javier García-Marco. Zaragoza, 1995. Pp. 7-25.

Farradane, Jason E.L. (1961). Fundamental fallacies and new needs in Classification. In *Theory of subject analysis*. Edited by Lois Mai Chan, Phyllis A. Richmond and Elaine Svenonius. Littleton, Colorado, 1985. Pp. 199-209.

Foskett, A.C. (1996). *The subject approach to information*. 5th edition. London: Library Association.

Fraser, W. J. (1978). Literary, user and logical warrants as indexing constraints. In *The Information Age in Perspective: Proceedings of the ASIS Annual Meeting 1978*. White Plains, Knowledge Industry Publications, New York, NY. Pp. 130-2.

Green, Rebecca, & Panzer, Michael (2014). The Interplay of Big Data, WorldCat, and Dewey. *Advances In Classification Research Online*, 24(1): 51-8.

Hallows, Kristen M. (2015). It's All Enumerative: Reconsidering Library of Congress Classification in U.S. Law libraries. *Law Library Journal*, 106(1): 85-99

Hjørland, Birger (2002). The methodology of constructing classification schemes: a discussion of the state-of-the-art. In *Proceedings of the 7th International ISKO Conference:10-13 July 2002, Granada*, Edited by M.J. López-Huertas. Würzburg: Ergon Verlag. Pp. 450-6.

Hjørland, Birger (2005). *Core concepts in Library and Information Science*. [http://www.db.dk/bh/core%20concepts%20in%20lis/home.htm]

Hood, William W., & Wilson, Concepción S. (2001). The literature of bibliometrics, scientometrics, and informetrics. *Scientometrics*, 52(2): 291–314.

Howarth, Lynne C., & Jansen, Eva Hourihan (2014). Towards a typology of warrant for 21st century Knowledge Organization Systems. In *Knowledge organization in the 21st century: between historical patterns and future prospects : proceedings of the 13th international ISKO conference in Kraków, May 19-22, 2014*. Würzburg: Ergon. Pp. 216-21.

Hulme, Edward Wyndham (1911). Principles of Book Classification: Chapter III - On the

Definition of Class Headings, and the Natural Limit to the Extension of Book Classification. *Library Association Record*, 13: 444-9.

Hulme, Edward Wyndham (1950). *Principles of Book Classification*. London: Association of Assistant Librarians.

Hulme, Edward Wyndham (1980). *Principles of Book Classification*. Ann Arbor, Mich.: University Microfilms. Photofacsimile of ed.: London: Association of Assistant Librarians, 1950.

Huvila, Isto (2006). *The Ecology of Information Work: a Case Study of Bridging Archaelogical Work and Virtual Reality Based Knowledge Organisation*. Åbo: Åbo Akademi University Press. [https://oa.doria.fi/bitstream/handle/10024/4153/TMP.objres.83.pdf?sequence=1]

Lancaster, F. Wilfrid (1977). Vocabulary control in information retrieval systems. In *Advances in Librarianship*, 7. Edited by Melvin Voight and Michael Harris. London: Academic Press. Pp. 1-40.

Lancaster, F. Wilfrid (1986). *Indización y resúmenes*. Buenos Aires: EB.

Lee, Joel M. (1976). E. Wyndham Hulme: a reconsideration. In T*he variety of Librarianship: essays in honour of John Wallace Metcalfe*. Edited by W. Boyd Rayward. Sydney: LAA.

Martínez-Ávila, Daniel (2012). *DDC-BISAC switching as a new case of reader-interest classification*. Tesis doctoral. Madrid: Universidad Carlos III de Madrid.

NISO = National Information Standards Organization (2010).*Guidelines for the construction, format and management of monolingual controlled vocabularies: an American National Standard developed by the National Information Standards Organization: ANSI/NISO Z39.19-2005 (R2010)*. Bethesda: NISO Press. [http://www.niso.org/apps/group_public/download.php/12591/z39-19-2005r2010.pdf]

Nunns, Heather, Peace, Robin, & Witten, Karen (2015). Evaluative reasoning in public-sector evaluation in Aotearoa New Zealand: How are we doing? *Evaluation Matters*, 1: 137-63.

Olding, Raymond Knox (1968). *Wyndham Hulme's literary warrant & information indication*. Los Angeles: University of California.

Olson, Hope A. (2002). *The Power to Name: Locating the Limits of Subject Representation in Libraries*. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Rajaram, Shyama (2015). Principles for helpful sequence and their relevance in technical writings: A study. *Annals of Library and Information Studie*s, 62 Dec: 268-273.

Ranganathan, S. R. (1957). *Prolegomena to Library Classification*. 2nd. ed. Madras: Madras Library Association.

Rodríguez, Robert D. (1984). Hulme's Concept of Literary Warrant. *Cataloging & Classification Quarterly*, 5(1): 17-26.

Rosso, Mark A., & Haas, Stephanie W. (2010). Identification of web genres by user warrant. In *Genres on the Web: Corpus Studies and Computational Models*. Mehler, Alexander, Sharoff, Serge, Rehm, G, &Santini, Marina, editors. Heidelberg: Springer-Verlag.

Scott, Mona L. *Dewey Decimal Classification 21st edition: a study manual and number building guide*.

Singh, Sukhdev (2001). Development of terminology in Colon Classification: a researcher's view-point. *Herald of Library Science*, 39(3): 177-88.

Vizine-Goetz, Diane, & Beall, Julianne (2004). Using literary warrant to define a version of the DDC for automated classification services. In *Proceedings of the 8th International ISKO Conference, 13-16 july 2004, London UK*. Würzburg: Ergon Verlag. Pp. 147-52.

Ward, Martin (2000). Phenomenological warrant: The case for working from the user's viewpoint. *Managing information*, 7(9): 68-71.

Wood, Susan E. (2010). *The Subject Representation of Core Works in Women's Studies: A Critical Analysis of the Library of Congress Subject Headings*. Master's Thesis. Knoxville: University of Tennessee. [http://trace.tennessee.edu/cgi/viewcontent.cgi?article=1632&context=utk_gradthes]

Yee, Martha M. (2001). Two Genre and Form Lists for Moving Image and Broadcast Materials : a comparison. *Cataloging & Classification Quarterly*, 31(3/4): 237-95.

Richard P. Smiraglia and Joshua A. Henry

Facets Among the *Topoi*: An Emerging Taxonomy of Silent Film Music

Abstract

Cultural forces shape the conceptual content of knowledge organization systems by means that overlap or intersect. Knowledge organization for film music is one such example. Music in film, diegetic or non-diegetic, is used in excerpts or even smaller segments, which often are combined for non-diegetic use. Facets are component parts or "the faces" of a phenomenon. Used in KOSs facets provide dimensional context to otherwise two-dimensional conceptual representations. Taxonomies of musical cues from silent film arose and survive as pragmatic tools of working film musicians. Erno Rapée's *Encyclopedia of Music for Pictures*, a keyboard manual for working musicians, demonstrated the use of facets in explicit and implicit ways to express mood or performance difficulty.

Cultural diversity and epistemological intersections

Divergent cultural forces shape the conceptual content of knowledge organization systems (KOSs), often by means that overlap or intersect. Knowledge organization for film music is one such example. Traditional music libraries require organization of musical entities by iconic composer-classes sub-arranged by work and instantiation. Film music and the libraries that are used both in its creation and to store its products are more complex, requiring distinctions to be made between diegetic and non-diegetic music − that heard by the characters in the film versus that heard only by the audience, respectively. Music in film, diegetic or non-diegetic, often is used in excerpts or even smaller segments, which often are combined for non-diegetic use. The appropriate assignment, performance or use of segments of music must also be subject to schema for thematic organization. Music to accompany a mysterious occurrence is different in character from music performed at a royal ball or music that seemingly portrays a panoramic landscape

This paper is a report of an emergent research stream of this last type. Music that was used to accompany silent film was performed during screenings by organists or pianists in the theatre who performed along to quite precise cue sheets that identified both the kind of music to be played and the duration. This practice is seen to have evolved from the earlier practice of accompanying traveling stage shows. The musical cues themselves during the era of the silent film came from a list of musical terms that had achieved canonical status as the vocabulary of vaudeville. This evolution has been described by film music historians Rosar (2012) and Plebuch (2012). Film music history itself combines the methodologies and epistemic stances of musicology with those of film history. The object of investigation − the cues and potential taxonomies of them − arose and survive as pragmatic tools of working film musicians. The affective content of the taxonomies of film cues represents a rich interweaving of cultural

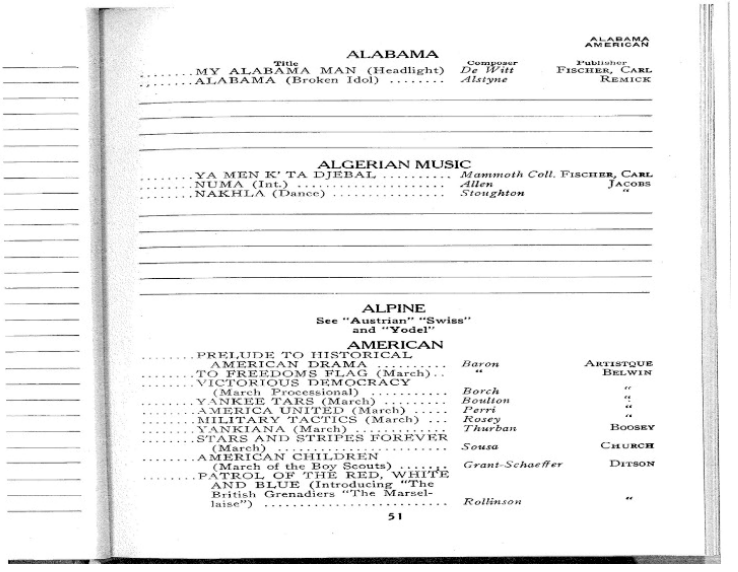
representations as enshrined in film and mirrored in music. The study of film music cues, then, presents the knowledge organization community with an opportunity to investigate a seemingly simple taxonomy that is, in fact, a temporal point at the intersection of a multiplicity of epistemological dimensions.

Cues to *topoi*

Our research begins with the era of silent film, roughly from the 1890s through the 1920s, for which brief segments of music were used to represent visual imagery in the films. It was thought to be a form of audio illustration, and although it was predominantly non-diegetic in intent it was not unheard of for a snippet to be used to accompany dance or the ringing of bells or some other potentially diegetic purpose. The repertoire of short musical segments was referred to as "cues," and studio film music librarians created cue sheets, or timed lists of segments that could be used to accompany the sequential scenes of a motion picture. Although specific compositions were used to generate the musical segments, the cues themselves often were expressed with musical terms such as "agitato" or "misterioso." It is these sets of terms that are thought to have formed a veritable, if not actual, taxonomy of film music cues (Rosar 2012). Plebuch (2012, 77) calls these "musical topoi" or characteristic musical expressions that are habitual and symbolic, yet flexible enough to admit a wide degree of interpretation and improvisation. Veritable taxonomies were contained in compendia produced by prominent musicians of the time. The most prominent were Erno Rapée's Encyclopedia ([1925], 1970) and Erdmann, Ecce and Brav's *Handbuch* (1927).

Erno Rapée's *Encyclopedia of Music for Pictures* was published in the form of a book that could serve as a keyboard manual for working musicians. Cues were centered in bold type together with lists of specific musical works by title, composer and publisher in which appropriate segments could be found. Dotted lines allow the musician to make a check mark next to music present in a local collection. Blank lines under each entry allow the musician to write in titles of music that isn't listed but that is acceptable for use with the given term. References, both "see" and "also see" occur throughout the list. As an example, Figure 1 is a reproduction from the book:

Figure 1. Alabama to American from Rapée's Encyclopedia page 51.



ALABAMA AMERICAN

ALABAMA

Title	Composer	Publisher
MY ALABAMA MAN (Headlight)	*De Witt*	FISCHER, CARL
ALABAMA (Broken Idol)	*Alstyne*	REMICK

ALGERIAN MUSIC

YA MEN K' TA DJEBAL	*Mammoth Coll.*	FISCHER, CARL
NUMA (Int.)	*Allen*	JACOBS
NAKHLA (Dance)	*Stoughton*	"

ALPINE

See "Austrian" "Swiss" and "Yodel"

AMERICAN

PRELUDE TO HISTORICAL AMERICAN DRAMA	*Baron*	ARTISTIQUE
TO FREEDOMS FLAG (March)	"	BELWIN
VICTORIOUS DEMOCRACY (March Processional)	*Borch*	"
YANKEE TARS (March)	*Boulton*	"
AMERICA UNITED (March)	*Perri*	"
MILITARY TACTICS (March)	*Rosey*	"
YANKIANA (March)	*Thurban*	BOOSEY
STARS AND STRIPES FOREVER (March)	*Sousa*	CHURCH
AMERICAN CHILDREN (March of the Boy Scouts)	*Grant-Schaeffer*	DITSON
PATROL OF THE RED, WHITE AND BLUE (Introducing "The British Grenadiers "The Marseillaise")	*Rollinson*	"

51

The taxonomic elements of the compendium can be seen on this one page. In the upper right corner are the first and last terms from the page—corner summaries for flipping quickly through the book to locate a specific term. In this case "Alpine" is not a preferred term, so the user is directed to "Austrian," "Swiss," and "Yodel," as potential representatives of Alpine musical concepts.

In two papers Smiraglia (2015) and Smiraglia and Henry (2016) have begun to describe an attempt to match the musical *topoi* derived from Rapée's *Encyclopedia* with the evidence of their use in Dutch silent film as recorded in the Eyl-Van Houten collection now housed in the EYE Filmmuseum in Amsterdam. The specific terminology is baldly representative of the social realities of the silent film era and is replete with terms that today might be considered politically-incorrect. These social realities are described in some detail in Smiraglia and Henry (2016).

Facets among the *topoi*

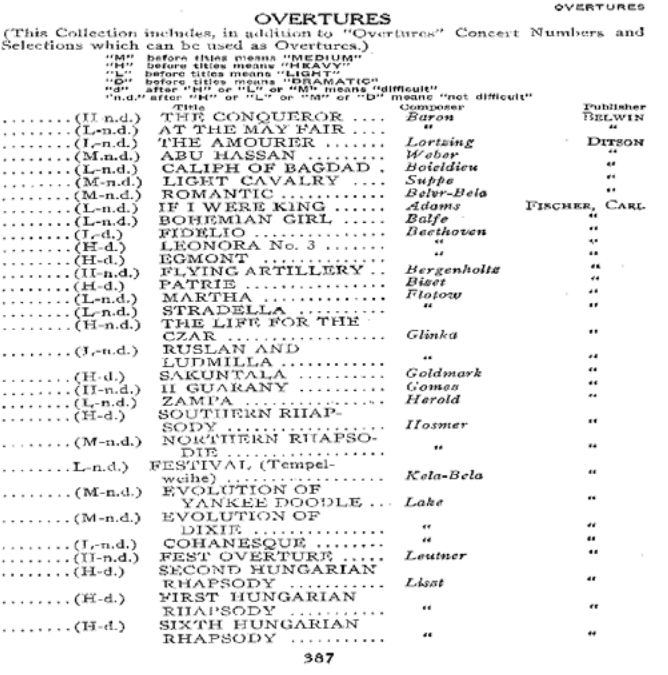
Facets have a rich history of use in KOSs because of the flexibility they bring to the expression of complex conceptual strings. In Smiraglia (2014, 82) facets are defined simply as "component parts[s] … [and] all of the facets together completely describe the domain," or (p. 44) as "the faces—the presentation characteristics … of a phenomenon." Facets are component parts or "the faces" of a phenomenon. Used in KOSs facets provide dimensional context to otherwise two-dimensional conceptual representations. Rapée used a technique that we can describe as faceting, in both explicit and implicit ways to express mood or performance difficulty. For example, six

of Rapée's preferred term headings—Agitatos, Dramatic, Galops, Love Themes, Mysteriosos, and Overtures—have explicit facets. These are: Heavy (H), Medium (M), Light (L), Dramatic (DR), and Dramatic-Neutral (D), which were used to further clarify the mood of the cues. Not all of these facets appear for each term. In fact, Dramatic-Neutral only appears under the term Dramatic. It is possible that Rapée used Dramatic-Neutral rather than just Neutral (N) as a way to differentiate between the facet and other headings that use the word Neutral. Yet, it is peculiar that he used Dramatic-Neutral at all, because there is only one cue designated as such.

There also seems to be use of secondary facets for the term Overtures. There are two of these secondary facets, difficult (d.) and not difficult (n.d.), which indicate the performance level. It is possible that these secondary facets are only included for the term Overtures, because theater music directors often didn't have time to rehearse an overture, a conventional necessity during this time, meaning they needed to know which overtures were easy to perform without rehearsal. This facet would not be as useful for the other cues.

Figure 2 below shows how Rapée indicated the facets. Directly under the heading for Overtures, Rapée indicated the meaning of the letters placed before each cue, like "M" for "MEDIUM." The letters are in parentheses and a hyphen separates the primary and secondary facets, for example (H-n.d.) for "HEAVY-not difficult."

Figure 2. Overtures from Rapee's Encyclopedia, page 387



OVERTURES

OVERTURES

(This Collection includes, in addition to "Overtures" Concert Numbers and Selections which can be used as Overtures.)

"M" before titles means "MEDIUM"
"H" before titles means "HEAVY"
"L" before titles means "LIGHT"
"D" before titles means "DRAMATIC"
"d" after "H" or "L" or "M" means "difficult"
"n.d." after "H" or "L" or "M" or "D" means "not difficult"

	Title	Composer	Publisher
.....(H n.d.)	THE CONQUEROR	*Byron*	BELWIN
.....(L-n.d.)	AT THE MAY FAIR	"	"
.....(L-n.d.)	THE AMOURER	*Lortzing*	DITSON
.....(M-n.d.)	ABU HASSAN	*Weber*	"
.....(L-n.d.)	CALIPH OF BAGDAD .	*Boieldieu*	"
.....(M-n.d.)	LIGHT CAVALRY	*Suppe*	"
.....(M-n.d.)	ROMANTIC	*Belvr-Bela*	"
.....(L-n.d.)	IF I WERE KING	*Adams*	FISCHER, CARL
.....(L-n.d.)	BOHEMIAN GIRL	*Balfe*	"
.....(L-d.)	FIDELIO	*Beethoven*	"
.....(H-d.)	LEONORA No. 3	"	"
.....(H-d.)	EGMONT	"	"
.....(H-n.d.)	FLYING ARTILLERY ..	*Bergenholtz*	"
.....(H-d.)	PATRIE	*Biset*	"
.....(L-n.d.)	MARTHA	*Flotow*	"
.....(L-n.d.)	STRADELLA	"	"
.....(H-n.d.)	THE LIFE FOR THE CZAR	*Glinka*	"
.....(L-n.d.)	RUSLAN AND LUDMILLA	"	"
.....(H-d.)	SAKUNTALA	*Goldmark*	"
.....(H-n.d.)	II GUARANY	*Gomes*	"
.....(L-n.d.)	ZAMPA	*Herold*	"
.....(H-d.)	SOUTHERN RHAPSODY	*Ilosmer*	"
.....(M-n.d.)	NORTHERN RHAPSODIE	"	"
.....L-n.d.)	FESTIVAL (Tempelweihe)	*Kela-Bela*	"
.....(M-n.d.)	EVOLUTION OF YANKEE DOODLE ...	*Lake*	"
.....(M-n.d.)	EVOLUTION OF DIXIE	"	"
.....(L-n.d.)	COHANESQUE	"	"
.....(H-n.d.)	FEST OVERTURE	*Leutner*	"
.....(H-d.)	SECOND HUNGARIAN RHAPSODY	*Liszt*	"
.....(H-d.)	FIRST HUNGARIAN RHAPSODY	"	"
.....(H-d.)	SIXTH HUNGARIAN RHAPSODY	"	"

387

The breakdown into the facets favors Medium. Out of the 123 listed cues under Agitato, 21 are Light, 53 are Medium, and 49 are Heavy. Under Dramatic, there are 3 Light, 52 Medium, and 40 Heavy from the 96 listed cues. For the 73 Galops cues, there are 24 Light, 41 Medium, and 8 Heavy. The heading Love Themes does not contain the Heavy facet. Instead, it has a facet for Dramatic (DR), and out of the 93 listed cues, 24 are Light, 51 are Medium, and 18 are Dramatic. Mysterioso has four facets, and from the 71 listed cues, there are 16 Light, 35 Medium, 9 Heavy, and 11 Dramatic. The heading Overtures is more complex with the addition of the secondary facet and some possible typos. This is the only term where the number of Medium cues is outnumbered by Light and Heavy cues. There are 102 listed cues, broken down into 43 Light, 10 Medium, 32 Heavy, 11 Dramatic, and 6 (N), a facet that possibly stands for Neutral and is not designated. There are also two cues with "l" as a secondary facet, but this could possibly just be a typo. The cues are further broken down into 10 (L-d.), 33 (L-n.d.), 2 (M-d.), 8 (M-n.d.), 24 (H-d.), 8 (H-n.d.), 2 (D-d.), 8 (D-n.d.), 1 (D-l.), 4 (N-d.), 1 (N-n.d.), and 1 (N-l.).

Agitato (p.32) (H-49), (L-21), (M-53) (123 total)
Dramatic (p. 182) (M:52), (H:40), (D:1), (L:3) (96 total)
Galops (p. 217) (H:8), (L:24), (M:41) (73 total)
Love Themes (p. 289) (DR:18), (M:51), (L:24) (93 total)
Mysterioso (p. 351) (M:35), (DR:11), (H:9), (L:16) (71 total)
*Overtures (p. 387) (M:10) [(M:d.-2) (M:n.d.-8)], (H:32) [(H:d.-24) (H:n.d.-8)], (L:43) [(L:d.-10) (L:n.d.-33)], (D:11) [(D:d.-2) (D:n.d.-8) (D:l.-1)], (N:6) [(N:l.-1) (N:d.-4) (N:n.d-1)] (102 total)

Implied facets

Implied facets occur for the terms Andante, Ballads, Chinese-Japanese, Italian, Oriental, Pizzicatos, Robbery, Russian, Spanish, and Suites. There are no headings for Andantes, Ballads, Pizzicatos, or Robbery, instead the headings appear with modifiers such as Andante (Happy), Andante (Neutral), Ballads (Old), Ballads (Neutral), Pizzicatos Mysterious, Pizzicatos Neutral, Robbery (Light), and Robbery (Serious). Pizzicatos is the only term from this list where the modifying term does not appear in parentheses. This could have been a typographical error, but it is not clear why all of the modifiers from this list do not appear as subheadings in the same manner as those found under Chinese-Japanese and others, where certain subheadings can be found under multiple headings. Formatting inconsistencies are not as interesting as the notion that these modifiers and subheadings are acting as facets.

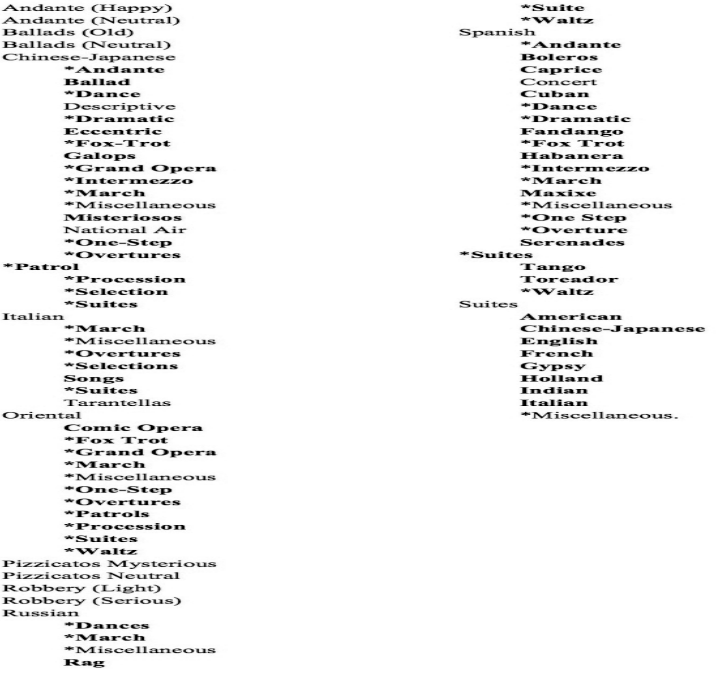
The modifier Neutral is the most obvious facet, because it appears for the terms Andante, Ballads, and Pizzicatos. It is also used as a possible secondary facet. The next obvious modifier acting as a facet is Light. The modifier Mysterious is the English form of Mysterioso, which is a heading that uses the facets examined above. So, although Mysterious only appears as a modifier for one term, it is also tied to a heading, making it a facet, although not in the conventional sense. The same situation applies for Happy and Old, because there are headings for Happy Content, Medleys of

Old Time Songs, and Old Time Songs. Serious is the only modifier that does not really seem like a facet, because it only appears with the term Robbery. Yet, due to all of the other modifiers that are like facets, the term Robbery has the potential to serve as a facet in the encyclopedia.

The subheadings found under Chinese-Japanese, Italian, Oriental, Russian, and Spanish are more interesting than the modifiers. Not only do many of the subheading appear under multiple headings, most of them are also headings found in the encyclopedia, like Ballad, Dance, Fox Trot, and Overtures. This feature certainly makes these terms seem like facets. The term Miscellaneous acts as a conventional facet, because it appears under multiple headings but is not a heading itself.

The heading Suites contains subheadings for countries and peoples, yet these subheadings are also headings. The fact that countries and peoples appear as separate headings and subheadings in the encyclopedia makes perfect sense for the document's purpose. Rapée was not creating an encyclopedia for music scholars interested in musical forms, nor was he creating it for ethnomusicologists who focus on music of all countries and peoples. He created the document for movie theater music directors who needed to match specific sheet music with particular scenes. Often this meant using music to represent countries or peoples, making it necessary to have separate headings for the countries and peoples as well as use them for subheadings. Figure 3 is a taxonomic representation of the faceted parts of the Rapée (bolded words indicate subheading that also appear as headings. * indicates subheadings that appear under multiple headings).

Figure 3. Facets in Rapée's taxonomy



Discussion and conclusions

Of course, Rapée's *Encyclopedia* was not a formal taxonomy nor was that ever the intention of its creator. Rather, it was a working tool for performing film musicians, and to succeed it had to be commercially viable as well. But like many other activity-based domain ontologies, the work-based manual reveals both a core ontology of musical *topoi*, and an underlying epistemology steeped in the cultural norms of early twentieth-century film. One limitation of this study is that we have not analyzed the citations for musical works Rapée included as examples for terminological agreement with the formal terms in the taxonomy.

A film music library manual of the time (Beynon 1921) has an extensive section of "classification" making use of *topoi*, and employing facet-like gathering devices for emotions, atmosphere and other characteristics. Rapée's facets as we have reported them here, follow Beynon's scheme by providing dimensional representations of mood and atmosphere, as well as an additional dimension related to musical difficulty. While the main core ontology represented covers a large gamut of musical styles, the facets provide a way to express both emotional shading for musical illustration, and at the same time a very pragmatic activity-based set of performing criteria.

As we pointed out in the introductory paragraphs, this research emanates from a

pivotal moment in silent film music history that crosses applications from musicians, film music librarians, film music historians, and our own lens from knowledge organization. In addition, the musical content described reveals a multi-tiered visualization of social reality in which musical terms are a lens for the visual content of the films, which like literature, comprise a mixture of representations of reality and fantasy at a certain point in time. A fuller historical analysis might turn to archival records of actual working copies of working Rapée's *Encyclopedia*, in which film musicians had annotated the particular musical works in their repertoires. Such further study would reveal with greater depth the cultural role of music as an aspect of social visualization through silent film performance.

References

Beynon, George W. (1921). *Musical Presentation of Motion Pictures*. New York: G. Schirmer.

Erdmann, Hans, Becce, Giuseppe, &Brav, Ludwig (1927). *Allgemeines Handbuch der Film-Musik*, vol. 2. Berlin: Schlesinger'sche Buch- und Musikhandlung Lienau.

Plebuch, Tobias (2012). Mysteriosos Demystified: Topical Strategies Within and Beyond the Silent Cinema. *Journal of Film Music*, 5(1/2): 77-92.

Rapée, Erno. [1925] (1970). *Encyclopaedia of Music for Pictures*. New York: Arno Press & The New York Times.

Rosar, William H. (2012). Knowledge Organization in Film Music and its Theatrical Origins: Recapitulation and Coda. *Journal of Film Music* 5(1/2): 207-15.

Smiraglia, Richard P. (2014). *The Elements of Knowledge Organization*. Cham: Springer.

Smiraglia, Richard P. (2015). Sounds of Yesterday: Case Study Taxonomy of Topoi from Dutch Silent Film Music. *Canadian Association for Information Science Annual Meeting 2015, Ottawa*.

Smiraglia, Richard P., & Henry, Joshua (2016). Film Music Cues: Visualizing Social Reality Through Music and Film. Submitted to *Canadian Association for Information Science Annual Meeting 2016*.

Van Houten, Theodore (1992). *Silent cinema music in the Netherlands: The Eyl, Van Houten Collection of Film and Cinema Music in the Nederlands Filmmuseum*. Buren: Frits Knuf Publishers.

Camila Monteiro de Barros, Lígia Maria Arruda Café and Audrey Laplante

Emotional Concepts in Music Knowledge Organization

Abstract

According to Peirce's semiotic, the construction of an emotional concept is primarily grounded on immediate experience, and is therefore a diffuse and less stable concept than, for example, a scientific concept. Emotional concepts are often used to express meanings pertaining music information. Knowledge organization theories are mainly concerned in making clear the processes of knowledge representation (both on concept mapping representation and on subject representation) and, thus, has adopted the idea of "concept" as a universe of homogeneous phenomenon. However, there are many specificities in the phenomenological construction of concepts, with certain impact on concept definition. It is in this niche that this discussion is proposed. For this sense, 17 reports of people's semiotic experience with music were collected. Taking music as the sign and emotion as the interpretant, we analysed the objects that are represented by music. Three categories of music objects could be outlined: 1) Vague/undefined, 2) Values/attitudes, and 3) Emotion. As the object is pivotal to the semiosis process, these categories indicate that for music Knowledge Organization it is indispensable to consider the semantic variations that are necessary to express emotion through words, as well as to access the user's semiotic experience report as central source to map terms and concepts within the music domain.

Introduction

According to Peirce, semiosis stems from three correlates: the sign, the object, and the interpretant. Music, which is the sign in this case, represents one or more objects, and creates a meaning in the mind that interprets the sign. The meaning of the sign − the interpretant − can take place at three levels: emotional, energetic, and logical. Following Peirce's phenomenological categories, these levels correspond to firstness, secondness, and thirdness. For example, imagine Beethoven's 5th Symphony. The feelings that this piece could evoke in a listener such as anguish, or the pure contemplation of the songs' movements, are manifestations of the interpretant at the emotional level. But if the listener starts to make a mental effort or to have a physical reaction, the interpretant is at the energetic level. When the listener is capable of constructing a relation between general concepts and the song, such as understanding the music structure, it is a logical interpretant. Amongst the levels of interpretants for music, the emotional level is most prominent, particularly in non-expert listeners, since music is primarily produced with the intention of creating an aesthetic experience.

Savan (1981) suggests that emotion is an unmediated and non-cognitive phenomenon, and that feelings are ill-defined and confused. According to him, the interpretant introduces an emotion as a simplifying hypothesis—which for the author is an "emotional concept". In Peirce's words, "when our nervous system is excited in a complicated way, there being a relation between the elements of the excitation, the result is a single harmonious disturbance which I call an emotion." (CP 2.643).

Thus, the construction of an emotional concept is primarily based on immediate

experience, and is therefore a diffuse and less stable concept than, for example, a scientific concept. Although it is a recurrent theme in music (cf. Martinez, 2001), the emotional level of music meaning has not yet been clearly defined within the field of Knowledge Organization (KO). Current KO theories deal mainly with pragmatic concepts from the point of view of language use, i.e., a concept is "general" that can be inductively tested in particular stances of use (such as literature), and thus its meaning can be reinforced or adjusted. Emotional concepts are not signs of convention; their meaning cannot be inductively tested, since what constitute this kind of meaning are ill defined elements, as explained by Savan above, or as defined by Søresen, Thellefsen and Thellefsen (2016, 5) "*it has no parts, it has no beginning, middle or end*". In this sense it is challenging to essay any fixed meanings for the terms that represent an emotional concept.

Smiraglia (2014, 26) states that "*if knowledge organization is the science of concept-theoretic then the natural first question is how to designate the essential concepts*". It is in this sense that we see a close relation with KO: in one hand, the concept definition process claimed by KO; on the other hand, the construction process of emotional concepts. By taking music as the sign and emotion as the interpretant, the aim of this study is to investigate what objects are represented by music, and to suggest ways of introducing emotional concepts into discussions on KO.

The object of music

Sound is an element of firstness, and thus—particularly in the case of non-specialist listeners—emotional interpretants will always present rather strongly, even when logical interpretants occur (Martinez, 1996). This does not necessarily imply that there will be a strong emotional response, but that the emotional interpretant will be prominent. For Cumming (2000) and Martinez (2001), the object of a musical sign is not clearly defined in the majority of semiotic analyses, although this does not mean that music is lacking references. The "*referential fragility [of the musical sign] is compensated for by its tremendous evocative power, [which] produces in us a kind of predisposition for the dominance of firstness*" (Santaella, 2009, 109). There are no direct references to the object when an emotional concept is created as the interpretant of musical signs (Ibri, 1992), which shows the suggestive potential of music.

As something intrinsic to firstness, the meaning of a musical sign is closely related to the immediate experience of listeners. An example is that a certain piece of music (the sign) evokes "joy" (the interpretant) in a given listener, as it is reminiscent of the success story by the artist (the object). From the perspective ofKO, it is possible to surmise that the concept of joy is part of the music domain, but this is not sufficient to state that this piece of music is joyful. However, we can see that the object of the sign may be described in the biographical information of an artist and perhaps complement the relationship between music information and emotional concept.

Knowledge Organization theories and semiotic approaches

In an overview of KO theories, Smiraglia (2014) states four most influential theories: Dahlberg's concept theoretical, Hjørland's domain analysis, Wilson's and Svenonius' bibliographical approaches. While concept theory is a structural approach to define the attributes of a concept, according to Smiraglia (2014) "*most domain analysis is informetrical, using combinations of citation analysis, author co-citation analysis, co-word analysis, and network analysis to compare visualizations of a domain*". (Smiraglia, 2014, 87). Smiragllia (2014) seems to make it clear that the bibliographical universe is the central point for mapping concepts.

Semiotic-based studies grounded on Peirce's theory have been addressed to discuss KO. Thellefsen (2004, 2002) proposes the *Semiotic Knowledge Organization* in which a knowledge profiling is developed according to epistemological principals of the analyzed domain. These epistemological elements are the basis to eliminate ambiguities in order to achieve terminological precision. The concepts of information and knowledge are also approached through a semiotic prism (cf. Thellefsen, Thellefsen & Sørensen, 2013; Raber & Budd, 2003), including their relation with emotion (Sørensen, Thellefsen & Thellefsen, 2016). Friedman & Smiraglia (2013) analyzed a corpus of 344 concept maps in a way to unveil their semiotic bases. Approximately half the conceptual maps were considered to have semiotic bases that were classified as peircean (triadic sign), saussurean (diadic sign), among others.

As concerning to indexing processes, Almeida (2007, 2011, 2012) moves toward the indexer logic of thought that would occur in a sequence of abduction, deduction, and induction. For Mai (2001, 2011) indexing processes are infinite semiosis. That is, there is an evolutive movement, as Mai (2001) explains: on the document analysis stage, the document is the sign, the ideas and meanings that are registered in the document are the object, and the interpretant will be a global understanding of the subject within the given document. This interpretant becomes the sign of the next stage, in which the indexer describes the subject, which will be the new sign in a new semiosis. After the indexing process is finished, the semiosis has continuity in the process of signification by its users, with subject representation as sign, and so on.

This brief presentation of KO theories and semiotic approaches indicate that KO as a field is concerned with clarifying processes of knowledge representation (both on concept mapping representation and on subject representation), adopting the idea of "concept" as an universe of homogeneous phenomena. When talking about concepts, authors do not address the many specificities that exist in the phenomenological construction of concepts and how they impact concept definition. The same theory cannot be applied for different kinds of concepts, such as scientific and emotional concepts. While the former is adjusted according to object definitions, the latter has no rules for occurrence; as Santaella (2009) states, emotions are as labels that we project onto music.It is within this niche that we propose an initial discussion.

Studies on the relationship between music and emotion

An initial survey in the field of Music Information Retrieval [1] (MIR) shows the importance of the emotional dimension of music within this domain. A search for the term *emotion* in the proceedings of the International Society for Music Information Retrieval Conference (ISMIR) results in 24 studies from the latest four editions of the event (2011 to 2014). Most of these studies deal with applications for automatic music emotion recognition for classification purposes.

According to Lee & Cunningham (2013) there are also numerous studies on emotions in the field of MIR that involve users.Song *et al.* (2013) asked 47 participants to provide emotion ratings for 80 musical excerpts, indicating both the perceived emotion (expressed in music) and the induced (felt) emotion. No significant difference was found between perceived and induced emotion. However, a similar study by Kawakami *et al.* (2013) yielded slightly different results. They asked 44 participants to provide emotion ratings (both perceived and induced) for excerpts and found that perceived and induced emotions did not always coincide: people could perceive music as sad and yet could feel positive emotions by listening to it. Hu and Downie (2007) found there were consistencies in the relationships between emotion/genre and emotion/artist, which points out the possibility of constructing an emotional interpretant through musical genre or artist.

These studies show different approaches, and this suggests there is still little consensus on the subject. According to Peircean phenomenology, emotion is a non-permanent meaning, and this makes generalization of research results more difficult.

Social tagging: the importance of emotional concepts in music Knowledge Organization

The emergence of social tags has helped in trying to understand the emotional dimension of music (Song *et al*., 2013; Kim *et al.*, 2010) as they are assigned directly by users. Lamere (2008) looks at the 500 tags most commonly applied to songs by Last.fm users and find that the top three categories of tags are music genre (68% of tags), artist nationality or song language (12%) and mood (5%).

Instead of looking only at the most frequent tags, Laplante (2015) analysed tags of a random selection (181 songs) in Last.fm, finding that mood/emotion tags accounts for 10.5% of tags, opinion tags for 15.5%, and self-reference tags (e.g., reminds me of you) for 3.9%, for a total of close to 30% of total tags. Bischoff *et al.* (2008) examine a stratified sample of tags in Last.fm and find that genre, artist names, and personal opinion (including mood and emotion) are the three most popular tag categories.

We can therefore suppose that studies on music tagging highlight the fact that emotions and personal context account for a significant proportion of tags, and consequently, for music domain representation purposes.

Methods

Patton's principle of purposeful sampling (2015) was used, interviewing 17 young adults aged between 18 and 29 who listen to music for recreational purposes, with no musical formal education. They were invited to describe an intense musical experience. From their responses, descriptions of semiosis context, sign, object, and interpretant were collected. In this paper, the objects of sign are analyzed. All the interviews were recorded and transcribed, creating a corpus of over 25,000 words. QDA Miner software was used for qualitative analysis.

Results

In some cases, the interviewee reported more than one interpretant and object. Focusing on emotional interpretant occurrences, three categories for the object of the sign were outlined: 1) Vague/undefined, 2) Values/attitudes, 3) Emotion.

Category 2 included cases where the object the music represented for the participant was related to the trajectory of musician or bands, which suggests that biographies could be used as a source for mapping music domain. Category 1 included cases where participants had trouble reporting external references, which indicates that the object is sometimes within the sign. Category 3 included cases where participants explained that the music represented some sort of emotion they could define more or less clearly. Categories 1 and 3 are also instances of music representing itself, causing a kind of immanence of some meanings represented on songs.

How is this relevant for Knowledge Organization purposes?

Knowing the nature of an object is fundamental to understand the nature of a concept, because the object is pivotal in all semiotic processes. As observed in the interviews, when the emotional interpretant occurs it is not possible to define the attributes of an object precisely. This indicates that the sign represents something that is not directly reflected by the object. In other words, it is the individual that interprets the sign and delineates the object.

This research reveals, in sum, three findings:

1) The mapping of music concept domain must consider semantic variations necessary to express emotion through words.
2) There are different levels of precision in the three categories of object. Consequently, some meanings can be more widely shared than others. In this sense, instead of dealing with the meaning (interpretant) itself, it is possible to have objects as evidences of *possible* meanings in music information. Objects from category 2, for example, can be used for concept mapping purposes as music-associated information that will give the user the possibility of choosing the best music according to their expected meaning.

3) Objects from categories 1 and 3 cannot be determined *a priori*, thus the report of user semiotic experience is the central source to map terms and concepts within music domain.

The main theoretical implication of these findings for KO is that objects from categories 1 and 3 (a total of 12 cases) imply that an emotional concept is constructed based on the entire phenomenological experience of a given listener. The emotional concept mostly *suggests* relations with objects which are constructed depending on the occurrence of semiosis. This suggestive (and not directly referential) role results in unstable conceptual relations for domain representation purposes. Further studies are still necessary for a methodological approach for KO purposes, but our theoretical contribution elucidates the nature of objects of emotional concepts and, subsequently its processes of construction and relation with music information as signs.

Note

[1] Music Information Retrieval is a multidisciplinary research area that seeks to develop tools for music management, access and usage (DOWNIE, 2004). The International Society for Music Information Retrieval (ISMIR) is the main representative body in the field, and it organizes yearly conferences. (http://ismir.net/).

References

Almeida, Carlos Cândido de (2007). Peirce e a Ciência da Informação: considerações preliminares sobre as relações entre a obra peirceana e a organização da informação. In *Proceedings of the 8thEncontro Nacional de Pesquisa em Ciência da Informação, Salvador (Brazil)*.[http://www.enancib.ppgci.ufba.br/artigos/GT2--200.pdf?origin=publication_detail]

Almeida, Carlos Cândido de (2011). Sobre o pensamento de Peirce e a organização da informação e do conhecimento. *Liinc em Revista*, 7(1): 104-120.

Almeida, Carlos Cândido de (2012). The Methodological Influence ofPeirce's Pragmatism on Knowledge Organization. *Knowledge Organization*, 39(3): 204-15.

Bischoff, Kerstin Bischoff, Firan, Claudiu S., Nejdl, Wolfgang, &Paiu, Raluca (2008). Can all tags be used for search? In *Proceedings of the 7th ACM Conference on Information and Knowledge Management,Napa Valley (USA)*.[http://dl.acm.org/citation.cfm?id=1458112]

Cumming, Naomi (2000). *The sonic self: musical subjectivity and signification*. Bloomington, Indiana University Press.

Friedman, Alon, & Smiraglia, Richard P. (2013). Nodes and arcs: concept map, semiotics, and knowledge organization. *Journal ofDocumentation*, 69(1): 27-48.

Hu, Xiao& Downie, J. Stephen (2007) Exploring mood metadata: relationships with genre, artist and usage metadata. In *Proceedings of the 7th International Society for Music Information Retrieval Conference, Vienna*.
[http://ismir2007.ismir.net/proceedings/ISMIR2007_p067_hu.pdf]

Ibri, Ivo Assad (1992). *Kosmos Noētos: a arquitetura metafísica de Charles S. Peirce*. São Paulo, Perspectiva & Hólon Press.

Kawakami, Ai, Furukawa, Kiyoshi, Katahira, Kentaro, & Okanoya, Kazuo (2013) Sad music induces pleasant emotion. *Frontiers in psychology*, June.
[http://journal.frontiersin.org/article/10.3389/fpsyg.2013.00311/full]

Kim, Youngmoo E., Schmidt, Erik M., Migneco, Raymond, Morton, Brandon G., Richardson, Patrick, Scott, Jeffrey, Speck, Jacquelin A., &Turnbull, Douglas(2010). Music emotion recognition: a state of the art review. In *Proceedings of the 9th International Society for Music Information Retrieval Conference,Utretcht*. [http://music.ece.drexel.edu/files/Navigation/Publications/Kim2010.pdf]

Lamere, Paul (2008). Social Tagging and Music Information Retrieval. *Journal of New Music Research*, 37(2): 101–114.

Laplante, Audrey (2015). Tagged at first listen: an examination of social tagging practices in a music recommender system. *Encontros Bibli*, 20(1): 33–54.

Lee, Jin Ha, & Cunningham, Sally Jo (2013). Toward an understanding of the history and impact of user studies in music information retrieval. *Journal of Intelligent Information Systems*, 41: 499–521.

Mai, Jens-Erik. (2001). Semiotics and indexing: an analysis of the subject indexing process. *Journal of documentation*, 57(5): 591-622.

Mai, Jens-Erik (2011). The modernity of classification. *Journal of Documentation*, 67(4):710-30.

Martinez, José Luiz (2001). *Semiosis in Hindustani music*. Delhi, Montilal Banarsidass.

Martinez, José Luiz (1996). Icons in music: a peircean rationale. *Semiotica*, 110(1/2): 57-86.

Patton, Michael Quinn (2015). *Qualitative research and evaluation methods*. 4. ed. Los Angeles, Sage Publications.

Peirce, Charles Sanders (1931-1958).*Collected Papers of Charles Sanders Peirce*. Ed. Hartshorne, Charles, Weiss, Paul,&Burks, Arthur W. Cambridge, Harvard University Press, 1931–1958.

Raber, Douglas,& Budd, John M. (2003). Information as sign: semiotics and information science. *Journal of Documentation*, 59(5): 507-22.

Santaella, Lúcia (2009). *Matrizes da linguagem e pensamento: sonora, visual, verbal: aplicações na hipermídia*. 3. ed. São Paulo, Iluminuras & FAPESP Press.

Savan, Davin (1981). Peirce's semiotic theory of emotion. In Ketner, K. L. et al. *Proceedings of the C. S. Peirce Bicentennial International Congress*. Texas Tech University Press.

Smiraglia, Richard P.(2014). *The elements of knowledge organization*. Zurich:Springer.

Song, Yading, Dixon, Simon, Pearce, Marcus, & Halpern, Andrea (2013). Do online social tags predict perceived or induced emotional responses to music? In *Proceedings of the 12th International Society for Music Information Retrieval Conference*. Curitiba https://www.eecs.qmul.ac.uk/~simond/pub/2013/Song-Dixon-Pearce-Halpern-ISMIR2013.pdf]

Sørensen, Bent, Thellefsen, Torkild,& Thellefsen, Martin (2016). The meaning creation process, information, emotion, knowledge, two objects, and significance-effects: Some Peircean remarks. *Semiotica*, 2016(208): 21-33.

Thellefsen, Torkild (2004). Knowledge profiling: the basis for knowledge organization. *Library Trends*, 52(3): 507–14.

Thellefsen, Torkild (2002). Semiotic knowledge organization: theory and method development. *Semiotica,* 142(1/4): 71-90.

Thellefsen, Martin, Thellefsen,Torkild,& Sørensen, Bent (2013). A pragmatic semiotic perspective on the concept of information need and its relevance for knowledge organization. *Knowledge Organization*, 40(4): 213-23.

Tesla Coutinho Andrade and Vera Dodebei

Traces of Digitized Newspapers and Born-Digital News Sites: A Trail to the Memory on the Internet

Abstract

The emerging field of digital memory, alongside the preservation tasks to protect documents originally produced by digital media (born-digital) or scanned from printed paper (digitized), are discussed, based on the research results of the thesis entitled "Digital impressions: journalism and memory in the 21st century". Newspapers and their digital *simulacra* (news sites) are the empirical field for a comparative analysis on information description, storage and retrieval in both media (digitized and born-digital). The analysis corpus for this research was the news coverage of the tsunami disaster in Indonesia, occurred in the last days of December 2004. Two sites were chosen for the empirical study: Folha de São Paulo and The New York Times. Based on the graphic and editorial elements (text, photos, composition and selection) that represent the discursive system of newspaper and news site, we conclude that the main differences from printed to digital shows a context organized by four main categories: time, space, search strategy and information value. The results show that digital memories are possible to be tracked by a narrative discourse related to digitized newspaper, but it is almost impossible when the context offers only a number of isolated conceptual item (memory cell) represented by database information structure.

Background and objectives

In the early modern era, newspapers appear as communication agents between groups, within the space of the growth of cities (Habermas, 2003). The technological development resulting from the industrial revolution changes the landscape of social relations, redraws the geopolitical boundaries of the world and consolidates the newspapers as an organizer of the public sphere. For nearly two hundred years, print newspapers ensured their place as transmitters of everyday information to the population of Western cities. In the 20th century, pictures and color were added to its basic black-and-white printing sheet structure of news fragments, illustrations and advertising, but little has changed to this day.

This study discusses the social life of newspapers and its sustainability as a source of social memories, specifically the transition of its printed editions to news sites in digital mode, seeking to emphasize this issue at the borders of knowledge organization and social memory research fields. Under the prevailing view of the reader / user who wants to retrieve a narrative more than to find isolated and disconnected keywords when searching data (Dodebei Orrico, 2014), the following objectives are underlined: a) identifying the impact of digital technologies on social memories through a newspapers production analysis, after the advent of the internet; b) relating studies that associate newspapers and the construction of memories in Western societies; and, c) comparing the news search strategies and their results in digitized newspapers and born-digital news sites.

Research context

Our hypothesis about the life cycle of daily newspapers is that there is a great change not only in shape but also in their informational structure of registration, storage and particularly in the search strategies of information retrieval in digital media. The theoretical and methodological approaches, found in the scientific literature to answer this preservationist question, relate more to the digital curation in libraries and archives concerned to newspaper collections indexing (Skinner et al. 2012) than to the quality/precision of a search strategy designed to retrieve 'interdisciplinary knowledge' (López-huertas, 2015), considering news published daily in print newspapers and news sites.

Skinner presents a diverse set of newspapers initiatives to organize collections of their digitized editions, as well as the preservation of born digital journalistic production. He notes the lack of preservation standards associated with the varied use of methods, techniques, languages and supplier systems. On these issues, our goal in this article is to discuss the memory preservation of journalistic material, considering the three phases of memory process: registration, storage and retrieval.

Started at the end of the 20th century, studies on cultural memory led some researchers to direct their works to the journalism, the press, whether printed or digital, and to the influence of the newspapers in the construction of groups's memory. Olick, in the essay "Reflections on the underdevelopment relations between journalism and memory studies", defines the relevance of this study field as follows:

> As journalism continuous to function as one of contemporary society's main institutions of recording and remembering, we need to invest more efforts in understanding how it remembers and why it remembers and why it remembers in the ways what it does (Olick, 2014, p.17).

By studying the characteristics of newspaper modernizations in the 20th century, Ribeiro (2007) notes that newspaper is not only made of words. It is an integrated system of text and images (photographs, illustrations, cartoons) that can be associated through visual resources (layout, pagination). This set will reflect the representation of what the newspapers record and how they record the facts in a balance between 'narrative and information', as Walter Benjamin said, when he discusses the loss of narrative in favor of journalistic news. (1994).

Marlene Manoff (2006), when dealing with the textuality of objects and their print and digital modeling, considers that technology itself, used in the creation of an object conveys different meanings, which, together with the narrative, whether visual or sonorous, will provoke a change in the informational content of the object. To highlight the relevance of the relationship between support and materiality of the text, Manoff cites Katherine Hayles:

> She claims that the meaning of a work, whether print or electronic, cannot be separated from its physical manifestation. A reader, viewer or listener's experience is shaped by its material characteristics. […] the structure and organization of hypertext documents are central to what they mean and how the meaning is conveyed. (Manoff, 2006)

Thus, text and content do not have a pure existence, free of the materiality that

conforms them. Both paper and the press, as the chip and digital platforms have a material nature, a "shelf life" and participate in the cyclical gear of the environment. The question that drives us is to know how much we can interfere in this process, in an attempt to maintain the balance between permanence and dissolution of memories in cyberspace.

The choice of newspaper (digitized and born-digital) as an empirical object to discuss memory traces on the internet is supported by three main theoretical axes: the *historical trajectory* of the printed newspaper medium; the changes of its materiality after the arrival of *digital technologies*; and the existing *information* strategies (technologies) to protect and to access knowledge.

Collective memory, in the history of mankind, is the result of the experience of communicative interactions of human groups and the emergence of technological processes for organizing this knowledge. As shown Le Goff (2012), these processes are born by the association of ideas and mnemotechnics in societies without writing; are organized through documents from writing, consolidating the production of artificial memories (outside the individual bodies), increasingly accumulated after the invention of printing. The accumulated external memories have its birth in archives and knowledge organization techniques: indexing, cataloging, abstracting. With digital culture, these techniques, already configured as databases, exponentially increase their processing capabilities and information storage.

Methodology

Research methodology consisted of mapping the printed newspaper in the digital environment, as well as news sites, and of observing how the journalistic record of certain recent past event is seen retrospectively from the internet. The media coverage of the tsunami disaster in Indonesia, which occurred in the last days of December 2004, was the case study for two reasons. It happens in a period in which the online versions of printed newspapers has already been established nearly a decade ago and, secondly, because it was the moment in which printed newspaper industry gave the first global signs of weakness (Pew research center, 2016).

The digital edition of Brazilian newspaper *Folha de São Paulo*, the site *Folha Online* and the site of the US newspaper *The New York Times* represent the research sample, taking into account: the longevity; the national relevance for the public, and because all of them present online journalism projects for over a decade and digitized collections of its print collection.

The observation proposal was to perform, analyze and describe the search for past news considering the two versions of the newspaper; to identify the impact of fragmentary information on memory construction of the event; and, to indicate which attributes were prominent in the transition from the printed digitized format to online versions.

Tsunami was the keyword chosen to delimitate the research - period from December

1st, 2004 to January 31th, 2005. In the digitized collection of the *Folha de São Paulo* the first pages were considered, alongside the pages dedicated to the subject - period from December 27th, the day after the tragedy. From this sample, it was possible to observe the amount of space, the hierarchy and visual resources dedicated to the subject day after day. On the site of *Folha Online*, the same research resulted in an ordered list of the most recent to the oldest file published, totalizing 81 links. The research in each of these files led us to relate links that not appear in the initial search and to recognize a larger universe of publications on the subject, as we shall see.

Data discussion

Regarding the impact of technology in the construction and preservation of social memory, four categories are highlighted in this research: time, space, search strategy and information value. The transition from print newspapers to digital media observes the coexistence of two models opposed: in print, we see that the whole issue is understood as a value to be preserved. This value is the result attributes as selection, hierarchy, temporal and spatial determination, linear narrative, permanence and continuity.

In another way, in the news sites, value is the present time, the instant fragmentation, temporal and spatial indeterminacy, non-linear narrative, impermanence and discontinuity. This inversion of values, however, is still not perceived in the record production context. The function of recording facts and events is maintained by the two models and keeps alive our link with the journalistic information in the middle of all the information coming in through the electronic media.

The breakdown of narrative (the newspaper front page) into memory cells that represent attributes assigned to object in order to facilitate the search strategy purposes (keywords) can impact the construction and preservation of memories. Who will build the narrative is the reader / user and not the editor (Dodebei, 2016).

Technologies move forward bringing new publishing standards for the web, including new graphic designs, but if there is no specific guidance in order to preserve the old pages they are simply disconnected from the new set and disappear as shown by Dantas in the website Buscas.br [1] (2014; 2016).

The structure of news sites shows some attributes suggested by Hand (2008), through a scheme of the impact of new technologies on objects: fragmentation, decentralization, indeterminacy and movement. A news site is a dynamic composition of pieces of information organized through links in a space of indeterminate dimension. It is no longer the date that delimits an issue, nor is the number of pages. Although newspapers continue to produce news in a linear time, the consumption of these reports will happen at any time, according to our access to them. The structure of the sites follows this non-linear model. The route we follow from a record, from today's news or from the past can lead us to multiple paths in and out of a time frame, of a theme and even the site itself. The temporal and spatial limits to the pursuit of information are

dissolved.

On the other hand, we observed that each fragment, each news item may have more than one version. The news is formatted and reformatted and more than one version is published without the editorial concern of a final version. It is noteworthy that indeterminacy, both temporal and spatial, is an observable attribute especially through the search process. In a news site, the present time is always indicated by a clock. All pages of a news site, including those related to the results of a past event, have a header in which the date and time exposed are from the present moment. When we seek and find a news from the past, it will be integrated into a graphic structure where most of the elements refer to the current time. In this context, we can say that, inevitably, the present time overlaps the past.

This present time in the news sites is also perceived as volatile. The composition of the top news at any given time on the cover of news sites is dynamically modified at each moment. The set of highlights or different news is about ten times greater on a news site cover than on a first newspaper page. We have also seen that record temporality is flexible. A news site can stay on the cover for more than one day, or for only a few minutes.

The question that arises is the value of time in relation to the record value. If time is dominant in the structure of news sites and it is indicated in the header of every page, its value is not determined by what it demarcates: the news. Instead, the value is related to the present time that is gone to become past. The covers of news sites are not archived. The value for years attributed to newspaper headlines and front page news ceased to exist on news sites, showing that it is not anymore a memory investiment goal to achieve from those who produce them.

Regarding the search strategies of digitized and born-digital news, we found some pages built in different structures from those presented nowadays. From these evidences, at least some of these older sets pages did not appear in the original search result. By following some links, the search also led us to records that showed how the sites were graphically built in 2004. This discovery shows that the redesigned sites tend to preserve only the titles and texts. The previous layouts and links that lead to it are discarded. Printed newspaper, besides the news, shows its time production characteristics: printing methods, types and paper formats, graphic composition, the use or not of photographs, commercials, among other narratives. The newspaper printed object speaks of himself while the news sites encapsulate their content in new formats, erasing the previous ones.

In addition to the association of ideas from pre-coordinated attributes and keywords, news sites have links that relate news with each other. The use of these links however is restricted and poorly explored in order to compose a range of options to the subject published. What we see are small changes in the scenario explored in this research. The structure of sites has changed since then, as noted above, but in reverse of what we are

emphasizing. Links that relate news to each other, for example, have decreased and, in some cases, have disappeared.

The seek result of a past event from news sites records shows an environment where time and space change its configuration in relation to printed newspapers. Until the advent of the internet, the newspaper consumption experience had a time determined: the date of the edition. This date would serve as reference and time limit to anchor an investigation about the past. The printed newspaper is a dated object, finite in its temporality. Although composed of fragments, it is an object of determined size and volume. It has a beginning, a middle and an end so it is also spatially limited.

The analysis of the first pages set of Folha de São Paulo at the first fifteen days of the tsunami tragedy brought together indicators of the size of the event both in volume - pages per issue - and pages ongoing - consecutive days of exposure of the first page. Such an analysis is not possible in news sites, since the first pages are not archived.

Conclusion

The born-digital newspaper tends to shape the dynamics of the medium and conveys news without: information selection; hierarchical narratives layout; and, linear readings. Digital model, instead of been read, is explored (Lévy, 1993). So, the resulting narrative of selecting and editing issues in particular hierarchy and sequence characteristic of printed editions, is less and less presented in the news sites.

The files of the material produced by newspapers and published directly on the Internet, unlike the material organized editorially in print newspapers, have few pre-established links and this fact reduce retrieval pertinence of what could have been an original narrative. In more organized sites like The New York Times, we can set filters by date, by author, by media, but the narrative varies according to the route of each researcher.

We must invest in preservation models in the digital world (for selected objects) and on their basis are metadata, labels or inscriptions embedded in objects with definitions of attribute (origin, date, author or format) and data about their life cycle. As suggested by Dodebei (2015), metadata are memory structures, virtually prepared to be a new narrative criated by de user.

The organization of information in born-digital environment should not be done only *a posteriori*, as in the case of indexing and archiving by editing the printed collections. Considering that, in the digital world, links among informational objects must be provided also *a priori* through metadata because they will be fundamental to understand, in the future, the records of the present. It is necessary, therefore, to invest in models of organization of information in the origin, at the time of the creation of those objects.

Technically, there is nothing to prevent the retrieval of something produced to be broadcast digitally. But the research suggests that the logic of the original digital production goes through new features of value. The issue understood as a narrative set

as the first page of a newspaper, or even the selection of news on a particular day in a paper copy, seems to be lost. Metaphorically speaking, it is like an unifying layer - the narrative - which gather the particles of a disintegrating newspaper object, dissolving itself in its transition to digital media. How these memory fragments will be gathered in the future to serve as a source of information about the past? This is the task on which we must reflect and work at the present time.

Note

[1] This web collection preserves captures from Brazilian web engines. The snapshots were archived by the Internet Archive, and curated to be part of this collection by the Federal State University of Rio de Janeiro. The web collection was built as part of a research project, which investigates born-digital heritage. Since the collection focuses on the History and Culture categories of the search engines, its crawlers were programmed to capture all the available in-links of those categories. In addition, the web collection is also an attempt to ensure that these fragments became easily accessible. For further information about the collection, also available in Portuguese, please follow the link of the Federal State University of Rio de Janeiro's web site: https://www.archive-it.org/collections/4266.

References

Andrade, Tesla C. (2016) I*mpressões digitais: jornalismo e memória no séc.XXI*: [http://www.memoriasocial.pro.br/dissertacoes-teses.php]

Benjamin, Walter (1994). O narrador: considerações sobre a obra de Nikolai Leskov. In *Magia e técnica, arte e política: ensaios sobre literatura e história da cultura*. São Paulo: Brasiliense. Pp. 197-222.

Dantas, Camila Guimarães (2014). *Criptografias da memória: um estudo teórico-prático sobre o arquivamento da web no Brasil*: [http://www.memoriasocial.pro.br/documentos/Teses/Tese42.pdf]

Dantas, Camila Guimarães (2016). *Buscas. Br: Brazilian web engines (1997-2013)*.
[https://www.archive-it.org/collections/4266]

Dodebei, Vera & Orrico, Evelyn (2014) Knowledge in Social Memory: empirical experiment for domain conceptual-discursive mapping. *Proceedings of the Thierteenth International ISKO Conference, Krakóv, Poland, 9-22 May*. Pp. 65-72.

Dodebei, Vera (2016). Ensaio sobre memória e informação. *Revista Morpheus: estudos interdisciplinares em memória social*, 9(15): 227-45.

Folha online. [http://www.folha.uol.com.br/]

Habermas, Jürgen (2003). *Mudança estrutural da esfera pública*. Rio de Janeiro: Tempo Brasileiro.

Hand, Martin (2008). *Making Digital Cultures. Access, interactivity, and authenticity*. England; USA, Ashgate Publishing.

Le goff, Jacques (2012). *História e Memória*. Campinas: Unicamp.

Lévy, Pierre (1993). *As tecnologias da inteligência: o futuro do pensamento na era da informática*. São Paulo: Ed. 34.

López-huertas, María J. (2015). Domain analysis for interdisciplinary knowledge domains. *Knowledge Organization*, 42(8): 570-80.

Mannof, Marlene (2006). The materiality of digital collections: theoretical and historical perspectives. *Portal: Libraries and the academy*, 6(3): 311-25.

Nytimes.com. [http://www.nytimes.com/]. Accessed 13 January 2016.

Olick, Jeffrey (2014) Reflections on the underdeveloped relations between journalism and memory studies. In: Zelizer, Barbie (Ed). tenenboim-weinblatt, Keren. *Journalism and memory*. New York: Palgrave Macmillan.

Pew research center. *State of the News Media*. [http://stateofthemedia.org/] Accessed 15 Jan. 2016.

Ribeiro, Ana Paula Goulart (2007). *Imprensa e história no Rio de Janeiro dos anos 1950*. Rio de Janeiro: E-papers.

Skinner, Katherine, Schultz, Matt, Halbert, Martin, &Phillips, Mark Edward (2012). Digital preservation of newspapers: findings of the Chronicles in Preservation Project. *UNT Digital Library*. [http://digital.library.unt.edu/ark:/67531/metadc109727]

Giulia Crippa, Deise Sabbag and Márcia Regina Silva

The Bibliographic Gesture in Knowledge

Abstract

Based on *Institutione Divinarum Litterarum* (IDL), the present text reflects on the bibliographic practices in a dimension of the bibliographic gesture, aiming to connect with contemporaneity. The bibliographic method was used: a primary source (Cassiodorus' text) and texts in the field of Knowledge Organization were analyzed. As he brings to light the principles of his bibliographic proposal, Cassiodorus constructs a catalogue *raisonné* that offers all the extent of the territory and of the geography of the Christian knowledge of that time, which is to be explored in that specific trajectory of knowledge. In the field of Knowledge Organization, the ontologies and the semantic links which construct concept maps of specific knowledge can also be considered examples of bibliographic gestures, in the sense that they create systems for knowledge organization and information retrieval. It follows that the historical reconstitution andthe conceptual issues of the field of Information do not depend on the circumstances of the times; they connect to problems more adequately, whereas the intervention of technology solves those problems more efficiently; therefore, the nature of the "bibliographic gesture" is basically preserved in contemporaneity.

Introduction

This is a reflection on the bibliographic practices in a dimension of the bibliographic gesture based on Cassiodorus' *Institutione Divinarum Litteratum* (IDL) (1), a treatise from the fourth century A.D. about a "cartography" of knowledge by means of its records.

We consider that the discussion comprising the bibliographic production from before the printed book is sparse, even though the definitions of bibliography available do encompass the bibliographic accomplishments relative to the set of manuscripts already available in the High Middle Ages. In fact, one of the definitions of bibliography, one provided by Cunha and Cavalcanti (2008, 46), is "The systematic production of descriptive lists of records of knowledge".

In this sense, we cannot help but consider the work of Cassiodorus as fitting those authors' definition, since the treatise constitutes one of the first examples of bibliographic production with a st rong component of systematization.

Another definition that underlies this work states that bibliography is the

> discipline that studies the methods, the techniques and the theories for creating and compiling bibliographies, as well as the structural analysis and the description of literary resources. That broad definition comprises the study of books as products of a certain society's technique, typography and culture, for the purpose of analytically reconstructing and assessing its physical characteristics (Bibliografia, s.d., s.p.).

Assuming that there is no specific theorization of bibliography in the sixth century, which is when Cassiodorus lived, we intend to develop our discussion, trying to depict bibliographic practices in a dimension of "bibliographic gesture", by which we mean an accomplishment that is not limited to an organization of individualized space in a library, but which comprises a broader idea of "Knowledge Organization". In this sense, we consider the bibliographic gesture as being the foundation of a knowledge

apparatus, assembled by the systematic elaboration of a list based on clear, explicit, critical selection criteria, and by its materialization in medieval libraries, based on the definition of apparatus provided by Agamben (2014, 25):

> a heterogeneous set that includes virtually anything, linguistic and nonlinguistic, under the same heading: discourses, institutions, buildings, laws, police measures, philosophical propositions, and so on. The apparatus itself is the network that is established between these elements.

The current technologies, the diverse sources of information and the international Information Organization (IO) standards characterize the practices of epistemological nature of information processing work as different from those adopted by Cassiodorus, which focused on reading guidance, in order to provide a structured apparatus of materials. It seems to us that the IDL is the materialization of the ideals and of the instruments of a Knowledge Organization centered around specific interests and possible records.

Even though contemporary bibliographic practices, connected to practically endless repertoires composed of technological networks, are not considered bibliographic gesture, in fact there is a need to elaborate on this concept in the current scenario, across the elements that arise through the superimposed layers of technology for information storage, reproduction and circulation, so we can realize the permanence of gesture in its sense of *techné*, "pragmatic" construction of ways, a journey of knowledge through selected records.

Therefore, we will discuss some elements that can depict this idea of bibliographic gesture as an occurrence that is rooted in practices much earlier than the printed book and which surpass the limits currently imposed by the definitions concerning new technologies.

This text aims to reflect on these bibliographic practices in a dimension of the bibliographic gesture based on *Institutione Divinarum Litterarum (IDL)*, in order to connect with contemporaneity.

In addition to that introduction, we will seek to map the notion of bibliographic gesture. Next, we will observe some aspects of this bibliographic gesture in times prior to printing. Finally, we will observe how the idea of bibliographic gesture may be transposed to informational environments generated by the new technologies.

This work was created using the bibliographic method: a primary source (Cassiodorus' text) and texts in the field of Knowledge Organization were analyzed

Bibliographic gesture: itineraries

What we intend to depict as bibliographic gesture is individualized in some concepts that point to a new pragmatic dimension of bibliography, as argued by Menezes (2015) in his article entirely devoted to bibliographic gestures.

According to Agamben (2008, 12-14)

> What characterizes gesture is that, in it, one neither produces nor acts, but admits and withstands. That is, gesture opens the sphere of ethos as the sphere which is the most characteristic of man [...] Gesture is the exibition of a mediality, of making visible such a medium. [...} Gesture is [...] Communication

> of a communicability. It does not have anything of its own to say, because that which it shows is the being-in-the language of man as pure mediality. [...] exposure of man's being-in-the-language: pure gestuality.

The sense of "bibliographic gesture" is structured from a concept pf Bibliography that "provides a compelling advantage to study the interaction between classification, rhetoric and knowledge building" (Paling, 2004, 588), crossing concepts (reflections on tasks, needs, actors that the bibliography satisfies and for whom it operates) and actions ("acts" of selection, registration and organization of materials). Gesture is therefore defined as the moment when the bibliographer establishes his or her meanings for the data, orienting them within a framework of socially shared knowledge to whose development he or she contributes, by developing techniques and by choosing technologies, as he or she manipulates the knowledge produced.

Bibliography as a joint activity of "documentation collection and information organization" (Balsamo, 1995, 8) pervades the institution of the modern library from Gesner and Naudé to Otlet in the 20th. By that statement we mean to further specify the idea of "bibliographic gesture": not a mere description of materials haphazardly given to the bibliographer, but an active attitude of inquisitiveness and guidance of users. It seems clear to us that the bibliographer is not "neutral", whether or not he transfers the problem to the digital environment.

Notes on the *Institutione Divinarum Litterarum*

Flavius Magnus Aurelius Cassiodorus was the creator of the monastery Vivarium. He built a monastery dedicated to Saint Martin, intended as a Christian school centered around the study of the Scriptures, albeit with the help of texts by pagan authors (Courcelle, 1948).

To structure the school, Cassiodorus implemented a library of Bible commentators, as well as well some classical authors, thus making it rich in choice and variety.Preservation, copying, restoration and transcription of manuscripts were performed on a daily basis. Vivarium was always considered a significant place for book production in the sixth century.

The *IDL* treatise, due to the way it divides the disciplines and due to its proposal of an order (re)produced in the library, becomes an example of a true "Information policy", intended to provide a thorough Christian education, by means of a detailed text filled with titles. However, it is not just a list of books contained in the library founded by its author: Cassiodorus' proposal is, indeed, a selection of titles for the monks to follow a path of enthralling knowledge. It combines suggestions and recommendations, and it indicates the availability of materials or the need to look for them.

Reading the *IDL* transports us to the universe of the medieval Christian manuscript: a manual to use the library, a catalogue and, using an seemingly anachronistic term, a bibliography *raisonnée* in which we recognize and encounter the "bibliographic gesture" as something predates its modern (re)invention.

In the "bibliographic gesture" made by Cassiodorus at the moment of the collapse of the structures of Late Antiquity, we observe the trajectory of a reorganization of Christian and pagan records and knowledge in search of an itinerary of the knowledge considered not only legitimate, but also structural for society.

Cassiodorus' IDL is presented as a bibliographic catalogue in the condition of an *apparatus* of organization and retrieval of that information that shifts the borders between disciplines with its materials, potentially assigning them a spatial *locus* in the philosophical/physical architecture of a library.

An apparatus is thus constructed by the library at Vivarium, by its practices of guidance (by means of the textuality connected to its management, administration and logistics), by its selection and organization, by way of a catalogue of the manuscripts that reflect the access to and the use of information in the technical and technological context of the time. From that point of view, we see an evident cross between the construction of the *locus* of the catalogue and the discourses, institutions and philosophical propositions of the time.

In the *IDL*, the mission of the library is to provide a coherent Christian education marked by a bibliography that aims to chart an innovative course in the chaotic political, social and cultural reality of the Western High Middle Ages. The IDL possesses the scope to concretely show the choices for the collection of a library oriented towards Christian education, indicating the characteristics that lead to the choice of a book and to its conceptual and/or spatial "position" (Halporn, 1981).

It possesses a system in its exposition of its structure, which is organic and consistent with a discipline-oriented architecture and with the bibliographic importance of the author. This leads to an itinerary through the knowledge developed by the catalogue and built from the foundations up, in a library that is not a mere set of documents, nor a collection of volumes: the *IDL* is the materialization of the ideals and instruments of a knowledge organization oriented toward specific interests (Christian civilization) and possible records, which is made evident by the affirmation (parchment, papyrus).

The IDL was written for the sole purpose of providing the monks at Vivarium with a complete bibliography which would subsequently spread to other monastic centers (O'donnell, 1979). Throughout the nine chapters of the *IDL*, Cassiodorus provides the foundations of Bible study, through commentators and through classical literature, selecting texts that would enable the monks to follow the program of Christian education.

Cassiodorus' bibliography contains a careful description of the utility of some authors, taking care, in many cases, to point out heretic opinions or, simply, stating what parts should be read and which parts should not. The *IDL* also indicates those authors whose works should be transcribed in the Vivarium *scriptorium* and those authors whose works the monastery already had copies of. In sum, we cannot consider

IDL as a catalogue of only one library but, instead, a work that seeks to delineate the field of Christian studies through pondered indications of texts, some of which were at Vivarium. We do not know for a fact what texts Vivarium did have, for Cassiodorus chose an itinerary of readings that combines suggestions, recommendations, references to texts the monastery had and *desiderata*.

In this sense, the *IDL* represents two aspects of the bibliographic gesture made by its author: first, he establishes his monastery as a benchmark for Christian studies by Cassiodorus provides its library with manuscripts. Next he exposes the principles of that education through the titles and the content that he considers essential, specifying throughout the text which books should be read to accomplish the program and, among them, which were contained in the library. We consider his work as being within the definitions of bibliography and gesture, since it is not limited to only one space, but encompasses a broader idea of "knowledge organization".

Bibliographic gesture and contemporaneity

In the present day we do not identify the bibliographic gesture as manifested by Cassiodorus in the catalogue of Vivarium, but traces of bibliographic gestures may characterize bibliographic practices related to the practically infinite repertoires of the technological networks.

By bringing to light the principle of his bibliographic proposal, Cassiodorus constructs this early catalogue *raisonneé* which covers the whole extension of the territory and of the geography of Christian knowledge of that time, which were to be traveled by that specific trajectory of knowledge.

In the field of Knowledge Organization, the ontologies and the semantic links that create concept maps of specific knowledge may also be considered examples of bibliographic gestures in the sense that they create systems for knowledge organization and information retrieval.

These two knowledge organization systems create links between terms and concepts by means of complex relationships that effectively present the reader with a vast repertoire, a way, an itinerary that extrapolates conventional systems of knowledge representation, such as controlled vocabularies and thesauri.

The literature of this field of knowledge mentions various conceptualizations and definitions of ontology (Schiessl & Shintaku, 2012; Currás, 2012) which may be understood as structures that congregate standardized sets of basic concepts (Dahlberg; 1983a, 1978b), terms, definitions and the relationships between them, that is, the common vocabulary of a community that needs to share information within a certain field of knowledge, information that is to be interpreted by a machine (Campos, 2010).

Semantic links connect informational resources by means of interconnected open data. That is the proposal of Linked Open Data to interconnect open data that are made available by Web semantics, RDF and URI (Bizer, Health & Berners-Lee, 2009). That might be the future of the accomplishment of the bibliographic gesture in the context of

Knowledge organization since Linked Open Data may enable the large scale integration of the contents of archives, libraries and museums (Marcondes, 2012). By breaking the physical space limits of informational environments, interconnected open data will enable bibliographic data to be accessed thus creating an itinerary, a map for the reader/user. We can consider that this follows the model of network and of rhizome, exceeding the hierarchical limitations of a databank (Allison-Cassin, 2012), thus allowing for a semantic map of knowledge, much like Cassiodorus' work did.

Final Considerations

What emerges over the long time of a cultural history that sets off to research the permanence and the changes of the concept of library throughout the ages is a reflection on the practices and creations of information models from the past, based on which it becomes possible to accurately measure the processes of technological innovation.

However, when it comes historical reconstruction, the conceptual questions in the field of Information often stand out as being independent from the circumstances of the times. Instead, they connect more adequately to the problems that the intervention of technology solves more efficiently. Basically, they preserve the nature of the "bibliographic gesture" in contemporaneity.

In that sense, the *IDH* is much more than just a library catalogue, a manuscript already void of meaning, except for High Middle Age historians. Instead, it is a text that reveals the multiple facets of an apparatus of manuscript culture, which is concerned with its physical immanence in the library and with the philosophical dimension of knowledge, shedding light on the descriptive and semantic aspects of the documents and of the knowledge therein, thus contributing to a history that becomes anthropology as it studies the bibliographic gesture within the conceptual dimension of bibliographic discipline.

Note

[1] The title Institutione Divinarum Litterarum (IDL) was taken from the original document titled retirado do documento original intitulado "Opera Omnia in duos tomos distributa" by Flavius Magnus Aurelius Cassiodorus, volume 70. The reader may find on the Internet diferente titles for that part of the document, such as "Institutiones divinarum et saecularium litterarum". It is important to stress that this difference does not imply a mistake, it just means that, for the purposes of this study, the primary source written by Cassiodorus was used.

References

Agamben, Giorgio (2014). *O amigo & O que é um dispositivo*? Chapecó: Argos.

Agamben, Giorgio (2008). Notas sobre o gesto. *Artefilosofia*, 4: 9-14.

Allison-Cassin, Stacy (2012). The Possibility of the Infinite Library: Exploring the Conceptual Boundaries of Works and Texts of Bibliographic Description. *Journal of Library Metadata*, 12: 294-309. [http://www.tandfonline.com/doi/pdf/10.1080/19386389.2012.700606] .

Balsamo, Luigi (1995). *La bibliografia: storia di una tradizione*. Firenze: Sansoni.

Bibliografia (s.d.). *Nuovo soggettario thesaurus*. [http://thes.bncf.firenze.sbn.it/termine.php?id=5401&menuR=1&menuS=2]

Bizer, Christian, Heath, Tom, & Berners-Lee, Tim (2009) Linked data: the story so far, In T. Heath, M. Hepp, C. Bizer (eds.), *Special Issue on Linked Data, International Journal on Semantic Web and Information Systems (IJSWIS)*. [http://tomheath.com/papers/bizer-heath-berners-lee-ijswis-linked-data.pdf]

Campos, Maria Luiza de Almeida (2010). O papel das definições na pesquisa em ontologia. *Perspectivas em Ciência da Informação*, 15(1):220–38.

Cassiodoro, Flavio Magno Aurelio. (s.d.) Opera Omnia in duos tomos distributa. In Migne, J.P. *Patrologia Latina Tomus 70*. Gareth, J. (ed.). Brepols: s.e.

Courcelle, Pierre (1948). *Les lettres grecques en occident: de Macrobe a Cassiodore*. Paris: E. de Boccard.

Cunha, Murilo Bastos, & Cavalcanti, Cordélia Robalino de Oliveira (2008). *Dicionário de biblioteconomia e arquivologia*. Brasília: Briquet de Lemos.

Currás, Emilia (2010). O*ntologias, taxonomies e tesauros*. Brasília: Thesaurus.

Dahlberg, Ingetraut (1983). Conceptual compatibility of ordering systems.*Internacional Classification*, 10(2): 5-8.

Dahlberg, Ingetraut (1978). Teoria do conceito. *Ciência da Informação*, 7(2): 101-7.

Halporn, James W. (1981). Methods of reference in Cassiodorus. *Journal of Library History*, 16(1): 71-91.

Marcondes, Carlos Henrique (2012). "Linked data": dados interligados e interoperabilidade entre arquivos, bibliotecas e museu na Web. *Encontros Bibli*, 17(34): 171-92.

Menezes, Vinícios (2015). O gesto bibliográfico e a Modernidade. *Informação e Informação*, 20 (2). [http://www.uel.br/revistas/uel/index.php/informacao/article/view/23129/pdf_64]

O'Donnell, James L. (1979). *Cassiodorus*. Berkeley: University of California Press.

Paling, Stephen (2004). Classification, rhetoric, and the classificatory horizon. *Library Trends*, 52(3): 588-603.

Schiessl, Marcelo,& Shintaku, Milton (2012). Sistemas de organização do conhecimento. In. Alavares, Lillian (Org.). *Organização da informação e do conhecimento*. São Paulo: B4 Ed.

Gustavo Silva Saldanha and Naira Christofoletti Silveira

The Treasure of Tesauro: Knowledge Organization, Rhetoric and Language

Abstract

Considering the studies of Ranganathan's theory and the post-ranganathanian studies on faceted analysis and documentary languages in Knowledge Organization, such as the logical analysis of the Indian mathematician carried out by Birger Hjørland (2013), we hereby suggest an epistemological-historical discussion on the potentials of rhetorical analyses about the same phenomenon. In his "The Infinity of Lists", Umberto Eco draws attention to Emanuele Tesauro's work, a XVIIth century erudite intellectual, author of Il Cannocchiale Aristotelico, published in 1670, essential work for the modern understanding of the metaphor as a possibility of understanding of the world through language. Tesauro regards his work as a relationship among oratoria, lapidaria, and simbolica, based on the principles of the Aristotelian Rhetoric. The main focus of the present reflection lies in the clear potentiality of comprehending a complex theoretical approach in Knowledge Organization, potentiality which, placed in the present time, becomes an "epistemic enigma": the metaphoricity in Tesauro (the man of the Italian Six Hundreds) directs thinking to the most recent discursive constructions and empirical challenges in studies of Knowledge Organization, like those in search of descriptions of conceptualizations and the exploration of dynamic relationships of a syntactic, semantic, and pragmatic character in the context of digital networks. With theoretical and methodological approach of a historical epistemology, based on Wittgenstein's thought, this paper intends to recognize the proposal of Emanuele Tesauro and correlates it with contemporary approaches to Knowledge Organization.

Preliminary considerations

Knowledge Considering the studies of Ranganathan's theory and the post-ranganathanian studies on faceted analysis and documentary languages in Knowledge Organization, as the analysis of logical intakes of the Indian mathematician carried out by Birger Hjørland (2013), we hereby suggest an epistemological-historical discussion on the potentials of rhetorical analyses about the same phenomenon. Therefore, we pose the following question: How can the practices of Knowledge Organization be reconsidered from the point of view of the metaphorical construction of the representation of documental reality?

In this sense, the method of study suggested here is founded on a theoretical-conceptual basis and it aims at the understanding of historical and conceptual elements which can contribute to the theoretical innovations applied in Knowledge Organization in our contemporary context. Specifically, the methodological proposal seeks to foster discussion about the philosophy of language and the relationship between Rhetoric and Knowledge Organization, having the support of Emanuele Tesauro; thinking, a scholar of the XVIIth century. Repercussions of his thinking are approached, in the philosophy of language, starting from Umberto Eco. In the scope of Information Science, Ranganathan and Hjørland thinking is adopted as a source for the discussion of the rhetorical impact on Knowledge Organization.

In his "The Infinity of Lists", Umberto Eco draws attention to Emanuele Tesauro's work, a XVIIth century scholar, author of *Il Cannocchiale Aristotelico*, published in

1670, essential work for the modern understanding of the metaphor as a possibility of understanding of the world through language. Tesauro regards his work as a relationship among *oratoria*, *lapidaria*, and *simbolica*, based on the principles of the Aristotelian Rhetoric. Tesauro's ideas anticipate key issues in the Philosophy of Language, existing, for instance, in authors like Charles Peirce, Nietzsche, Wittgenstein, and Ricoeur, between the XIXth and the XXth centuries. Umberto Eco (2010) identifies in Tesauro a proposal of a model of the metaphor as a mode to discover unprecedented relationships among knowledge data. In Eco's (2010) view, for Tesauro, finding metaphors means, in an Aritotelian sense, getting to know new determinations of things, that is, all that could be said about an object.

The main focus of the present reflection lies in the clear potentiality of comprehending a complex theoretical approach in Knowledge Organization, potentiality which, placed in the present time, becomes an "epistemic enigma": the metaphoricity in Tesauro (the man of the Italian Six Hundreds) directs thinking to the most recent discursive constructions and empirical challenges in studies of Knowledge Organization, like those in search of descriptions of conceptualizations and the exploration of dynamic relationships of a syntactic, semantic, and pragmatic character in the context of digital networks.

Therefore, it is a theoretical research, based on the approach of a historical epistemology of the studies of Knowledge Organization. Here we aim at a bibliographical-conceptual study in the construction of the concept of "thesaurus" starting from a rhetoric perspective, in its condition either as a theoretical approach or as documentary language tools, having for central focus the ideas of the Italian thinker Emanuele Tesauro. In general, the immersion in Aristotle's thinking – like Hjørland's (2013) analysis about Ranganathan – done in Knowledge Organization, initiates with his categorizing principles, for example, the Stagirist's five predicates (gender, kind, difference, property, and accident), concentrated in the reading of Organon. However, what is forgotten is that Aristotle's influence in the West happens either via "logicism", or via "discourse", that is, via consolidation of a reflection on language beyond Logic.

Emanuele tesauro: theoretical-conceptual issues

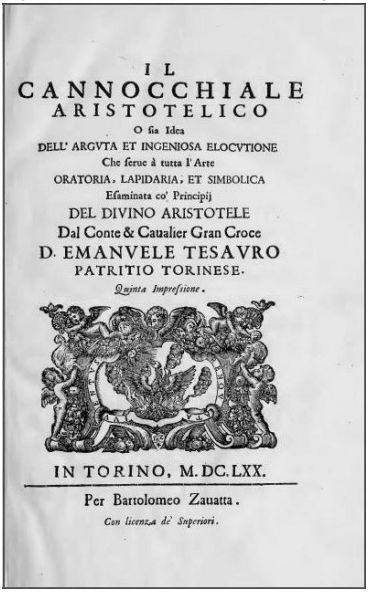
In a historical sense, we know that the term "thesaurus" has different meanings and uses. It appears as dictionary synonym or as linguistic tool similar to the vocabulary repertoire of a given tradition or context, like the *Thesaurus linguae graecae*, cited by Otlet (1934: 141) and *Thesaurus linguae latinae*, indicated by Peignot (p. 253) and the *Thesaurus linguarum orientalium, turcicae, arabicae, persicae …* (p. 346), and besides as grammar descriptions, which is the case of the *Thesaurus linguae armenicae antiquae et hodiernae* (p. 353), or compilation of medals (the example of the *Thesaurus Morellianus*) (p. 346).

Despite being mentioned, from time to time, in the foundations of studies on theoretical and applied approaches on the development of documentary languages, Emanuele Tesauro did not win any prominent chapter in the philosophical reflection on Knowledge Organization. His ideas appear, for example, in the scope of Information Science, in Almeida & Crippa (2009) and in Monteiro & Giraldes (2008). The latter say that it is possible to identify in Tesauro, the Aristotelian thinker, the creation of the thesaurus, an instrument of documentary language; in Almeida & Crippa (2009), Emanuele appears as a prominent baroque thinker, representing the lineage that points to the machinable potential of the metaphor to transform objects. In other cases, like in Dodebei (2002) and Gomes, Tesauro's pioneering spirit is not restricted, neither historically, nor theoretically, least of all instrumentally. According to Gomes (1990), the "term 'thesaurus' has its origin in Peter Mark Roget's analogical dictionary, entitled 'Thesaurus of English Words and Phrases', published, for the first time, in London, in 1852", historical view shared by Dodebei (2012). Apart from informational studies, Emanuele Tesauro's thoughts appear, recurrently, in discussions about the philosophy of language, as in Eco (2001), and about philology and linguistics, as in the works of Molina Cantó & Chiuminato (2004) and Proctor (1973).

Tesauro's thinking anticipates central issues of the Philosophy of Language, present in the works of authors like, for example, Charles Pierce, Nietzsche, Wittgenstein, and Ricoeur, between the XIXth and the XXth centuries. Umberto Eco identifies in Tesauro's ideas the proposal for a metaphor model as a mode to discover unprecedented relationships among knowledge data. According to Eco, for Tesauro, finding metaphors means, in an Aristotelian sense, getting to know new determinations for things, that is, all that could be said of an object. The curious and/or historical "co-incidence" of the proper name "Tesauro" and the term of the instrument of documentary languages (entitled "thesaurus") is only a marginal clue to this study.

A obra *Il cannocchiale Aristotélico*, é dividida em 19 capítulos, a saber: 1. Delle argutezze e suoi parti; 2. Cagioni efficcienti delle argutezze Iddio, Spiriti, Natura, Animali et Huomini; 3. Cagioni Instrumentali delle argutezze oratorie simboliche et lapidarie; 4. Cagion formale dell'argutia circa le figure; 5. Delle figure poetiche o concertative; 6. Delle figure ingeniose; 7. Trattato della metáfora; 8. Delle metafore continuate: et prima delle propositioni metaforiche, lequali comprendono i più bei motti arguti et l'allegoria; 9. Degli argomenti metaforichi et dei veri concetti; 10. Causa finale: et materiale dell'argutezza; 11. Teoremi pratici per fabricar concetti arguti; 12. Trattato dei ridicoli; 13. Trattato delle inscritioni argute; 14. Passagio dalle argutezze uerbali a quelle dei simboli in figura, ò in fatti; 15. Idea delle argutezze heroiche vulgarmente chiamate Imprese; 16. Trattato degli Emblemi; 17. Dei reuersi delle medaglie; 18. Deffinitione, et essenza di tutti gli altri simboli in fatto. 19. Insertivarii et ingegnosi di tutte lê specie simboliche fra loro: et dell'arte lapidaria com la simbólica. (Moraes, 2010).

Figure 1. Title page of Il cannocchiale Aristotélico (Tesauro, E. Il cannocchiale Aristotélico. Berlin: Verlag Gehlen; Zürich: Bad Homburg v. d. H., 1968)



I L
CANNOCCHIALE
ARISTOTELICO
O fia Idea
DELL' ARGVTA ET INGENIOSA ELOCVTIONE
Che ſerue à tutta l' Arte
ORATORIA, LAPIDARIA, ET SIMBOLICA
Eſaminata co' Principij
DEL DIVINO ARISTOTELE
Dal Conte & Caualier Gran Croce
D. EMANVELE TESAVRO
PATRITIO TORINESE.
Quinta Impreſsione.

IN TORINO, M. DC. LXX.
Per Bartolomeo Zauatta.
Con licenza de' Superiori.

His purpose, as seen before, is to prescribe a mode of "writing", which we can deal with as "mode of representing" what is real. As in traditional models of rhetorical knowledge, it is possible to observe here, an immediate belief in language, not only as a tool, but as an object for the construction of what is real as well.

> As a treaty, the Cannochiale follows the model of discourses, mainly appealing to the use of proofs, examples, and with them constructing its argumentative process. While discussion advances in the horizontal axis (that encompasses knowledge of all aspects of rhetoric which Aristotle worked on) and the vertical axis (with a historical and cultural overview of the events privileged by the author), the metalinguistic character of the work consolidates the discussion using the Aristotelian model itself. (Moraes, 2010, page not registered, translated by the author of the article)

As it occurs with Rhetoric, language is produced in the construction of the fold of metalanguage: "things" are created from the relationship of the overlap of languages, of discourses, of comments, of resumption of dialogues.

> Tesauro [practices] his Aristotelian rhetorical exercise to defend the current usage of the reflections of the Greek thinker. This usage is guided by aspects of development and maturity in each language, attained through balance, in Ancient Times obtained from reflections concerning remote practices, systematized through debates, ruptures, and conclusions recorded in the texts of thinkers, from which Aristotle stands out, while for Tesauro, it is prescribed through the experience of approximately fifteen centuries that document phases of Latin writing and later on of Romance languages (particularly Italian in the phase discussed in the Cannochiale Aristotelico). (Moraes, 2010, page not registered, translated by the author of the article)

Tesauro's works, based on reflections about poetry and rhetoric, aim at presenting a modus operandi, a technique, and a teaching method, for the practice of writing.

Introduced in the “catalog” arrangement, it discusses a large array of examples of possibilities of imitation/emulation in order to elaborate an “astute” text. The text is characterized by metalanguage (derived from Rhetoric), by trying to construct rules, combining particularly the relationship between teaching-imitating and teaching practice by means of imitation. The metalinguistic character of Tesauro's text evokes not only a way to understand and restore the Aristotelian rhetoric, but of establishing a writing practice as well, along with the presentation of the method of such practice, full of the intertwining of information, concepts, practices, in the form of comments. All these possibilities, at the same time, sound like a “prescription”, in the sense of exploring ways of writing, speaking, and representing as it should be done, a source that exists for every great.

The “problem” in Ranganathan's thinking that Hjørland (2013) discusses is of direct interest to us, as it deals, prior to an “issue of Logic”, with an “issue of Trivium”, that is, the relations among Grammar, Logic and Rhetoric in the Indian philosopher's texts. More specifically, the key to a more critical and open interpretation of faceted classificationism would be in an immersion in the idea of reality representation starting from rhetorical assumptions. Here we meet Tesauro (the scholar) in order to rethink the thesaurus (the theoretical input and the instrument of thematic representation).

Rereading Ranganathan and his successors it is necessary to notice not only the presence of Logic in his classifying discourse, but the rhetorical potential (from the rhetorical science presented in it) as well. First and foremost, we realize, pointing to the presence of Tesauro's thinking, the importance of the metaphor to knowledge organization. Despite Hjørland's (2013, p. 555) assertionson the primacy of Logic in faceted analysis, it may also receive “pragmatic reading”, based on Rhetoric, which can reset the metaphor at the head of the studies on knowledge representation.

> Facet analysis is primarily a logical approach to classification and knowledge organization. Although the methodological principles also sometimes mention empirical elements (such as examining a representative sample of texts) and pragmatic criteria (such as producing the most helpful classification), these elements are so vaguely peripherally described that they do not change the general conclusion of FA as a rationalist approach based on a priori knowledge, not on empirical knowledge or on historical or pragmatic methods.20 When concrete classifications are produced (such as the single volumes in the BC2 system) the classifiers do, of course, consult libraries and terminology lists. This part of the methodology is not well described, however. It is not described what differences it makes whether the empirical work is done one way or another. There are in the tradition clear assumptions about ‘‘discover the very nature and order of things, an order based on principles which are eternal, unchanging, and all-encompassing’’ (cf., the above quote from Miksa, 1998). (Hjørland, 2013, p. 555)

The assertions stated above only bring to us the point of view of Ranganathan, the “mathematician” and, then, of his trajectory as a logical classificationist. If we understand the interpretational potentialities of the modern approachesof knowledge organization and representation from Rhetoric, such as Emanuele Tesauro did in the specific context of “representation”, we can have a very different scope as we read Ranganathan's ideas and the changes that occurred after him.

Final considerations

Our "rereading" contains a critique of the distinctions between the thesaurus (the instrument and product of indexation) and the ontologies (systems of conceptualization of reality). Tesauro's potentialities (dealing with Emanuele), demonstrate the great and defying linguistic relations open by rhetorical discourses, whether through the notion of discovery, the notion of memory, the notion of metaphor, or many not explored yet.

The challenge of subjectivity present in these notions proves that a purely logical analysis needs to be also co-constituted from a rhetorical point of view, one establishing equal analytical possibilities of pragmatic variation of signified and signifier, that is, the socio-historical production of discourses is manifested in artifacts which are appropriate to knowledge organization.

Therefore, this study recovers, theoretically, epistemological issues related to Emanuele Tesauro's works published as early as the XVIIth century, with an influence on instruments lasting until nowadays. This reflection strengthens the historical epistemological identity of Knowledge Organization, its vast potential to conduct comprehensive studies, and it makes evident more than three centuries of knowledge organization in the trajectory of thinkers, in approaches, and instruments.

References

Almeida, Marco Antônio de,& Crippa, Giulia (2009). De bacon à internet: considerações sobre a organização do conhecimento e a constituição da ciência da informação. *Ponto de Acesso*, 3(2) Ago.: 109-31.

Dodebei, Vera Lucia D. L. de M. (2002). *Tesauro: linguagem de representação da memória documentária*. Rio de Janeiro: Intertexto.

Eco, Umberto (2001). *Semiótica e Filosofia da Linguagem*. Lisboa: Instistuto Piaget.

Gomes, Hagar Espanha,& Campos, Maria Luiza de Almeida (2004). Tesauro e normalização terminológica: o termo como base para intercâmbio de informações. *DataGramaZero*, 5(6).

Gomes, Hagar Espanha (1996). *Classificação, tesauro e terminologia: fundamentos comuns*. [http://www.conexaorio.com/biti/tertulia/tertulia.htm

Gomes, Hagar Espanha (1990). *Manual de Elaboração de Tesauros Monolingues*. Brasília: Ministério da Educação.

Gruber, Thomas R. (1993). Toward principles for the design of Ontologies used for knowledge sharing. *International Journal Human-Computer Studies*, 43, revised August 1993. Substantial revision of paper presented at the International Workshop on Formal Ontology, March, 1993, Padova, Italy. Available as Technical Report KSL 93-04, Knowledge Systems Laboratory, Stanford University.

Hjørland, Birger (2013). Facet analysis: the logical approach to knowledge organization. *Information Processing and Management*, 49: 545-57.

Mollina Cantó, Eduardo, & Chiuminatto, Pablo (2004). Sobre la agudeza. un capítulo del catalejo aristotélico de emanuele tesauro. *Onomázein*,9: 27-49.

Monteiro, Silvana D.,& Giraldes, Maria Júlia C. (2008) Aspectos lógico-filosóficos da organização do conhecimento na esfera da ciência da informação. *Informaçao & Sociedade: Estudos*, 18(3): 13-27.

Moraes, Carlos Eduardo M. (2010). Italiano versus latim: Il cannocchiale aristotélico, capítulo VI. *Philologus* 12 (36). [http: //www.filologia.org.br/revista/36/05.htm]

Moura, Maria Aparecida (2009). Informação, ferramentas ontológicas e redes sociais ad hoc: a interoperabilidade na construção de tesauros e ontologias. *Informação & Sociedade: estudos* 19 (1) Jan./Abr.: 59-73. [http://www.ies.ufpb.br/ojs/index.php/ies/article/view/2396]. Acessed 16 Febrero 2016.

National Information Standards Organition (c2005 reaffirmed 2010). *Guidelines for the Construction, Format, and Management of Controlled Vocabularies. ANSI/NISO Z39.19-2005 (R2010)*. Baltimore: NISO. [http://www.niso.org/apps/group_public/download.php/12591/z39-19-2005r2010.pdf]. Acessed 16 Febrero 2016.

Otlet, Paul (1934). *Traité de documenatation: le livre sur le livre: théorie et pratique*. Bruxelas: Editiones Mundaneum.

Peignot, Gabriel (1802a). *Dictionnaire raisonné de bibliologie, tomo I*. Paris: Chez Villier.

Peignot, Gabriel (1802b). *Dictionnaire raisonné de bibliologie, tomo II*. Paris: Chez Villier.

Proctor, Robert E. (1973). Emanuele Tesauro: A Theory of the Conceit. *MLN*, 88 (1): 68-94.

Sales, Rodrigo, & Café, Ligia (2009). Diferenças entre tesauros e ontologias. *Perspectivas em Ciência da Informação* 14(1) Jan./Abr.: 99-116.

Simões, Maria das Graças (2008). *Da abstração à complexidade formal: relações conceptuais num tesauro*. Coimbra: Almedina.

Aline Elis Arboit and José Augusto Chaves Guimarães

Searching for a Metatheoretical Mapping of the Process of Socio-Cognitive Institutionalization of the Knowledge Organization Domain

Abstract

This study presents a meta-theoretical matrix representing the socio-cognitive institutionalization process of Knowledge Organization (KO) as a proposal to map the domain. Based on the official information provided by the International Society for Knowledge Organization (ISKO), combined with the reflection on the development of the domain's main theoretical axes and the analysis of the Classification Scheme for Knowlegde Organization Literature (CSKOL) main categories, already discussed in a previous study, the matrix was designed with four dimensions: internal intellectual, oriented to the study of internal cognitive elements; internal social, representative of internal institutions; external social, concerning external historical contextual factors; and external intellectual, related to cognitive factors from other areas.

Introduction

Considering the widely accepted definitions of domain in Knowledge Organization (KO) (Hjørland and Albrechtsen, 1995 Hjørland, 2002, 2004; Smiraglia, 2012), it is possible to identify an agreement on the notion of group or individuals who act together rather than the collectively constructed products, whether they are applied or theoretical. Thus, it is understood that such products, even those considered fruit of a predominantly abstractive process, reflect what the subject and his group were, are, or have been in certain spatiotemporal contexts.

KO, therefore, can be thought of as a domain rather than a scientific discipline, as it is conceived that the involved subjects are fundamental and inseparable part of its scientific products. Hence, it is conceived that the ways of thinking and acting of the subjects, which are, in turn, socially constructed, are materialized in discourses and, then, the scientific products are developed, carrying the ideological options adopted by the subjects inserted in a collectivity. Scientific knowledge is registered discourse, a way of thinking and acting, which, depending on the intersubjective load cannot be considered purely objective nor neutral or absolute (Arboit, Guimarães, 2015). The same idea follows the elements of this knowledge, such as concepts, nomenclatures, terms as they are supported by those discourses.

From the recorded knowledge in KO literature and its intersubjective consensus, theoretical and methodological principles guiding the domain have been outlined, as well as the purposes and actions taken by ISKO, considered in this study as the most representative entity of institutionalization processes in our area. In this sense, ISKO can be seen as a driving social institution in the development of cognitive institutions in the area, since it offers a concrete framework for researchers to develop and validate the domain fundamentals in collaboration.

It is observed that, within the institutionalization process of knowledge, institutions are only constituted and renewed if the interrelation of cognitive structures with social structures is observed. It is assumed, therefore, from Whitley (1974), that KO - such as any scientific institutionalized domain - has been built from the reciprocal relationship and inseparability of social and cognitive approaches.

Since the Renaissance, scientific societies have been responsible for much of the advance and communication of scientific knowledge. In the beginning, according to Burke (2000), they were characterized by marginalization and innovation, seen as a refuge for free thinkers to interact with each other, while universities - although they continued to play their traditional role of teaching, due to "institutional inertia" that permeated them - prioritized their corporatist traditions, isolated from new ideas. Because of scientific advances promoted by societies, universities formalization processes and financial support was transferred from the State to societies, hence they were gradually forced to reorganize, reforming their curricula and regulations, so that today a movement of integrating scientific societies to universities has been observed (Burke, 2000).

Scientific societies have been consisted of groups of scholars involved and concerned with both formal and informal scientific communication from the beginning of their formation. They have also held regular meetings, where information dissemination was encouraged, and members reported their research, made demonstrations and exchanged experiences and knowledge. According to Meadows (1974) editorial programs were thus established, in which their papers were assessed and publicized, transmitting what had been developed in the society to a larger group of people, including subsequent generations playing the role of main vehicle of scientific communication.

Given the above, it is verified that, in addition to the importance played by scientific societies in the scientific and educational development, their consolidation process was the result of intellectual and social conflicts that occurred during the knowledge production process. This argument reinforces the idea advocated by Whitley (1974) on the relationship of coexistence between social and cognitive dimensions. ISKO, as a scientific society, was also consolidated from intellectual and social conflicts between groups of researchers representatives of distinct currents of thought which differed on the direction the domain should take. As reported by Dahlberg (1995, 2006), tensions occurring within the scientific research community in the former Society for Classification established in Germany in 1977, are seen as determining factors for ISKO's foundation, and consequently and simultaneously, the development of a theoretical base for KO "discipline" grounded on the Concept Theory. Dahlberg points out that divergences among the researchers that composed the former society, especially in relation to a dominant position focused for an essentially mathematical view of classification, were key for the formation of their conceptions, oriented to

conceptual issues applied to classification and other KO systems, investing against mathematical and statistical conception on classification, which, at that time, was prevailing in the group.

Developed by Dahlberg (1993) within ISKO, the Classification Scheme for Knowlegde Organization Literature (CSKOL) can elucidate the reciprocity relationship and mutual penetration between cognitive and social dimensions of the institutionalization processes of KO as a domain, for it is an instrument designed for the systematic control and publicity of publications considered relevant to the area, and at the same time, it is a KO intellectual map since it was built with its main concepts and theoretical framework. Despite the wide acceptance of the scheme among researchers in the area, to the extent it is considered reference for KO conceptual framework, we identify an intellectual and social mapping of the area as necessary, with the aim of reaching an understanding over its socio-cognitive institutionalization process as well as the understanding on the development and consolidation process.

Objective and Methodology

Based on the description about the creation and development of ISKO, especially on the entity's official information available on the webpage (ISKO 1989, 2016a, 2016b), in addition to the reflection on the development of the main theoretical axes and analysis of main CSKOL categories developed in Arboit (2014), the aim of this study was to design a meta-theoretical matrix representing the domain as a proposal for mapping the KO socio-cognitive institutionalization process.

As methodology for constructing the matrix, we adopted the metatheory method described by Ritzer (1991). The metatheory is defined by the author as an analysis method for a better understanding of a theory or set of theories. It seeks to achieve a thorough understanding of the concepts postulated by the examined theories. In addition to mental satisfaction, the metatheoretical method allows to represent, organize and explain in which ways an area has been constituted, and thus recreate knowledge.

Ritzer (1991) characterizes metatheory into three types: one that aims to understand, systematize and explain the theories of a domain; one that produces a theoretical basis to support a new theory; and one whose product is to draw a transcendent perspective to the studied area. The last two types, according to the author, are dependent on the first, as once deep understanding of a domain's theories is reached, the researcher is able to identify social and intellectual connections among the scholars, as well as the relationship with their production contexts and thus explain and draw a complete theoretical trajectory of the area. It is noteworthy that any researcher, either the one proposing a new theory or perspective or the one explaining existing theories, or even the one using empirical data, depends on the contact with his/her predecessors studies to support his/her own ideas. To strengthen this position, the ideas by Voloshinov (1973) and Bakhtin (1993) regarding the argument that the individual does not produce

any discourse isolatedly; intellectual creation occurs as the individual consciousness absorbs the linguistic signs of others, i.e., it is a result of interindividual interaction. Knowledge is, therefore, constantly recreated via human beings coexistence in society.

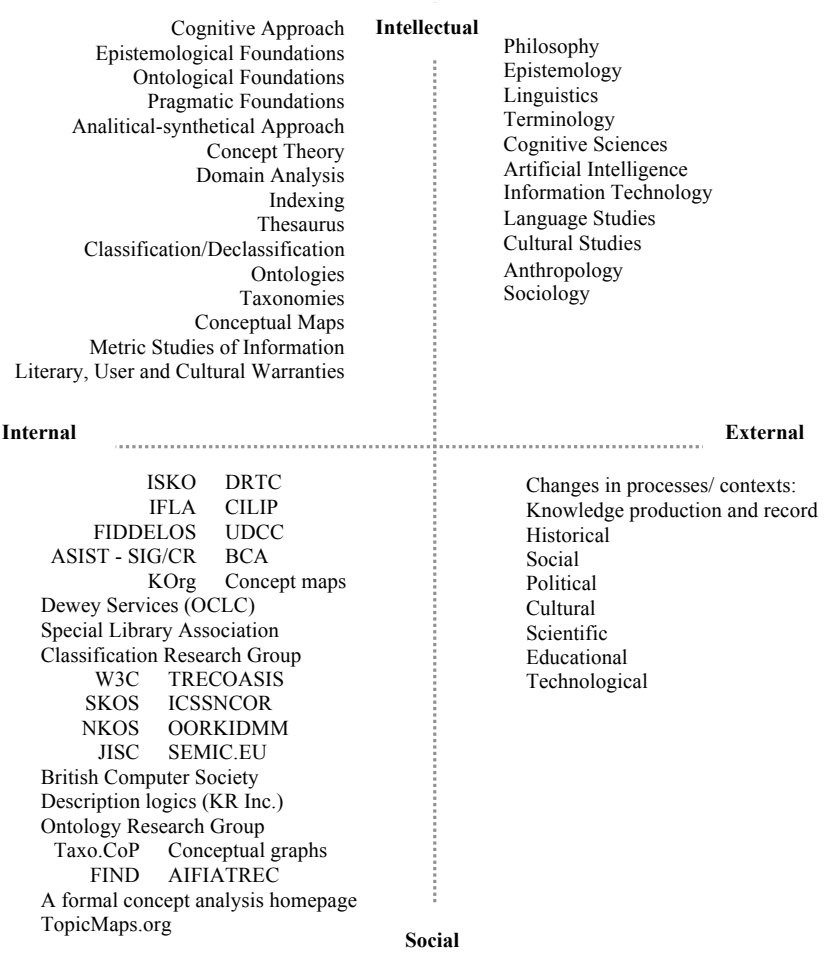
Still according to Ritzer (1991), the metatheoric method may include four dimensions: internal intellectual, oriented towards the study of cognitive elements developed within a given domain; internal social, which seeks to identify groups of theoretical influence building a genealogy of intellectual connections occurring within the domain; external social, which seeks to identify historical contextual factors external to the discipline that influenced in the theoretical constitution of the area; and external intellectual, which seeks to disclose cognitive factors from other disciplines that theoretically impacted the area.

Results

While Ritzer represents the four metatheoretical dimensions in four completely delimited quadrants (1991, p. 18), we conceive that they do not develop in isolation. The lines dividing the four dimensions should not be so impermeable as shown, i.e., each dimension is necessarily mutually affected by the other. Back to Whitley (1974), it is assumed that the internal cognitive/intellectual contexts, internal social, external cognitive/intellectual and external social are permeable; they influence and are continuously influenced by what occurs in each of its spheres. Therefore, they are inseparable both in relation to internal and external interrelationships and social and cognitive interrelationships.

Based on the socio-cognitive perspective we adopted, we chose to represent the line "separating" the quadrants of the matrix (as shown below) using a dashed line, understanding that each dimension influences and, at the same time, suffers the influence of one another.

Figure 1 – Metatheorical Matrix of KO



In the inner intellectual quadrant, we placed the main philosophical orientations and theories and methods developed "in" the KO. This does not mean that these elements were not also the result of social development internal and external to the area or that KO has not absorbed them from discoveries in other areas of knowledge. In addition, the quadrants of the proposed matrix are a way of representing reality, and as any representation, the matrix is also reducing, temporary and artificial.

Regarding the institutionalization process of KO, from the analysis of the theoretical framework (Arboit, 2014), it was found that, in the beginning, the interests were mostly permeated by technical, idealistic and cognitive perspectives of knowledge as a result of embracing the ideas defended mostly by Ranganhatan (1951, 1967, 1989) and Dahlberg (1993, 1995, 2006).

After a change of direction for genealogical, pragmatic, cultural and ethical approaches, mainly represented by authors such as Hjørland (1992, 2002, 2003, 2008, 2009, 2012), Frohmann (1990, 2004), Olson (2001, 2002), Beghtol (2002, 2005), Guimarães et al (2008), and García Gutiérrez (2002, 2014), but not entirely abandoning idealism due to the defense of naturalism and ontological approach advocated by Gnoli (2004, 2011).

The internal social quadrant is responsible for listing the main institutions related to KO. These institutions are constituted by researchers, their discourses, the clashes and consensus resulting from ideological options accepted by the members of such institutions, not completely dissociated from the internal intellectual dimension. The institutions mentioned in the quadrant are listed on ISKO's official website (2006b) as entities related to the KO. The external intellectual quadrant brings the areas of knowledge that have influenced the theoretical and practical development of KO, and the external social quadrant brings a list of general historical facts that have determined the course of both KO and other areas of knowledge. Such perspectives can be visualized on ISKO's official website (2016a), on which the entity is described as "an *interdisciplinary* society (...) brings together professionals from many different fields" and "counts about 600 members all over the world, from fields such as *information science, philosophy, linguistics, computer science*, as well as special domains such as medical informatics", as well as its statute, which contains the society's goals:

> to promote research, development and application of all methods for the organization of knowledge in general or of particular fields by integrating especially *the conceptual approaches of classification research and artificial intelligence*. The Society stresses *philosophicological, psychological and semantic approaches* for a conceptual order of objects (ISKO, 1989).

The quadrants representing the external aspects to KO are more general than the internal ones as they are responsible for representing a wider reality, and therefore cannot present, in the limits of this matrix, all specific historical events that influenced the socio-cognitive development of KO. We then mention examples, such as the creation of computers, databases and the Internet, which have impacted discussions in both KO and Information Science, especially concerning the Information Theory.

Conclusions

The institutionalization of KO domain is here considered as a process characterized by incompleteness. As a condition to ensure its own existence, it is understood that the institutions are not isolated from the acts carried out in social life. Such acts are sociocognitive par excellence because they are carried out by the subjects in their collective ambience, where they are influenced and influence others, according to the variations that occur cronotopically. Hence, we conceive that the institutions and scientific domains are, although they materialize relations among the superstructures and infrastructures, groups of people who communicate, dialogue, debate, negotiate, enter into agreements, clash, dispute, defend points of view, and polemicize.

References

Arboit, Aline (2014). *O processo de institucionalização sociocognitiva do domínio de Organização do Conhecimento a partir dos trabalhos científicos dos congressos da ISKO*. Doctoral thesis in Information Science, UNESP (Brazil).

Arboit, Aline & Guimarães. J.A.C. (2015). The ethics of Knowledge Organization from a Bakhtinian perspective. *Knowledge Organization*, 42(5): 324-331.

Bakhtin, M.M. (1993) *Toward a Philosophy of the Act*. Austin: University of Texas.

Beghtol, Clare (2002) A proposed ethical warrant for global knowledge representation and organization systems. *Journal of Documentation*, 58(5): 507-532.

Beghtol, Clare (2005) Ethical decision-making for knowledge representation and organization systems for global use. *Journal of the American Society for Information Science and Technology*, 56(9): 903-912.

Burke, Peter (2000). *A social history of knowledge: from Gutenberg to Diderot*. Polity.

Dahlberg, Ingetraut (1993) Knowledge organization: its scope and possibilities. *Knowledge Organization*, 20(4): 211-222.

Dahlberg, Ingetraut (1995). Current trends in knowledge organization. In Garcia Marco, F. J. (org.). *Organización del conocimiento en sistemas de información y documentación*. Zaragoza: Universidad de Zaragoza.

Dahlberg, Iingetraut (2006). Knowledge organization: a new science? *Knowledge Organization*, 33(1): 11-19.

García Gutiérrez, Antonio L (2002). Knowledge organization from a culture of the border: towards a transcultural ethics of mediation. In: López-Huertas, M.J.(Ed.). *Proceedings of the Seventh International ISKO Conference*. Würzburg: Ergon. Pp. 516-522.

García Gutiérrez, Antonio Luis (2014). Declassifying Knowledge Organization. *Knowledge organization*, 41(5): 393-409.

Gnoli, Claudio (2004) Naturalism vs pragmatism in knowledge organization. In: McIlwaine, Ia (Ed.).*Knowledge organization and global information society: Proceedings of the Eighth International ISKO Conference*. Würzburg:Ergon. 263-68.

Gnoli, Claudio (2011). Ontological foundations in knowledge organization: the theory of integrative levels applied in citation order.*Scire*, 17(1): 29-34.

Guimarães, José A.C., Fernández-Molina, Juan C., Pinho, Fabio A., Milani, Suellen O. (2008). Ethics in the Knowledge Organization Environment: an overview of values and problems in the LIS literature. In Arsenault, C., & Tennis, J. T. (eds.). *Cultural and Identity in Knowledge Organization*. Würzburg: Ergon. Pp. 340-346.

Frohmann, Bernd (1990) Rules of indexing: a critique of mentalism in information retrieval theory. *Journal of Documentation*, 46(2): 81-101.

Frohmann, Bernd (2004) *Deflating information: from science studies to documentation*. Toronto; Buffalo: University of Toronto.

Hjørland, Birger (1992). The concept of subject in information science. *Journal of Documentation*, 48(2): 172-200.

Hjørland, Birger (2002). Epistemology and the socio-cognitive perspective in Information Science. *Journal of the American Society for Information Science and Technology*, 53(4): 257-270.

Hjørland, Birger (2003). Fundamentals of knowledge organization. *Knowledge Organization*, 30(2).

Hjørland, Birger (2004). Domain analysis: a socio-cognitive orientation for Information Science research. *Bulletin of the American Society for Information Science and Technology*, feb./mar.

Hjørland, Birger (2008). What is knowledge organization (KO)? *Knowledge Organization*, 35(2/3).

Hjørland, Birger (2009) Concept theory. *Journal of the American Society for Information Science and Technology*, 60(8): 1519–1536.

Hjørland, Birger (2012). Knowledge Organization = Information Organization? Neelameghan, A., & Raghavan, K. S. (eds.). *Categories, contexts and relations in knowledge organization: Proceedings of the Twelfth International ISKO Conference*. Würzburg: Ergon. Pp. 8-14.

Hjørland, Birger, & Albrechtsen, Hanne (1995). Toward a new horizon in information science: domain analysis. *Journal of the American Society for Information Science*, 46 (6): 400-425.

ISKO (2016a). *About ISKO*. [http://www.isko.org/about.html].

ISKO (2016b). *KO institutions*. [http:// http://www.isko.org/inst.html].

ISKO (1989). *Charter and preamble*. [http://www.isko.org/charter.pdf].

Meadows, A. J. (1974). *Communication in science*. Butterworths.

Olson, Hope A. (2001). Patriarchal structures of subject access and subversive techniques for change. *Canadian Journal for Information and Library Science*, 26(2/3): 1-29.

Olson, Hope A. (2002). *The power to name*. Dordrecht: Kluwer.

Ranganathan, S. R. (1951). *Documentation*. New Delhi, Vikas.

Ranganathan, S. R. (1967). *Prolegomena to library classification*. Bombay: Asia.

Ranganathan, S. R. (1989). *Philosophy of library classification*. Bangalore: S. R. Endowment for Library Science.

Ritzer, George (1991). *Metatheorizing in sociology*. Lexington: Lexington Books.

Smiraglia, Richard, P. (2012). Universes, dimensions, domains, intensions and extensions: knowledge organization for the 21st century. In Neelameghan, A., & Raghavan, K.S. (eds). *Categories, relations and contexts in knowledge organization*. Würzburg: Ergon. Pp. 1-7.

Voloshinov, V. N. (1973). *Marxism and the Philosophy of Language*, 48-49, Harvard University Press.

Whitley, Richard (1974). Cognitive and social institutionalization of scientific specialities and research áreas. In *Social processes of scientific development*. London: Routledge and Kegan. Pp. 69-95.

Cynthia Maria Kiyonaga Suenaga, João Batista Ernesto de Moraes and Natália Bolfarini Tognoli

Metatheoretical Introduction of Discourse Analysis and the Theory of Speech Acts for Knowledge Organization Improvement

Abstract

The conception that Discourse is originated from speech, whether oral or written, leads us to try understanding the theory of Speech Acts and Discourse Analysis aiming to comprehend possible discoursive formations, which is our interest to study within the Knowledge Organization field. Therefore a metatheoretical study was undertaken to confront these theories to enlighten if and how they are related to each other and to Knowledge Organization. Partial conclusions point out that the substances comprising studies in Discourse Analysis and Speech Acts Theory are in a broad sense the same: statements uttered by subjects in communication. As a possibility of enunciation analysis the Theory of Speech Acts can be an additional analytical tool, to be used in discourse analysis. What turns out to be more important, however, is the idea that through enunciations we provoke actions, or even act by means of language.

Introduction

One can say that a theory represents a phenomenon or an event, in the same way as words represent objects or events. The problem of Knowledge construction or theoretical elaboration is no different from the problem of representation of things through language. Both our knowledge of the world and the theoretical elaboration of the facts are permeated by language, not understood as a set of signs that serves to communication, but as an element that mediates the world and human knowledge.

Theories attempt to fill up a space in human knowledge through a more or less coherent set of hypotheses, seeking to clarify a problem. The ability to scientifically explain the world comes from human ability to represent, and to make these representations of data manageable, in order that hypotheses can be elaborated and confirmed or refuted. Theories bring together the representative elements of a domain over which researchers aim to increase understanding.

Regarding theories many biases can be found. Even though the act of theorizing intent to clarify facts or phenomena, theories created around specific corpora may be modeled on preconceived ideas, and preconceived judgments since it is nothing more than a representation of these and not properly the corpora under study. That might be a proper reason for researchers to focus on discoursive concerns when organizing knowledge, since the discoursive view widens the field of investigation to involve the registries, the context of creation and the context of use altogether.

Following this line of thought a theory is a form of representation operated, directly or indirectly, by language, through which researchers interact with the world. Thus, a theory is artificially put in the place of what it wishes to explain and becomes a representation that can be communicated, shared and/or challenged by a specific scientific community.

Metatheory is a sociological field dedicated to the study of sociological theories, that seeks a greater understanding of the theories themselves, or going beyond, seeks from the metatheoretical studies to promote the development of new theories or also raises corpora to serve as a source of new sociological perspectives (Ritzer, 1991).

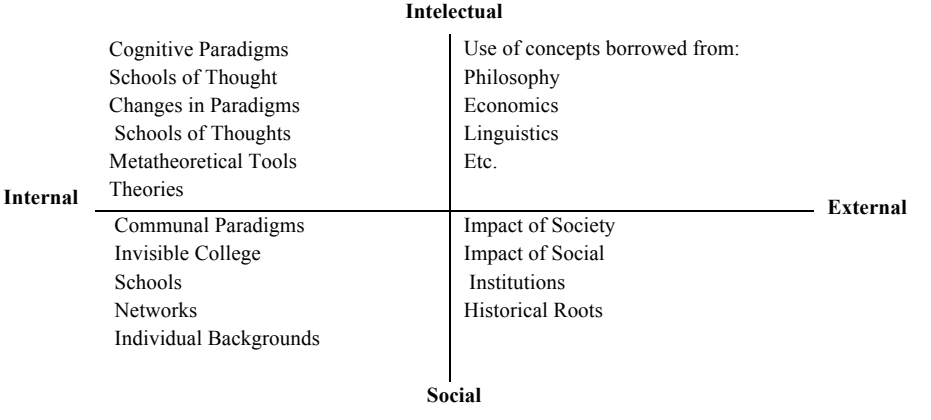
It is important to highlight that metatheory has been used within Knowledge Organization, as demonstrated by the studies of Vickery (1997), Svenonius (2004), Bates (2005), Tennis (2008) and Tognoli (2013). The works of these authors demonstrate a need to consciously analyze the presuppositions present in our field of research and in related fields to promote new ways of thinking.

Metatheory is an important field of study because the metatheoretical perspective of each person shapes the way in which they understand theory and how theory is understood influences how we create, validate, test, select and apply theory to optimize the human condition (Wallis, 2010).

Among other approaches to metatheorizing, Ritzer (1997) brings an interesting development presenting three types of metatheory, defined in general terms by differences presented in the final products of each one: 1) metatheorizing as a means to gain a deeper understanding of the theory (Mu); 2) metatheorizing as a prelude to the development of new theory (Mp); 3) metatheorizing as a source of new prospects covering sociological theories.

Regarding Mu, Ritzer points out subdivisions arising from the categorization of researches in internal or external and intellectual or social perspectives related to the sociology field, resulting in quadrants that help visualizing and comprehending the nuances present in the analysis (Figure 1).

Figure 1 – Mayor types of Mu by Ritzer (1991, p. 1)



The internal perspective refers to existing elements within sociology, while the internal to phenomena that are found outside of sociology, but has a decisive influence in it. The intellectual perspective refers to everything that is related to the cognitive

structure of sociology as theories, metatheoretical tools, ideas borrowed from other disciplines and so on. The social aspect is related to the sociological structure of sociology, the effect of individual background factors, and the impact on society, among others.

The fourfold table and its quadrants assist the understanding of the various types of work occurring in metatheory focused on deepening theoretical understanding. It is important to say that the boundary lines between the dimensions (internal-external; intellectual-social) are not rigid and may involve one or indeed several aspects.

Ritzer (1993) begins most of his writings on metatheory explaining that this is not an exclusive activity of sociologists as philosophers (Radnitzky, 1973), psychologists (Gernen, 1973, 1986; Schmidt et al, 1984), historians (White, 1973) and social scientists (Fiske and Schweder, 1986), among others, are some examples, named by Ritzer, of studies related to the understanding of the discipline itself.

What will effectively distinguish between metatheoretical researches, which are the systematic study of theories, is the final product resultant of the analysis. Thus the perspective provided by a metatheoretical study of two theories that may possibly provide elements to increase understanding of the knowledge building process within Information Science is what motivates this study.

In the following topics Discourse analysis and the Speech Acts Theory will be presented aiming to enlighten the elements of these theories that can be groundwork to Knowledge Organization studies and practice.

Discourse Analysis

Harris' work (1952), in which the statements become analyzed instead of phrases, can be considered the starting point of Discourse Analysis, but his approach is also, in a way, a continuation of linguistics, as it applies analysis procedures to language units without taking into account the socio historical conditions for these statements. Benveniste, in contrast, emphasizes the enunciation process and the subject who emits it, according to the position the speaker occupies and its relationship to the statement.

Thus it seems that Discourse Analysis in its beginnings, developed in at least two parallel forms, which are an extension of the language. One corresponding to the American perspective which considers the text and the phrase as isomorphic elements whose analyzes differ only in degree of complexity, concerned not with forms of production of meaning, but with forms of internal organization; and another that seeks solutions to the statements by extrapolating the language and limitations of semantics, hitherto practiced (Brandão, 2004).

In the XIX century successive proposals of authors, who considered language primarily as a specific social activity or speech activity, gained strength. The basis of this new vision are present in Humboldt’s work in which, as Bronckart (2008) states, language produces sense objects that are the representative units of human thought and is a social activity.

Discourse can be broadly considered as all kinds of spoken interaction, formal and informal, and written texts of all kinds (Perryman and Wildemuth, 2009).

The crucial issue of language is to be significant activity manifested by speech or discourse, its essence is the speech, it is an activity that takes place through language and is therefore historically determined in time and among particular groups which we can call discourse communities as defined by Hjørland and Albrechtsen (1995). Language seen from this point of view is a self-adaptive system, and fundamentally a significant activity in which the signs are not merely produced to mean something already given (we don't look for meaning only in dictionaries), but are the very creation of content and expression, that is, they are constructed discursively, within a context.

From this perspective Discourse Analysis studies knowledge production practices or meaning in concrete texts and institutions, be it a library, a data interface, the Information Society strategies, or classification systems systematizing different ways to "talk" to make visible the prospects and starting points on the basis of which knowledge and meaning are produced in a particular historical moment (Talja, 1999). It is common for groups that share a common space, or a purpose to also share interpretative repertoires, where individuals select knowledge and information as part of the communication process and identify these repertoires is the initial task of discourse analysis.

Foucauldian discourse analysis does not study the rules and conventions of worldly speech, but the serious speech acts, or institutionalized practice of speech (Talja, 1999; Frohmann, 2008). Language, for discursive concerns, becomes a phenomenon to be studied in its internal system, while language practice, as well as in its external system and ideological formations, taken in a total sense, admitting not only explicit ideological positions, but also implicit ones.

The historical and social context is considered as to respond how ideology is manifested in language and how language expresses the ideology. The meanings that can be read in a text are not necessarily in it, once the meaning of a text pass through intertextual relationship with other texts and their contexts (Orlandi, 2007; 2008). Discourse analysis works with the speech as socio- historical element in that language acts as conjecture, and does not work with the history and society highlighted in their meanings, but considers that, together, they build the discourse.

To Foucault (2008), the discourse is constructed through signs, representations, discourse is not merely what reveals the desire or conceals it, but also the object of desire, as the discourse translates struggles or domination systems and therefore, who expresses the discourse expresses the dominant power and can influence their surroundings. Understood in this way the discourse is unifying, in the sense that unites words or similar thoughts, leading to the development of discourse communities.

As Talja (1999) states, discourse analysis is part of the linguísitc turn toward social

sciences and humanities, which emphasizes the role of language in the construction of social reality. It is a tool that focuses on the analysis of the formations of knowledge, which organize institutional practices and social reality on a large scale and can be used to discover meanings that are not apparent in the literal sense of the words. It is an approach that goes beyond the dichotomy of subjective meaning and objective reality, as the dichotomy between user-centered research and focused on the system research.

As a research approach in communication, sociology, psychology and social psychology, it is one of the methodological theories that positions itself in the mainstream of these areas of study. Although several articles have discussed the implementation of Discourse Analysis in information studies (Budd & Rober, 1996; Frohman, 1994, 2001; Talja, 1997) it has been seldom used as a concrete method of research (Talja, 1999).

Martinez-Ávila (2012) discusses the problems and characteristics of Foucauldian Discourse analysis and identifies, in the studies of diverse authors, uses of discourse analysis as a critical theory, where researches can find elements to critically analyze knowledge and information from a discoursive point of view. Here he identifies to that some developments occurred leading to revisions of the concept of discourse analysis (eg Laclau and Mouffe, 2001) originating new methods that differs from Foucault's (although Foucault himself didn't leave a structured method of analysis).

In another work Martinez-Ávila (2015) does a discourse analysis of the construction of ontologies aiming at the constitution of the concept and its dialogue with discourse analysis figuring as an element that adds critical discussions on classification systems in the detection of epistemological stance within the discourse to the ontologies in computer science or LIS.

Speech Acts Theory

The Theory of Speech Acts originated from the philosophy of language, in the early sixties, and was later appropriated by Pragmatics. From the Analytical School of Oxford, John Langshaw Austin (1911-1960), followed by John Searle, among others, conceived language as a way of acting which is achieved by language through "speech acts". Austin says that not all sentences are descriptions, and for that reason he preferred to use the word 'constative' (Austin, 1962). In the constative sentence, something is stated and can therefore be analyzed in relation to the character of truth or falsity. As Costa (2002) states Austin starts his analysis from the enunciations (utterances) which are acts of issuing statements made by speakers to listeners in concrete situations.

Austin wanted to classify the various types of speech acts and innovates in that he does not consider the enunciation statements as privileged forms in which propositions and the real world articulate, but that the speech acts are not tied exclusively to truth conditions, which would apply only to constative speech acts with which we find something.

By examining the utterances Austin certifies that part of them does not communicate something about facts, they are not true or false, they are performative as they are directly linked to the achievement of an action. According to Austin (1962) the term performative is derived from the English verb 'to perform', indicating that a performative sentence is meant to perform an action, or an act. Contractual sentences like 'I bet' or declaratives like 'I declare war' are performing, as well as the 'yes' in a marriage, the 'promise', 'I baptize', among others.

One of the most important contributions to the theory of meaning and the philosophy of language is the fact that the statements, classified before by Austin as constatives are acts of Discourse (Araújo, 2004). In fact to Austin a statement is not a proposition announcing a real fact but an act of speech, and the various speech acts would be logical constructs drawn from the acts of discourse, and not the opposite.

During the series of conferences that originated the work How to do things with words, Austin expands what was restricted to a constative/performative distinction to the theory of speech acts or acts of discourse (Araújo, 2004). For us to succeed in enunciating something three simultaneous acts occur: 1) the locutionary act; 2) the illocutionary act; and 3) the perlocutionary act.

The locutionary act is the act of saying something, it involves complete elements of speech with meaning and reference. Composes the locutionary act: a) a phonetic act (issuance of a sequence of sounds or phonemes); b) one factual act, issuing a sequence of words in a vocabulary organized in line with a grammar; and c) a Rheticus act, which is the sequence of words wich 'say something' about 'something', hence with sense and reference.

Austin calls illocutionary act the act we perform to say something. The illocutionary act, occurs when a locutionary act acquires, by convention, illocutionary value. When a person utters the phrase ‘Tomorrow I will come’ he/she could be informing, making a threat, a. promise etc. The different meanings given to utterances by each of these verbs were called by Austin “illocutionary forces”.

The perlocutionary act occurs when a locutionary act produces an illocutionary act (with the power of saying) and also causes an effect on the listener. For example, a warning may cause fear, deter, harassment, among other effects.

Searle develops and refines the theory of speech acts, it is this , times when the concepts of locutionary act and illocutionary act overlap, caused an obstacle in theory since the abstraction of the meaning of an utterance will, according to Searle (1968) abstract the illocutionary force wherever this force is included in the meaning.

Austin and Searle`s approach for speech acts occur in very different ways. By changing the elements of speech acts Searle changes the focus of the theory, Austin considered the illocutionary act as pragmatic and not semantic. Searle finds close ties between the meaning and strength of illocutionary acts. The semantic approach of Searle limited the units of speech act to statements, while as to Austin unit would be

the enunciation. One of the strengths of Searle's approach is that it allows an enunciation to have several illocutionary forces (Montheith, 2010). The differences in the approach of these authors have resulted in rather contrasting developments.

Universal pragmatics developed by Habermas, which seeks, essentially, to clarify the conditions that need to be met in communicative actions in natural language, was constructed in connection with the acts of speech, but does not integrate in-depth reflections on the status of language signs and does not consider the level of texts and/or discourses, that are the main means of organizing the acting in language (Bronckart, 2008).

For Habermas there are four classes of speech acts: the communicative speech acts, constative speech acts, regulative speech acts, and the representative speech acts. Institutionally bound speech acts are left out because it is assumed that institutions are random and thus not appropriate to reflect universal communication conditions (Costa, 2002). Issues related to the validity claim of communicative action of Habermas are based on the discourse.

The Theory of Speech Acts states that certain utterances issued, whether in speech or in writing, as a result produce an action that leads to a fact. It is based on the premise that in certain circumstances an utterance may indicate the intentions of its creator and have an effect on the receiver. That was the foundation for studies developed on Document Acts from researcher Barry Smith (2014). By ‘document act’ the author refers to what humans (and other agents) do with documents. Actions that go from signing, stamping, depositing in registries, to its use to grant or withhold permission, and so on. To this theory, documents are utterances performed in acts of promising or commanding not merely of epistemic significance, the concern is with document acts in general where the action itself reflects the status of the documents, rather than as, for example, mere pieces of paper.

Smith (2014) speaks of a collective intentionality related to the document acts, that are acts in which people use documents not only to record information, but also to produce further ends, thereby extending the scope of what human beings can achieve through the mere performance of speech acts.

Discussion

In undertaking is a Mu kind of meta-theoretical analysis directed to two theories as complex as the Discourse Analysis and the Theory of speech Acts the first conclusion that is reached is that it is an enterprise that is just beginning, many authors of both parts still need to enter this dialogue, as well as other approaches to discourse analysis besides the Foucauldian.

The development of these theories is interlaced seemingly randomly bumping into various subject areas such as linguistics, philosophy, psychology, history and even perhaps other contributions and also refutations are to be identified.

Visibly from Austin’s speech act idea many researchers, from different philosophy

of study approaches, developed their own theories, as Searle who was a student of Austin, and following his line of thought, changed what judged flawed in Austin's theory to the point that some authors do not consider him heir of Austin's theory (Montheith, 2010).

The pragmatics of Habermas, which also had its basis in speech acts, or rather speech that triggers action and consequently communication is one of the theories that came from the idea of speech acts as the Document acts of Smith. There must be a much wider range of developments, but to this study it is enough to identify not all, but some ways to better comprehend discursive formations and their registry to seek for the knowledge present in documents.

Speech and discourse are not different terms that refer to the same thing, but to study discourse we must investigate speech and its effects (the acts is originates). Author Araújo identifies constatives as acts of discourse.

It seems the Speech Acts Theory can provide different ways to analyze discourses, although this is not the focus of the studies presented by Austin or Searle. But, as raised in the theoretical framework of this study, Foucault also refers to serious speech acts as the object of study of discourse analysis.

As Monteith (2010) sugestsk, the theory of Speech Acts provides a way to approach how documents perform communicative acts. The so-called perlocutionary effects of documents, that is, the linguistic act of producing effect on the listener (or the receipt of a document), can be revealed by observation of the components of an utterance.

Both the theory of Speech Acts as its development to the document Acts reinforce the value of the discursive approach to Knowledge Organization, since they are concerned about the representations of intentions that culminate in documents, as in registered knowledge, and in the reflections that these registered "acts" can have on society, institutions and individuals. In the Theory of document acts what generates the action is a document.

At the same time discourse analysis can be of use to “dismantle the ideals of homogenization and universalization in knowledge organization” (Martinez-Ávila 2012, 108) and bring emphasis to a certain modularity in the construction of Knowledge as pointed by the concept of discourse communities being the constitutive element of domains of knowledge (Hjørland and Albrechtsen, 1995). By considering this view in knowledge organization, and knowledge organization systems (KOS) projects, the possible general ideologies (one of Foucault’s main concerns) loose its strength, once each discourse community develops their own discourse, and although influenced by some major ideology, can produce and empower its own.

Conclusions

As a possibility of enunciation analysis the Theory of Speech Acts can be an additional analytical tool, to be used in discourse analysis. What turns out to be more important, however, is the idea that through enunciations we provoke actions, or even

act by means of language. This idea promotes the importance and strength that discourse has in the lived reality, hence the importance of this type of analysis at all levels of reflection about knowledge of what we are, where we come from and what do we do with our knowledge.

Many differences of approach were identified and both theories have led to other studies and theories that need to be analyzed further, but it is suggested by the resulted discussion of this paper that a dialogue between these approaches could be fruitful to the knowledge organization studies, especially the ones focused in language and knowledge representation, once the intention of the discourse is not always in the text or the speech.

Discourse analysis is being frequently used and is considered an important analytical and critical tool in Information Science. Speech acts might as well occupy a similar space or even help in unveiling the discoursive formations that lead to the knowledge representations we study. As demonstrated by the document acts theory. There seems to be a great number of possibilities of use to both the theory of speech acts and discourse analysis in the knowledge organization field.

References

Araújo, Inês Lacerda (2004). *Do signo ao discurso: introdução à filosofia da linguagem*. São Paulo: Parábola.

Austin, John Langshaw (1962). *How to do things with words*. Glasgow: Oxford University.

Bates, Marcia (2005). An Introduction to Metatheories, Theories, and Models. In Fisher, K. E., Erdelez, S. &McKechnie, L., Eds. *Theories of Information Behavior*. Medford: Information Today, pp.1-24.

Brandão, Helena Nagamine (2004). *Introdução à análise de discurso*. São Paulo: UNICAMP.

Bronckart, Jean-Paul. (2008). *O agir nos discursos: das concepções teóricas ás concepções dos trabalhadores*. São Paulo: Mercado das Letras.

Costa, Claudio (2002). *Filosofia da linguagem*. Rio de Janeiro: Jorge Zahar.

Foucault, Michel (2008). *A ordem do Discurso*. São Paulo: Loyola.

Frohmann, Bernd (2008). O caráter social, material e público da informação. In *A dimensão epistemológica da ciência da informação e suas interfaces técnicas, políticas e institucionais nos processos de produção, acesso e disseminação da informação*, organized by Fujita, M., Marteleto, R., & Lara, M., p. 19-34. São Paulo: Cultura Acadêmica,

Harris, Zellig S (1952). Discourse Analysis. *Language* 28: 1-30.

Hjørland, Birger, & Albrechtsen, Hanne (1995). Toward a New Horizon in Information Science: Domain-Analysis. *Journal of The American Society for Information Science: J*ASIS, 46(6): 400-425. DOI: 10.1002/(SICI)1097-4571(199507)46:6<400::AID-ASI2>3.0.CO;2-Y.

Martínez-Ávila, Daniel, & Melodie J. Fox (2015). The Construction of Ontology: A Discourse Analysis. In *Ontology for Knowledge Organization*, ed. Richard P. Smiraglia, & Hur-li Lee. Würzburg: Ergon. Pp. 9-26.

Martinez-Avila, Daniel (2012). Problems and Characteristics of Foucauldian Discourse Analysis as a Research Method. In *Cultural Frames of Knowledge*, edited by Richard P. Smiraglia, & Hur-Li Lee. Wiirzburg, Germany: Ergon. Pp 99-110.

Monteith, Peter. (2010). Can Records Speak for Themselves?" *Journal of the Society of Archivists*, 31(2): 119–133.

Perryman, Carol, & Barbara M Wildemut (2009). Discourse analysis. In *Applications of social research methods to questions in information and library science*, edited by Wildemuth, Barbara M, Westport: Libraries Unlimited.

Ritzer, George (1991). *Metatheorizing in Sociology*. Lexington, MA: Lexignton Books.

Ritzer, George (1997). *Teoria sociologica contemporanea*. México: McGraw-Hill.

Searle, John R. (1968). Austin on locutionary and illocutionary acts. *Philosophical Review* 77(4): 405-424.

Smith, Barry (2014). Document Acts. In A. Konzelmann-Ziv, H. B. Schmid (eds.), *Institutions, Emotions, and Group Agents: Contributions to Social Ontology* Dordrecht: Springer, 19-31.

Svenonius, Elaine (2004). The epistemological foundations of knowledge representation. *Library Trends*, 52(3) Winter: 571-587.

Talja, Sanna (1999). Analysing qualitative interview data: the discourse analytic method. *Libray and Information Science Research*, 21(4): 459-477.

Tognoli, Natália B. (2013). *A construção teórica da diplomática: em busca de uma sistematização de seus marcos teóricos como subsídio aos estudos arquivísticos*. Dissertation Universidade Estadual Paulista.

Tennis, Joseph T. (2008). Epistemology, Theory, and Methodology in Knowledge Organization: toward a classification, metatheory, and research framework. *Knowledge Organization*, 35(2/3): 102-112.

Vickery, Brian (1997). Metatheory and information science. *Journal of Documentation*, 53(5):457–476. [http://dx.doi.org/10.1108/EUM0000000007206.]

Wallis, Steven (2010). Emerging Perspectives of Metatheory and Theory: A Special Issue of Integral Review. *Integral Review: A Transdisciplinary & Transcultural Journal for New Thought, Research and Praxis*, 6(3): 73-120.

Leila Cristina Weiss, Marisa Bräscher and William Barbosa Vianna

Pragmatism, Constructivism and Knowledge Organization

Abstract

It examines whether the pragmatic approach in Knowledge Organization (KO) is compatible, and/or may be combined with constructivism. This is a comparative exploratory study, of a theoretical nature, between the pragmatic approach in the KO and the constructivist perspective. It presents theoretical aspects of research that deal with KO and Information Retrieval (IR) with the pragmatic epistemological stance. It was identified that constructivism is an epistemological stance consistent with the pragmatic approach in KO. The incorporation of the constructivist learning perspective has the potential to optimize IR from the expansion of pragmatic epistemological border. It appears that the need for information is related to the completion of a task, the fulfillment of the goal of a decision maker, and this need can exceed what was initially wanted or expected to find in a search for information.

Introduction

In Knowledge Organization (KO), as well as in other areas of knowledge, an important theme of epistemological research is the classification and identification of paradigms and epistemological currents or like Tennis (2008) suggests, the epistemic stances present in the area. We highlight in this direction the researches developed by Birger Hjørland, which points to the existence of the empiricist epistemological, rationalist, historicist and pragmatic positions in KO. Hjørland (2007) considers that different approaches for KO imply different views on semantics and from the Peregrin research (2004), which highlights the pragmatic and positivist paradigms as dominant in semantics, it is possible to distinguish approaches in KO in pragmatic and positivist. Hjørland (2007) adds further that the semantics of the study have been neglected in the KO. Hjørland (2007), as well as Francelin and Kobashi (2011), points out that researches on KO do not explain in which theory or philosophical currents they are based. Still, the authors agree that there is an inclination to positivism and in that respect, Francelin and Kobashi (2011) state that research oriented to the study of the concept using foundations of this theoretical current uncritically, such as formal logic and the categories and Aristotelian logic. Hjørland (2009;2007) considers the pragmatic and historicist views as the most productive for the development of KO and states that the pragmatic perspective on concepts, meaning and semantics may be able to solve fundamental problems in KO from a promising new angle. Hjørland (2007) emphasizes the theoretical point of view of the American philosopher Hilary Putnam Whitehall, which, unlike the traditional semantic theory (positivist), takes into account the contributions of the real world and society to determine the meaning. However, the same author recalls that Putnam is known as a philosopher of the pragmatic tradition and his philosophy is based on three aspects: the relationship between the meaning and the real world (realism); nature of practical and functional sense (Pragmatism); development of meaning in a social context (Historicism). In another study, Hjørland (2003) states that social constructivism is related to the pragmatic approach in KO,

although differing in the realistic approach, because many researches done under the banner of social constructivism are deeply relevant to the understanding of the structure of the various knowledge areas. From this, the question of the nature of the relationship between epistemological positions arises, as well as the possibility of combining two distinct epistemological positions, constructivism and pragmatism.Thus, this article analyzes if the pragmatic approach in KO is compatible, and/or may be combined with constructivism. Therefore, we discuss the pragmatic and constructive approaches in KO and present a systematization proposal of epistemological positions based in Hessen's classification (1999).

Methodological procedures

The analysis of the relationship between the pragmatic approach in KO and constructivism was carried out through literature review on the pragmatic in KO and constructivism. From the theoretical aspects identified in the literature that show the relationship between the pragmatic approach in KO and constructivism, subsidies were sought in the epistemology literature area to a greater understanding of this relationship. It was possible with the proposal of Hessen (1999) for the classification of epistemological currents. He organizes the epistemological currents from three criteria, the possibility of knowledge, the source of knowledge and the essence of knowledge. In his proposal, pragmatism, along with dogmatism, skepticism, subjectivism/relativism and criticism, are in the category of "possibilities of knowledge." Rationalism, empiricism, intellectualism, apriorism and the critical position are in the category of "sources of knowledge." And realism, idealism and phenomenalism, among others are in the category of "essence of knowledge." This proposed classification developed by the philosopher Johannes Hessen (1889-1971) and explained on his book "Theory of Knowledge", contributes to the understanding of that a particular theory can be related to pragmatism, according to his guesses as to the possibilities of knowledge; empiricism, according to his assumptions concerning the sources of knowledge; and realism, according to his guesses as to the essence of knowledge. As the division of criteria varies from category to category, each type of current is mutually exclusive only in the respective categories. Hessen (1999), when indicating different division criteria, allows the combination in aconsistent form of some epistemological positions, which proved useful for achieving the objective of this research, which is characterized as a comparative exploratory study of a theoretical nature, between the pragmatic approach in KO and the constructivist perspective.

Results

From the literature review, it appears that constructivism is one of the theoretical currents involved in explaining how human intelligence is developed considering the mutual actions between the individual and the environment and its influence on decision-making, from interpretation of objective and subjective elements. It also

represents the influence of the environment, ie, the external stimuli that act on the subject in the process of building and organizing his or her knowledge, so more and more elaborated, integrated and horizontal: subject-object-context.

Pragmatism considers that the utility verification is the truth test function, as it is the use of propositions that determine the truth conditions in pursuit of overcoming the dualisms of classical philosophy. The conception of truth as what is useful is considered able to overcome dualities such as thought and matter, soul and body, ideal and real, freedom and necessity, history and nature. And beliefs are rules of action and to develop the meaning of a thought is essential to determine what conduct it is able to produce, that is, the result is what its only meaning is. According to Hjørland (2003), the pragmatic paradigm in KO falls within the realistic side, because if the research in KO produces only "social constructions", reality can show that these constructions are inconsistent and may be challenged by empirical arguments. Moreover, according to the author, the pragmatic method is not opposed to aspects of empiricism, rationalism and historicism, as isolated evidence is not sufficient and the final truth criteria is linked to objectives and to human activities. Hesse (1999) considers that pragmatism, as well as skepticism, abandons the concept of truth as an agreement between thought and being. "However, it does not stop in this denial, but it puts another concept of truth in place of the one that was abandoned. True, according to this view, means the same as useful, valuable, life promoter "(Hesse, 1999, 40). The author goes on explaining that not everything that is true is useful because experience shows that truth can have harmful effects and not be useful. These objections, however, do not affect the positions of Friedrich Nietzsche and Hans Vaihinger, two important advocates of pragmatism according to the author, because they retain the concept of truth in the sense of agreement between thought and being, but this agreement would never be reached by us. There is no true judgment, unlike our knowing consciousness works with knowingly false representations and that view is identical to skepticism (Hesse, 1999).

When you try to approach the pragmatic approach to the constructivist approach in KO, similar discussion occurs, but in this case, the question is no longer about the nature of truth (what we can know), but if there is a reality outside the human mind that can be known. Hjørland (2003) believes that social constructivism is related to the pragmatic view in KO and claims to have found several studies made under the relevant constructivist flag for understanding the structure of various areas of knowledge. He stresses, however, that social constructivism is anti-realist and anti-realism is not well accepted in the pragmatic view. On the other hand, Frohmann (2008, 275), citing Latour (2005), helps us to understand how social constructivism can be compatible with realism.

> The absurdity of supposing that to show something is constructed is to diminish its reality or to show it is a fake is excoriated by Latour in his 2005 book, where he says: “In all domains, to say that something is constructed has always been associated with appreciation of its robustness, quality, style,

> durability, worth, etc. So much so that no one would bother to say that a skyscraper, a nuclear plant, a sculpture, or an automobile is 'constructed.' This is too obvious to be pointed out"; [...]"When we say that a fact is constructed, we simply mean that we account for the solid objective reality by mobilizing various entities whose assemblage could fail"

By analyzing the different constructivist currents, Castañon (2005) understands this issue differently and discusses the issue of realism in the various constructivist currents and with the question of the nature of the language - if it is representation or convention. From this point of view, regarding the understanding of the nature of the language, he considers Wittgenstein and Rorty Latour as members of the conventionalist strand. In Conventionalism, as defined by Gergen (1985, 1994 see Castanon, 2005), based on Wittgenstein (1975) and Rorty (1989), the meaning is not based on objects, on the mental process or ideal ones, but it is acquired by through social contact in the context of a particular culture. Representationalism is the doctrine that advocates that it exists or could exist a stable relationship between words and the world they represent.

To Castañon (2005), what divides the different interpretations of constructivism are the positions that each current assumes before the ontological status of the object of knowledge. According to the author, nor an idealistic neither a relativist stance is needed when we reject objectivism, and the best example of this is the critical rationalism of Karl Popper, which responsibility is assigned to, for the end of logical positivism.

Popper's philosophy, critical rationalism, is primarily concerned with issues related to the theory of knowledge, to epistemology. In 1934, he published his first book, Logic der Forschung (Popper, 1985) which constitutes a critique to logical positivism of the Vienna Circle, when he defended the view that all knowledge is fallible, correctable, and therefore, provisional.

> rationalism is an array of attitude to hear critical arguments and to learn from experience. It is fundamentally an attitude of admitting "I could be wrong and you could be right, and, by an effort, we can get closer to the truth." (...) In short, the rationalist attitude (...) is very similar to the scientific attitude, the belief that in the pursuit to the truth, we need cooperation and that with the help of argument, we may, in time, achieve something like objectivity. (Popper, 1987, 232).

Castañon (2005, 46) considers that the main idea of critical rationalism, which is also central to constructivism (Piaget), is that "there is no neutral, objective reality observation, for every observation is the light of a theory." We note that this is also the view of Birger Hjørland that advocates the fallible and provisional nature of knowledge. Their research is developed according to the pragmatic approach in KO. Hjørland (2003) explains the influence of different points of views on the establishment of semantic relations and other processes in KO through the description of the evolution of scientific knowledge. He presents, as an example, the classification of animals. Whales live in water and can be classified as aquatic animals; they are mammals, not fish. The classification requires that similar properties among the items be sorted and then be grouped. These similar properties can also establish relationships

between items. Hjørland (2003, 102) points out that "The history of all natural sciences documents the discovery that certain entities that share immediate properties nonetheless belong to different kinds."

When information professionals classify documents, meanings and relevant properties are only available on the basis of some description. This consideration is opposed to the prevailing implicit assumption that all relevant properties of objects are obvious to experts in information, which would accompany certain established principles and would provide a better classification: objective, neutral and universal, hence, technically efficient (Hjørland, 2003). As well as the different areas of knowledge are not neutral and have only a part of all possible descriptions on a particular topic, Knowledge Organization Systems (KOS) are not either."It is not possible to be neutral, but is absolutely unacceptable to hide different views and to suppress the users' ability to develop their own points of view." (Hjørland; Pedersen, 2005, 593).

We realize that, in his work, Birger Hjørland, besides pointing out the pragmatic epistemological stance in KO as the most productive one, he also points out that the elements of a KOS, terms, concepts and relationships, should be identified especially in literature, the order to minimize bias. In our view, the observations of Blair (2003) on IR meet Birger Hjørland's proposal on the importance of the elements of KOS be identified in the literature to be indexed. Blair (2003), IR process is seen, traditionally as one in which the researcher has something in mind, the supposed need for information, which is translated into a search query. However, based on Wittgenstein's statements, he explains that the way you think the need for information is conditioned by the retrieval language available. To the extent that this language is limited, so is thinking about the need for information. Thus, KOS or information retrieval languages cannot be based primarily on what users expect, but as Birger Hjørland and David Blair argue, in the literature to be indexed and retrieved. Traditionally, assessment of information retrieval systems (IRS) takes into account primarily and perhaps only, precision and recall levels. These indices are calculated from the relevant, understood as what the user would already have in mind, what the user wanted to find with the search. Thinking of the difference between want and need can help you understand this issue. What the user wants, which is related to the traditional concept of relevance cannot be what he needs. The need has a direct link to functional and practical aspects while want is something more subjective. Thus, the need seems to be consistent with the definition of relevancy by Hjørland and Christensen (2002, 964) - "*Something (A) is relevant to a task (T) if it increases the likelihood of accomplishing the goal (G), which is implied by T*."

Frohmann (. 1990, 98) also addresses this issue and asks the following question: "Does text retrieval fulfil a need, or does it satisfy a want?" Wishes, as explains Frohmann (1990), are explicitly recognized and accepted, they reflect objectives,

purposes and intentions of agents. But the needs are not always explicitly recognized. For example, not everyone knows what they need to prevent AIDS, and not everyone wants what they need. Identification of needs depends on a conception of human nature and the social world. Wishes may be identified by means of a questionnaire. If only the satisfaction of desire is considered as the purpose of information retrieval, most of the indexation rules for recovery practices will serve as the predominant social organization form. Among the important indexation rules to the satisfaction of desire in consumer capitalism, for example, are those that effectively represent goods for consumption. On the other hand, if the text retrieval must meet the requirements, then the rules for its practice may not only be inconsistent with the objectives of the dominant social order, but also be antagonistic to them. (Frohmann, 1990).

These and other issues have required a much more extensive analysis or political analysis as the author suggests. In any case, this issue makes clear is that “differing conceptions of the social role of text retrieval will determine the kinds of indexing rules we construct.” (Frohmann, 1990, 98). This clarification also contributes to the understanding that it may be limiting that IRS are geared primarily to "desires" and to what the user already has in mind, what he wants to find. This practice can hide the different points of view present in literature.

With the analysis of such research one can see, for example, that the need for information is related to the completion of a task, the fulfillment of a goal, this need may exceed what was wished or expected to find in a search for information. Therefore, Knowledge Organization System should provide the means for the user to realize that there are different perspectives and make the choice that he considers most appropriate. Thus, the RI process can be compared to learning, because as described, the result of RI may exceed what the user had expected to find, and constructivism, as developed epistemological position and widely used in the field of education is shown especially useful also for the pragmatic approach in KO, which should provide the necessary subsidies for this learning in RI. The information environment, influenced by the intensive use of technology presents new contexts to explore in the KO, which requires the overcoming of the subject-object fragmentation by a holistic approach of further consideration and incorporation of the context element. This is a particular appeal of epistemological change in which it is important to look into possible contributions of constructivist epistemology.

From the analysis of examined theoretical aspects and based on the way to classify the epistemological positions developed by Hessen (1999), it is considered that there is a theoretical body in KO related to pragmatism, according to their guesses related to the possibilities of knowledge, constructivism, according to his assumptions regarding the origins of knowledge and critical realism, according to his guesses related to the essence of knowledge.

Table 1: Epistemological Framework of pragmatic approach in KO (based on Hessen,1999)

Pragmatic approach in KO		
Is the subject able to grasp the object?	Possibility of knowledge	Pragmatism
Is the source of knowledge the reason or experience?	Source of knowledge	Construtivism
Does the object determine the subject or does the subject determine the object?	Essence of knowledge	Critical Realism

Although Hessen (1999) does not refer to constructivism, from his remarks on the "source of knowledge", it was inferred that constructivism would fit into this category, as one of the other approaches that lie between rationalism and empiricism. Rationalism according to Hessen (1999) sees in thought, in the reason, the main source of human knowledge (logic and mathematics), as empiricism, sees the experience as the only source of human knowledge. Between these two opposing positions on the origin of knowledge, in an attempt to mediate these, there are intellectualism, apriorism and critical position. Apart from these, according to the adaptation developed and presented on Table 1, we can mention constructivism as epistemological positions regarding the origin of knowledge that would be between the two poles of this category, rationalism and empiricism. In pragmatic approach in KO, the most evident epistemological position and already mentioned by authors of the area, is pragmatic, but constructivism and critical realism show to be consistent with this position because they respond to different questions on the development of knowledge in KO, forming a theoretical body we can call a pragmatic approach.

Final considerations

It was found in this research that constructivism is an epistemological stance compatible with pragmatic epistemological stance in KO. In this theoretical body of KO, called pragmatic approach, we consider knowledge, and therefore information, a phenomenon that is built in different circumstances and involves different actors. This construction, because it depends on the individual activity, which is fallible, also implies fallibility and provisional knowledge. Besides being fallible, each individual is inserted in a context and builds knowledge in circumstances and specific objectives. So, in addition to temporary (variation in time), knowledge of the same object of reality can also vary according to the context in which it was built (variation in space).

From this evidence KOS should take into account the different perspectives, views, and the domain to which it is intended. These different perspectives should be identified in literature to be indexed by KOS, to meet the possible information needs of users. These needs may exceed what users wished or wanted to find, and so the information retrieved with the help of KOS can also be a learning element.

KOS, whose purpose is to optimize IR, would have a central role in this process that can be compared to learning. In a way that constructivist theories widely used in education are shown useful also for KO. For these theories present contributions to the optimization of the learning process. The constructivist perspective in KO, when contributing to the understanding of the structure of the various areas of knowledge, by being one of the theoretical currents involved in explaining how human intelligence is developed considering the mutual actions between the individual and the environment and under the influence of the decision making from the interpenetration of objective and subjective elements, has the potential to expand and optimize IR.

References

Blair, David C. (2003). Information retrieval and the philosophy of language. *Annual Review of Information Science and Technology*, 37(1): 3-50.

Castañon, Gustavo (2005). Construtivismo e Ciências Humanas.*Ciências & Cognição*,5(1):36-49

Fainburg, Linda I. (2009). Information seeking and learning: a comparison of Kuhlthau's information seeking model and John Dewey's problem solving model. *New Library World*, 110(9): 457-466.

Francelin, Marivalde M. & Kobashi, Nair Y. (2011) Concepções sobre o conceito na organização da informação e do conhecimento. *Ciência da Informação*, 40(2): 207-228.

Frohmann, Bernd (2008). Subjectivity and Information Ethics. *Journal of the American Society for Information Science and Technology*, 59(2): 267–277.

Frohmann, Bernd (1990). Rules of indexing: a critique of mentalism in information retrieval theory. *Journal of Documentation*, 46(2) June: 81-101.

Hessen, Johannes (1999). *Teoria do conhecimento*. São Paulo: Martins Fontes.

Hjørland, Birger (2009). Concept theory. *Journal of the American Society for Information Science and Technology*, 60(8): 1519-1536.

Hjørland, Birger (2003). Fundamentals of Knowledge Organization. *Knowledge Organization*, 30(2):87-111.

Hjørland, Birger (2007). Semantic and Knowledge organization. *Annual Review of Information Science and Technology*, 41(1): 367-405.

Hjørland, Birger, & Christensen, Frank S. (2002). Work Tasks and Socio-Cognitive Relevance: A Specific Example.*Journal of the American Society for Information Science and Technology*, 53(11): 960-965.

Hjørland, Birger, & Pedersen, Karsten Nissen (2005). A substantive theory of classification for information retrieval.*Journal of Documentation*, 61(5): 582-597.

Popper, Karl (1985). *A Lógica da pesquisa científica*. São Paulo: EDUSP.

Popper, Karl (1987). *A sociedade aberta e seus inimigos*. São Paulo: EDUSP.

Tennis, Joseph (2008). Epistemology, Theory, and Methodology in Knowledge Organization: Toward a Classification, Metatheory, and Research Framework. *Knowledge Organization*, 5(2/3): 102-112.

Renata Cristina Gutierres Castanha, Fábio Sampaio Rosas and Maria Cláudia Cabrini Grácio

The Complementarity of Hjørland's and Tennis's Proposals to Domain Analysis under Bibliometrics

Abstract

This paper presents a dialogue between the methodologies proposed by Hjørland and Tennis for Domain Analysis under Bibliometric studies. According to Tennis (2003; 2012), before starting any Domain Analysis, one must define the area to be studied, specify the scope of this analysis by the two axes and establish the ultimate purpose of the analysis. For a better analysis of the knowledge field through the use of Bibliometric studies, it is necessary that these studies also consider social, historical and epistemological aspects, as pointed by Hjørland (2002). It is concluded that the methodologies presented by Tennis (2003; 2012) are a complement to Hjørland's (2002) and both bring important contributions to the development of scientific research using Bibliometric studies as Domain Analysis approach.

Introduction

Scientific knowledge construction results from a process involving individuals' social and work relationships in a discourse community, and in this context, it is constructed and disseminated through documental records. This practice has been the core study of Knowledge Organization under the aegis of Information Science.

The increasing growth of scientific knowledge production records has encouraged the mapping of this disseminated knowledge. The publication of these scientific results allows knowledge socialization, and, in a helical cycle between science production and communication, knowledge is the effect of these social relationships (Guimarães, 2000). In this context of scientific production growth, an analysis and assessment of this production has become essential to create instruments for identifying a science's behavior. We highlight the role of Bibliometrics to bring significant contribution to provide a quantitative analysis of the communication processes, the nature and development of scientific domains, that associated to historical and epistemological studies allows an objective and broad view of a scientific domain.

Metric Studies (including Scientometrics, Bibliometrics, Informetrics, among others metrics) are quantitative studies focused on the development of appropriate methodologies for formulating adequate indicators for every measurement level of a scientific domain (Spinak, 1998). However, over the past years, a strong quantitative trend has been observed in metric studies, which do not consider the historical and social context and the relation within the community where knowledge was created. Macias-Chapula (1998) considers that science is a social process in which its actions and behaviors are closely related to its context, and its quantitative aspects arising from bibliometric analysis need to be interpreted and contextualized.

In this line, Hjørland and Albrechtsen (1995) claim that the best way to understand and interpret information and scientific dynamics is by studying the knowledge

domains in which they are inserted in relation to their discursive communities sharing similar theories of thought, language and knowledge.

In this context, the scientific production of a domain configures as the main propeller element of information and knowledge development, as the publication is an intrinsic activity to scientific research of a domain. For the authors, Bibliometrics constitutes a scientific approach that provides valuable information on a domain, as well as on the relations across disciplines, thematic, authors, among others, revealing social patterns of scientific communication.

Considering that Metric Studies have been characterized mostly by their quantitative aspects, this paper aims to highlight the importance of the social paradigm contribution of Hjørland's Domain Analysis to aggregate a social and contextual perspective to Bibliometrics, allowing wider and deeper objective visualizations, by taking into account social, contextual and quantitative aspects of a domain in its analyses.

Domain Analysis in Information Science

Hjørland and Albrechtsen (1995, p.400) were the first authors to use the concept of domain, theory and methodology in Information Science, defining domains as "thought or discourse communities, which are parts of society's division of labor", hence establishing their social and cultural foundations. Communication patterns and language, working structure, cooperation standards, knowledge organization, information systems and relevance criteria are reflections of work objects of a community (domain) and their role in society (Hjørland, 2002). All scientific labor reflects the social context in which it is inserted, its historical moment, changes and complexities, and so it is liable to the current hegemonic theoretical currents.

The concept of a Domain can be understood as a field of study in its different specialties, a set of literature on a particular subject or group of people working together in an organization, comprehending the study of a discourse community, and the role this community plays in science (Mai, 2005; Hjørland and Albrechtsen, 1995).

Thus, a domain can be a scientific discipline, a scientific knowledge area or a discourse community related to a political party, religion or any other group. In this context, the notion of knowledge domain encompasses both the conceptual universe and the way that a given discourse community is formed (Thellefsen & Thellefsen, 2004; Mai, 2005; Oliveira & Grácio, 2013).

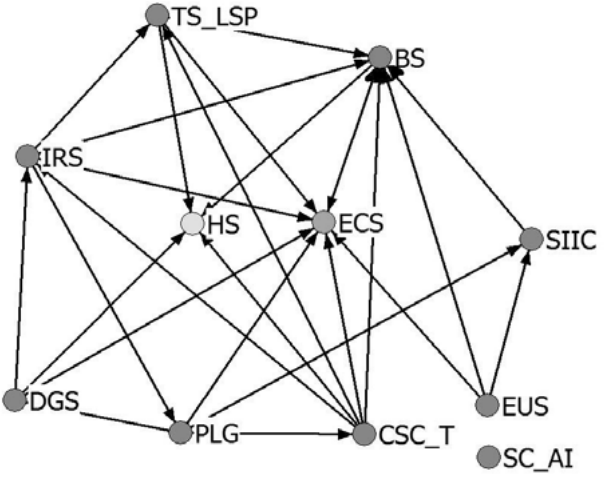
Domain Analysis seeks the integration between individual and social context of the communities, where they are inserted and the concepts of information become meaningful when sharing occurs between these different communities and their members. Hjørland (2002) presents 11 approaches to Domain Analysis, highlighting that the combined use of more than one of these approaches enriches the analysis and understanding of a domain (Figure 1).

Hjørland points that complement empirical approaches, among them Bibliometrics analysis, with other approaches, especially the epistemological and historical

approaches, provide a broader and deeper knowledge on the studied domain.

As seen in Figure 1, bibliometric studies are reinforced by epistemological studies. It is noteworthy that, in Hjørland's view, there is a symmetrical and complementary relationship between these two approaches when analyzing a domain, i.e., when working with one of these two approaches, one must associate or enrich it with the other.

Figure 1. Hjørland's 11 approaches to Domain Analysis and the co-operation among them.



Caption: PLG = Producing literature guides or subject gateways; CSC_T = Constructing special classifications and thesauri; IRS = Indexing and retrieving specialties; EUS = Empirical user studies; BS = Bibliometric studies; HS = Historical studies; DGS = Document and genre studies; ECS =Epistemological and critical studies; TS_LSP = Terminological studies, language for special purpose, database semantics and discourse studies; SIIC = Structures and institutions in scientific communication; SC_AI = Scientific cognition, expert knowledge and artificial intelligence.

Other researchers have studied the issues related to Domain Analysis. Specially in Knowledge Organization, Smiraglia (2011) highlights the importance of all researchers interact theoretically through geopolitical and cultural borders. Using Domain Analysis, it is possible to assess what is actually important or significant in a scientific field, so that aspects such as trends, patterns, processes, dominant thoughts, agents and their relationships can be identified and analyzed.

Domain Analysis is an outstanding theoretical and methodological approach to characterization and evaluation of science, typically represented by scientific literature or research community, identifying the conditions under which scientific knowledge is constructed and socialized. Through Domain Analysis, it becomes possible to verify what is actually significant in a particular area, such as trends, theoretical currents, patterns, processes, agents and its relationships can be identified and analyzed (Guimarães, 2015).

These studies also point to the constructed character of information, indicating the relevance of historical, cultural and social dimensions in which information flows are presented, involving the applied dimension of Information Science. Therefore, Domain Analysis opposes the original and classical studies that explains informational issues through "laws and generalizations", often in a static way (Araújo, 2009).

According to Capurro (2003), special emphasis is given to the study of discourse communities and their different perspectives, points of view, approaches, controversies or user communities in different fields of knowledge.

Methodological issues about Domain

Tennis (2003) observes that it is necessary to acknowledge the core and boundaries of certain knowledge to analyze a domain, regardless of its magnitude. The author focuses his studies on domain, especially on methodological issues, rather than on conceptual issues, i.e. those operationalization procedures meant to define domains. His methodological conception seeks the construction of a methodological approach that can be seen by other domain analysts in their different laterality.

In that direction, Tennis (2003) presents two axes and four parameters, which contribute to a better Domain Analysis; he does not aim to define domain, but to complement Hjørland's approaches (2002) with a methodology that precedes Domain Analysis, and which facilitates the work of domain analysts. In this context, the author discusses the importance to accurately outline a domain, and that it should be done prior to Domain Analysis. The author suggests establishing criteria for delimitating the area, with two axes, which he named modulation areas and degrees of specialization.

The first axis is the modulation area. This axis considers two parameters: the first one to indicate the total of what is covered in Domain Analysis, i.e., its extension; and the second parameter indicates the nomenclature of the domain.

Regarding the second axis, degrees of specialization, Tennis states that it can only be set after the definition of the domain extension limits (covered in the first axis).

This second axis also has two parameters: the first one must qualify the domain, which the author names focus; and the second parameter positions the domain in relation to other domains, their intersection. It is understood, therefore, that the greater the length of a domain, the lower its specialization will be, and, the greater the intension or specialization of a domain, the smaller its extent. In any study involving Domain Analysis, the methodological application of these two axes at the beginning of the studies is necessary. In other words, it is necessary to know the limits and depth of the domain under study. For a better understanding, Chart 1 presents a comparative view of the concepts in Hjørland's and Tennis's theories.

Figure 2. Main concepts involved in Hjørland's and Tennis's theories

Hjørland (2002)	Tennis (2003)
- Domains are "thought and discursive communities" as a discipline, a school field; - Analytical view of domain within the Information Science; - Eleven approaches for methodological application in Domain Analysis; - Domain Analysis is a key point in the development of efficient information systems; - Doesn't address definition of domain boundaries.	- Sets the operationalization of Domain Analysis; - Presents the need for domain analysts to provide a standardized definition of a particular domain and that it may be transferable to other researchers; - Complements Hjørland's approaches with a methodology of two application axes, with two parameters each, aimed at delineating the area for further analysis.

Scientific Production and Bibliometrics

Scientific publications are an important part of the science dynamic. One way to study this dynamic is through bibliometric indicators, which are used as indirect measures of scientific research activities and contribute to the understanding of scientific community frames, as well as the social, political and economical impact of the produced sciences. They are relevant for monitoring science and enable an estimate of how countries contribute to mainstream science (Vanz & Stumpf, 2010).

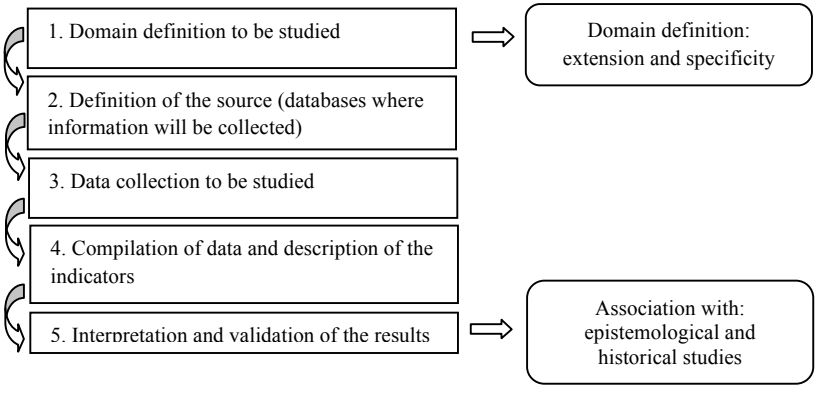
Bibliometric studies involve production, citation and relational indicators, which constitute a consistent and objective approach to analyze and characterize a scientific domain (Hjørland, 2002). These studies are based on detailed analysis of the connections between individual documents, and these details and connections highlight the explicit recognition of authors dependence through articles, research, courses, approaches and geographical regions. As an AD approach, bibliometric indicators contribute to the visualization of different aspects and characteristics of a domain, such as production indicators (language, the forms of communication and knowledge organization), citation indicators (relevance criteria, epistemic communities) and relational indicators (collaboration patterns, front of scientific research, epistemic communities). Hjørland supports an analytical view of domain in the scope of Information Science, understanding Bibliometric studies as a strong, objective and consolidated methodology for analysis and identification of a scientific domain and unanimous in pointing out the advantages of articulating Bibliometric studies with epistemological and historical approaches or others of qualitative nature.

Among the Bibliometric studies, Hjørland (2002) specially highlights the contribution of citation and cocitation analysis to visualize scientific knowledge areas. Citations and co-citations studies are relevant procedures for analyzing interlocution among researchers and their role in different areas of knowledge, as they contribute to the visualization of communicative and interactive process, as well as the underlying structure of a knowledge domain. The set of references of scientific papers can thus be analyzed as a reflection of a discourse community, so as to constitute a domain.

They are defined as quantitative methodologies that establish measures and

indicators to reveal the behavior of a scientific domain, specially through its knowledge production. Although Bibliometrics has been recognized as an efficient approach associated to other theoretical approaches within Information Science, it has rarely been used as suggested by the analytical approach of Domain Analysis, i.e., by adopting a social perspective in the study of informational practices (Hjørland, 2002). For Hjørland (2002), the best way to correctly analyze Bibliometrics indicators is by using qualitative Domain Analysis approaches such as historical, epistemological and critical studies, i.e., a contextualization is required for the obtained Bibliometric data. We illustrate the importance of associating epistemological and historical approaches to bibliometric studies, as cited by Hjørland, as well as the precise definition of the domain, through the illustration of steps taken for the methodological procedures of bibliometric studies, presented in Figure 2.

Figure 3. Steps in the methodological procedures in bibliometric analyzes. Inspired by McCain (1990)



The first step in all bibliometric analysis is the definition of the Domain under study. In this context, the two-axis proposition for defining a domain by Tennis (2003) makes a significant contribution, providing greater clarity to the exact definition of the area to be studied. This step takes place prior to the bibliometric analysis itself. In this first step, it is essential to accurately outline the area, since subsequent steps and procedures depend on this first step, as well as obtaining more robust and faithful results to the studied behavior. By defining the first axis - domain extension - the amplitude of the bibliometric study is established, that is, how comprehensive the results of the bibliometric study are, and the nomenclature of this domain. By defining the second axis - the degrees of specialization of the domain - in a Bibliometric study, one specifies the depth of the analysis within the extension (amplitude), considering the focus constraint (domain scope) and the intersection (position in relation to other areas). Tennis (2003) argues that describing an entire domain is not either desirable nor feasible, even if the domain has a name and an extension that can be set. Thus, we highlight the importance of defining and delimitating the domain to be analyzed by a bibliometric study. By decreasing the amplitude of a domain, its extension is restricted

and its specificity can then be increased. It is also conceivable that the degree of specialization limits the analysis to one single person. Thus, associating the methodology proposed by Tennis (2003) to the first step of a bibliometric study (domain definition) is significant and representative to bibliometric analysis because its use contributes to greater clarity regarding the definition and boundaries of the domain under study, contributes to the necessary procedures and security of more cohesive results in relation to the objectives proposed in the study.

In the second step of a bibliometric analysis, the sources constituting the main and most representative communication channels are defined for data collection, analysis and more expressive results in relation to the objectives of the study. Thus, the correct identification of these sources decisively contributes to a better Domain Analysis, conducted by bibliometric approach. In this context, we highlight the need for epistemological knowledge of the domain under study, considering that the forms of communication reflect the research objects of a domain.

In the last step of all bibliometric analysis, it is essential to analyze, contextualize and qualify the results in light of the epistemology, as well as its historical aspects of development so that the obtained results provide significant, real and useful contributions for the community, as well as point appropriate scientific policy proposals. In this sense, bibliometric studies are inseparable from epistemological, critical and historical studies. These approaches complement each other bringing a contextual data treatment profile that reflect a more complete analysis of the collected data, including the four factors the author described in bibliometric studies (Hjørland, 2002), namely:construction of a map, the authors' citation behavior; chosen methodologies by the ease of use and popular theories, as well as the socio-contextual factors from the historical, epistemological and critical studies that influence the final result of data or Domain Analysis.

Recently, Tennis (2012) presented a supposedly ideal methodology of Domain Analysis, separating it into two types: descriptive Domain Analysis, that would fit Bibliometric studies; and instrumental Domain Analysis, focused on the construction or review of an information system. The author claims that the supposed ideal form of Domain Analysis implies disclosing the configuration of the device, i.e., the basic elements of DA must be clear before proceeding to the analysis. Thus, according to the author, one must first define the scope and form of analysis, specify the scope and range (Tennis, 2003) and finally establish the purpose of this Domain Analysis, if it is a descriptive or instrumental one.

Conclusions

For better understanding a domain, we highlight the evidence in both Hjørland's (2002) and Tennis's (2003, 2012) studies for their concern regarding the domain context across its analysis and the importance of studies that consider epistemological, critical and historical aspects along with Bibliometric studies approach. In practice,

according to Tennis (2012), prior to starting any Domain Analysis, it is necessary to have the definition of the studied domain, specify its scope and analysis reach through the two axes (Tennis, 2003) and establish the ultimate purpose of the analysis, descriptive or instrumental. Then, the review process begins with the approaches proposed by Hjørland (2002), and when it comes to descriptive Domain Analysis, with the use of Bibliometric studies approach, to achieve a more reliable result, it is essential to associate it with another qualitative approach such as epistemological, critical and historical studies. The methodologies presented by Tennis (2003; 2012) are complementary to Hjørland's (2002) and bring important contributions to the development of scientific research using bibliometric studies as a Domain Analysis approach. Therefore, the implementation of Hjørland's and Tennis's methodologies in Domain Analysis through Bibliometric studies contributes to a better understanding of the evolution and dynamic of science.

References

Araújo, Carlos Alberto Ávila (2009). Correntes teóricas da Ciência da Informação. *Ciência da Informação*, 38(3): 192-204.

Capurro, Rafael (2003). Epistemologia e Ciência da Informação. In *V Encontro Nacional de Pesquisa em Ciência da Informação, Belo Horizonte (Brasil)*.

Guimarães, José Augusto Chaves (2015). Análise de domínio como perspectiva metodológica em organização da informação. *Ciência da Informação*, 43(1).

Hjørland, Birger (2002). Domain Analysis in information science: eleven approaches – tradional as well as innovative. *Journal of Documentation*, 58(4).

Hjørland, Birger, & Albrechtsen, Hanne (1995) Toward a new horizon in information science Domain Analysis", *Journal of the American Society for Information Science*, 46(6).

Macias-Chapula, Cesar A. (1998). O papel da informetria e da cienciometria e sua perspectiva nacional e internacional. *Ciência da Informação*, 27(2): 134-140

Mai, Jens-Erik (2005). Analysis in indexing: document and domain centered approaches. *Information Processing and Management*, 41.

McCain, Katherine W. (1990). Mapping author intellectual space: a technical overview. *Journal of the American Society for Information Science*, 41(66).

Oliveira, Ely Francina Tannuri, & Grácio, Maria Cláudia Cabrini (2013). Studies of author cocitation analysis: a bibliometric approach for Domain Analysis. *Iris*, 2(1).

Smiraglia, Richard P. (2011). Domain coherence within knowledge organization: people, interacting theoretically, across geopolitical and cultural boundaries. In*Procs of the 39th annual CAIS/ACSI conf*, University of New Brunswick, Fredericton, NB Canada.

Tennis, Joseph T. (2003). Two Axes of Domains for Domain Analysis. *Knowledge Organization*, 30(4).

Tennis, Joseph T. (2012). What does a Domain Analysis look like in form, function, and genre?.*Brazilian Journal of Information Science*, 6(1).

Thellefsen, Torkild L., & Thellefsen, Martin M. (2004). Pragmatic semiotics and knowledge organization. *Knowledge Organization*, 31(3).

Vanz, Andréa de Souza, & Stumpf, Ida R.Chittó (2010). Procedimentos e ferramentas aplicados aos estudos bibliométricos. *Informação & Sociedade*, 20(2).

Leilah S. Bufrem, Ely F. Tannuri Oliveira and Bruno H. Alves

Seminal Theoretical References and Their Contributions to Knowledge Organization (KO) from Citation Analysis of ISKO Ibérico Communications (2005/2015)

Abstract

This study aims to highlight the most productive group of ISKO-Ibérico researchersin its editions from 2005 to 2015. From these productions, we seek to consider the most cited authors from the citing corpus to highlight and analyze the group of authors considered seminal in Knowledge Organization. As research procedure, we retrieved 379 complete papers from the proceedings of the last five ISKO-Ibérico editions. We built up a list of the 30 most productive researchers with at least four papers each. Then, we proceeded to citation analysis for these 30 researchers. We considered the cited authors in at least seven papers, totaling 25 researchers, with the analysis of theoretical lines of these researchers. We concluded that, from citation analysis, taking ISKO-Ibérico as source, it was possible to identify the seminal theoretical references for Knowledge Organization.

Introduction

Considering that knowledge of a scientific domain requires an understanding of its origins, its objects and research themes recognized in its historical development, it is argued that one way to achieve this understanding is by acknowledging the domain's seminal theoretical frameworks. Through the identification of texts, founding authors, and explicit links among the most significant representatives of a domain, researchers not only expands their research opportunities and scientific production in the area, but are also able to recognize the relationship between their research problematic and the underpinnings guiding their peers' production.

This study considers seminal scientific texts the ones whose merit lies in the importance they have had to their "descendants". They become essential in certain areas or themes, either because they cause a break or insight, enabling new and generating synthesis of ideas or because their impact would cause a change in the concepts already accepted on certain aspects of reality or on themes related to a knowledge domain; they may either propose models or alternative transforming structures of a conceptual or methodological framework or present original research methods. Due to these characteristics, these articles may become classics in a knowledge domain - or even canonical - becoming necessary in order to learn a domain in its entirety.

By defining what it considered a seminal article, Lussky refers to the one presenting a new theory that may be accepted by the community:

> This seminal paper influences the scholarly community's thinking and ultimately, the body of knowledge. The seminal paper stimulates the writing of other scholarly papers. Last, the novel thinking, expressed in the seminal paper and subsequent scholarly papers, is organized into new patterns of thinking which can be recorded in subject heading schemes and then applied to the subject

indexing of newly published scholarly papers" (Lussky, 2004, p. 4-5).

However, when asked how it would be possible to identify seminal papers, the author claims that they are often the most cited ones among those addressing the research area in question. Thus, this type of article has been called "seminal" or "influential" or "core" or "classic", therefore indicating the central importance of the study that reports to a research field.

The founding authors are the producers of seminal papers, which are recognized as the most fertile not only by the repercussions in their own knowledge domain, but also by the recognition that authors from other areas assign to them, making them anthological.

Among the possibilities for identifying these articles, citation analysis is a tool to identify articles often cited in the literature. Considering citation as the remissive act a text performs to other texts, illustrating the paths taken by a researcher in knowledge construction (Meadows, 1999), it is believed that citation analysis allows scientific communication mapping.

With the production mapped through citation analysis, it is possible to identify evidences of how communication in an area of knowledge has been taking place, and as a result, contribute to the construction of relationship networks in order to reveal communication and relations among researchers. For this communication, citation and content analysis were adopted to recognize socio-generative aspects as well as the perspective of cognitive institutionalization of scientific research evidenced in the corpus, in order to relate the most representative citing authors of the corpus in relation to the domain and the most representative authors in the cited references.

Thus, citation analysis of scientific production has been recognized by any area of knowledge as a valuable tool as it can represent the dynamic, social and historical process of the relations among seminal authors, also considered founding authors, and those they influenced, as well as the thematic and institutional relations disclosed within a scientific context. Furthermore, we consider the emphasis given by Glänzel (2003) to citations as indicators of community paradigms formed in the history of a domain, its ethical and methodological procedures, their groups of scientists, their publications. They also disclose researchers of greater impact of an area, conception that converges to what was stated by Smiraglia (2011, p.181) that "citations define the domain" and that the identification of the most cited authors constitutes the front researchers, or the core of researchers in an area.

By linking citation studies to the Mertonian sociology of science, Thelwall (2008) reinforces the argument that the set of citations made by researchers translates the recognition of the influence of previous studies. This set, endowed with scientific prestige, can be identified, according to Bourdieu (2013, p. 111), by the recognition granted in the scientific area.

With this perception, we argue that the study and historical rescue of defined scientific production, considering its theoretical and methodological structure favor, as

argued by Lloyd, the understanding of a scientific area (1995, p. 38). This understanding would include the cultural and anthropological, the historical and social aspects, related to an area of study in its different specialties, a set of literature on a particular subject or a group of people working in an organization, including the study of a discourse community and the role it plays in science (May, 2005; Hjørland & Albrechtsen, 1995).

For this research, we defined Knowledge Organization (KO) as the area of study concerned with the nature and quality of processes involved in them: the description of documents, indexing, classification, processes that are carried out by librarians, information professionals or by computational algorithms (Hjørland, 2007). With support on Hjørland (2002, p. 432), we highlight the contribution of citation and cocitation studies, especially regarding the construction of bibliometric mappings or visualization of scientific knowledge areas. Citation studies are based on analysis of citation frequencies, either of authors or documents, enabling the visualization of a domain. Citation is considered an objective and clear indicator of scientific communication for allowing the identification of groups of scientists and their publications in order to show researchers the greatest impact of an area, pointing their paradigms and relevant methodological procedures, as well as vanguard researchers that build new knowledge in the area. This investigation then turns to the corpus constituted by the literature produced in Knowledge Organization (KO), recorded in ISKO Ibérico proceedings (2005/2015), aiming to represent and analyze the relationships among the most fruitful authors in the period and the seminal authors they cited in the considered corpus.

Based on these assumptions, the study aims to highlight the most productive group of ISKO Ibérico researchers, in its editions from 2005 to 2015. From these productions, we seeks to identify the most cited authors through the citing corpus in order to highlight and analyze the group of authors considered seminal in KO.

Considering the principle that authors and papers regarded as seminal generate and communicate methodological and epistemological foundations, built and consolidated in scientific domains, the recognition of their importance leads to knowledge construction and transformation. Thus, the corpus constituted by scientific papers in an knowledge domain becomes a suitable locus for consideration and disclosure of these authors and their seminal papers through references in the communications. We justify the choice for ISKO congresses by considering that they represent concrete evidence of current ideas, built by a group of researchers who are the elite of KO domain.

Research Procedure

We retrieved 379 full papers from the proceedings of ISKO-Ibérico editions, as follows: 43 papers from Barcelona,2005; 60 from León,2007; 87 from Valencia,2009; 38 from Ferrol,2011, 92 from Oporto,2013, and 59 from Murcia,2015.

The co-authorships of this corpus was unfolded building up a list, in descending

order, of 30 authors with at least four papers each, resulting in 5.2% of the total of 573 researchers. The value is representative according to Price's Law of elitism, considering that "n" represents the total number of researchers in a discipline, then "√n" represent the elite of the studied discipline. We proceeded to the citation analysis of these 30 researchers responsible for 129 papers. We justified the cut we performed, as the citation analysis turned to 34% (129) of the 379 total papers in analysis, percentage considered representative. We excluded self-citations in order to avoid bias, such as the position achieved by citing authors of seminal or founding authors.

We found 2,690 references related to the 1,675 cited authors. We considered the authors cited in at least seven papers, totaling 25 researchers, and the researchers withmore than one reference in the same article were counted only once. We built an asymmetric 30 X 25 matrix, with the most productive researchers (30 citing) and the most cited ones (25 cited). We generated a two-mode relationship network between the two variables, using Ucinet software. The most cited authors were considered seminal authors, i.e.; those who constitute the fundamental theoretical reference for Knowledge Organization domain, recognizing their different theoretical perspectives.

Data analysis and presentation

Table 1 presents the 25 most cited authors, highlighting Hjørland, the researcher of greatest relevance, with 30 papers presented from 2005 to 2015. He conducts studies on Knowledge Organization, which in his conception, refers to activities related to information organization in bibliographic records (2003, p. 87). The author also addresses issues related to Information Architecture, Information Retrieval and Information Behavior; and has the KO studies on Domain Analysis as one of his greatest contributions, in the 1990's; and the user-centered perspective together with Albrechtsen, which was cited 12 times. These studies have served as foundations to the epistemology of scientific knowledge construction in Information Science, as they have been developed and socialized.

Dahlberg, the second most cited author, contributed to the foundation of the International Society for Knowledge Organization (ISKO), contributed to the creation of Classification Scheme for Knowledge Organization Literature (CSKOL), which groups the main KO concepts into ten categories. As pointed out by Guimarães (2008), Dahlberg, Henry Evelyn Bliss and Dagobert Soergel conceive knowledge organization as an autonomous area in the sciences system.

Guimarães and García Gutiérrez conduct studies on document analysis, knowledge organization and representation, epistemology of information science, and professional ethics in Information Science; and especially Garcia Gutierrez studies non-epistemic communities, mediation and culture. We add, to these researchers, Barité, who conducts research in the areas common to both cited researchers in addition to Terminology. For this author, knowledge is embodied in documents and is expressed through concepts organized into systems which are useful to science, literature and

documentation. Highlighting the epistemological studies, the presence of Buckland and Capurro is noted, as they were cited in nine and ten papers, respectively, and whose studies about document reinforce the theoretical foundations of the area.

Lancaster, Fujita and Foskett develop research in Information Retrieval, especially with emphasis on indexing languages, whose concerns about the construction of such languages are influenced by the research of the Classification Research Group.

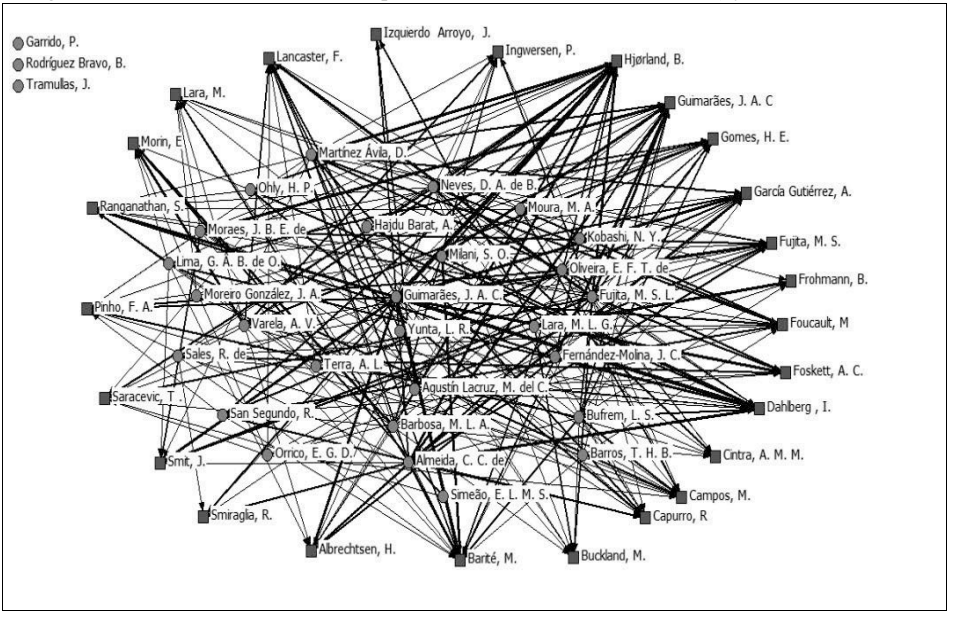
Table 1 - Most cited authors in ISKO Ibérico

Most cited researchers	# of papers they were cited
Hjørland, B.	30
Dahlberg , I.	26
Barité, M.	17
Guimarães, J. A. C	17
Lancaster, F.	16
García Gutiérrez, A.	14
Albrechtsen, H.	12
Gomes, H. E.	11
Foucault, M	11
Capurro, R	10
Fujita, M. S.	10
Lara, M.	10
Buckland, M.	9
Campos, M.	9
Frohmann, B.	9
Morin, E	9
Pinho, F. A.	8
Ranganathan, S.	8
Saracevic, T.	8
Cintra, A. M. M.	7
Foskett, A. C.	7
Ingwersen, P.	7
Izquierdo Arroyo, J.	7
Smiraglia, R.	7
Smit, J.	7

Although in a position of less emphasis among the most cited, Ranganathan can be regarded a seminal author, due to the application of the principles of classical philosophy and logic, and the rigor of mathematical sciences in the organization of conceptual fields. Smiraglia, cited in seven papers, highlights the construction of tools for storage and retrieval of what he denominates documentary entities. Guimarães (2008) considers the influence of this current in important studies and research carried out in Brazil by different groups of researchers in the area, especially in IBICT and UNB graduate programs. There are also, among the most cited authors, those who provide sociology and philosophy foundations, such as Morin and Foucault, especially due to the interdisciplinary character of IS.

Figure 1 completes the analyzes we performed, presenting the most cited authors in a network configuration, represented in blue squares: the citing authors in the central region of the network, in red; and the intensity of connections represented by the vector segments, with thicknesses related to higher frequencies of cited papers. The network is presented in one single component.

Figure 1 - Two-mode network: most productive and most cited authors (by Ucinet Software)



Hjørland, the most cited researcher is related to several other researchers visualized by the segments, coming from Guimarães (10 citations), Almeida (6 quotes), Fujita (5), Martínez-Ávila (4) and others, totaling 19 authors,from the total of 30 who cite the author in their research. Dahlberg, the second most cited researcher is related to several other researchers, by the segments, originating from Guimarães (7 citations), Almeida (6 citations), Ohly (3 citations), among others, with a total of 15 citations, developing research in Epistemology, Knowledge, Classification, Ontology and other themes.

A more careful analysis of the network shows three researchers among the most productive (citing), but also among the most cited, namely: Guimarães, Fujita and Lara, suggesting, in the context of ISKO Ibérico, the representativeness of this group of Brazilian researchers. We highlight Guimarães as the most central researcher, for being both citing and cited author, and therefore presenting the highest centrality in the resulting configuration.

Completing the analysis of Table 1 and Figure 1, other researchers present in the table and in the figure, but not highlighted in this analysis, bring their contribution to Knowledge Organization area. Within their specialty and perspective, they offer

elements that cause the area to advance and rethink their theoretical and methodological aspects.

Conclusion

From the presented, systematized and analyzed data, resulting in 129 papers in the 2005-2015 period in ISKO Ibérico, we highlight the representativeness of the papers presented at this conference, especially because it is the group of the 25 most cited researchers coming from different countries with a prevalence of 68% of renowned foreign researchers.

If we consider the evolution of 'Subject Treatment of Information' (TTI), according to Guimarães (2008), from the three distinct lines of approach, subject cataloging (predominantly American), indexing (predominantly English) and *analyse documentaire* (predominantly French), it is possible to note the great English line influence, concluding therefore that, from citation analysis, considering ISKO Ibérico as source, it was possible to identify the seminal theoretical references for the area of Knowledge Organization.

References

Bourdieu, Pierre (2013). *Homo academicus*. Florianópolis: EDUFSC.

Glänzel, Wolfgang (2003). *Bibliometrics as a research field: a course on theory and application of bibliometric indicators*. Bélgica: [s.n.].

Guimarães, José Augusto Chaves (2008). A dimensão teórica do tratamento temático da informação e suas interlocuções com o universo científico da ISKO. *Revista Ibero-americana de Ciência da Informação (RICI)*, 1(1): 77-99.

Hjørland, Birger, & Albrechtsen, H. (1995). Toward a new horizon in Information Science: domain-analysis. *Journal of the American Society for Information Science*, 46(6): 400-425.

Hjørland, Birger (2002). Domain analysis in information science: eleven approaches-traditional as well as innovative. *Journal of Documentation*, 58(4): 422-462.

Hjørland, Birger (2003). Fundamentals of knowledge organization. *Knowledge Organization*, 30(2) April: 87-111.

Hjørland, Birger (2007). What is Knowledge Organization (KO)?.*Knowledge Organization*, 35 (2/3): 86-101.

Lloyd, Christopher (1995). *As estruturas da história*. Rio de Janeiro: Zahar.

Lussky, Joan P. (2004). *Bibliometric patterns in an historical medical index: using the newly digitized index catalogue of the library of the surgeon general's office, U.S.Army*. Drexel University.

Mai, Jens-Erik (2005). Analysis in indexing: document and domain centered approaches. *Information Processing and Management*, 41(3): 599-611.

Meadows, Arthur J. (1999). *A comunicação científica*. Brasília: Briquet de Lemos.

Smiraglia, Richard P. (2011). Isko 11´diverse book shielf: an editorial. *Knowledge Organization*, 38(3): 179-189.

Thelwall, Mike (2008). Bibliometrics to webometrics. *Journal of Information Science*, 34(4): 605–621.

Sonia Troitiño Rodriguez, Mariângela S. Lopes Fujita and Dulce A. de Brito Neves

Indexing in Records Management

Abstract

This paper aimed to present a study on indexing in the record management of archives. From a theoretical/practical perspective, an indexing process was applied to a documentary itemin progress followed by an analysis proposed in a documentary reading modelin order to determine the subject and indexing of records. As a result, it proposes a new method for indexing records by considering the formal characteristics of the document so as to explore the textual structure combined with the identification of terms. This method allowed the representation of the meaningful content of the document by means of subject determination, which contributes to the management, use and access to records.It concludes that the application of the indexing process is more adequate at the moment of the record registration.

1 Introduction

Any change in the processes of production, organization and access to information must reflect the historical moment and so it requires adjustments to meet the social demands from which it originates. Knowledge Organization (KO) is much like societies in the sense that it should be dynamic because, according to Barité (2001, p.41), socialized knowledge is its object of study. Considering an institutional environment and the concept of recorded knowledge, Hjørland (2008) makes a distinction between a narrow sense of KO, meaning a cognitive or intellectual knowledge organization and, in a broader sense, a social KO.

In the narrow sense, KO comprises activities such as "[...]document description,indexing and classification performed in libraries,bibliographical databases, archives and otherkinds of 'memory institutions' by librarians, archivists,information specialists, subject specialists" (Hjørland, 2008, p.86). According to the author,KO as a field of study is concerned "with the nature and qualityof such knowledge organizing processes (KOP)as well as the knowledge organizing systems (KOS) used to organize documents, document representations,works and concepts".

Thus, indexing as a KO process has a methodology for information representation that ensures the access and retrieval of documents. In this context, indexing can contribute to Archival Science in relation to record management in order to organize and represent information in an organizational setting. However, theoretical and methodological proposals for dealing with subject determination, which is the first stage of the indexing process aimed to represent documents contents, were not found in the literature of Archival Science under analysis.

Considering the need to evaluate the contents of a documentin order to point out its representation aimed at subject information retrieval for archival purposes, this study

put forward a proposal for analysis and subject determination of records.Indexing for records management from a theoretical and normative perspective as well an adaptation and application of the methodbased on the documentary reading model of analysis for subject determinationwere presented in this study.

2 Theoretical and normative assumptions

Topics such as record management and access to contextualized information are widely discussed in Archival Science both from theoretical and practical perspectives. A great number of studies range from the establishment of policies for administrative rationalization to development and system operationalization focusing on information control and retrieval either for strategic or identitary purposes. Distinguished scholars of international contemporary Archival Science, such as José Ramón Cruz Mundet (2006, 2011), Joaquin Llansó Sanjuan (2006), Luciana Duranti (1996), Terry Cook (1991, 2005), Geoffrey Yeo (2003, 2015), Antonia Heredia Herrera (2011, 2013), Heloisa Liberalli Bellotto (2004), among many others, have been working on issues such as record production and management, archivesorganization and information representation in order to understand the creation, maintenance and access torecords in today's society.

However, a lack of studies on the use of indexing in the record managementprocess was observed. Recognized as a procedure traditionally linked to records description, the literature seldom deals with the issue directly. In fact, it is often addressed tangentially, as in the studies by Javier Barbadillo Alonso (2007), Alicia Barnand Azamorrutia (2002) or in normative parameters for the treatment of permanent records (ISOs, ISADg, Nobrade etc.) marked as *access point*.

In an information system, methods and representation tools are essential for the retrieval of the information contained in archives. In its "Glossary of Archival and Records Terminology" (2005), the "Society of American Archivists"seeks to reduce the impact of terminological issues on archives and archivology, as well as the transition from paper documents to digital ones. This publication points out and emphasizes the indexing types that can be used in access points to documents.

In this sense, it was observed that some institutions, such as the National Archives of Torre do Tombo, in Portugal, stresspresential research by providing fonds and collection catalogs whose inquiries are made by accessing these materials on their websites. The National Archives of the United States of America offer online access to a vast index from which users retrieve information not only by the description of the document, but also by the index of nominal authorities and topical subjects.

ISAD(G)-General International Standard Archival Description (1999)in its Brazilian edition (2000)does not explicitly mention indexing, but in the Introduction it acknowledges the need for the development of controlled vocabularies and indexes. However, ISAD(G) leaves the decision of building and using every archiveup to each country according to its particular needs.

ISAD(G) also indicates ISO specific standards for the development of vocabularies (ISO 2788 Documentation - Guidelines for the establishment and development of monolingual thesauri) and the standards required to perform indexing (ISO 5963 Documentation - Methods for examining documents, determining their subject and selecting indexing terms), as well as the elaboration of indexes (ISO 999 Information and Documentation - Guidelines for the content, organization and presentation of indexes,1999, f.9).

This research showed that the second part of ISO 15489 - Information and Documentation - Records Management (2001)aiming toestablishguidelines for the implementation of the management system of archival documents,seeks to demonstrate the relationship between records both of administrative activities and business processes, especially as todecision-making on documentary production, incorporation, control, archiving and access. It also points out some important products resulting from the necessary study about the origin, use and custody of the documents:

> The analysis provides the basis for developing records management tools, which may include:
>
> – thesaurus of terms to control the language for titling and indexing records in a specific business context, and
> – a disposition authority that defines the retention periods and consequent disposition actions for records. (ISO 15489-2, 2001(E), p. 4).

Moreover, when outlining the design and implementation of a management policy ofarchival documents, the standard indicates two essential tools for the management plan: vocabulary controland indexing.

> 9.5.3. Vocabulary controls
>
> Classification systems and indexes may be supported by vocabulary controls that are suited to the complexity of the records of an organization. Such vocabulary controls should explain organization-specific definitions or usage of terms.
>
> 9.5.4. Indexing
>
> Indexing can be done manually or be automatically generated. It may occur at various levels of aggregation within a records system. (ISO 15489-1:2001(E), p. 14)

This understanding widens the applicability of indexing beyond the content of individualized documents, and makes it possible to understand that the document groups in a multi-level classification perception also have content that can be indexed.

Prior knowledge of the possible different levels for the application of indexing will guide the construction of controlled vocabulary. Thus, indexing effectiveness also largely depends on specialized documentary reading, whose content representation is linked to a cognitive methodology of the indexer's knowledge organization.

From this perspective, the discussion about indexing incorporation during document management is extremely important and useful from the standpoint of the representation of the information recorded, especially at the time of classification and description, which certainly contributes to the assessment and retrieval of documents.

According to Javier Barbadillo Alonso (2007), indexing will also contribute to avoid overloading the classification plan by using the hierarchical level subseries in its elaboration at risk of losing its internal coherence when subject determinationis based

on thematic rather than typological criteria. In this way, classification detours and the emergence of mixed documentary units are avoided, as well as the control over documentary production (Barbadillo Alonso, 2007, p. 19-20).

However, the most recurrent aspect observed in management systems is the use of indexing at the *documentary unit*level, which is usually applied at the time of document recording (protocol). Barnand Azamorrutia (2002) states that the record and description of an archival document serve to ratify access and retrieval systems and to facilitate their control, physical location and development of tools for information control and access to information. So, in his *Guía for Organización y Control del Expedient of Archives* (2002), the author establishes thatthe field "related terms"must be filled and supplied with descriptors related to the recorded document.

3 Methodology

Indexing is a record representation operation of the thematic contents of documents and its ultimate goal is retrieval, whose process basically consists of two steps: recognition and extraction of information concepts and translation of these concepts into a documentary language (Chaumier, 1986, p.28). The first step, known as document analysis to determine the subject (Cleveland & Cleveland, 1990, p.104, the Brazilian Association of Technical Standards, 1992, May 2000, p.277), is performed during the documentary reading by the indexer, who needs a methodology that guarantees the identification and selection of concepts for the representation of the contents of documents with different types, structures and subject domains.

The “Documentary Reading Model” is a methodology of contents analysis that can contribute to the use of textual structure exploitation combined with questioning for concept identification.This model, originally proposed by Fujita (2003) for scientific texts (Table 1), basically combines the systematics of concept identification-conceptual analysis (first column) and the systematic approach of the Brazilian Association of Technical Standards (ABNT) NBR12676 (second column) - with the localization of concepts in the parts of the textual structure(third column).

Table 1:Documentary Reading Model for scientific texts: concept identification by questioning about the parts of the textual structure

CONCEPT (CONCEPTUAL ANALYSIS)	**QUESTIONING** (ABNT NBR 12676)	**PART OF THETEXTUALSTRUCTURE**
Object	Does the document have in its context an object under the effect of an activity?	Introduction (objectives)
Action	Does the subject contain an active concept (e.g., an action, operation, process, etc)?	Introduction (objectives)
Agent	Does the document have an agent that performed this action?	Introduction (objectives)
Methods of the agente	Does the agent refer to specific means to perform the action (e.g., special instruments, techniques or methods)?	Methodology
Place or environment	Are all these factors considered in the context of a specific place or environment?	Methodology
Cause and effect	Are any dependent or independent variables identified?	Results; discussion of the results
Author's point of view; perspective	Was the subject considered from a point of view normally not associated to the field of study (e.g., a sociological or religious study)?	Conclusions

Thus, the reading model methodology focuses on two aspects: the combination of the textual structure with the identification of concepts by means of a systematic concept identification. Table 1 provides the foundation for further proposals for adaptations aimed at its understanding and improvement, as suggested by Fujita and Rubi (2006). It also shows the different types of documentsas well as the tasks involved in subjectanalysis as presentedby Fujita's "Documentary representation in the indexing process using the documentary reading modelfor the subject cataloging of scientific papers and books: a cognitive approach with verbal protocol. *Access point*" (Fujita, 2013). What ensures uniformity in the subject analysis proposed in the documentary reading model of scientific papers, books or records is the representation by concepts (first column) because they are considered universal and do not change when a document contains different types of textual structures.

Table 2 is an exampleof an improvementput forward by Fujita and Rubi (2006). The authors considered that the same procedure used for scientific texts and books couldbe applied to a document with a textual structure composed of Initial Protocol, Body of the Text and Final Protocol or Eschatocol, according to Tomás Marín Martínez (2001) as follows:

Table 2: Documentary reading model for document indexing

CONCEPTS	QUESTIONINGFOR CONCEPT IDENTIFICATION	PARTS OF THE TEXTUAL STRUCTURE
OBJECT AND PART(S) OF THE OBJECT (something or someone studied by the author)	Does the document have an objectinits context under the effect of this action?	Body of the Text
ACTION (process performed by something or someone)	Does the subject contain an action (implying an operation, process, etc.)?	Body of the Text
AGENT (something or someone that performed an action)	Does the document have an agent that performed this action?	Body of the Text
METHODS (methods used in the research)	Does the document cite and/or describe specific means, such as: special instruments, techniques, methods, materials and equipment) used to study the object or action implementation?	Body of the Text
PLACE ORENVIRONMENT (physical place where the research was carried out)	Are all these factors considered in the context of a specific place or environment?	Initial Protocol
AUTHOR'S POINT OF VIEW; PERSPECTIVE	Was the subject considered from a point of view, normally not associated to the field of study (e.g., a sociological or religious study)?	Final Protocol
CAUSE AND EFFECT Cause (action+object) /Effect	Considering that action and object identify a cause, what is the effect of such cause?	Body of the Text

Considering the analysis of studies on records management indexingit is advisable to apply indexing at the classificatory level of the documentaryunit used to record the documentat the moment of its inclusion in the system. Therefore, a composite documentary unit was selected, in this case *an evaluation process and allocation* of university recordsfor the application ofthe methodology in "A documentary reading model for indexing"adapted from the proposal by Fujita and Rubi (2006).

4 Results

The application of the methodology proposed in "A documentary reading model for indexing" was carried out by an archivist with familiarity and professional knowledge of the typology and the textual structure of records. The archivist was previously taught how to use the methodology. After the application, the result showed the identification and allocation of the terms in the textual structure of the recordchosen comprising Initial Protocol, Body of the Text and Final Protocol, as follows:

Action (Text) + Object (Body of the Text): Discard + funding processes
Methods (Body of the Text): Analysis by the Court of Accounts of the State of Sao Paulo; List of Document elimination;
Location or Ambience (Initial Protocol): Experimental Campus of the State of Sao Paulo Coastline;
Author's point of view (Final Protocol): Analysis by the Executive Coordination of the Experimental Campus of the State of Sao Paulo Coastline;
Cause and effect (Body of theText): Analysis by the Central Commission of Documentary Assessment.

The application of the documentary reading model for indexing allowed the subject identification and selection of the following terms to represent the contents of the document: Discard, Funding processes, List of document elimination, Court of Accounts of the State of Sao Paulo, Experimental Campus of the State of Sao Paulo Coastline,Analysis by the Central Commission of Documentary Assessment.

5 Final considerations

Throughout this study, it was possible to observe that indexing plays a prominent role in document management process. On the one hand, while its role in information representation contributes to document retrieval, on the other hand it directly acts upon the rationalization of the documentary system organization since it collaborates in the application of the classification scheme without concealing the thematic contents.

Thus,considering the multilevel perspective adopted by the archival classification plan, regardless ofthe method used (functional or structural), the most general classification level is represented by the fonds/archive; the intermediate levels, by the functions or institutional structures and their subdivisions;and the most particular level, by the document itself. Therefore, the appropriateclassification level forindexing applicationis the document itself, because it presents anadequate textual structure for identifyingthe concepts that are useful to information management and retrieval for access purposes.

Moreover, it was found that, due to the typical dynamism of records, determining the moment for indexing application should be carefully considered.This means that the application of indexing atthe protocol stageis a priority, since this is thetime to include the document in the record management system.

This way, thenew method for records indexingproposed in this study consideredthe formal characteristics of the document by adopting the textual structure borrowed from Diplomatics for the analysis and determination of the subjects of the records registration in the process of theirmanagement.

In order to do this, adocumentary unit was considered appropriate for indexing at the time of its registration. Then, the application of the documentary reading modelenabled the determination of the subjectby identifying and selecting the terms combined with the exploration of the textual structure. However, it was alsoobserved that by

monitoring the documentaryprogression particularly in its first stage, a review of the descriptors defined is periodically necessarytaking into account the documentary types similar to the ones studied.

References

Associação Brasileira de Normas Técnicas (1992). *NBR 12676:* Métodos para análise de documentos: determinação de seus assuntos e seleção de termos de indexação. Rio de Janeiro.

Barbadillo Alonso, Javier (2007). Apuntes de clasificación archivística. *Legajos. Cuadernos de Investigación Archivística y Gestión Documental*, Publicación del Archivo Municipal de Priego de Córdoba, n.10, pp. 27 – 50.

Barité, M. G. Organización del conocimiento: un nuevo marco teórico-conceptual en bibliotecologia y documentación. In: *Educação, universidade e pesquisaIII Simpósio em Filosofia e Ciências Marília.* Held at São Paulo State University, 2001. São Paulo: Unesp-Marília-Publicações.

Barnand Azamorrutia, Alícia (2002). *Guía para organización y control del expediente de archivo.* México, DF: Archivo General de la Nación.

Bellotto, Heloísa Liberalli (2004). *Arquivo permanente: tratamento documental*. 2.ed. Rio de Janeiro: Fundação Getúlio Vargas.

Chaumier, Jacques (1986). *Analisis y lenguajes documentales*: el tratamiento lingüístico de la información documental. Barcelona: Editorial Mitre. 172p.

Cleveland, Donald B. & Cleveland, Anna D. (1990). *Introduction to indexing and abstracts*. 2. ed. Englewood: Libraries Unlimited.

Cook, Terry (2004). Macro-appraisal and functional analysis: documenting governance rather than government 1. *Journal of the Society of Archivists*, 25 (1): 5-18.

Cook, Terry (2005). Macroappraisal in theory and practice: origins, characteristics, and implementation in Canada, 1950–2000. *Archival Science*, 5 (2-4): 101-161.

Cook, Terry (1991). *The archival appraisal of records containing personal information*: a RAMP study with guidelines. Paris: Unesco, General Information Programme and UNISIST.

Cruz Mundet, José Ramón (2011). *Diccionario de archivística*. Madrid: Alianza Editorial.

Cruz Mundet, José Ramón & Mikelarena Peña, Fernando (2006). *Información y documentación administrativa*. 2. ed. Madrid: Editorial Tecnos.

Duranti, Luciana (1991). Diplomatics: New Uses for an Old Science, Part V. Archivaria, 32. [http://journals.sfu.ca/archivar/index.php/archivaria/article/view/11758/12708].

Fujita, Mariângela Spotti Lopez (2003) *A leitura documentária do indexador*: aspectos cognitivos e linguísticos influentes na formação do leitor profissional. Postdoctoral dissertation. Marília: São Paulo State University.

Fujita, Mariângela Spotti Lopez (2013) A representação documentária no processo de indexação com o modelo deleitura documentária para textos científicos e livros: uma abordagem cognitiva com protocolo verbal. *Ponto de Acesso* (UFBA), 17: 42 - 66.

Fujita, Mariângela Spotti Lopez &Rubi, Milena Polsinelli (2006). Modelo de lectura profesional para la indización de textos científicos. *Scire* (Zaragoza). 12: p.47 – 69.

Heredia Herrera, Antonia (2011). *Lenguaje y vocabulario archivísticos*: algo más que un diccionario. Andalucía: Junta de Andalucía/Consejeria de Cultura.

Heredia Herrera, Antonia (2013). *Manual de archivística básica*: gestión y sistemas. Puebla, México: Benemérita Universidad Autónoma de Puebla/Archivo Histórico Universitário.

Hjørland, Birger (2008). What is knowledge organization? *Knowledge Organization*, 35 (2/3): 86-101.

International Council on Archives (1999). ISAD(G): *General International Standard Archival Description*.2. ed. Ottawa: Stockholm, Sweden.

Llansó Sanjuan, Joaquin (2006). *Buenas prácticas en gestión de documentos y archivos*: manual de normas y procedimientos archivísticos de la Universidad Pública de Navarra. Pamplona: Universidad Pública de Navarra.

Mai, Jens-Erik (2000). Deconstructing the Indexing Process. *Advances in Librarianship,* 23: 269-298.

Marín Martinez, Tomás (2011). *Paleografía e diplomática*. 5. ed. Madrid: Ed. UNED.

Norma Internacional ISO 15489: *Information and Documentation* – Record Management - Part 1 (2001). General. Ginebra, Suiza: ISO.

Norma Internacional ISO 15489: *Information and Documentation* – Record Management - Part 1: Guidelines [Technical Report] (2001). Ginebra, Suiza: ISO

Pearce-Moses, Richard (2005).*Archival Fundamentals Series II*. Chicago: Society of American Archivists.

Yeo, Geoffrey (2015). Proporcionar o Acesso à Informação no Domínio do "Records Management". In *Gestão de documentos e acesso à informação: desafios e diretrizes para as instituições de ensino e pesquisa*. 2015. Rio de Janeiro: Fundação Casa de Rui Barbosa. Pp. 39-57.

Yeo, Geoffrey & Shepherd, Elizabeth (2003). *Managing records*: a handbook of principles and practice. London: Facet Publishing.

Renato Rocha Souza and Isidoro Gil-Leiva

Automatic Indexing of Scientific Texts: A Methodological Comparison

Abstract

We are aiming at establishing a comparison between two information retrieval systems: SISA and PyPLN, regarding their performance when indexing the same set of documents. To this end, we took a corpus of a hundred scientific articles on the field of Agriculture and have them processed by both tools. The index produced by each tool was stored in two different databases. Subsequently, seven queries with information needs were prepared, based on the document contents, in order to establish which set of documents would be relevant for each tool. With the result set, the index and precision indexes were calculated and it was possible to highlight each tool's strengths and weaknesses.

1 Introduction

Research on automating indexing began in the late fifties. Since then, there have been numerous and varied proposals to undertake the intellectual process that involves indexing. The terminology used in the literature to refer to the process of making indexing automatic is varied: we can find names as "Automated assisted indexing", "Automated indexing", "Automated supported indexing", "Automatic support to indexing ", "Computer aided indexing ", "Computer assisted indexing ", among others, whereas the most used is "Automatic indexing ". The definition of automatic indexing must be derived from three perspectives: a) Computer programs that assist in the process of storing indexing terms, once obtained intellectually. (Computer Aided Indexing during storage); b) Systems that analyze documents automatically, but the indexing terms proposed are validated and published - if necessary – by a professional (Semiautomatic Indexing); and c) programs without any further validation programs, i. e., the proposed terms are stored directly as descriptors of that document. (Automatic Indexing).

The methodologies used in automating indexing through the decades have changed until nowadays. In the early days, indexing documents was made almost exclusively from statistics based on terms frequency; but from the eighties on, they incorporated techniques as natural language processing to get the roots of words (stems), morphological taggers and parsers (POS taggers), among others. It is, though, usual that the proposals or prototypes submitted by researchers include a combination of both approaches, i. e, calculating the frequency and tools, more or less complex, for automatic processing of texts (Gil Leiva, 2008).

In spite of all this years of work and research, the use of automatic indexing software is still rare in libraries and documentations centers. Nevertheless, since manual indexing was found impossible for some activities in most of the digital information environments, given the massive amounts of documents to be processed,

researchers seek alternatives to represent documents' subjects automatically; using statistical and/or rule based computational linguistic techniques. The oldest and most common process seek to determine documents' subjects solely through the analysis of words' frequencies, but that can lead to poor indexing and erroneous assumptions, as the context can be lost when the collocations are broken into single words. In the last decades, many other techniques were developed, either trying to capture corpus structure with statistical methods, as the TfIDf methodology (Spärck Jones, 1972); Multiword expressions (Silva & Souza, 2014); Latent Semantic Indexing (Deerwester et al., 1988); Latent Dirichlet Allocation (Blei et al., 2003); Word2vec (Mikolov et al., 2010); or aiming at the extraction of the deep semantic structure of the texts (i.e. Extraction of Noun Phrases, Souza & Raghavan, 2006). Also, the use of each technique presents some advantages and drawbacks over the others, as language dependencies (as the case of Noun Phrases), the need of huge computational structures to process the documents timely and the quality of the results. So far, there is no rules of thumb on the techniques and strategies, and it is very common to observe ensembles of these in automatic indexing systems.

In this paper, we are aiming at comparing two indexing systems, each of them using different sets of techniques for indexing documents: the first, named SISA was developed by Gil-Leiva (1999 and 2008); the other, named PyPLN, was developed by the Applied Mathematics School from Fundação Getulio Vargas.

2 The information retrieval systems

In this section, we will present the main characteristics of both SISA and PyPLN.

2.1 SISA

SISA is designed to be used as a semiautomatic system (users can edit the result of the process by adding terms not proposed by the system or browsing the embedded controlled vocabulary of the system to assign additional terms or as a fully automatic system without user intervention once the configuration set.

The system has been developed in JAVA, and different libraries have been used to:

- extract information from documents in PDF, TXT or XML;
- read the controlled vocabulary (SKOS);
- remove the roots of words.

On the other hand, it has been used as a MySQL database to store fonts, documents, results and a retrieval module can be used for system evaluation.

The SISA main features are as follows: It is a system designed for indexing journal articles on web platform implemented with ease of use through a web browser. It works with various file formats such as HTML, PDF, XML and plain text. It also processes documents in Spanish, Portuguese and English, using stopwords and controlled vocabularies in these languages. It makes use of stemming and is based on

heuristic and statistical methods with a set of rules that mark the extraction parameters or weighting of terms.

SISA has used a stopword list in Portuguese composed of 586 words and a controlled vocabulary with 9,588 1,122 descriptors and non-descriptors. This vocabulary has only the relationship of synonymy (USE). The vocabulary used by SISA comes from Thesagro, a thesaurus prepared by the National Agricultural Library (BINAGRI) of the Ministry of Agriculture of Brazil. In SISA the following parts of the article are labeled: title, abstract, keywords, authors, headings, first paragraph, conclusions and references with tags such as # ITI # and # FTI #, # CRE # and # FRE #, to delimit the title, starts and ends for many articles parties. If the source texts in txt or PDF formats are not labeled they can be labeled when items are loaded into SISA. Finally, SISA has handled a set of 41 rules that can be grouped into a) positional heuristic rules: If a word is not an empty word, is in a particular combination of tags and appears in the controlled vocabulary, it is presented as descriptor; b) statistical rules: if a word is not a stopword and exceeds a certain frequency or, if a word exceeds a certain TFIDF, it is presented as a descriptor; and c) mixed rules: if a word is not empty word, is in one or more tags or appears above a certain threshold frequency, it is proposed as a descriptor.

Successive tasks for indexing an article with SISA are: label items, process (apply stemming apply, calculate and record TFIDF the place in which they appear words and phrases) and index them according to the configured rules.

SISA is installed on a Proliant server with 32GB RAM ML310E and a CentOS 7.0 operating system. It has been developed in JAVA and different libraries have been used to extract information from documents, read in SKOS format controlled and remove the roots of vocabulary words. On the other hand, it has been used Cascading Style Sheets for application design and MySQL as a database for storing fonts, documents, results and a retrieval module that can be used for system evaluation.

2.2 PyPLN

The PyPLN platform is a research project in active development. Its main goal is to make available a scalable computational platform for a variety of language-based analyses. Its main target audience is the academic community, where it can have a powerful impact by making sophisticated computational analyses doable without the requirement of programming skills on the part of the user. Among the many features already available, we can cite: Simplified access to corpora with interactive visualization tools, text extraction from TXT, RTF, HTML and PDF documents, encoding detection and conversion to utf-8, language detection, tokenization, full-text search across corpora, part-of-speech tagging, word and sentence level statistics, n-gram extraction and word concordance. Many more features are in development and should become available soon, such as: semantic annotation, sentiment and text polarity analysis, automatic social network information monitoring, stylistic analysis

and the generation of Knowledge Organization Systems such as ontologies and thesauri. PyPLN aims for unrivaled ease of use, and wide availability, through its web interface and full support to Portuguese language. Besides being a free, uncomplicated research platform for language scholars capable of handling large corpora, PyPLN is also a free software platform for distributed text processing, which can be downloaded and installed by users on their own infrastructure. It was developed using the Python programming language and can be deployed in a single server or in a cluster of servers, for fast parallel processing of documents. It exposes a REST and a Python APIs (Application Program Interface) for ease of embedding its functionalities within other applications.

2 Materials and Methods

To carry out this experiment we have used the two indexing systems (SISA and PyPLN) described in the preceding paragraphs and a corpus of one hundred items in the field of agriculture, published in the Brazilian Journal of Fruticultura, between 2006 (vol. 28, No. 1) and 2007 (vol. 29 , No. 1).

SISA have used a Portuguese stopwords list composed of 586 words and a controlled vocabulary composed by 9,588 1,122 descriptors and non-descriptors. This vocabulary has only the relationship of synonymy (USE). The vocabulary used by SISA comes from Thesagro, thesaurus prepared by the National Agricultural Library (BINAGRI) of the Ministry of Agriculture of Brazil.

The main tasks for the performance of this test were as follows:

1. Build two databases of documents, in each of the tools;
2. Index a hundred documents using both SISA and PyPLN;
3. Choose seven examples of user information needs;
4. For each information need, establish the relevant documents;
5. Convert the information needs into seven search terms and query the database;
6. Apply tests to measure recall and precision of each information need and for each platform;
7. Use these measures the recall and precision to compare SISA and PyPLN.

To measure the rates of recall and precision, we have been used traditional formulas:

- Recall = Number of relevant items retrieved / Number of relevant items in the collection
- Precision = Number of relevant items retrieved / Number of items retrieved.

SISA is composed of three integrated modules which allow the following tasks: processing and indexing of documents; storing metadata items as title, data source magazine, abstract, keywords, descriptors A (descriptors assigned by SISA) and descriptors B (descriptors assigned from another indexation system); and a third information retrieval module. This retrieval module allows searches on the metadata of the stored items.

Once the documents were collected and stored, indexing was automatically triggered by SISA (without human action) for a hundred articles on Agriculture, and we proceeded to manually enter the descriptors also obtained automatically by PyPLN in the field descriptors_B. Thus, we stored SISA and PyPLN indexing results in the database. Indexing in PyPLN was made using a Part of Speech Tagger and an automatic Noun Phrase extractor at first. After extracting the Noun Phrases, the most frequent are considered for assigning descriptors. No sophisticated stopword removal was done in this experiment, because the system does not provide this functionality yet – though it can be easily done in an after processing fashion.

The retrieval module was used to perform information searches in both databases and to apply Recall and Precision formulas with the results obtained.

Fig. 1: SISA Interface

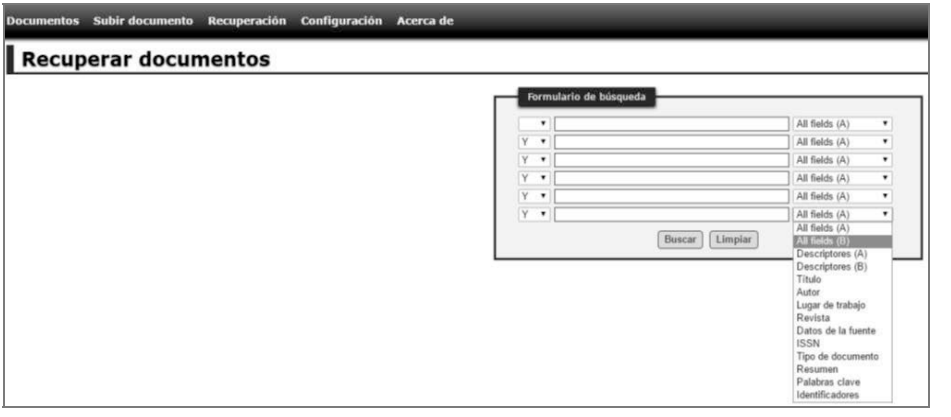


Fig. 2: PyPLN Interface

The queries ran against SISA have used all fields available, such as title, abstract, keywords proposed by the authors of papers and indexing terms obtained by the tool. Queries against PyPLN have used only the terms in the noun phrases automatically attributed by the platform. We also present in appendix 2 the index terms attributed to the documents.

3 Results and discussion

The following tables present the results from the indexing and retrieval process:

TAB. 1: Recall for SISA and PyPLN

SISA Recall			**PyPLN Recall**		
	Searched in all fields	Searched in Descriptors Field		Searched in all fields	Searched in Descriptors Field
Searched 1	0,85	0,71	Searched 1	0,71	0,14
Searched 2	0,75	0	Searched 2	0,75	0
Searched 3	1	1	Searched 3	0	0
Searched 4	1	1	Searched 4	1	0
Searched 5	1	1	Searched 5	1	0
Searched 6	1	0,75	Searched 6	1	0
Searched 7	0,83	0	Searched 7	0,83	0.16
Average	**0,91**	**0,59**		**0,75**	**0,04**

As we can see, recall is lower in PyPLN because it does not make any distinction between descriptors' position in the text, whilst SISA uses this information when indexing. The same occurs when we are comparing the precision measures. The fact that only the most frequent noun phrases were used in the PyPLN indexing process takes a toll in its results, making the results not as good as it would be expected:

TAB. 2: Precision for SISA and PyPLN

SISA Precision			**PyPLN Precision**		
	Searched in all fields	Searched in Descriptors Field		Searched in all fields	Searched in Descriptors Field
Searched 1	1	1	Searched 1	1	1
Searched 2	1	0	Searched 2	1	0
Searched 3	1	1	Searched 3	0	0
Searched 4	1	1	Searched 4	1	0
Searched 5	1	1	Searched 5	1	0
Searched 6	1	1	Searched 6	1	0
Searched 7	1	0	Searched 7	1	1
Average	**1**	**0,75**		**0,85**	**0,28**

Regarding the limitations identified in the operation of SISA and possible improvements, it can be noted that most of the effort and time spent on SISA has been to insert labels to documents. In future experiments, XML format should be prioritized for the scientific papers, since SISA is already implemented to automatically tag documents with certain structures. On the other hand, the controlled vocabulary is an important tool in the operation of SISA, therefore it's necessary to use a large vocabulary of preferred terms and non-preferred terms for enhancing the results. Although the controlled vocabulary used in this experiment has nearly eleven thousand terms it has been observed that there is room to incorporate new terms and to introduce a greater number of synonyms. Finally, it is necessary to continue working on other ways to combine rules SISA.

In the PyPLN side, speed (the whole processing took only three minutes) and the absence of human interaction is key for numbering its advantages. In addition, the use of high frequency Noun Phrases can add a bit of semantics. The lack of stopwords and of any TfIDf weighting procedure, though, has set a penalty in the results. By design, it does not discriminate of the parts of the document in which the extracted words reside. Incorporating these features can truly enhance the performance of the platform.

4 Conclusions

This paper aimed at comparing two automatic indexing platforms; SISA and PYPLN. The results has shown advantages from both of them, with clearly better results presented by SISA, although PyPLN took less time to process the documents. The researchers are planning to incorporate the best features of both tools in new versions of their software, to achieve even better results.

References

Blei, David M., Andrew Y. Ng & Michael I. Jordan (2003). Latent dirichlet allocation. *The Journal of machine Learning research*, 3: 993-1022.

Deerwester, Scott et al (1988). Improving Information Retrieval with Latent Semantic Indexing. In Proceedings of the 51st Annual Meeting of the American Society for Information Science 25. Pp. 36–40.

Gil Leiva, Isidoro (1999). *La automatización de la indización*. Gijón: Trea.

Gil Leiva, Isidoro (2008). *Manual de indización. Teoría y práctica*. Gijón: Trea.

Mesquita, L.A., Souza , R.R., & Porto, R.M.A.B. (2014). Noun phrases in automatic indexing: A structural analysis of the distribution of relevant terms in doctoral theses. *Advances in Knowledge Organization*, 14: 327-34.

Mikolov, Tomas, et al. (2013). Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301-3781.

Silva, Edson Marchetti & Souza, Renato Rocha (2014). Fundamentos em processamento de linguagem natural: uma proposta para extração de bigramas. *Encontros Bibli*, 19 : 1.

Souza, Renato Rocha & Raghavan, Koti S. (2006). A methodology for noun phrase-based automatic indexing. *Knowledge Organization*, 33 (1): 45-56.

Spärck Jones, Karen(1972).A Statistical Interpretation of Term Specificity and Its Application in Retrieval. *Journal of Documentation*, 28: 11–21.

Lorena Tavares de Paula and Maria Aparecida Moura

Nanopublication and Indexing: Semantic and Pragmatic Interchanges in Methodological Applications

Abstract

This article presents results from the research: "Nanopublication and indexation: semantic, pragmatic, and discursive dialogues in methodological applications." This research is inserted in the context of knowledge organization in digital environments. It establishes an interface between Linguistics, Information Science, and Websemantic technologies. Its theoretical and methodological bases come from interdisciplinary relations, with a view to offer new prospects for the area of Organization of Information and Knowledge in digital environments. The result of that search on the concept of nanodocument can be considered a product for the representation of information and knowledge, composed of enunciation elements referenced by semantic and pragmatic components, structured from nanopublication modelling. It should be noted that, bearing in mind the sight of the advances in knowledge and the resulting information explosion in network, organizational methodologies of information in dynamic contexts that support and boost the use of the increasing document production should be sought. In this aspect, the research has relevant aspects that corroborates this claim.

1 Introduction

In these last few years, the Organization of knowledge has become very complex, especially due to the agility in the processes of information circulation and to the active presence of users in the instances of information production and dissemination.

For Hjørland (2003), the organization of knowledge is materialized in activities such as the indexation, the bibliographical classification and its taxonomic structures for representing knowledge. That is a field of study dedicated to comprehending the nature of knowledge organization processes, along with the systems generated in that environment.

This paper presents the results of the research titled “Nanopublication and indexation: semantic, pragmatic and discursive relation in methodological applications”, part of the context of information organization and knowledge in digital environments. The theoretical and methodological foundations of the research are located in the interdisciplinary relations, aiming at offering new perspectives for the area of Information Organization and Knowledge in digital environments.

Nanopublication is a concept still little disseminated in the field of Knowledge Organization and it refers to “the smallest unit of information published in a formal document”: it is the textual extraction attributed to the “author’s voice”. According to Concept Web Alliance (CWA), nanopublications may be cited, because they are an exact reference of the ideas contained in a formal publication. They enable a communication based in contextual information of high quality, used for the dissemination, appropriation and organization of contextualized information according to an author and a contextual unit.

The nanopublication provides a content representation through the structuring of conceptual statements from a formal publication.

The process of indexation responds to two basic principles: conceptual analysis and translation to an indexation language. Lancaster (2004, p. 28) explains that indexation and abstract are "intimately related activities, because both imply the preparation of a representation for a thematic content of the documents". The indexer employs indexation terms that have relations with the intellectual content of the document and the probable questions that generate information searches by the user.

The indexations aim at recovering documents through the representation of its themes. According to Amar (2011), that indexation may be considered a discursive action on the documents and their content. Because of that, the document analysis on a given topic may be recognized within a discursive field. According to the author, indexation is an interpretational discourse. Amar (2011) questions the process of traditional indexation established from outside the text, departing from a topic analysis and posterior translation to a documenting language, and suggests that it is possible to establish topic describers from within the texts themselves.

Yet, the Discursive Indexation suggested by Amar (2000) takes into account, in addition to the classic extraction of describers that represent document themes, representative elements of the discourse established in the text. Such an act is materialized through identification and marking of phrases and words in their syntactic relations, which are themselves materialized in a relevant statement for charactering the themes discussed in the publication.

We argue that the characterization and application of nanopublications have a methodological relationship with the processes of topic indexation. Both nanopublications and indexation are processes of information representation based on language. In that sense, the linguistic aspects inherent to information representation in nanopublications are considered essential to its model.

2 Nanopublication: research approach

This study aimed at proposing a methodology based in experimentation for the development of information organization environments in a digital medium that enables the indexation of documents in context. The study started from the following question: how can a methodology of information organization, based in indexation and nanopublications, guide the construction of units of information and content that can be recovered and that present semantic, pragmatic and discursive coherence?

The proposal for elaboration of nanopublications, established by Concept Web Alliance (CWA) aims at the development of systems that delimit patterns for the representation of scientific information. The elaboration of nanopublications is done, simply speaking, through the mapping of representative "entries" for concepts in a given area of knowledge represented by digital publications.

Gorth et. al (2010) clarifies that the nanopublications is a group of annotations related to declarations and content turned to communities that propose the conceptual definitions. Nanopublications may serve some basic requisites, such as: conceptual and community identification capacity, compilation of concepts and declarations in a computational perspective, and interoperability permission with different computational formats.

Besides, it is worth noting that the nanopublications will always be attributed to authors and their respective publications – the aim is to render accessible the adequate conceptual knowledge from various areas of knowledge according to their own conceptual references and definitions.

The linguistic aspects guiding the experiments for nanopublications modelling are essential. In that sense, semantics and pragmatics have provided the structure and the experiment analysis established in this study.

3 Semantic and pragmatic aspects of nanopublications in interface with indexation

The devices for information organization have a strong influence in Linguistics. Such an influence is expressive in indexation processes.

Semantics has a fundamental role in the organization of information because it aids the comprehension a term's meaning, considering the adjustment of signification through instruments of indexation language.

In the context of indexation, the concepts may be considered knowledge units, identified through true statements about a reference item, represented by a term or word. In that sense, Dahlberg (1978, p. 102) emphasizes that,

> [...] the formation of concepts is the gathering and compiling of true statements about a given object.In order to fixate the result of that compilation, we need an instrument, which is constituted by the word or by any other sign that may translate and fixate that compilation. It is, then, possible to define the concept as the compilation of true statements about a given object, fixated by a linguistic symbol.

The concepts consist of a mental representation that allows us to categorize components. According to this perspective, we can verify that the formation of concepts implies processes that involve discrimination and grouping.

According to Ilari (200, p. 48), semantics and pragmatics are the two linguistic areas that research signification. The author highlights that interlocutors and the interaction between them are the essence of the pragmatic perspective.

Semantics has among its functions the task of organizing expressions in order to formulate "guidelines" to systematize correspondences among words, sentences and statements. With the aid of pragmatics, these analyses join the context, an essential element for interpretation.

Novellino (1998) defends the need for validating pragmatic questions in order to support content organization and proposes that languages of "information transfer" are

built, instead of the simple thematic representation of content. In that sense, the contextualization of concepts and meanings used in the definition is essential.

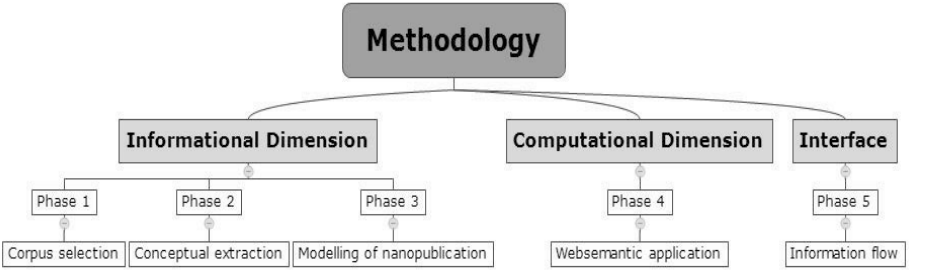
The interchange between indexation and nanopublications approaches the concepts in its discursive function in a given text that is presented to a specific scientific community.

Hjørland (2007) states that the meaning of concepts, from the perspective of scientific theories, is explicit from different theoretical positions. Variations may occur according to the degree of theoretical consensus in an area of knowledge. That gradation is related to the degree of agreement among the individuals in the definition of a document's topic.

4 Considerations on the developed method

The research methodology was established from the structure and data analysis in three dimensions, considered fundamental for the organization of information and knowledge in digital environments: the informational dimension, the computational dimension and the interface.

Figure 1 – Stages of methodology



The method's informational dimension was established through the mapping of corpus elements, conceptual extraction, and the application of indexation principles and nanopublications modelling.

A priori, it is necessary to clarify that the concepts, elements to be identified in this process, may be precise or vague and multifaceted. Hjørland (2001, p. 774), quoting Wilson (1968), highlights the most important elements in determining a document's subject for the extraction of relevant concepts: to identify the author's purpose when writing the document; to identify the domain and the subordination of different conceptual elements; to observe the concept group and the established references; to establish a set of rules for selecting the "essential" elements (in contrast with the non-essential) of the entire document.

The computational dimension was formulated from semantic classifiers (trope software) and websemantic languages. In order to model the RDF graph we opted to

mark the conceptual elements using the Tropes software. It can isolate and identify the main concepts through the semantic application levels. So, it is possible to mark, from the text, who says what to whom, who does what, when and where, and to what end.

The interface representative layer presents the adequate information flows for the distinct user profiles (end users and information managers). From that action, it is possible to access the perspective of establishing nanopublications and their authorship, concepts and citation networks in an interface that allows user direct interaction.

For validation of methodology, articles related to the concept of "netnography" and "innovation" were selected and analyzed from the selection taken from Google Scholar and Scielo in the three dimensions described in the methodology. The results from the developed experiment were analyzed under a linguistic perspective. They demonstrated a potential and efficiency for nanopublications in the construction of recoverable informational content units with semantic, pragmatic and discursive coherence from the methodological interchange with indexation principles. The experiments proposed in the research surfaced the concept of nanodocument, that can be considered an element of discursive mediation between the information and the user.

The nanodocument can be developed from indexation elements guided by subject analysis and materialized in statements, mediated by semantic realizations that manifest one or more quotable statements of nanopublications. It can also be considered a representation model of information and knowledge in a digital environment, guided by websemantic technologies for the recovery of information in context. Its main characteristics are reuse, quotable informational unit, and accuracy basis of representation for a document.

The nanodocument is composed by stated elements referenced by semantic and pragmatic components, structured from nanopublication modelling and it can be considered a product for the representation of information and knowledge.

5 Final Remarks

The elaboration of a nanodocument, from principles of indexation and nanopublications modelling, is presented as an alternative model for the representation of knowledge in context. This representation becomes possible from the three methodological dimensions for information and knowledge organization in a digital environment: informational and computational dimensions, along with the interface.

The knowledge represented by the nanodocument reflects the messages that the informational item may transmit. In addition, it enables the user to analyze its documental choice for a discursive order, not just by describers.

From the development of the proposed study, it is possible to state that the use of nanopublications, with the guidelines of indexation processes, has proved to be coherent and promising and revealed that the possibilities for exploration of dimensions suggested in the methodologies can be guided by variables that go beyond indexation and nanopublication. They may guide the experimentation of websemantic

computational elements for the organization of information; they can also explore linguistic aspects related to the construction of technologies.

Considering the visible advances of knowledge and the consequent informational explosion in network, the Organization of Knowledge requires consist methods of information organization that enable the construction of semantic tools for the recovery of information in dynamic contexts that support and propel the use of the growing document production. In that sense, the developed research brought relevant results that can contribute for the advance of informational mediation in context.

Acknowledgements: We thank CNPq for the support to the development of this study.

References

Amar, Muriel (2000). Indexation Discursive versus indexation lexicale: elements de definition. *Terminologies nouvelles*, 21: 71-9.

Amar, Muriel (1997). Lesfondements théoriques de l'indexation**:** Une approche linguistique. *PhD Thesis*. Lyon: ADBS.

Amar, Muriel (2011).Les langages documentaires. *In*:*Repere:* Ressources électroniques pour les étudiants, la recherche et l'enseignement. Villeurbanne: Enssib. Pp. 61-64.

CWA. Concept Web Alliance. (2012). Nanopublication Guidelines, Working Draft, [S.l]. [http://nanopub.org/guidelines/working_draft].

Dahlberg, Ingetraut. (1978). Referent-oriented analytical concept theory of interconcept. *International Classification*. v. 5, n. 3, Pp. 142-50

Groth, Paul & Gibson & Andrew; Velterop, Jan. (2010). The Anatomy of a Nanopublication. *Information Services & Use*, 30: 51-6.

Hjørland, Birger (2001). Towards a theory of aboutness, subject, topicality, theme, domain, field, content ... and relevance**.** *JASIST,* 52(9): 774-778

Hjørland, Birger (2007). Semantics and knowledge organization**.***ARIST* 41(1): 367-405.

Hjørland, Birger. (2003). Fundamentals of knowledge organization. *Knowledge Organization*, v.30, n.2.

Ilari, Rodolfo.(2000). Semântica e pragmática: duas formas de descrever e explicar os fenômenos da significação. *Revista de Estudos da Linguagem***,** 9(1): 109-162.

Lancaster, Frederik Wilfrid (2004). *Indexação e resumos*: teoria e prática. 2. Ed. Brasília, DF: Briquet de Lemos Livros.

Mons, Barent & Velterop, Jan.(2009). Nano-publication in the e-science era. *In*: *Workshop On Semantic Web Applications In Scientific Discourse* (SWASD). Washington, DC. *Proceedings...* Washington, DC: CEUR-WS.org, 2009.

Nanopub.Org. [http://nanopub.org/wordpress]

Novellino, Maria Salete F. (1996). A teoria da ação comunicativa e a representação da informação. *Informare*: caderno do programa de pós-graduação em ciência da informação, 2(2): 73-9.

Tamba, Irène (2009). *A semântica*. 2. ed. rev. São Paulo: Parábola.

Roberta C. Dal'Evedove Tartarotti and Mariângela S. Lopes Fujita

The Perspective of Social Indexing in Online Bibliographic Catalogs: Between the Individual and the Collaborative

Abstract

One of the challenges of the field of Knowledge Organization is to monitor the dynamics of the documentary treatment, given the technological and cultural changes in contemporary society. In the context of Web 2.0, emerges the social indexing or folksonomy as a new way to organize and share content on the Internet. Considering that the purpose of the online bibliographic catalogs is represented by subject and give access to the intellectual content of the documents, the work aims to make reflections on the social indexing as collaborative construction proposed in the indexing process in online bibliographic catalogs under university libraries. It appeared that the social indexing, together with the theoretical and methodological indexing foundations, set up as a new theoretical-practical paradigm in the representation and thematic retrieval, presenting potentially viable for deployment in online bibliographic catalogs, adding value to products and services offered by university libraries..

1 Introduction

The Internet evolution has enabled the creation of new interactive spaces in the collaborative context of Web 2.0 as a result of free indexing and staff of so-called tags or labels to digital objects, called folksonomy approach or "knowledge organization done by users" (Hjørland, 2008, p. 93).In this scenario, despite of an innovative informational setting, there are new social practices due to the dynamic and collaborative construction of digital objects, which favor contemporary research in the field of Knowledge Organization.

The connections with computational approaches are driven because of the social aspects of knowledge organization processes, as the way knowledge is acquired, represented, managed and exploited has changed with the connected world and the new features associated. In this sense, the issues around knowledge in the digital world, such as the advent of folksonomies, need further investigation in the field (Ohly, 2014, p. 328; David, 2014, p. 329).

In contemporary times, one of the main challenges of university libraries is the creation of integrated and continuous access to its online bibliographic catalogs, considering the variety of sources and formats of their informational items. In these information retrieval systems, searches are conducted primarily in three ways: by author or title (if the item is known); by keywords (if certain words of the title or particular author are known) and subject headings (item on a particular subject). Buckland (1992) calls bibliographic access the process to connect to the records of various types (textual, numerical, visual, musical, etc.) contained in different media (books, journals, microforms, computer files, etc.) and includes three main points: the identification of the documents, their location and physical access to the material.

In the field of Knowledge Organization, the thematic representation plays a key role in information retrieval systems. In online bibliographic catalogs of university libraries,

the proper use of the documentary language is critical because it enables the representation of documentary content that is compatible with users' search requests. The research aims to conduct a theoretical study around the ontologies as knowledge representation tools, aiming their applicability in the representation and thematic retrieval in online bibliographic catalogs. It is believed that the theoretical and methodological ontologies foundations are presented potentially viable for implantation in online bibliographic catalogs, adding greater value to products and services offered by university libraries in this new informational configuration of the semantic web.

Although the indexing quality in online bibliographic catalogs is related to the ability to reconstruct the subject matter in a document on concepts for later retrieval by the user, it is not still realized major advances in representation and thematic retrieval. For example, when performing the query to some of the main library catalogs available on the web, Dias (2006, p. 63) found that few changes were actually implemented and that “catalogs are still limited to much of the information that existed in the corresponding catalogs sheets”, displaying the same pattern, i.e., “preferring to delegate the search for people who have the skills and patience to make queries in information retrieval systems increasingly complex”. Thus, the subjacent problem is that there is a growing consensus that the online bibliographic catalogs of university libraries are no longer suitable for the role they should play as “a catalog, in definition, should allow communication, in other words, should avoid the risk of feeling autonomous, if not insensitive with respect to the needs, potential or actual users” (Rasmussen, 2011, p. 687; Friás, 2004, p. 234).

Starting from the premise that the purpose of the online bibliographic catalogs is represented by subject and give access to the intellectual content of the documents, this paper aims to make theoretical reflections on the social approach of indexing or folksonomy as collaborative construction proposal in online bibliographic catalogs in the context of university libraries.

2 The convergence of traditional indexing approach and social indexing approach in online bibliographic catalogs: some challenges

In thematic treatment of information approach, the catalog is a product of theoretical current of american influence called subject cataloguing. Both cataloguing (process) as the catalog (product) conduct mediation between informational items in a collection and information needs of users.In online bibliographic catalogs, individual cataloguing of informational items involves two main aspects: the physical description that reflects the extrinsic content (descriptive metadata) of a document; and the thematic description that reflects the intrinsic content, characterizing it through its issues (semantic metadata). For the experts on the field of Knowledge Organization, semantic metadata can be understood as concepts; for information scientists as descriptors; for taxonomists as taxonomies; and librarians as subject headings (Dahlberg, 2014, p 332).

The activity that most adds value to the online bibliographic catalog are the subject of accessing points of informational items, defined by subject analysis traditionally performed by professional indexers, called indexing. Conceptually, indexing is a process consisting of sub-processes or steps that aims to identify the contents of a document and express it in terms of indexing through a built metalanguage – the documentary language - in order to promote effective recovery information held by librarians or experts in a particular field of knowledge (Tartarotti, 2014, p. 21).

Although there is no consensus in the literature, we have as main indexing steps: document reading; subject analysis or concept identification; selection of concepts and translating concepts. According to Mai (1997a, p. 61; 1997b, p. 55), the indexing process can be deconstructed revealing three steps: document review process, the subject description and subject analysis process; and four elements: document, subject, the subject description and input the subject. In the first stage the analysis of the document is carried out aimed at thematic description, called the *document analysis process*. The first element is the *physical document* or *digital objec*t being analyzed. The second step is the formulation of an indexing phrase or subject description, called the *subjectdescription process*, a mental formulation or written matter by the indexer, with the second element the *subject* of the document, which may be present only in the mind indexer. In the third stage is the subject description of the translation in an indexing language or classification scheme, called the *subject analysis process*. The third element is the *formal description of the subject*. The fourth element, called the *subject entry*, is the product of translation of the formal description of the subject in a particular indexing language information retrieval system.

The indexing complexity lies in the subject analysis of a document held during document reading, early stage that triggers all other operations. In practice, the indexing process is analyzed under three theoretical conceptions: document-centered approach (emphasis in the document); user-centered approach (emphasis on users) and the domain-centered approach (includes the context, the document and users). In this sense, the domain-centered approach is the ideal in terms of indexing, considering factors other than the document or the user (Gil Leiva, 2008). However, as interfering factors include: the information retrieval system that is being used; the users profile; prior knowledge of the indexer, their professional experience and training in business analysis (Gil Leiva, 2008); professional guidelines (Mai, 1997a); the library indexing policy (Gil Leiva & Fujita, 2012) and, in a broader context, science policy and technology university (Tartarotti, 2014), among others.

As for the differences between the practice of subject cataloging and indexing in university libraries, the performance of catalogs as true databases is a trend due to two main factors: “the extent that the internet has given the catalogs of libraries, since now they are available without spatial and temporal boundaries, allowing users to access them from anywhere at any time” and “the demand increasing user on aspire to that

catalogs act as real databases, offering specificity, speed and hyperlinks to full texts". In this way, the term *indexing* is also used to describe the thematic treatment performed during cataloging in university libraries. The aim of the indexer is to achieve thematic representation according to the content of the documents (the author's term) and information needs of the user's online bibliographic catalog (Fujita, Rubi & Boccato, 2009, p. 31; 39).

Historically, librarians created order in the knowledge of the universe based on the analysis / understanding of the objects in the universe of knowledge and its use. On the other hand, folksonomies arise without the interpretive involvement of professionals, where the order of the objects is carried out collaboratively, in a socio-constructivist approach to represent and organize information (Mai, 2011, p.120; 118; 115).

> The organization of information as practiced in catalogs, indexing and abstracting databases, and other tools of bibliographic control is primarily based on traditional or Aristotelian logic. The result is a linear, hierarchical structure made up of mutually exclusive categories. […] Radically different from these information standards are the so-called folksonomies. They develop from social tagging —the naming of information by the user for the user. These tags are usually shared with other users. There is generally no controlled vocabulary and no hierarchy, or only a very shallow one. In fact, there is typically no structure at all imposed on tagging (Olson, 2007, p. 530; 2009, p. 135).

Mai (2011) presents a comparison of the values, success factors, challenges, naming and authority in traditional indexing approach and social indexing approach (Table 1):

Table 1: Comparison of two approaches in knowledge organization (Source: Mai, 2011, p. 121).

	Authoritarian, professional, expert-based	**Collaborative, democratic, everyone**
Values	Transparency, consistency, interoperability, stability, professionalism	Inclusiveness, openness, conversation, collaboration, interpretation
Success factors	Understand and match user's information needs, ability to reflect the domain's structure, ability to modify system accordingly to changes in domain	Involvement of users in meaning making, ability to facilitate collaboration among users, ability to accommodate diverse interpretations.
Challenges	Analyzing the domain and understanding it and its user's information needs	Getting people involved in sharing interpretation and collaborating on a shared goal
Naming	Information objects are named centrally by professionals	Information objects are named locally by users
Authority	Established through reference to external sources.	Established through autopoietic warrant [1]

In a democratic prospect of indexation, the phenomenon of collaborative or social tagging is essentially indexing by non-professionals without the benefit of a controlled vocabulary (Wolfram, Olson & Bloom, 2009, p. 1997). The term folksonomy, coined by Thomas Vander Wal in 2004, is derived from "folk" (people) and "taxonomy" (the study of the classification of things), or "classification made by people". The result is the free personal labeling of information or objects subjected to subsequent recovery in the social environment, shared and open Web 2.0. The author distinguishes between

two types of folksonomies: open folksonomy and restricted folksonomy. While the first allows anyone to create labels for the same object and uses terms of their own vocabulary, the tags available in the second type are previously defined by one or a few people, enabling indexing of more complex objects, such as images (Wander Wal, 2007).

According to Catarino and Baptista (2007, p. 13), there is no consensus on the definition of the term. While some authors understand folksonomy as the result of a process as a product (in the conception of the term of the creator), others refer to the folksonomy as a system, a methodology or approach or the process itself. In the literature, there are many concepts around the folksonomy: social indexing, collaborative tagging, social classification, cooperative classification, social tagging, ethnoclassification, among others, by allowing "the same item to be indexed by different individuals, resulting in a intersubjective description". However, it is worth noting that any of these terms show the social nature imbued with these concepts. Thus, it is justified the choice of the term social indexing and not social catalogingbecause we consider the importance of the theoretical- methodological foundations of the traditional indexing approach in the cataloging of subject in online bibliographic catalogs.

Originally, folksonomies arise in a pragmatic approach free of a theoretical foundation and create a new order in the landscape of knowledge organization (Mai, 2011, p. 116; 120). With the product a free indexing performed by the users, develop in a social and democratic environment, enabling the sharing for the recovery of information. The main advantages of the applicability of folksonomy in online bibliographic catalogs of university libraries are is noted: cost, time and investment reduction; rapid adaptation to language users; freedom of speech expression; collaboration; possibility of interaction between users; easy retrieval and low cost. On the other hand, the main challenges are related to the meaning and language, as the lack of vocabulary control, resulting in low accuracy rate and high level of recall, and polysemy, synonymy and ambiguity. However, this disorder is also the strength of folksonomies, especially considering that the disorder can be controlled and represent the plurality of views. It is believed also that this theoretical and technological approach can help improve semantic tools, classification, navigation taxonomies and methodologies for the construction of an indexing language in collaborative virtual environments (Mai, 2011, p. 117; Moura, 2014, p. 309).

To the field of Knowledge Organization, proposing information tools in the context of recovery without understanding the dynamics of discursive formation in a given field of knowledge has become even more complex, since the concepts refer to the description of a field where principles are defined. These are formed as a bundle of relationships (not an isolated object, an individual work or an area of knowledge at a given time), where the discursive context and coercion regularities, the theoretical

choices and historicity are taken into consideration, and events, transformations, mutations and processes are articulated (Moura, 2014, p. 305).

Although there is the possibility of a tension between thetraditional indexing approach (considered a universal approach) and the social indexingapproach (considered a contextualized approach), the theoretical and practical challenges lie in the convergence of both approaches in order to improve the representation and the information retrieval by subject in online bibliographic catalogs of university libraries.

3 Final considerations

In online bibliographic catalogs of university libraries, the purpose of indexing is to determine the subject content of documents aimed at effective information retrieval. In this context, the social indexing approach is configured as a new theoretical-practical paradigm of subject analysis, presenting potentially viable for deployment in online bibliographic catalogs.

In practical perspective, the application of the principles folksonomy in university libraries can bring benefits to the online bibliographic catalogs, improving thematic representation by allowing the users to organize information according to their perceptions. By allowing a more democratic documentary treatment, both among professional indexers and users in the indexing process, opens up a collaborative space despite of the decentralization of this activity librarians, historically experts and holders of knowledge/power to the documentary nature activities, especially in the thematic content analysis of the representation. In theoretical perspective, the traditional indexing approach and the social indexingapproach feature, when used together, new investigation glances for thematic representation in the field of Knowledge Organization, contributing to its establishment as a scientific field nowadays.

Considering that indexing can not be investigated separately from the social context and the area where it is held, the applicability of the social indexing approach in online bibliographic catalogs is even more complex than the traditional indexing approach performed only by professional indexers, due to the linguistic, cultural and contextual variability where the collaborative users in this discourse community are inserted.

Regarding future research possibilities in university libraries, it points to the need for investigations into the technological aspects for the implementation of social indexingapproach in online bibliographic catalogs; application of qualitative methodology of Verbal Protocol in libraries that use folksonomies for representation and thematic retrieval in online bibliographic catalogs; the intraindexing/interindexing consistency in assigning terms between librarians and users through the quantitative methodology of the Indexing Evaluation; comparative analysis of the matters assigned by users with the controlled vocabulary of the information retrieval system; the power psychological aspects to assign and design guidelines for the development/redesign indexing policy in this collaborative environment of folksonomies, allowing information to really be represented and recovered properly. It reiterates the need for

further research around the subject of analysis in this collaborative environment, as it is the first step in the indexing process and more complex, still lacking theoretical and methodological foundations that enable a recovery for quality issues and credibility in online bibliographic catalogs of university libraries.

In this transition in what we consider an individual approach (centered) for a collaborative approach (decentralized) of the indexing process, some questions emerge: Are Brazilian online bibliographic catalogs prepared to incorporate this new configuration in the process of determining subjects of digital objects? Would the combination of paradigmatic relations from controlled vocabularies and syntagmatic relations marking be the ideal way of indexing? For this to be achieved, two things are needed: a creative work between theory and practice to develop real tools and institutional willingness to sign and implement these innovations (Olson, 2009, p. 143; 2007, p. 536).

Finally, by incorporating *new*, *different* and *future voices* (McTavish, 2014, p. 330), considering the natural language combined with standardized language, it is expected that the document language in this new *collaborative*, *discursive* and *contextualapproach* as an alternative model actually allows the relevant document retrieval in online bibliographic catalogs of university libraries, contributing to the construction of collective knowledge in contemporary, purpose of yesterday, today and tomorrow's indexing.

Note

[1] The author realizes the autopoietic warrant as a guarantee in collaborative systems aimed at organizing information, where the informational subjects establish the terms and classes to be included and the system authority emerges from this interaction (Mai, 2011, p. 119).

References

Buckland, Michael (1992). *Redesigning library services*: a manifesto. Chicago, London, American Library Association.

Catarino, Maria Elisabete & Baptista, Ana Alice (2007). Folksonomia: um novo conceito para a organização dos recursos digitais na *Web*. *DataGramaZero*: Revista de Ciência da Informação, *8(3)*.

Dahlberg, Ingetraut (2014). In ISKO and Knowledge Organization's 25th anniversary: the future of Knowledge Organization and ISKO Panel Discussion. Reported by Rebecca Green. *Knowledge Organization*, 41(4):327-31.

David, A. In ISKO and Knowledge Organization's 25th anniversary: the future of Knowledge Organization and ISKO Panel Discussion. Reported by Rebecca Green. *Knowledge Organization*, 41(4):327-31.

Dias, E. W. (2006). Organização do conhecimento no contexto de bibliotecas tradicionais e digitais. In: *Organização* da *informação: princípios e tendências.* Brasília : Briquet de Lemos. Pp 62-75.

Frías, José Antonio (2004). De las tabillas sumerias ao acceso público em línea: la recuperación de la información a través del catálogo de la biblioteca. In: Magán Walls, J. A. (Coord.). *Tratado básico de Biblioteconomía*. 5. ed. Madrid: Editorial Complutense.

Fujita, Mariângela S.L., Rubi, Milena P., & Boccato, Vera R. C. (2009). As diferentes perspectivas teóricas e metodológicas sobre indexação e catalogação de assuntos. In: *A indexação de livros*: a percepção de catalogadores e usuários de bibliotecas universitárias. São Paulo: Cultura Acadêmica. Pp 19-42.

Gil Leiva, Isidoro (2008). *Manual de indización*: teoría y práctica. Gijón: Trea.

Gil Leiva, Isidoro & Fujita, Mariângela Spotti Lopez (Eds.). (2009). *Política de indexação*. São Paulo: Cultura Acadêmica; Marília: Oficina Universitária.

Hjørland, Birger (2008). What is Knowledge Organization (KO)? *Knowledge Organization*, 35(2/3): 86-101.

Mai, Jens-Erik. (1997a). The concept of subject: on problems in indexing. *Knowledge Organization for Information Retrieval*. Proceedings of the 6th International Study Conference on Classification Research, *6*. Pp 60-67.

Mai, Jens-Erik. (1997b). The concept of subject in a semiotic light. Digital Collections: Implications for Users, Funders, Developers and Maintainers.*Proceedings* of the ASIS Annual Meeting, *34*. Pp 54-64.

Mai, Jens-Erik (2011). Folksonomies and the new order: authority in the digital disorder. *Knowledge Organization*, 38(2): 114-22.

Moura, Maria Aparecida (2014). Emerging discursive formations, folksonomy and social semantic information spaces (SSIS). *Knowledge Organization*, 41(4): 304-10.

McTavish, J. In: ISKO and Knowledge Organization's 25th anniversary: the future of Knowledge Organization and ISKO Panel Discussion. Reported by Rebecca Green. *Knowledge Organization*, 41(4):327-31.

Ohly, H. Peter. In: ISKO and Knowledge Organization's 25th anniversary: the future of Knowledge Organization and ISKO Panel Discussion. Reported by Rebecca Green. *Knowledge Organization*, 41(4):327-31.

Olson, Hope (2007). How we construct subjects: a feminist analysis. *Library Trends*, 56(2):509-41.

Olson, Hope (2009). Folksonomies, syntagmatic relationships, & feminist research: alternative knowledge structures. In: *Memoria del XXVI Coloquio de Investigación Bibliotecológica y sobre la Información*, Held at Centro Universitario de Investigaciones Bibliotecológicas , October 1, 2 and 3, 2008.México: UNAM. Pp 131-43.

Rasmussen, E. (2011). Library systems. In: Baeza-Yates, R., & Ribeiro-Neto, B. *Modern information retrieval: concepts and technology behind search. 2.ed. Harlow:* Addison Wesley.

Tartarotti, Roberta C. Dal'Evedove (2014). Atuação bibliotecária no tratamento temático da informação em unidades informacionais: um estudo comparativo qualitativo-quantitativo. *Master Degree Thesis*. São Carlos: Universidade Federal de São Carlos.

Wander Wal, Thomas (2007). *Folksonomy coinage ad definition*. [http://www.vanderwal.net/folksonomy.html]

Wolfram, Dietmar, Olson, Hope & Bloom, Raina (2009). Measuring consistency for multiple taggers using vector space modeling. *Journal of the American Society for Information Science and Technology*, *60(10)*: 1995–2003.

Marcilio de Brito, Widad Mustafa El Hadi, Maja Žumer and Simone Bastos Vieira

Indexing with Images: The *Imagetic* Conceptual Methodology

Abstract

This proposal presents the methodology of indexing with images based on Peirce's semiotics. The indexing processes are analysed under the perspective of the sign both as a word and as image to show correlations between signs from literal and imaged universes. It is explained how the traditional indexing mechanisms are related to Peirce's semiotics in its simplified triad form of icon, index and symbol. Imaged indexing leads to more intuitive interfaces, an imaged KOS (iOPAC), improving the communication process throughout information systems. iOPAC would be a solution for supporting information access andnavigation, language and concepts fostering better understanding and could pave the way towards semantic, social and cultural interoperability. Furthermore, the indexing methodology is analysed in line with the FRSAD model providing the entity-relational frame of reference for relating images to *nomens*.

Introduction

The pervasive power of digitization causes scientific, educational, economic and cultural communities to change their modes of accessing, sharing and disseminating knowledge. Besides, complexity and uncertainty from compulsive modernisation drive people to rely on information to perform their professional activities or to simply exercise their citizenship. In this social flow, the world became a great consumer of images because they are more effective in mass communication than written-only documents (Social Bakers, 2014; Asthon, 2015). Nowadays, it is extremely easy to find things on the web except for illiterate individuals. Initiatives supporting privileged contacts between people and books are valuable strategies for developing reading habits according to the International Federation of Library Associations and Institutions (IFLA).

> For many deaf people, mastery of the oral and written idiom is a particular challenge. Libraries should strive to acquire general materials that may be understood by as many of their clientele as possible. Additionally, libraries should build and actively maintain a collection of high interest materials which are written purposefully with direct and simple vocabulary and which are heavily illustrated where appropriate so that they may be easily understood by people who have yet to gain full mastery of the local oral and written language, including many deaf people as well as people from other linguistic minorities.(IFLA, 2000,18-19)

Semiotics shows that image is a communication language and people have natural competencies to assimilate linguistic properties from images to engage in a communication process, what corroborates with prior linguistics theories. In fact, the grammar codes that govern imaged communication are easier understood by a larger public, although structurally more complex and less explicit.

Libraries are social spaces characterized by the gathering of informational objects in synchronicity with users' demands. In the process of information organization, and specifically in the step of content description, the (written) language is the code largely used to represent concepts.

Indexing with images consists of using images instead of key-words or descriptors, to represent and organize information. Profundity in discourse is a characteristic of the written language, yet its representation is only possible with elements of cognition. It is a complex process to make use of words to represent knowledge. Yet, text contents have been embodied by textual elements themselves. Admitting the multidimensional characteristics of images to broadcast messages, this work proposes a different point of view in the communication perspective. It certainly means a shift on knowledge organization systems (KOS) and leads towards the achievement of social and cultural interoperability as defined by Mustafa El Hadi (2015). For example, in terms of knowledge organisation, the brain of a born deaf individual (Marschark et al., 2000; Mcevoy et al., 2004) does not create the same connexions as hearing individuals. They generate alternative logical networks to understand and interact with the textual (oral and written) society. Deaf people do not naturally classify information as our current systems of classification do. Despite their normal intellectual abilities, they do not develop the same aptitudes of reading and writing that the majority of students at (normal) school. However, concerning the pervasive imaged world, deaf people attest far more awareness of images then hearing people. These observations lead to significant understandings of image perceptions. Evidences are that accuracy in image details identification is related to the ability of decoding imaged messages. Because messages are codes, frequently converted to textual systems, the decoding methods for imaged messages are the key for interoperability between written and imaged communication. Regardless of the predominant competence of deaf people in recognising images, society still cannot give them the freedom to evolve independently in our written world. There is potentially a bridge connecting messages from these distinct universes, establishing an authentic two-way communication process.

What hinders semantic and consequently cultural interoperability is that the degree of success that can be achieved in the integration of multiple knowledge representation systems or knowledge organization schemes is constrained by limitations on the universality of human conceptual systems. In contrast, considering that images are powerful communication skills, imaged KOS are potentially valuable "intercultural" interfaces for semantic interoperability which is one of the core elements towards cultural interoperability. In this sense, we support the idea that imaged KOS support translatability and additionally promote navigation within social and cultural diversity through their iconic interfaces. We will develop in the following sections the concept of the imaged online public catalogue (iOPAC) [1] as a KOS for accessibility in libraries, archives and museums.

To go further in developing the imaged KOS we will consider implementing it as a user-focused mechanism compatible with FRSAD (Functional Requirements for Subject Authority Data).FRSAD supports the idea that a work has subjects (*thema*), and a *thema* has one or more appellations *nomen*. *Nomen* is any sign or combination of

signs (alphanumeric characters, symbols, sound, images, etc.) that a *thema* is known by, referred to or addressed as. FRSAD is focused on aboutness to provide a clearly defined, structured frame of reference for relating the data that are recorded in subject records and tailored to meet the needs of the users of these records, and to assist in an assessment of the potential need for a global information share and use of subject data, both within the library sector and beyond. While associated with topics, images can be interpreted as *nomens* within FRSAD conceptual model.

2 Linguistic elements

The information retrieval process contains a complex relationship between communicating and describing the language to express thoughts. In fact, what is needed is to represent materialized knowledge in a text by acts of language. This sort of representation applies methods to canalize intrinsic proprieties from language and from its grammar. Content transposition to a meta-language that uses images as representatives of this controlled (meta) language is which we are looking for.

Intuitively we know that inside a text there is a signification (objects inferring knowledge). What is verified, however, is that those objects can be expressed by natural operations, which are difficult to represent artificially, such as reading. The essential criteria of modernity in information processing methods reside in innovative ways of solving problems, not necessarily in a sophisticated technology. In practice, indexing as it is known is a translation of lexical units drawn from language, or a syntactic translation reflecting the relationships between parts of the speech, the ones describing contents, the descriptors. It aims to represent the objects that the document is talking about, or in other words, what is said in the speaker's message. Although words are used to index contents, the words from the natural language or from dictionary (morphemes) point only to their signifiers, not to referents. They do not have these characteristics of designing objects.

Words from the lexicon, the list of words from a documentary system, do not have the status of words from the language nor from the discourse. Words from lexicon do not design objects neither; they refer to an open set of objects with common characteristics. It is not possible to perceive the borders of this set of objects, and so it is difficult to identify referents, and connexions between lexicon words and the objects from the reality (extra-linguistic reality or imaginary). What separates lexicon words from words in terminology is very subtle understanding that explains why there is often confusion between them.

Words in terminology [2] introduce a notion of a boundary or "terminus" in Latin that gave the word "term". The French linguist Michel Le Guern (1989) explains that in the lexicon as in terminology there are words on both sides, but they are not the same words. The object "word" from the lexicon is a distinct reality, the lexicon processes words disconnected from the objects, while in terminology the words are connected to things. Words in lexicography are considered as nouns but in reality they

are predicates. They talk about qualities, not about substances; they refer to proprieties, not to substances; to qualities, not to objects. The presence of an object calls for a term to be admitted into the discourse.

3 Semiotic contributions

For a long time the advantages of using images to transmit and receive messages have been investigated in semiotics. Advances on information technologies encourage important considerations about the information media. Literacy is essential, nonetheless we all read differently. The registered information is not only made by text, image has (re)conquered its place in communication. Interpreting imaged signs is as to reopen the gates of ancient temples reading stories trough cultures and ages.

In his pragmatism and logic, Charles Sanders Peirce (1839-1914) assumed that there is no immediate (without signs) perception of reality, so in the perception process everything is a sign including thoughts (Nöth, 2012). He claims that a sign consists of three elements, one of them is the sign, the second is an object to which a sign refers and the third, the most significant, is an interpretant (Lefebvre, 2007). In its simplified triad form of icon, index and symbol, Peirce's logic explains that an icon represents an acquired experience from the past, a reminder from that experience, a portrait of a concluded moment. An index apprehends its experience from the present, it points to a thing, without giving any information about it. The symbol is the real fact that will be experimented in the future, at the very moment when the message arrives to the receiver and the formulated intensions are to be re-established.

For Peirce, language and image compose the signs whose functions are embedded in the relations between the icon, the index and the symbol. When the roles of those signs are in balance, there is fulfilment. Peirce's theory is largely employed for image description purposes and in archival practices. The focus is on the description of an 'image-document' for further retrieval.

Instead of describing images, we are interested in using them as a sign, replacing key-words with images for content representation. This imaged representation presupposes a different navigation approach for the KOS and a subjacent (lying under or below Peirce's theory) theory to support the whole. For Peirce each object is in relation with a number of other objects from the same universe, directly or indirectly, in such a path that each element carries within itself an undetermined indicial potential. How then a sign can represent an object and reveal it itself? The iconic representation is a quality that emerges from the object in order to identify it as it is. Signs leading to inferences over true realities are the ones that make possible indexing with images principles as explained by De Brito and Caribé (2015).

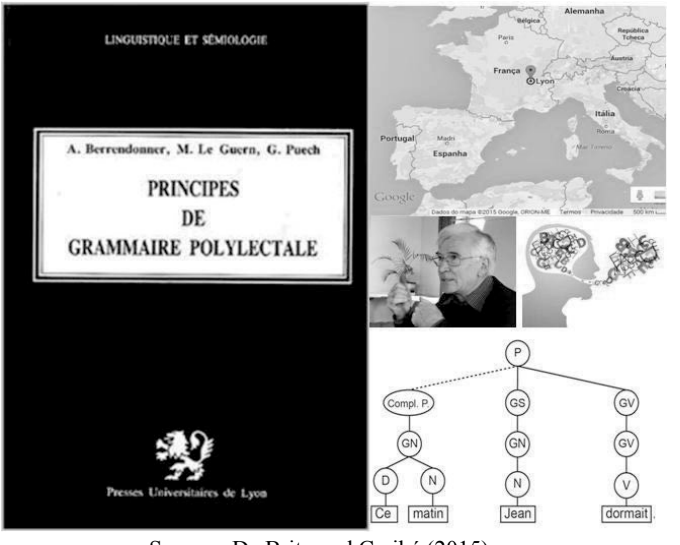
The emerging 'key-image' concept means substitution or equivalence with the key-word concept commonly used to describe subject contents. The procedure of intentional construction of images of this model is inspired from Jacques Bertin's (*apud* Dantier, 2008; Bertin, 1970) works in which he demonstrates that image

composition follows the rules of linguistic semiotic. For this author, every thought is expressed throughout a system of signs imitating a natural codification. The verbal language is a code of audible signs, the writings of a language are another kind of code, and so it is a graphic representation. So, if a graphic representation is a transcription of information from a graphic system of signs, then it has to be considered semiotic matter.

For Bertin a graphic representation might have three basic functions: register, communicate and process information. Indexing with images anticipates a moment of lecture and another of creation. It results in building an image (a chart) able to communicate a thematic message corresponding to the document's contents. In comparison to the key-words, the result of indexing with images is not a simple selection of images subjacent one to another, but a chart of significative images intentionally composed to become the key-image of the document. These indexing images are made of a set of iconic, indicial and symbolic proprieties (transmitted or inherited) to represent the document. This new composed image, gathering multiple semiotic traits, has itself a new symbolic interpretation.

The example below shows a key-image collage concerning a book in linguistics. The reader can notice that even if this example, which is in a foreign language, it is still possible to capture some information about the book's subject, its origin and knowledge domain.

Figure 1- BERRENDONNER, Alain; LE GUERN, Michel; PUECH, Gilbert. Principes de grammaire polylectale. Presses universitaires de Lyon, 1983



Source : De Brito and Caribé (2015)

4 Functional Requirements for Subject Authority Data (FRSAD)

IFLA proposed in 2010 a new bibliographic infrastructure to support global sharing and reuse of subject authority data, the FRSAD model. Aboutness is the FRSAD focus to provide a clearly defined, structured frame of reference for relating the data that are recorded in subject records. Žumer, Zeng and Salaba (2012), show how this structure is tailored to meet the needs of the users of these records and to assist in an assessment of the potential need for a global information share and use of subject data both within the library sector and beyond.

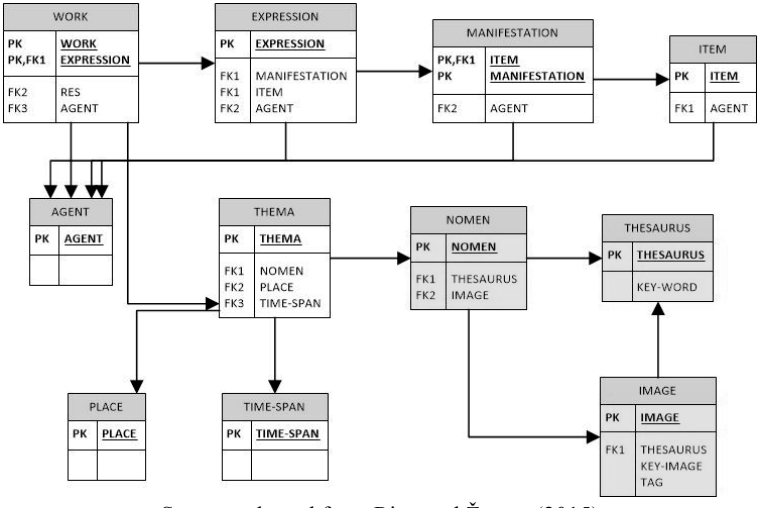
The scope of the FRSAD was defined in the following terms of reference:

- To build a conceptual model of Group 3 entities within the FRBR framework as they relate to the aboutness of works;
- To provide a clearly defined, structured frame of reference for relating the data that are recorded in subject authority records to the needs of the users of that data;
- To assist in an assessment of the potential for international sharing and use of subject authority data both within the library sector and beyond (Salaba, Žumer and Zeng, 2011, 9).

According to Gemberling (2016) one of innovations is to extend FRSAD's element *nomen* to apply to Groups One and Two as well as subjects. *Nomen* is "any sign or arrangement of signs by which an entity is known." The author reminded that *nomens* are the "symbols," in Peirce's sense, which we use to represent things.

In FRSAD model a work has a subject *thema*, and a *thema* has a appellation *nomen*. *Nomen* is any sign or combination of signs (images inclusive). FRSAD presents four subject authority data user tasks. Find to find an entity (*thema* or *nomen*) or set of entities corresponding to stated criteria. Identify to identify an entity (*thema* or *nomen*) based on certain attributes or characteristics. Select to select an entity (*thema* or *nomen*). And Explore to explore any relationships between entities (*thema* or *nomen*), correlations to other subject vocabularies and structure of a subject domain. A *nomen* can be human-readable or machine-readable. *Nomen* is a superclass of the FRSAD entities name, identifier, and controlled access point.

Figure 2- FRSAD Relationships



Source: adapted from Riva and Žumer (2015)

FRSAD Entity-Relationship conceptual model (Riva and Žumer, 2015; Salaba, Žumer and Zeng, 2011,15) postulates that"has appellation/ is appellation of" relationship is in general many-to-many. A *thema* has one or more *nomens* and there may be a *nomen* referring to more than one *thema*. In addition the diagram above shows how *nomens* are related to images (key-image descriptors) and traditional relationships with thesaurus. Images can be tagged with entries from thesaurus (Gheorghita, 2011; Aitchison and Clarke, 2004) or with external social tags, such as from folksonomies.

As a result the relational model illustrates that *key-images* (tailored made)are tagged as well as original images explaining inheritance from concepts embedded in images, what ensures wider semantic fields for imaged descriptors and sense integrity throughout the indexing process. Moreover, the same integrity improves KOS functions and information retrieval quality in interoperability perspectives.

5 The imaged online public access catalogue (iOPAC)

To reformulate OPACs, under this imaged indexing initiative, this will involve not only IT solutions but also giving answers to epistemological questions inherent to the nature of images. The procedure of using images to describe documents is not free of impacts. Besides, Papy (2016, 57) reminds us that it is imperative to meet the users' needs by rethinking the design of devices and stopping privileging the technological orientation.

Structurally, while looking for a specific document using the iOPAC interface, the user's behaviour first tends to recognize the conceptual relationship between images and objects (image-subject/document-subject), then he uses this cognitive mechanism

to find the object (a book) or set of objects he is looking for. In a second case, the user has an information need, but ignores that there is an object (or a set of objects) that could respond to his demand. He tries first to match needs to images, and then to verify inside the collection of documents if there are connexions of the same kind, repeating the first case above.

Implementing the imaged navigation in OPACs denotes multiple advantages derived from this. The iOPAC has a greater visual attraction pushing users towards the catalogue; a more intuitive comprehension of indexing codes, a larger conceptual portability of descriptors (as images), and a better interoperability between discourse codes and indexing competences affecting positively social and cultural interoperability.

Applying the imaged concept to *nomen* in the FRSAD model is rethinking the catalogue anew, since we are looking forward to sharing concepts within the subject authority data. This happens when images, carrying linguistic objects, permeate inter-social and cultural concepts. In practice it includes translated metadata, symmetrical multilingual thesaurus, or any traditional indexing tools. iOPAC embodies efforts focused on conceptual levels as expected from librarians. Also, model implementations have encountered challenges during its validation regarding methodology of mapping concepts in images or establishing conceptual relationships among subjects and classification systems. Yet, these are nothing but imminent issues for future research.

6 Final considerations

Indexing with images brings us to reconsider the current paradigms about using key-words to describe document contents. The key-images announce a legitimate approach to index documents with multiple perspectives in technical, professional and social areas. We can perceive changes in documentary retrieval fields, when enhanced with universal KOS based on imaged communication, and no longer contained by a specific written language. This contribution to knowledge representation is supported by semiotic theories and presents a new approach for documentation which needs to be tested in large scale of documents and users.

Web search engines may use these alternatives of indexing techniques to support KOS performance. A broader area for interface development is now available with effective benefices for handicapped users, such as deaf people, groups with functional illiteracy in general or in the sense of multicultural interoperability. In short, the use of images to create bonds between people and documents meet KOS challenges in a prosperous scientific ground.

The Imaged model, in addition, can be implemented as a user-focused mechanism compatible with FRSAD. We believe that the Imaged methodology can offer new possibilities for considering semantic, social and cultural interoperability when using OPACs.

Notes

[1] Work presented at ISKO-BRAZIL CONFERENCE (2015).

[2] Latin, boundary marker, limit − more at term. First Known Use: circa 1617. Source: Merriam-Webster's Learner's Dictionary.

References

Aitchison, Jean and Clarke, Stella Dextre (2004). The Thesaurus: A Historical Viewpoint, with a Look to the Future. *Cataloging&Classification Quarterly,*37(3/4): 5-22.

Asthon, Danny (2015). *10 reasons why you should care about visual content marketing*.[http://neomam.com/blog/13reasons/]

Bertin, Jacques (1970). La graphique. *Communications,* 15(1): 169–85.

Caribé, Rita de C.V and de Brito, Marcílio (2015). Indexação por imagens: via OPACs imagéticos. In: III *Congresso Brasileiro de Organização e Representação do Conhecimento.* Held at Universidade Estadual Paulista. Marília - SP: ISKO-Brasil. Pp 425-48.

Dantier, Bernard (2008). La représentation et l'étude visuelles des informations. In.: *Semiologie graphique, les diagrammes, les reseaux, les cartes*.4.ed. Paris. Ehess.

De Brito, Marcílio and Caribé, Rita de C.V (2015). Princípios da indexação por imagens. In: XVI *Encontro Nacional de Pesquisa em Ciência da Informação* (XVI ENANCIB), João Pessoa. Pp. 1-2

Gemberling, Ted (2016). FRSAD, Semiotics, and FRBR-LRM. *Cataloging & Classification Quarterly*54(2): 136.

Gheorghita, Inga (2011). Methodologie de construction automatique du thesaurus pour l'indexation et la recherche des images. *RECITAL,* 2, June:221–8, Jun. [https://hal.archives-ouvertes.fr/hal-00695722]

IFLA (2000). *Guidelines for Library Services to Deaf People*. 2nd edition. Hague: IFLA

Le Guern, Michel (1989). Sur les relations entre terminologie e lexique. *META Journal de traducteurs.* Montréal.

Lefebvre, Martin (2007). The art of pointing: on Peirce, indexicality, and photographic images. *Photography Theory*: 220-44.

Marschark, M. et al. (2000). Understanding theory of mind in children who are deaf. *Journal of Child Psychology and Psychiatry,* 41(8): 1067-73.

Mcevoy, C. et al. (2004). Organization and use of the mental lexicon by deaf and hearing individuals. *American Annals of the Deaf*, 149 (1): 51-61

Mustafa El Hadi, Widad (2015). *Cultural Interoperability and Knowledge Organization Systems*. Keynote address. In *Proceedings of the 3rd Brazilian ISKO-Conference,* Held atUniversity of St Paulo at Marilia.

Nöth, Winfried (2012). Charles S. Peirce's theory of information: a theory of the growth of symbols and of knowledge. *Cybern. Hum. Knowing*19 (1/2): 137-161.

Papy, Fabrice (2016). Digital libraries: interoperability and uses. Oxford: Elsevier.

Riva, P. and Žumer, M. (2015). *Introducing the FRBR Library Reference Model*. Paper presented at: IFLA WLIC 2015 - Cape Town, South Africa in Session 207 - Cataloguing.

Salaba, A; Žumer, M and Zeng, M (2011). IFLA Working Group on the Functional Requirements for Subject Authority Records (*Functional Requirements For Subject Authority Data (FRSAD): A Conceptual Model* [e-book]. Berlin: De Gruyter Saur.

SOCIAL BAKERS (2014). *Photos cluttering your facebook feed? here's why pictures account for majority of brand facebook posts.* April 21. Advertising & Marketing. [http://www.emarketer.com/Article/Photos-Cluttering-Your-Facebook-Feed-Herersquos-Why/1010777/1]

Tanja, Merčun, Žumer, Maja, and Aalberg,Trond (2012). *Presenting and Exploring the Complexity of Bibliographic Relationships. S*pringer-Verlag Berlin Heidelberg;H.-H. Chen and G. Chowdhury (Eds.): ICADL 2012, LNCS 7634, 63–6.

Tufte, Edward R (1998). *Envisioning information*. Connecticut, USA: Graphics Press.

Žumer, Maja, Salaba, Athena and Zeng, Marcia Lei (2007). Functional Requirements for Subject Authority Records (FRSAR): A Conceptual Model of Aboutness. In: *Proceedings of the 10th International Conference on Asian Digital Libraries (ICADL*), Hanoi, Vietnam, December 10-13, 2007. (Lecture Notes in Computer Science Series, Volume 4822) Berlin: Springer. Pp 487-92.

Žumer, Maja; Zeng, Marcia Lei and Salaba, Athena (2012) FRSAD: Conceptual Modeling of Aboutness. Third Millennium Cataloguing. Santa Barbara: Libraries Unlimited.

Jun Deng and Dagobert Soergel

Concept Maps to Support Paper Topic Exploration and Student-Advisor Communication

Abstract

This paper presents an ontology for organizing information about metrics and its potential application to defining and mana.

1 Aims and Introduction

A concept map is a node-link-diagram with nodes representing concepts and links representing relationships between concepts (Novak and Cañas, 2008; Novak, 2010; Cañas et al., 2005, CMC 2004 – 2016). This study explores the use of concept maps in formulating topics through the following research questions:

1 Do concept maps support students' exploration and definition of paper or thesis topics?

2 Do concept maps support communication between student and advisor in exploring and defining a thesis topic?

The short answer is yes to both questions, especially when concept maps are combined with a thesaurus (a source of additional concepts and relationships) and with conversation that stimulates thinking.

Sensemaking and thinking in general are supported by external representation of internal cognitive structures, conceptual structures held in the mind. The visual language of concept maps provides a powerful tool for clear and efficient external representation. "A concept map is a picture of the ideas or topics in the information and the ways these ideas or topics are related to each other. It is a visual summary that shows the structure of the material the writer will describe." (Crandell et al., 1996). Th external representation provided by a concept map can be manipulated and edited by the individual learner / sensemaker to improve his or her own understanding, and it can be used as a tool to direct search for further information and integrate the new information found. It can also be used as an effective means for communicating the structure of complex topics: "Concept mapping is a technique to let one person convey meaning to another in a visual format, and concept maps have been shown to foster a joint understanding between two individuals viewing the same map" (Freeman, 2004). If information in the mind is stored as an interrelated network of concepts and propositions, then the visual network representation might indeed be the most effective and efficient means to transfer such a network to the mind of another. The external representation can also be used for collaborative editing of knowledge structures.

Concept maps were used to display concept relationships in thesauri (Doyle, 1961, EURATOM, 1967) and later introduced to education by Novak (see references above)..

There is a vast literature on the *use of concept maps in education* for learning, including collaborative learning, assessment, curriculum and lesson planning, and more; for succinct reviews see Simone 2007 and Hay et sl. 2008.

Concept maps have been used in *assisting writers*: Crandell et al., 1996 studied the use of concept maps as an aid in revising documents and found that the resulting technical documents were easier to understand than documents revised without concept map assistance. Concept maps have been The entire session (the computer actions and screens and the conversation, but no visual image of the participant) was recorded as a video file. An observer took notes during the session. After the conclusion, we conducted a semi-structured interview with the participant. The combination of these data found *useful for collaborative thinking and negotiating meaning*. Freeman, 2004 tested concept maps in simulated interviews for the elicitation of user requirements; participants found concept maps to be useful in arriving at a shared understanding. Positive results are also reported for systems engineering teams at Boeing (McCartor and Simpson, 1999) and for lesson planning by teachers (Mackinnon and Kappel, to give just two examples. Basque and Lavoie, 2006 review collaborative concept mapping in education.

2 Methods

2.1 Sample Selection

Convenience sample: Through student listservs we recruited 10 PhD and master students in a Graduate School of Education who work on defining a thesis or term paper topic.

2.2 Procedures

Participants explored their own thesis or term paper topic using a software environment consisting of CMap (concept mapping), MS One Note, and MS Explorer. (Zhang, 2010)

Each participant (P) worked in one or more sessions with an information specialist S (project staff) who operated the software. P emailed a paragraph describing their topic beforehand. P and S discussed the concepts involved in the topic, and S, in discussion with P, drew a concept map showing the relationships among these concepts. S then showed P relevant sections of the ERIC Thesaurus to provide more information on concepts and/or relationships that could be added to the concept map under P's control. S also helped P to conduct literature searches based on the concepts in the map. In one case, the student and her adviser participated jointly in the session.

2.3 Data Collection

The collection methods provided a complete picture of the users accomplishing tasks with the assistance of concept maps.

2.4 Data Analysis

We transcribed and summarized the interview recordings. The first author coded all materials and completed a case report t for each participant: how they used the concept map, the thesaurus information, the results of literature searches, and the conversation.. Emerging patterns and themes were noted and added to the coding scheme.

3 Results

3.1 Use of the ERIC Thesaurus

The ERIC Thesaurus helped users discover new concepts to develop new directions of the topic. The concept relationships (NT, BT, RT) help expand the term pool, which in turn helps with elaborating and rethinking the topic. They also help in laying out the structure of the topic more logically and completely. For example, P2's concept map included the concept *early intervention;* in the thesaurus she found *early intervention* RT *at-risk persons*, so she added the concept *at-risk* to her map with a link to *early intervention*. NT relationships in the thesaurus help develop the detailed structure of the topic. The thesaurus helped with literature searches.

3.2 Use of the concept map

All participants agreed that the concept map was useful in defining their topic. The user progresses through a series of improving visual displays of the hierarchical and associative structure of the topic, developing a progressive and finally clear understanding of the structure of the topic. The concept map helps users think about the topic in a visually comprehensive way rather than in a linear way. It made it easier for users to figure out what should be modified and what should be elaborated.

Figures 1a-c show the progression of P5's topic development. When P5 came into the study room, all she thought of was the core of the topic *performance*. But by the end of the session she had developed the structure of the whole topic. Each time P5 developed the sub-points of the topic, she could find where the topic should be expanded, seeing clearly which part is short of organization. P9 developed the structure of the parts t*eacher perception of technology* and *administrator perception of technology* that she was not very clear about before the session, see Figures 2a - b.

The concept map serves as a concise outline of the topic and helps to choose a focus on a specific facet of the topic and limit the scope of the paper or thesis to something that is manageable.

CMap helps modify the structure flexibly and easily since it is kind of an electronic whiteboard with Post-its. CMap allows students to easily reposition a concept or a whole group of concepts with their links. All participants were interested in further use.

The concept map is a good starting place for literatures search: it gives a whole picture of the topic, and all the concepts are well-developed and useful as search terms. The concept map supports communication between a student and his or her advisor. Both can communicate ideas by modifying the map.

3.4 Use of literature search

Most participants were pleased to have help with searching for literature. They searched the ERIC database, other data bases suggested by the project staff, and Google. Search terms came from the concept map and the ERIC Thesaurus. Most searches were quite successful. For example, P5 found the article *Teacher as Performer: Unpacking a Metaphor in Performance Theory and Critical Performative Pedagogy*, which was exactly what she had been seeking since she started her research on the topic, and 16 additional relevant articles in the same journal issue. The references found helped with the definition/explanation of a topic and in elaborating the topic, adding new concepts, and revising the structure. Descriptors listed with ERIC abstracts were used as a source of concepts to be added.

3.5 Use of conversation

The communication between a project staff and the participant proved beneficial. The staff put the ideas and terms coming from the participants and from the thesaurus into the map and discussed placing and other aspects with the participants. This conversation helped bounce ideas back and forth and encouraged thinking about the topic more deeply and more comprehensively, resulting in many concepts and terms being added. For P03, P06, and P17 the majority of the terms came from conversation.

3.6 Use of OneNote

Participants could add a note to a map node with a few clicks. They used this function extensively to save literature references (with abstracts) and add explanatory notes.

4 Conclusions and implications. Original value of the paper

The detailed traces of the user's thinking and the combination of several tools in addition to concept mapping into an integrated software environment coupled with the use of a thesaurus as a knowledge source are novel. That this environment engendered both user success and user satisfaction and supported student-adviser communication should be of great interest to educators in particular.

Recommendations:

- Introduce students to concept mapping early in their studies.
- Make available an integrated environment for concept mapping, literature searching working from the concept map (Cañas, 2006), collaboration, and note taking.
- Support concept mapping by thesauri and other KOS, with the system making active suggestions for concepts to be added.
- Make it a practice to use concept maps as a tool in student-adviser conversations

References

Basque, Josianne & Lavoie, Marie-Claude (2006). Collaborative Concept Mapping in Education: Major Research Trends.*CMC* 2006. p. 79-86. [http://cmc.ihmc.us/cmc2006Papers/cmc2006-p192.pdf.]

Cañas, Alberto J. et al. (2005). Concept Maps: Integrating Knowledge and Information Visualization. In *Knowledge and Information Visualization*. Heidelberg: Springer. Pp. 205-219 http://cmap.ihmc.us/Publications/ResearchPapers/ConceptMapsIntegratingKnowInfVisual.pdf

CMC 2004 – 2016. *Concept Maps: Theory, Methodology, Technology. Proc. International. Conference on Concept Mapping (CMC).*. Tallahassee, FL: Institute for Human & Machine Cognition. [http://cmc.ihmc.us/cmc/CMCProceedings.html]

Crandell, Thomas L., Kleid, Naomi A. & Soderston, Candace (1996). Empirical Evaluation of Concept Mapping: A Job Performance Aid for Writers. T*echcomm Technical Communication,* 43 (2): 157-163.

Doyle, Lauren B. (1961). Semantic road maps for literature searchers. *Journal of the ACM.* 8 (4), p. 553-578.

EURATOM 1967: *EURATOM-Thesaurus. Pt. 1: Indexing terms used within EURATOM*s nuclear documentation system*. 2nd ed. Brussels: EURATOM 1966.12,90 p. Pt. 2: Terminology charts used in EURATOM*s nuclear documentation system. 2nd ed. Brussels: EURATOM CID 1967, 57 p. EUR500e. Newer edition 1969

Freeman, L. A. `(2004). The power and benefits of concept mapping: measuring use, usefulness, ease of use, and satisfaction. In *CMC 2004*. http://cmc.ihmc.us/cmc2004Proceedings/cmc2004%20-%20Vol%201.pdf.zip

Hay, David; Kinchin, I. & Lygo-Baker, S. (2008). Making Learning Visible: The Role of Concept Mapping in Higher Education. *Studies in Higher Ed.* 33 (3): 295-311.

Mackinnon, Gregory R. & Keppell, Mike. (2005). Concept Mapping: A Unique Means for Negotiating Meaning in Professional Studies. *Journal of Educational Multimedia and Hypermedia* 14 (3): 291-315.

McCartor, Mary; Morgan; Simpson, Joseph J .(1999). *Concept Mapping as a Communications Tool in Systems Engineering*. Presented at the INCOSE '99 Brighton Symposium. [www.systemsconcept.org/static_files/1999/INCOSE99_CMCTSE.pdf]

Moon, Brian M. (2011). *Applied Concept Mapping*: *Theory, Techniques, and Case Studies in the Business Applications of Novakian Concept Mapping*. Boca Raton, FL: CRC Pr.

Novak, Joseph D. (2010). *Learning, Creating, and Using Knowledge: Concept Maps as Facilitative Tools in Schools and Corporations.2. ed.* Routledge

Novak, J. D. & Cañas, A. J.(2008). *The Theory underlying concept maps and how to construct them*. Tech. Report IHMC Cmap Tools, Florida Institute for Human and Machine Cognition, 2006-01, rev. 2008-01, 33 p. Simone, Christina de (2007). Applications of Concept Mapping. *College Teaching.* 55 (1): p. 33-6

Zhang, Pengyi 2010. *Sensemaking: conceptual changes, cognitive mechanisms, and structural representations. a qualitative user study.PhD Dissertation*, University of Maryland.[http://drum.lib.umd.edu/handle/1903/10371].

Figure 1 Participant 5 Concept map

Figures 1a ad b just show the progression in complexity. Figure 1c shows terms legibly

Figure 1a. Participant 5 Concept map at the beginning of the session

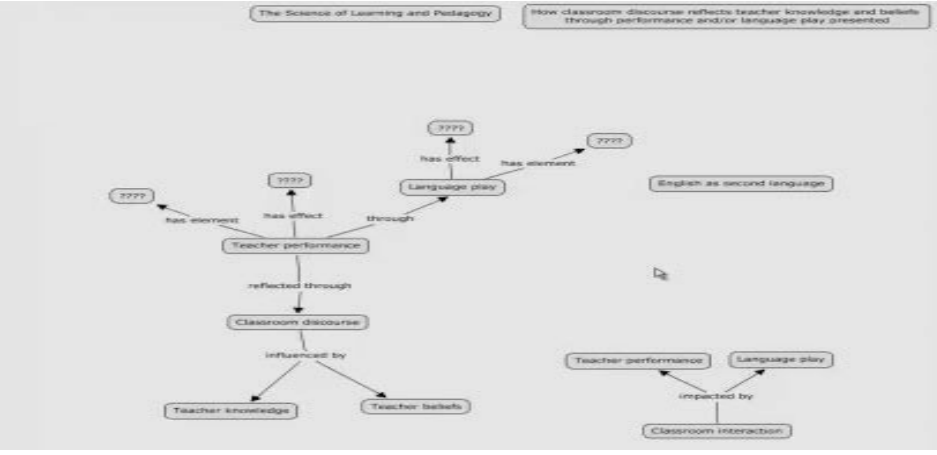


Figure 1b. Participant 5 Concept map at the middle of the session

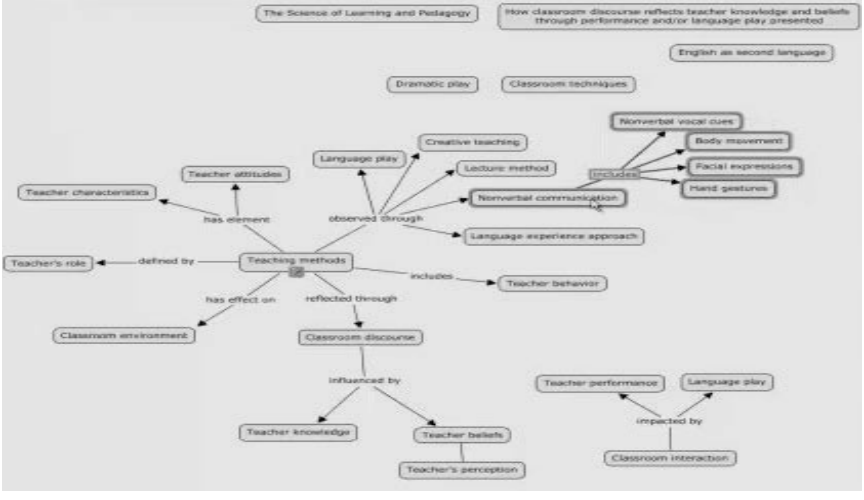


Figure 1c. Participant 5 Concept map at the final stage of the session

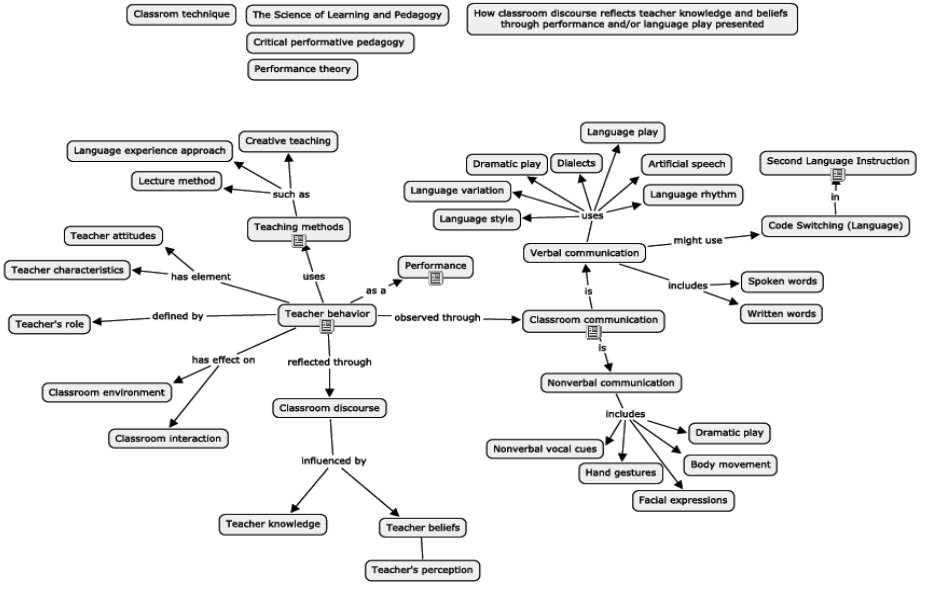


Figure 2a. Participant 9 Concept Map. Part "Teacher perception of technology"

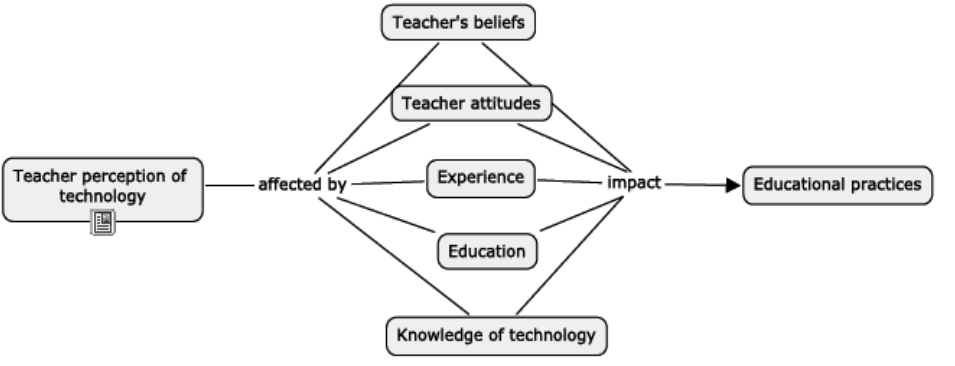
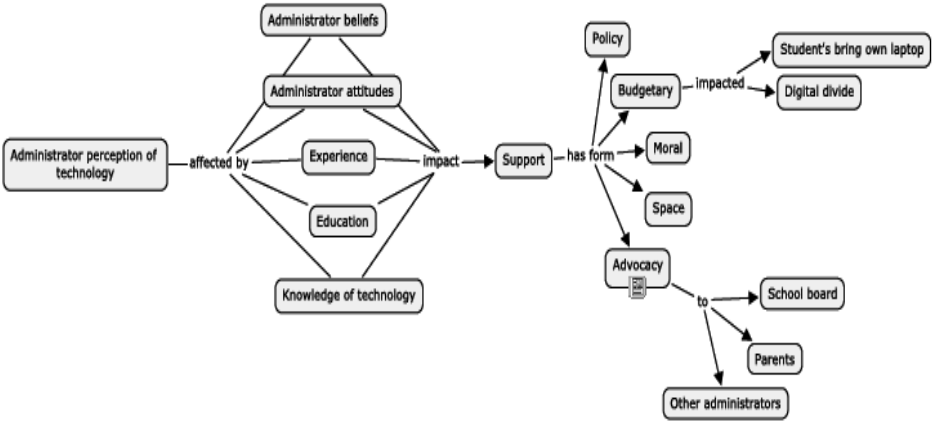


Figure 2b. Participant 9 Concept Map. Part "Administrator perception of technology"



Marisol Solis, Renata Wassermann and Vânia Mara Alves Lima

On the Use of Ontologies for Search in a Collaborative System for Architectural Images

Abstract

In this work, we have proposed an ontology for the Arquigrafia, a social network for sharing architectural images, based on a controlled vocabulary of architectural domain and tags created by users. We have followed the systematic approach for the ontology development that consists of four steps: purpose identification, capture, formalization and evaluation. We concluded that the Ontology can be used as a conceptual basis to relate the represented knowledge with computational processes such as, for example, information retrieval in a collaborative system, as is the case of Arquigrafia.

1 Introduction

Arquigrafia (www.arquigrafia.org.br) is a social network for sharing architectural images, housing both institutional collections, such as the set of slides belonging to the School of Architecture and Urbanism of the University of São Paulo (FAU-USP), as well as private collections. Arquigrafia was developed as a collaborative environment on the web, in which users participate describing and entering data either using tags, assigning a title or assigning subjects to a particular element of the system.

The images belonging to institutional collections are indexed according to the documentary standards used by the institution for the standardization of: (i) descriptive representation of fields as authors, title, date, following the AACR (Anglo American Catalog Rules) second edition (2002) and (ii) thematic representation, i.e., subject attribution according to the Controlled Vocabulary of the Integrated Library System of the University of São Paulo (VOCAUSP) (2001). Furthermore, institutional images are also indexed using a standardized tag list categorized by experts in architectural elements, materials and type, according to reference works in the area, such as the Architecture Experimental Thesaurus (1982).

On the other hand, in images of private collections added by users, author information, title, date, tags and subject are freely assigned. This freedom is typical of collaborative systems, and is often contrasted with formal ontologies, that are imposed by experts, not by users(Halpin, Robu and Shephard, 2006).In a collaborative environment there is no control on user defined tags, by consequence, in this data there may occur synonyms, homonyms and lexical anomalies, which may produce noise in the retrieval process (Macgregor; McCulloch, 2006). Thus, on one side we have what is commonly called folksonomies, structures generated from user data, which are easy to obtain but may be unreliable for image retrieval, while on the other hand we have standardized vocabularies that make retrieval more precise, although creating a controlled vocabulary remains a process depending on highly trained information professionals (Macgregor; McCulloch, 2006).

In the usual retrieval by tags, the system does not always find all images on the same subject, as some images may have been indexed by different tags, but with a similar semantic meaning, such as “school” and “college”. Similarly, the descriptors used in image indexing may contain ambiguous meanings, such as “banks”, that generate a distortion in the retrieval. In Arquigrafia, the list of descriptors and tags is available to users, but in order to protect the collaborative nature of the system, its use is not mandatory.

In an attempt to solve the distortion in information retrieval caused by user defined tags, Halpin, Robu and Shephard (2006) use tag co-occurrence networks for a sample domain of tags to analyze the meaning of particular tags given their relationship to other tags and automatically create an ontology. The ontology works well when the corpus is small or in a constrained domain, the objects to be categorized are stable, and the users are experts. In this work, we present the construction of ontology to relate authors, titles, tags and keywords with structured concepts in the area of architecture.

In the area of Knowledge Representation, ontology was defined by Gruber (1993, 199) as “an explicit specification of a conceptualization”, that is, a description of concepts of a domain and relationships that exist between these concepts. The basic components of an ontology are concepts (organized in a taxonomy), relationships that represent the kind of interaction between the domain concepts, axioms used to model always true statements and instances that are used to represent individuals described by the concepts and relationships (Almeida and Bax 2003, 9). Ontologies can be used as reference tools like vocabularies, providing a controlled and standardized terminology for indexing, but enriched with inference rules and axioms representing connections between different concepts.

A controlled vocabulary is a documentary language whose purpose is to represent the affairs of a particular area for information retrieval in the information system, either physical or virtual (ANSI/NISO 2005, 1). This instrument performs several functions as controlling the use of synonyms and grammatical variations; discriminating between homonyms and establishing relationships between terms.

In this work, we propose an ontology for the architectural domain, based on a controlled vocabulary and tags created by users; the first is used for constructing the class hierarchy and the second is used for the analysis and definition of properties and relationships between classes. The union of this information allows the inference of knowledge, which is useful to enrich the ontology construction. Furthermore, the ontology will be used to add terms related to the original search query posed by the user, creating a new extended query (Bhogal, Macfarlane and Smith, 2007) which we claim to provide better retrieval results.

2 Constructing the Ontology

For the development of our ontology, we have followed the systematic approach suggested by Falbo, Guizzardi and Duarte (2002, 351-358), which consists of the following four steps: purpose identification, capture, formalization and evaluation.

2.1 Purpose identification and requirements specification

The purpose of the development is to be able to deal with user queries in Arquigrafia in a satisfactory manner. For the specification of the requirements, we have used a set of competency questions, as proposed by Grüninger and Fox (1995). Competency questions delimit the minimal set of "competencies" that the ontology must address, i.e., the knowledge that must be represented in order to answer the questions.

Our competency questions were obtained by interviewing domain specialists and also processing the actual queries posed by users of the system.Examples of the competency questions are:

- Who designed the Brasiliana Library?
- Who is the author of the project of the Praça do Relógio?
- What is the typology of the Brasiliana building?

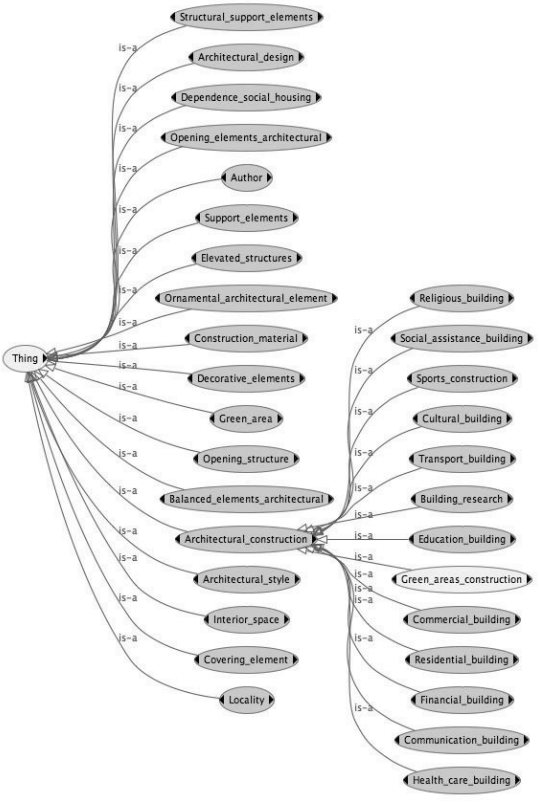
2.2 Ontology capture

Several sources were used during the construction of the ontology: the architecture area of the Controlled Vocabulary of the University of São Paulo (VOCAUSP), the Architecture Experimental Thesaurus and the data in the system, such as the descriptive fields such as title, author; and thematic fields as tags. The analysis of all this information allowed to define the concepts, properties and relationships of the ontology in four steps:

- The first step was the creation of the class taxonomy, in which classes and subclasses are connected by the relationship "is-a". Part of the class taxonomy of the proposed ontology is shown in Figure 1. We used the controlled vocabulary and the thesaurus, together with the titles and tags found in the images present in Arquigrafia.
- The second step concerned the creation of properties and attributes. There are object properties, which connect two individuals (class instances) and data properties, which connect an instance with a literal (values such as numbers, strings, dates, etc.). We used the history of queries performed by users in Arquigrafia.
- The third step was the creation of instances, which allowed us to validate the ontology using queries. This information was obtained from the tags related to each image. Example:for the "*Metallic_material"* class, we have instances as aluminum, steel, brass, copper.

- The fourth step was the creation of relationships, which allowed to use inference in the ontology. We had to create the relations between the instances, obtained from the query history. For example: the class "*Museum"* has as a property its "localization". The instance "*Jewish_museum*" is located in the instance "*São_Paulo"*.

Figure 1: Class hierarchy of the ontology



2.3. Ontology formalization

Our ontology was described in the Web Ontology Language (OWL), which is the standard recommendation of the World Wide Web Consortium (W3C).We used the ontology editor Protégé, which provides a graphical interface for editing and visualizing the ontology, as well as inference mechanisms for verifying the consistency of the ontology.

2.4. Ontology evaluation

The competency questions were formalized in SPARQL, the standard language for querying ontologies and the query results were analysed by domain specialists. For example, the question "What is the typology of the Brasiliana building?" is formalized as:

```
SELECT distinct ?individual ?typeClass ?classF ?classGF
where{          ?individual rdf:type ?typeClass.
        ?typeClass rdfs:subClassOf ?classF.
        ?classF rdfs:subClassOf ?classGF.
FILTER (regex(str(?individual)," Brasiliana_building ","i" ) ) }
```

When the query is applied to the ontology, the result in figure 2 is obtained, and the type of the building is *Library*.

*Figure 2:*Result for the query "What is the typology of the Brasiliana building?"

individual	typeClass	classF	classGF
Brasiliana_building	Library	Cultural_building	Architectural_construction

The ontology is considered to be ready when all the competency questions are answered in a satisfactory way.

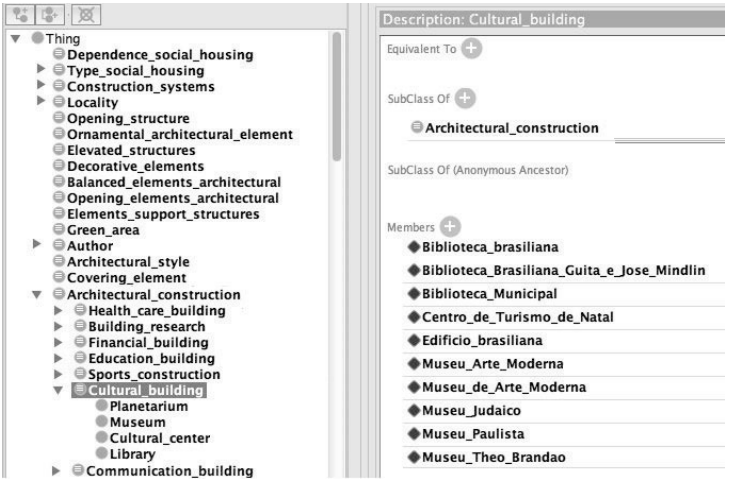
3 Results: Examples of queries with and without the use of the ontology

In Figure 3, we see the taxonomy of the class "*Architectural_construction*". If the user searches for images of "*Cultural Building*", he will get no results. This happens because there is no image with this explicit description or tag.

Figure 3: Information for the class " Cultural_building"

However, the search using the ontology will return instances of related classes, such as libraries, cultural centers, museums and so on. The corresponding classes are linked to "*Cultural_building"* through the "is-a" relationship. Hence, applying an inference mechanism such as HermiT allows the system to infer that since "*Library"* is a sub class of "*Cultural_building"*, instances of "*Library"* are also instances of "*Cultural_building"*, as shown in Figure 4.

Figure 4: List of instances of the class " Cultural_building " after inference.



The ontology has been evaluated according to the set of competency questions using a small set of images. We are currently running experiments with a large part of the Arquigrafia dataset (3720 images) in order to have a quantitative measurement of the effect of adding the ontology as background knowledge to the system.

4 Conclusion

Ontology is a formal structure, which not only can represent knowledge in a particular field, but also be shared and reused. Furthermore, it can be used as a conceptual basis to relate the represented knowledge with computational processes such as, for example, information retrieval in a collaborative system, as is the case of Arquigrafia. The ontology allows us to handle ambiguous terms and terms with similar semantic meaning, but with different written representation found in the tags or titles. The system uses an inference engine on the concepts, properties and relationships that allows finding similar terms to be used in addition to the terms in the user query, generating an extended query. This extended query can be used to improve the retrieval of relevant images. Thus, the construction of an ontology in the field of architecture was developed and tested in the system in order to improve the image retrieval.

References

Almeida, Maurício B. & Bax, Marcello P. (2003). Uma visão geral sobre ontologias: pesquisa sobre definições, tipos, aplicações, métodos de avaliação e de construção. *Ciência da Informação*, Brasília, 32(3):7-20.

ANSI/NISO (2005) *Z39.19-2005(R2010): guidelines for the construction, format, and management of monolingual controlled vocabularies.* Bethesda, Ma: NISO. Press, 2005. 184p. [http://www.niso.org/apps/group_public/download.php/12591/z39-19-2005r2010.pdf>] Acessed on 30 January 2016.

Bhogal, J., Macfarlane, A. & Smith, P. (2007). A review of ontology based query expansion. *Information Processing and Management,* 43(4) July: 866-86. DOI=http://dx.doi.org/10.1016/j.ipm.2006.09.003

Código de Catalogação Anglo-Americano: revisão 2002 2.ed. (2010). São Paulo, FEBAB : Imprensa Oficial.

Costa, Eunice Ribeiro R. & Douchkin, Tatiana (1982). *Thesaurus experimental de Arquitetura*. São Paulo: FAUUSP.

Grüninger, Michael & Fox, Mark S. (1995). Methodology for the Design and Evaluation of Ontologies. In *Workshop on Basic Ontologies Issues in KnowledgeSharing*, *IJCAI-95*, Montreal.

Gruber, Thomas R. (1993). A translation approach to portable ontology specifications. *Knowledge acquisition*, 5 (2):199-220.

Falbo, Ricardo de Almeida, Guizzardi, Giancarlo & Duarte, Katia Cristina (2002). An ontological approach to domain engineering. In *Proceedings of the 14th international conference on Software engineering and knowledge engineering*, ACM. Pp 351-8.

Halpi, Harry, Robu, Valentin & Shepherd, Hana (2006). The dinamics and semantics of collaborative tagging. In: *Proceedings of 1st Semantic Authoring and Annotation Workshop* (SAAW 2006)

Macgregor, George & McCulloch, Emma (2006).Collaborative tagging as a knowledge organisation and resource discovery tool. *Library Review* 55 (5): 291 - 300

OWL 2 [http://www.w3.org/TR/owl2-primer/#What_is_OWL_2.3F] Acessed 30 June 2015.

SPARQL [http://www.w3.org/TR/2008/REC-rdf-sparql-query-20080115/] Acessed 30 June 2015.

Vocabulário Controlado do Sistema Integrado de Bibliotecas da USP (VOCAUSP) [http://143.107.154.62/Vocab/Sibix652.dll/] Acessed 23 June 2015.

Carlos Guardado da Silva

Knowledge Organization in Portuguese Public Administration: From the Functional Classification Plan to the Creation of an Ontology from the Semantic Web's Perspective

Abstract

This paper presents a functional classification plan supported on business processes for the Portuguese public administration as a tool to promote semantic interoperability. The author initiates discussion by presenting the classification of functional information, briefly reviewing literature to justify the classification of systems in archival information systems. Then, he presents the business plan classification and how it was constructed, to later conclude that it is a new approach not only in the organization, representation and retrieval of information/knowledge, but also in the management of archival information, making it a matrix model that links functions to business processes. Also, despite the importance of this tool, he recognizes the need to develop the business plan classification tool to an ontology based on WOL (Web Ontology Language), a language for knowledge representation, which has been proposed by W3C as a 'standard' to codify ontologies from the semantic web's perspective.

Introduction

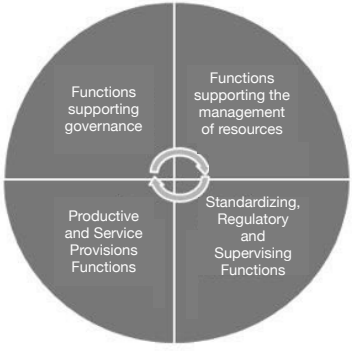
Considering the framework of European policies and strategies for interoperability, for the promotion of information access and for its reusability, as defined by the Decision No. 922/2009 and by the Directive 2013/37/EU of the European Parliament and of the Council, Portugal defined a structure of information classification for its entire public administration. The DGLAB (*General-Administration of Book, Archives and Libraries*), the coordinating body for the national archival policy, conceived this structure while working alongside with more than two hundred bodies of public administration (central, regional and local), over the last five years.

Regarding the Program for Electronic Government and Interoperability, DGLAB created Meta-information for Interoperability (MIP), «a set of meta-information elements with the purpose of supporting semantic interoperability within an electronic government's information production» (Silva, Guardado da, 2013, 4), as well as the Functional Macro-Structure (MEF) for Public Administration (version 2.0), which «is the standardization of the MIP element *classification code*», with the purpose of «identifying the significance of the information asset within the corporate body's functional context, which has to be posited transversally from an inter-organizational perspective» (Penteado, 2013, 4).

The Functional Macrostructure for public administration defines the classes for the 1st and 2nd levels of public administration functions, indicating, for each represented unit, a code, a name, a description, execution notes and exclusion notes. The aim is to support the conception of an incremental classification plan for public administration. It is based on a consolidated list of business processes that may be materialized at different levels in the classification plan, depending on the activities undertaken by the

different organizations. The Functional Macro-Structure in grounded on a conceptual model rooted on the establishment of four domains for functions, from which the 19 functions (F) for the Portuguese Public Administration were defined. Therefore, it is characterized by a functional structure that can best precise not only the identity of the administration's identity, but of society itself.

Fig. 1 - Functional Macrostructure – Portuguese Public Administration
Conceptual model



Functional Macrostructure
(1st level classes)

100 LEGAL AND REGULATORY FRAMEWORKS
150 PLANNING AND STRATEGIC MANAGEMENT
200 IMPLEMENTATION OF EXTERNAL POLICY
250 ADMINISTRATION OF WORK RELATIONS
300 ADMINISTRATION OF RIGHTS, GOODS AND SERVICES
350 ADMNISTRATION OF FINANCES
400 SERVICES PROVISION IN IDENTIFICATION AND REGISTRY
450 ACKNOWLEDGEMENTS AND PERMISSIONS
500 SUPERVISION, CONTROL AND ACCOUNTABILITY
550 IMPLEMENTATION OF SECURITY, PROTECTION OR DEFENSE OPERATIONS
600 ADMINISTRATION OF JUSTICE
650 SERVICES PROVISION IN PROTECTION AND SOCIAL INCLUSION
700 PROVISION OF HEALTH CARE
710 SERVICES PROVISION IN HYGIENE AND PUBLIC WHOLESOMENESS
750 SERVICES PROVISION IN TEACHING AND TRAINNING
800 SERVICES PROVISION IN TECHNICAL, SCIENTIFIC, RESEARCH AND DEVELOPMENT SERVICES
850 IMPLEMENTATION OF PROGRAMS AND ENCOURAGEMENT INITIATIVES
900 DYNAMIZATION AND INSTITUCIONAL COMMUNICATION
950 ADMINISTRATION OF CIVIC PARTICIPATION

The classification of functional information

The selection of a classification scheme that lays its foundations both on functions and sub-functions, which can be regarded as activities, and on business processes is increasingly becoming a prerequisite for the conception of organizational information systems. Firstly, as it is our belief, it's the functional nature of information that justifies a functional approach since such information is the result of a function and activity, according to the diplomatic concept of "function" proposed by L. Duranti, i.e., "the set of activities necessary to accomplish a goal, posited in abstract terms*"* (1998, 90).

Such approach is not recent, since it has at least been observed in the *Registratur* system in Prussia, during the sixteenth and seventeenth centuries, where classification was already based on functions and subjects. During the twentieth century, the British archivist H. Jenkinson demonstrated the alignment between function and structure, typical in the first bureaucratic organizations, so that archival series should report to a specific administrative function necessary for their existence. Likewise, he showed that the highest-level class in a classification scheme should match the division of the organizational unit or service that produced it (Jenkinson, 1937, 1965, 111; Jenkinson, 1943, 1980, 201).

When R. Schellenberg formulated a set of principles for the classification of North American records, he bolstered functional analysis by creating a hierarchical structure of functions, actions and transactions. He considered *the action* (the function) as the first and most relevant criteria for records creation, since most *public records* are the result of an action, i.e., a function, therefore, they should be classified as such (Schellenberg, 1956, 53, 62-63). Schellenberg is commonly praised by bibliography for this innovation, although the idea that records result from a function can already be found in E. Campbell (1941), in the context of the National Archives of the United States.

The '80s of the twentieth century witness the first attempts in devising a functional classification in classification systems developed in order to promote interoperability under the *Administrative records classification system* (ARCS) and the *Operational records classification system* (ORCS), in the Canadian provinces of British Columbia and Nova Scotia, respectively. By maintaining the main goals of information classification, regardless of dealing with hierarchical or enumerative and multifaceted classification systems, the systems brought on some benefits, such as the relation between classification and appraisal and retention, at the lowest level in the classification plan, with the indication of administrative retention schedules as well as the final destination, in order to favor the management of the complete life cycle of information.

By the end of the '90s, the former National Archives of Canada initiated a new project that endeavored to review the information classification system based on a methodology of functional appraisal, known as macro-appraisal, which led to the

creation of the Business Activity Structure classification system (BASCS). As a consequence, information is now arranged according to the structure of the activity (mentioned in the acronym BASCS), a functional structure conceived as a principle of original order through the decomposition of functions and activities, hierarchically and sequentially, down to the level of transactions that generate informational processes (Foscarini, 2010, 48).

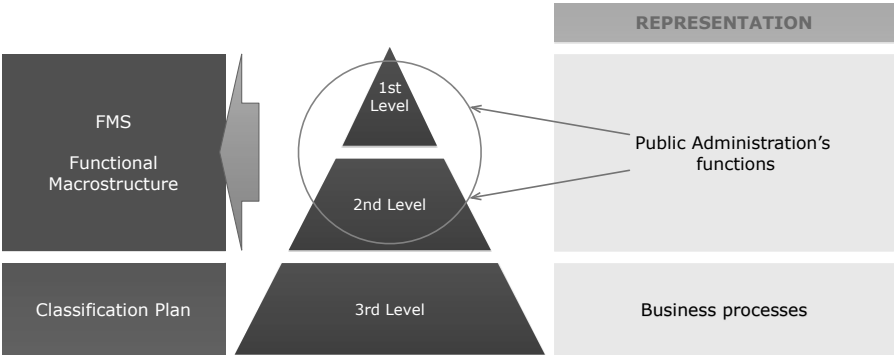
In such context, the archival discipline grants appraisal a major role, as opposed to bibliographic classifications. Despite the fact that appraisal is also useful for the organization, representation and recovery of information, it is mostly crucial for information management as it provides the grounds for administrative efficiency and effectiveness (Silva, 2015, 8) ‹‹since it promotes the organization and management of information›› (Simões & Freitas, 2013, 99). As a result, archival classification plays a significant part in the permanent management of information and knowledge that allows it to maintain the original, necessary and incremental bond — the organic nature that L. Duranti defined as the archival bond (1997), present in every organizational information, bonding records and data because they were created as a consequence of the same function, activity, or business process. Its purpose is to determine the initial network of relations that each informational unit has with other informational units and with the activity and function that produced it. This refers to the original principle of organization that must be maintained, and that is ensured by the classification of archival information, justified by the relevance and up-to-dateness of classification systems in archival information systems. However, we also recognize the added value of taxonomies and ontologies under the perspective of the semantic web. In this topic we second B. HjØrland's reply to his own question: is classification necessary after Google? (2012). Despite the fact that automated classification is possible and desired, there are multiple ways to classify information produced by public administration. However, collaborative appraisal still shows an insufficient level of quality. In other words, no matter the possibilities of classification, organizational information still relies on classification to guarantee that certain information 'belongs' to a class that ascertains its *archival bond*. Nevertheless, we recognize the semantic web's high potential, accomplished not with order and hierarchy, but with integration, collaboration and cooperation (San Segundo & Martínez-Ávila, 2012, 420).

We believe to have demonstrated the role that classification plays in the organization of archival information, as well as its significance for management. Its preponderance justifies the fact that classification is, on par with archival theory, the most discussed topic in the journals *American Archivist* and *Archivaria* over the last twenty years (Barros, 2012, 165), owing the most relevant revisions on classification and, more particularly, on functional classification to T. Eastwood and L. Millar.

Business Classification Plan

Following the legacy of the Functional Macro-Structure, a third product for information/knowledge organization and information representation, retrieval and management is under development. It is an information classification plan for Portuguese public administration (PCI-AP), with a multi-level hierarchical structure, elaborated according to three levels, so that the first and second levels match the Functional Macro-Structure's functions and sub-functions, respectively, while the third level relates to business processes. This is a process that replicates the theories proposed by archivists that have leaned towards information classification systems that rely on functions and business processes (Bak, 2010, 59, 71).

Fig. 2 - From the Structural Macrostructure to the Classification Plan



Considering that 'business process' is a polysemic concept, we revisit the definitions proposed by Thomas Davenport as a « (…) specific ordering of work activities across time and space, with a beginning and an end, and clearly defined inputs and outputs: a structure for action» (1993), and by Michael Hammer and James Champy, for whom business process is a « (…) collection of activities that takes one or more kinds of inputs and creates an output that is of value to the customer» (1995). We deconstructed this concept in order to establish the set of requirements for the profiling of a business process, namely:

- The identification in the framework of a Function and Sub-function (which we'd call 'respect for the function');
- The definition of input and output; identification of an output with a service or product;
- The understanding of a structured set of actions, tasks and transactions;
- The identification of the participants, regardless of their nature (owner or participant);
- The inexistence of a link between business process and work business or procedure;

- The existence of legal support, although the relation between law and process is not necessarily unambiguous;
- And finally, the observation of mutual relationships (for instance, if one pays, other receives; if one purchases, other sells) (Grupo de Trabalho para a elaboração do Plano de Classificação para a Administração Local, 2012, 10).

The creation of the classification plan had the following purposes:

1. To expand classification to the third levels, based on the Functional Macrostructure (MEF);
2. To elaborate a single Plan that could be used as a common tool for the entire Portuguese Public Administration;
3. To identify and represent the Business Processes (BP) carried out by the Public Administration (PA) throughout their duration (principle of wholesomeness).
4. To create a tool able to promote semantic interoperability in services and in e-government.
5. To standardize the classification of information in Portuguese public administration.
6. To include appraisal (administrative retention schedules and final destinations) in the classification plan.
7. To facilitate the creation of digital preservation plans; and
8. To promote accountability.

The project was initiated with an analysis of the law, in addition to research on the organizational context of the participating institutions. Once the concept of business process was consensual, the different processes, which would later be represented and integrated in the corresponding function in the conceptual model, were identified and described. Simultaneously, the business processes were classified as specific, common or overarching, in order to identify the owner and the participants in each of them but, mostly, in order to identify the nature of their participation, so that the descriptions of the identified common and overarching business processes could be harmonized.

Table 1 – Representation of a business process

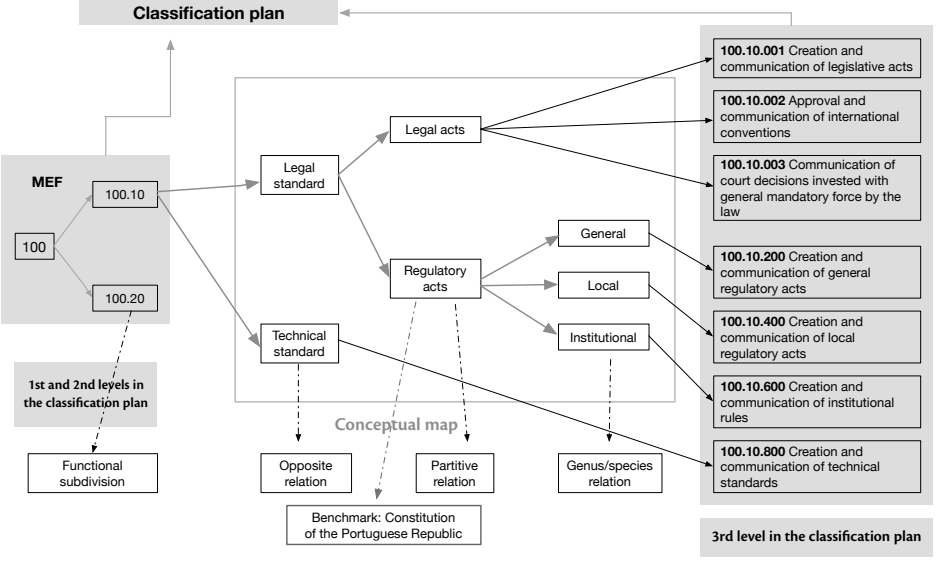
REFERENCE CODE	TITLE	DESCRIPTION
350.30.001	Revenue collection and expenditure	**R**eception and payment of any financial amount. **B**egins with the emission of a revenue or expense document and ends with the collection or payment of funds. **I**ncludes payment authorization, transfer of funds or issuing of cheques, confirmation of funds reception.

In their representation in the classification plan, we adopted a hierarchical and multilevel structure, from Function (F) to Sub-function (SF), and from sub-function to Business Process (BP). In the Macrostructure, each business process is represented by a

numeric code, a description (that defines what it is, not what it is used for; where it begins and ends; and stages of transmission), execution and exclusion notes. Finally, it is also represented by information concerning appraisal.

In the next step we created conceptual maps, according to function and sub-function, that would contribute to the identification and perception of granularity at the third level, with implications on the representation of the business processes. Amongst the range of available theories for the establishment of division principles applied specifically to the creation of the conceptual maps, we adopted I. Dahlberg's theory (1978, 101-107) that suggests the following types of semantic relations:genus-species relations (all elements in the subdivision have identical features, but each of them has one more feature than the root-element where it comes from that specifies it); partitive relations (between a whole and its parts or a product and its constitutive elements); opposite relations (contradiction); and functional relations (a subdivision created according to functional deconstruction). Lastly, we clarified the rules for coding and representation of the third levels.

Fig. 3 - Class 100 - Legal and Regulatory Frameworks
100.10 – Preparation of legal and regulatory diplomas and technical standards



The understanding of the conceptual map paved the path for codification upon three basic rules that explain the structure of the classification plan:

1. divide 999 by the number of branches obtained in the subdivision of each function and sub-function (999/x);
2. round up in the hundreds;

3. begin the first branch in 001 and the following in 100, 200, 300, etc. depending on the number of branches.

One of the main achievements of the project can be considered to be the creation of different tools that define a new system of information classification in the Portuguese public administration (Meta-information for Interoperability, Functional Macrostructure and Classification Plan), based on a functional structure and a approach to business processes. These promote semantic interoperability and are essential for the organization, representation, retrieval and management of information within the framework of e-government services that reflect European and national directives for interoperability. The research endeavored by the project has the potential to benefit the entire public administration in its several levels: central, regional and local. The classification plan already includes a Consolidated List with more than a thousand business processes that is managed by DGLAB, the body that coordinates the national archival policy. It is responsible for codification, which offers the various Portuguese public administration bodies a set of advantages, such as:

- The production and use of a single classification tool at the disposal of public administration for the classification of organizational information, leading to an economy in resources;
- The availability of a standardized functional classification plan, which is particularly significant when considering the vast number of bodies that have none;
- Simplification in when preparing other information management tools, such as preservation plans;
- Assistance in appraisal and selection of archival information;
- Contribution to the development of projects in business processes' reengineering (Millar, L.; Roper, M. & Stewart, K., 1999, 6);
- Improvements in the efficiency and effectiveness of public administration;
- Optimization in the management of internal resources by each body in public administration;
- Improvements in the internal and external mobility of resources;
- Support and anticipate decision making;
- Enhancement of horizontal and vertical interinstitutional communication.
- Assistance to bodies undergoing restructuration regarding the permanent management of information;
- Promotion of information reuse;
- The possibility of integration with performance metrics.

Conclusion: from the functional classification plan to the creation of an ontology

Overall, regardless of the need of improvement, both the Functional Macro-Structure (MEF-AP) and the Information's Classification Plan for Portuguese Public Administration (PCI-AP) contribute significantly to the emergence of a new paradigm concerning the management of archival information and documentation within the framework of public administration. In this new paradigm, functions are matched with business processes, both transversely and supra-institutionally. The public administration bodies are posited as open systems, according to the analytic paradigm (von Bertalanffy, 1973; Crubellate, 2007, 201). Likewise, an organization is considered to be an open system, in line with the phenomenological school (Gherardi and Nicoli, 2003) and with the Organizational Theory (Scott, 1992).

Simultaneously, the public administration and, more specifically, the archival community, also gain a new standardized tool for information management that is useful for the classification, appraisal and selection of information. It is also currently being developed an ontology based on WOL (Web Ontology Language), a language for knowledge representation, which has been proposed by W3C as a 'standard' to codify ontologies from the semantic web's perspective.

The paradigm suggested by Portugal represents a new approach not only in the organization, representation and retrieval of information/knowledge, but also in the management of archival information, making it a matrix model that links functions to business processes. Such change definitively places the manager of the information system at the elaboration, planning and development of the information system, granting him a leading role in the organizational management centered around the asset information, perceived as an object, process and product.

References

Bak, Greg (2010). La clasificación de documentos electrónicos: documentando relaciones entre documentos. *Tabula*. 13: 59-78.

Barros, Thiago (2012). A classificação funcional em Arquivística: uma análise da colaboração científica nos periódicos *Archivaria* e *American Archivist* = The functional classification in Archival Science: an analysis of the scientific collaboration in the Journals *Archivaria*and *American Archivist*.In *20 Años del Capítulo Español de ISKO. Actas del X Congreso ISKO Capítulo Español (Ferrol, 2011)*. Held at Universidade da Coruña, España. Pp. 157-69.

Crubellate, João Marcelo (2007). Três contribuições conceituais neofuncionalistas à teoria institucional em organizações. *RAC: Revista de Administração Contemporânea*, 199-222.

Dahlberg, Ingetraut (1978). Teoria do conceito [online]. *Ciência da Informação*. Rio de Janeiro, Instituto Brasileiro de Informação em Ciência e Tecnologia, 7(2): 101-107. [http://revista.ibict.br/ciinf/index.php/ciinf/article/viewFile/1680/1286] Duranti, Luciana (1998). *Diplomatics. New uses for an Old Science*. Lahnam and London: The Society of American Archivists and Association of Canadian Archivists in association with The Scarecrow Press.

Duranti, Luciana (1997). The archival bond. *Archives and Museum Informatics*. 11: 213-8.

Foscarini, Fiorella (2010). La clasificación de documentos basada en funciones: comparación de la teoría y la práctica. *Tabula*, 13: 41-57.

Gherardi, Silvia & Nicolini, Davide (2003). The sociological foundations of organizational learning. In *Handbook of organizational learning & knowledge*. Oxford: Oxford Universityu Press. Pp. 35-60.

Hjørland, Birger (2012).Is classification necessary after Google?In *20 Años del Capítulo Español de ISKO. Actas del X Congreso ISKO Capítulo Español.* Held at Universidade da Coruña, España. Pp. 19-30.

Millar, Laura, Roper, Michael & Stewart, Kelly (1999). *Glossary* [online]London, International Records Management Trust.

Penteado, Pedro, Lourenço, Alexandra & Henriques, Cecília (2013). *Macroestrutura Funcional (MEF)* [em linha]. Versão 2.0. Lisboa, Direção Geral do Livro, dos arquivos e das Bibliotecas, 2013.

San Segundo, Rosa & Martínez-Ávila, Daniel (2012). El orden de los saberes y la organización digital = The order of knowledge and the digital organization. In *20 Años del Capítulo Español de ISKO. Actas del X Congreso ISKO Capítulo Español.* Held at Universidade da Coruña, España. Pp. 413-21.

Schellenberg, Theodore R. (1956). *Modern Archives: Principles and techniques*. Chicago: University of Chicago Press.

Scott, W. Richard (1992). *Organisations: rational, natural and open systems*. 3rd ed. Thousand Oaks: SAGE.

Silva, Carlos Guardado da (2015). *Sumário pormenorizado da lição para obtenção do título de agregado*. Coimbra: Universidade de Coimbra.

Silva, Carlos Guardado da (2013). A classificação da informação arquivística da administração local nos países ibéricos: uma análise comparada. [*pen disk*]. In *Jornadas Ibéricas de Arquivos Municipais: Políticas, sistemas e instrumentos*. Held at Câmara Municipal de Lisboa, June 4-5, 2013.

Simões, Maria Graça & Freitas, Maria Cristina Vieira (2013). A classificação em arquivos e em bibliotecas à luz da teoria da classificação: pontos de convergência e de divergência. *PontodeAcesso*, 7(1): 81-115.

Von Bertalanffy, Ludwig (1973). *General system theory: foundations, development, applications*. Harmondsworth: Penguin.

Benildes C. M. S. Maculan, Gercina A. B. Oliveira Lima and Elaine D. Oliveira

Conversion Methods from Thesaurus to Ontologies: A Review

Abstract

This article presents the results of a literature review covering the use of the structured knowledge contained in thesauri for conversion into domain ontologies. These conversions can be semiautomatic or created through intellectual labor. Sixteen different models are presented; each includes a brief description of the model and the results. In the end, general considerations about the results are presented.

1 Introduction

Knowledge organization systems (KOS's) are used to produce more semantic support for information retrieval systems (IRS's). The intention is to minimize two basic deficiencies: (1) an IRS is not able to decode the user's needs, and (2) an IRS cannot interpret the content of the documents. This is because the keywords used to represent documents or the user's search may have different meanings, and without attributes and semantic links that represent the data domain's knowledge, it is impossible for the machine to interpret its meaning. Without these semantic properties, an IRS can only measure the similarity between the words used in a document with those of the user query. But this is not enough for the retrieval of results relevant to the users.

Thesauri are able to represent the knowledge of a domain through a set of concepts and have semantic properties identifying relationships between them: equivalence (USE and Used For: UF), hierarchical (Broader Term: BT and Narrower Term: NT), and associations (Related Term: RT). Traditionally, the number of relationships can be considered limited, sometimes causing ambiguous relationships between concepts. As a result, these ambiguous relations can lead to problems, especially in the digital environment, when there is the need for the use of intelligent applications that require inferences. In this context, it is important that the representation of connections between concepts have richer semantics, i.e. explicit and formalized. This feature is a property of ontologies.

In information science, studies about ontologies have been receiving a lot of attention since the 1990s. Ontologies are tools that assign greater formality in information representation and allow the inference of intelligent applications. Domain ontologies represent the explicit knowledge of the particular domain, intelligible in machine languages, with an unlimited number of relationships between concepts. Thus, they become an instrument with very rich semantic properties, minimizing problems of ambiguous relationships between concepts and enabling the reuse of information across different systems.

Due to its complexity, the construction of domain ontologies is a time-consuming job. To reduce development time, there are studies that have developed methods for the reuse of representations of the knowledge of various fields that already exist in pre-structured thesauri into domain ontologies. From this perspective, this article presents a literature review of the methodologies for the reengineering of thesauri, converting them into domain ontologies. It one part of the results of Maculan's (2015) thesis,

which applied the model of Soergel et al. (2004) and Lauser et al. (2006), which were developed for the semiautomatic conversion of the AGROVOC thesaurus into an ontology, including improvements at the semantic and syntactic levels.

2 Methodology

A literature search [1] was conducted in the area of information science to explore studies for converting thesauri into ontologies. The literature review was performed with a temporal cut from publications between 1991 (when the theme of ontologies began to be more frequently discussed) and July 2013. As a result of a bibliographical research, publications were picked up only from the year 2000 on. These publications focused on the use of different tools for the enrichment of terminology or its conversion into ontologies, such as thesauri, glossaries, dictionaries, bibliographic classification systems and folksonomies.

In addition, it was necessary to take a thematic approach, selecting works that dealt specifically with converting thesauri into ontologies, using semiautomatic or intellectual processes. Studies were excluded that covered 1) only the terminology reuse of thesauri, with no reuse of the conceptual framework; 2) only automatic thesaurus conversion into ontologies; and 3) only the assessment of the conceptual structure of the thesaurus for its conversion into ontologies, without actually applying a methodology. At the end, the selection resulted in sixteen works, with only one Brazilian work.

The articles were described using the following parameters: a) purpose of the study; b) conversion type: automatic, semiautomatic or intellectual; c) description of the model and of the conversion procedures; d) use of the relationships represented in the thesaurus and the explicit semantics; and e) results of the study. In the end, general considerations about the results of the described studies are presented.

3 Thesauri conversion into ontologies

In this paper, the Brazilian study is described first and then the remaining fifteen foreign articles are delineated in chronological order of publication.

The Brazilian study, Campos et al. (2008), present a semiautomatic conversion of the Thesaurus of Folklore and Brazilian Popular Culture into a domain ontology. The objectives were to verify if the existing relationships gain expression in the refinement and if new associations between terms, using their settings, could be discovered. The method includes the: 1) establishment of classes; 2) creation of hierarchical and associative relationships; 3) creation of partitive relations; 4) creation of inverse relationships; and 5) organization of non-hierarchical classes. The results demonstrated ease in modeling hierarchical relations. For associative relationships, it was necessary to create a property called “isAssociated” to connect the classes. This was mainly due to lack of clarity in the definition of the type of relationship. A more generic class was created, <partOf>, to group non-existent terms in the original structure. The ontology allowed inferences about the domain, which facilitated the search and enhanced the visualization of the relationships between concepts.

Qin and Paling (2001) develop a method for semiautomatic conversion of the Portal Thesaurus for Educational Materials (GEM) in an ontology. The method includes: 1) a detailed description of the objects; 2) allocating more refined semantics for classes,

subclasses and relations; 3) expression of concepts and relationships in Ontolanguage; and 4) reuse of knowledge in heterogeneous systems. Eight new classes were restructured in the thesaurus. The results showed that the printed Ontolanguage impressed a greater formality in relationships and that the records were more easily manipulated by the recovery mechanisms.

Wielinga et al. (2001) present a model for the semiautomatic conversion of the Art and Architecture Thesaurus (AAT) into a domain ontology, with the thematic focus "antique furniture" added to the existing knowledge capture on art objects. The subset is depicted as a general class and the set of descriptors has been described as standard slot. Slots accounted for the properties in RDFS and qualifiers represented sub-properties, considering only the relationship "subClassOf". The method includes: 1) building a model description; 2) binding properties of movable objects to specific subsets of the thesaurus; 3) adding an additional knowledge description of the area; and 4) expansion of relationships for fields not yet covered by the thesaurus. In the end, a lightweight [2] ontology, consisting of classes, attributes, and relationships was produced.

Hahn and Schulz (2003) and Hahn (2003) present a method for semiautomatic conversion of the Unified Medical Language System thesaurus - the UMLS Metathesaurus - into an ontology of anatomy and medical conditions. The method includes: 1) automatic generation settings; 2) automatic verification of the integrity of the hierarchies (with the classifier LOOM); 3) the manual removal of inconsistencies; and 4) the manual refinement of the knowledge base. There is also an automatic conversion of the relationships, using: <partOf/hasPart>, <isA>, <siblingOf> and <associativeWith>. It was relatively simple to restore the consistency of the thesaurus knowledge base. However, it was almost impossible to reach a high degree of adequacy and scope, due to the amount of intellectual work required. As a result, it was possible to extract the conceptual knowledge of the thesaurus, allowing its conversion into a formal logical description in LOOM, with classes and relationships, and the allocation of greater semantic meaning to the biomedical specialty area.

Goldbeck et al. (2003) propose the semiautomatic conversion of the thesaurus of the National Cancer Institute, NCI Thesaurus, in the field of bioinformatics, into an ontology in OWL language. The method includes: 1) a list of editing concepts and relationships; 2) the export of the lists to a content manager; 3) identification of issues by different modelers; 4) returning the lists issued by modelers to the manager; 5) analysis of issues, fixes, and validation; 6) weekly release to the modelers of updates, validating and consolidating changes; 7) release of the list of candidates under the concepts for automated testing, to identify and resolve conflicts and consistency problems; and 8) publication of the final version. The results allowed the mapping and defining of classes and properties in the ontology, facilitating user access to information.

Van Assem et al. (2004; 2006) describe a method for semiautomatic conversion of thesauri for the RDFS and OWL formats of an ontology using the Medical Subject Headings (MeSH) and WordNet [3] thesauri. The method includes: 1) thesaurus analysis and identification of standards for the allocation of relations; 2) syntactic conversion, RDFS; 3) semantic translation (class expansion and properties) using constraints of RDFS and OWL; and 4) adoption of international standard (SKOS or

other) in order to facilitate cross-language interoperability. The results showed the clarification of existing relationships and the creation of new relationships: <subClassOf> and ad hoc. A lightweight ontology was created in the RDFS/OWL language, containing classes, attributes, and relationships.

Soualmia, Golbreich, and Darmoni (2004) published a method for the first phase (definition of modeling principles) of the MeSH thesaurus conversion (in French, CISMeF [4] project) into a formal ontology in OWL-DL [5]. The model, based mainly in syntactical conversions, includes: 1) formal terminology in OWL; 2) an import of an ontology into theOWL editor for Protégé with the verification of classes and their consistency in the use of an inference engine; 3) distinction of the hierarchical relations <isA> and <isPartOf>; 4) concepts of distinction (specialty, descriptor and qualifier); 5) assessment of the characteristics and other domain attributes to support automatic conversion; and 6) translation of qualifier properties and application restrictions (for consistency). As a result, a partially heavy-weight ontology was created, and the authors intended, in the second phase, to include the definition of all resources. Since 2008, a multi-universe terminology is used to index resources.

Chrisment et al. (2006) present a method (TERMINAE) that converted the semi-automatic International Astronomical Union (IAU) thesaurus into a domain ontology. There was the enrichment and updating of the existing terminology from the literature of the discipline. The method includes: 1) specification of requirements; 2) specification of concepts and terms (and lexical variants); 3) grouping of terms in larger classes; 4) determination of hierarchical and associative relationships; 5) formalization of the ontology into a language interpretable by machine; and 6) validation by domain experts. The results showed that the use of the thesaurus allowed the automation of step 3 into a faster printing process. A lightweight ontology was created in the OWL language.

Hepp and Bruijn (2007) present a method (GenTax) for the semiautomatic conversion of SOCs (bibliographic classification systems, thesauri, and taxonomies) into ontologies in OWL and RDFS, supported by SKOS. The method includes: 1) informal specification of hierarchies; 2) determination of the context; 3) intellectual property verification, anomalies, and creation ofconversion script; and 4) generation of the ontology. Two classes in each category were created: a class of generic concepts (context) and an other class of broader taxonomic concepts (preservation of the original hierarchy). There were not instances or additional information mappings, which generated little semantics and few axioms. As a result, lightweight ontologies were created (in RDFS and OWL-DL) based only on the relations of <rdfs: subclasseDe> (subclass of) type.

Hyvönen et al. (2008) publish a method for the semiautomatic conversion of the FinnishGeneral Thesaurus(YSA) into an ontology. The method includes: 1) syntactic conversion (OWL and SKOS); 2) definition of the structure (with DOLCE [6] principles), with hierarchical subclasses in three levels: abstract, tough, and persistent, avoiding multiple inheritance; 3) disambiguation of BT and NT relations; 4) reorganization of concepts and allocation of transitivity, particularly in BT/NT relations; 5) disambiguation and definition of concepts; and 6) alignment of the ontology. In the end, the conversion was in the syntactic and semantic levels and the enrichment of relations in the ontology (classes, attributes, and relationships). There

was the clarification of the relations <subclassOf> and <partOf>, generating a heavy [6] ontology.

Villazón-Terrazas, Suárez-Figueroa, and Gómez-Pérez (2009) submit a semiautomatic method of converting thesauri into ontologies. It was applied in the *European Training Thesaurus* (*ETT*) and includes: 1) conversion of terms in instances of metaclasses; 2) identification of concepts without superordination and adding relationships; 3) transformation of BT and NT relations <rdf: type>, <subclassOf> or <rdfs: partOf>; 4) inclusion of the terms identified in step 2 in existing classes or new classes; 5) reorganization of the hierarchy; 6) identification of equivalent terms; and 7) application of specific relationships: <synonymOrEquivalent>. As a result, a lightweight ontology is created, which can be implemented in any language.

Villazón-Terrazas (2011) present methods for automatic conversion or semiautomatic conversion of non-ontological resources (bibliographic classification systems, thesauri, and dictionaries) into domain ontologies. For semiautomatic conversion of thesauri into ontologies, themethod includes two distinct procedures (guidelines, activities, and tasks): 1) a polynomial TBox transformation (description of concepts, roles, and properties, with an intensional approach), using WordNet to explain the associative relationships; 2) a linear ABox transformation (assignment of specific properties for the domain, from the extensional knowledge), with some changes: descriptors in classes and instances; hierarchical relationships in <subclassOf>; equivalences and labels. It is possible to produce a set of conversion standards, allowing the creation of a network of ontologies. The results showed gains in building ontologies, mainly as a guide for novice modelers.

Kless et al. (2012) propose a method to convert the semiautomatic AGROVOC thesaurus, with the thematic focus of agricultural fertilizers, into a domain ontology. The method includes: 1) initial verification and refinement; 2) syntactic conversion; 3) criteria for terminology inclusion (core modeling); 4) choice and alignment of superordinate classes and formal relations; 5) formal specification classes (core modeling); 6) elimination of poly-hierarchies; 7) dissociation of independent entities; and 8) normalization classes and adjustments to the labeling of properties. The results indicate the need for more than one syntactic conversion and the use of formal language; additionally, steps 3, 4, 5, and 7 require technology that is not yet available.

Ping and Yong (2012) present a semiautomatic conversion of a thesaurus into a domain ontology, using cereal crops as the parameter. The descriptive logic the method was based on includes: 1) purpose definition; 2) selection of the terminology of the field; 3) description of properties of the concepts; 4) addition of instances; 5) determination of the relations (equivalence: <equivalenceClass>; hierarchical: <subclassOf> and associative: <hasAssociativeRelation>. The results showed an ontology with more semantics; however, it indicated a need for a greater intellectual intervention of domain modelers and experts.

Sanaa et al. (2013) proposed the intellectual conversion of the *Thesaurus of Tourism and Recreation* into a domain ontology. The method includes: 1) change of the terms into concepts (with synonyms, lexical variations, and settings); 2) a creation ratio of equivalence between preferred and non-preferred descriptors; 3) determination of the non-preferred descriptors as labels; 4) structuring of the concepts (class and subclass hierarchies) and determination of relations <isGenderOf><isSpeciesOf>, <isPartOf>

and <isInstanceOf>; 5) poly-hierarchies management (test “some/all”); 6) redundancy removal; 7) exploration and explanation of associative relationships (literary warrant); and 8) enrichment of terminology. The results showed that the thesaurus allowed the enrichment of the conceptual structure of the ontology (which is still a prototype).

This review demonstrates that the international literature since 2000 offers studies on different methods for converting thesauri into ontologies, using semiautomatic or fully intellectual methods in the remodeling process. Some of the models were restricted to the syntactic conversion of thesauri (Van Assem et al., 2004; Soualmia, Golbreich, & Darmoni, 2004), resulting in lightweight ontologies. In the semantic and syntactic levels, Soergel et al. (2004) and Lauser et al. (2006) present a semiautomatic approach that produces one ontology in OWL-DL compose with classes, attributes, and relations components. They propose the rules-as-you-go (rules for semantic refinement on the thesaurus) approach to build an inventory of patterns for the agricultural domain. It was supposed that the rules and patterns are identified through intellectual work and the refinements will occur automatically.

But more recent models (Kless et al., 2012; Sanaa et al., 2013) are most complete and noteworthy works because create heavy ontologies. They also address solutions to avoid and prevent problems when integrating the knowledge from the thesauri into the ontologies. It was clear that in all models present in this paper were more expressive semantic in the relationships, which always brought gains in information retrieval.

It should be emphasized that the re-engineering of the thesauri and the formalization of the concept definitions, with the goal of conversion into a domain ontology, requires the use of international norms and standards in the creation of the thesaurus.

4 Final considerations

The similarities and differences between thesauri and ontologies have been described by different authors (Moreira, 2003; Sales & Café, 2009; Currás, 2010; Kless, 2012) in an attempt to determine the nature and identity of each of these instruments. In these studies, it is recognized that the theoretical grounds of the thesauri construction are already consolidated in IS. Therefore, the use of thesauri in creating ontologies is justified because “they can serve as theoretical substrates for building ontologies” (Boccato, Ramalho, & Fujita, 2008, p. 207). Thus, they can be considered a potential knowledge base to be reused in the creation of domain ontologies.

Villazón-Terrazas (2011) corroborate this statement with the assurance that the research field of thesauri conversion into ontologies has been strengthened since 2007, a result of the adoption of a new paradigm in building ontologies, which is the reuse of already-structured knowledge. In this perspective, a contribution of this review is to describe different models of reusing and converting already-structured knowledge from thesauri into ontologies, which can be adapted to different contexts and goals.

Notes

[1] The search was done in the Portal Digital Library of Scientific Journals (CAPES), Annual Review of Information Science and Technology, Journal of the American Society for Information Science and Technology;Knowledge Organization, Advances in Knowledge Organization, the meta-search engine of Hannover University, Educational Resources

Information Center, E-Prints in Library and Information Science, Referential Database of Scientific Article and Journals in Information Science, and Google Scholar.

[2] Soft or lightweight ontologies include concepts, taxonomies of concepts, relationships between concepts, and properties that describe the concepts.

[3] WordNet is considered a dictionary that is based on the principles of Roget's dictionary.

[4] CISMeF (Catalog and Index of French-speaking resources) was a project originally initiated by Rouen University Hospital (RUH). CISMeF began in February 1995 and in 2000 created a generic searchtool using the CISMeF semi-informal ontology. The CISMeF terminology encapsulates the French MeSH thesaurus. The objective of CISMeF is to assist the healthprofessional and lay people during the search for electronic information available on the Internet. Is important to say that National Library of Medicine released the initial beta version of MeSH RDF on November 17, 2014.The product was a true beta version and not finalized. In addition to MeSH thesaurus, the concepts of metaterms (a medical specialty or a biological science) and resource types (a generalization of the Medline publication types) were added.

[5] A OWL-DL: a language with a medium level of expressiveness, based on logical description, a fragment of first order logic, capable of automatic reasoning (Soualmia, Golbreich & Darmoni, 2004).

[6] DOLCE (*Descriptive Ontology for Linguistic and Cognitive Engineering*) is an ontology with the purpose of effective cooperation and the establishment of consensus.

[7] Heavy ontologies include axioms and restrictions.

References

Boccato, Vera Regina Casari, Ramalho, Rogério Aparecido Sá & Fujita, Mariângela S. L. (2008). A contribuição dos tesauros na construção de ontologias como instrumento de organização e recuperação da informação em ambientes digitais. *Ibersid*, 2: 199-209.

Campos, Maria Luiza Machado et al. (2008). O uso de tesauro como base terminológica para a elaboração de ontologias de domínio: uma experiência com o domínio do Folclore e Cultura Popular. *In Encontro Nacional de Pesquisa em Ciência da Informação.* São Paulo: USP.

Chrisment, Claude et al. (2006). D'un thesaurus vers une ontologie de domaine pourl'exploration dun corpus. *AMETIST,* INIST, 59-92.

Currás, Emilia (2010). *Ontologias, taxonomia e tesauros em teoria de sistemas e sistemática*. Brasília: Thesaurus.

Goldbeck, Jenifer et al. (2003). The National Cancer Institute's Thesaurus and Ontology. *Journal of Web Semantics*, 1(1).

Hahn, Udo (2003). Turning informal thesauri into formal ontologies: A feasibility study on biomedical knowledge re-use. *Comparative Functional Genomics*, 4(1): 94-7.

Hahn, Udo & Schulz, Stephan (2003). Towards a broad-coverage biomedical ontology based on description logics. *Pac Symp Biocomput*, 577-88.

Hepp, Matin & Bruijn, Jos de (2007). GenTax: A generic methodology for deriving OWL and RDF-S ontologies from hierarchical classifications, thesauri, and inconsistent taxonomies. In *The Semantic Web: European Semantic Web Conference*, Springer, 2007, Innsbruck, Austria. Innsbruck, Austria: ESWC. v. 4519, pp. 129-44.

Hyvönen, Eero et al. (2008). Building national semantic web ontology and ontology service infrastructure: The FinnONTO approach. In *European Semantic Web Conference*, June 1-5, Tenerife, Espanha. Tenerife: ESWC. Pp. 95-109.

Kless, D. (2012). Quality ontology engineering based on thesaurus. In *3rd International Conference on Biomedical Ontology*, Graz, Austria, July 21-25. Graz: ICBO.

Kless, Daniel et al. (2012). A method of re-engineering a thesaurus into an ontology. In *Formal Ontology in Information Systems. International Conference* - FOIS, 2012, Amsterdam. Amsterdam: IOS Press. pp. 133-46.

Lauser, Boris et al. (2006). From AGROVOC to the Agricultural Ontology Service: Concept Server, an OWL model for creating ontologies in the agricultural domain. In *International Conference on Dublin Core and Metadata Applications*, 2006, Colima. México: DCMI.

Maculan, Benildes C.M.S. (2015). Estudo e aplicação de metodologia para reengenharia detesauro: remodelagem do THESAGRO. *PhD Thesis.* Belo Horizonte: Universidade Federal de Minas Gerais.

Moreira, Alexandra (2003). Tesauros e ontologias: estudo de definições presentes na literatura das áreas das Ciências da Computação e da Informação, utilizando-se o método analítico-sintético. *Master degree thesis.* Belo Horizonte: Universidade Federal de Minas Gerais.

Ping, Li & Yong, Li (2012). On transformation from the thesaurus into domain ontology. In *International Conference on Computer and Information Application*, 8-9 December 2012, Taiyuan, China. Taiyuan, China: ICCIA.

Qin, Jiang & Paling, Stephen (2001). Converting a controlled vocabulary into an ontology: The case of GEM. *Information Research*, 6(2).

Sales, Rodrigo & Café, Ligia (2009). Diferenças entre Tesauros e Ontologias. *Perspectivas em Ciência da Informação*, Belo Horizonte, 14(1): 17-98.

Sanaa, Mouhim et al. (2013). A methodological approach for converting thesaurus to domain ontology: Application to tourism. *International Journal of Engineering and Innovative Technology* (IJEIT), 3(5).

Soergel, Dagobert et al. (2004). Reengineering thesauri for new applications: The AGROVOC example. *Journal of Digital Information*, 4(4).

Soualmia, Lina F., Goldbreich, Christine & Darmoni, Stephan J. (2004). Representing the MESH in OWL: Towards a semi-automatic migration. In *International Workshop on Formal Biomedical Knowledge Representation*, June 2004, Whistler, Canada. Whistler, Canada: KR-MED. Pp. 81-7.

Van Assem, Market al.(2004). A method for converting thesauri to RDF/OWL. In *International Semantic Web Conference*, November 7-11, 2004, Hiroshima, Japan. Hiroshima, Japan: ISWC2004.

Van Assem, Mark et al. (2006). A method to convert thesauri to SKOS. In *European Semantic Web Conference*, 11-14 June, 2006, Budva, Montenegro. Budva, Montenegro: ESWC'06. v. 4011, pp. 95-109.

Villazón-Terrazas, Boris Marcelo (2011). Method for Reusing and Re-engineering Non-ontological Resources for Building Ontologies. *PhD Thesis.* Madrid: Facultad de Informática, Universidade Politécnica de Madrid.

Villazón-Terrazas, Boris Marcelo, Suárez-Figueroa, Mari Carmen & Gómez-Pérez, Assunción (2009). A pattern for re-engineering a term-based thesaurus, which follows the record-based model, to a lightweight ontology. In *Workshop on Ontology Patterns*, October 25, 2009, Washington: WOP.

Wielinga, B. J. et al. (2001). From thesaurus to ontology. In *International Conference on Knowledge Capture*, 2001, Victoria. Victoria, Canada: ICKC. Pp. 194-201.

G. Arave and Elin K. Jacob

Evaluating Semantic Interoperability across Ontologies

Abstract

A defining characteristic of the Semantic Web is the ability of software agents to draw inferences by combining information gathered across multiple resources, which depends on the *semantic interoperability* of those resources: on the extent to which the conceptual definitions of the elements that make up those resources can be related to one another. With the proliferation, both within and across disciplines, of the ontologies that provide those definitions, there is increasing need for metrics to quantify the extent to which individual ontologies share conceptualizations. Because the structure of an ontology is a network, network analysis techniques can be applied to analyze and quantify both the internal and external relationships of the elements that comprise an ontology.

Knowledge organization (KO) relies on formal systems of representation, but these traditional systems are being challenged by the changing technological landscape and the expanded demands for representation in an environment of linked digital resources.Giunchiglia, Dutta and Maltese (2014) argue that traditional KO systems have limited expressivity and that ontologies offer the power and flexibility necessary to meet the demands of knowledge representation in the digital environment. Ohly (2013) sees ontologies as part of "the new knowledge organization" (p. 807), and Herre (2013) notes the increasing prevalence of ontology research in areas as diverse as e-commerce, information integration, natural language processing, and knowledge engineering.

With the proliferation of ontologies in a wide range of domains, there is increasing need for evaluation metrics that can assess the utility of competing ontologies (Vrandečić, 2009). Gomez-Perez (2004) argues that the evaluation of ontologies consists of two distinct activities: verification activities that assess the internal coherence of an ontology and validation activities that analyze the alignment of an ontology with the particular conceptualization of the domain it represents. However, while verification and validation may provide insights regarding an ontology's internal structure of relationships and its representational validity, they do not address interoperability between ontologies.

The concept of interoperability has various interpretations ranging from the ability of systems to interface at a physical level to the capacity for exchanging information between systems with minimal loss of semantic meaning. Several authors have posited that interoperability can be divided into various "flavors" (Miller, 2000) or "levels" (Tolk, 2006). Ouksel and Sheth (1999) present a classification of interoperability consisting of system interoperability, syntactic interoperability, structural interoperability and semantic interoperability, the latter supporting "context-sensitive information requests over heterogeneous information resources [while] hiding system, syntax, and structural heterogeneity" (p. 6). Semantic interoperability is the higher-

level form of interoperability reflecting the ability of two or more systems to exchange data without loss of the original semantics.

Semantic interoperability is at the heart of the semantic web envisioned by Berners-Lee, Hendler and Lassila (2001) and is thus part of the new KO proposed by Herre (2013). However, no metrics have been proposed for assessing semantic interoperability across ontologies. Semantic interoperability is relative to the ontologies involved: The semantic interoperability between ontologies A and B can be better or worse than the semantic interoperability between ontologies A and C. For this reason, it is important to be able to quantify semantic interoperability in order to compare one ontology with another or to evaluate successive iterations of the same ontology.

Although there are various, sometimes conflicting definitions of *ontology* (Guarino & Giaretta, 1995), an ontology is functionally understood to be a machine-readable metadata schema that defines categories of entities (i.e., classes) that are of interest in some domain, the attributes (i.e., properties) of those entities, and the relationships that inhere among entities and attributes (Gruber, 1995; Guerrini & Possemato, 2013).In the digital environment, ontologies are tools used to represent resources and data in order to facilitate machine processing of digital information and to enable semantic web technologies (Staab & Studer, 2009).Because an ontology is a partial conceptualization (Gruber, 1993; Guarino & Giaretta, 1995) of a particular worldview (Francq, 2011), an ontology is selective in what it includes; and, because multiple worldviews of a domain are possible, multiple ontologies may actually represent the same domain from different perspectives. Furthermore, domains often overlap, and ontologies representing different domains may define the same entities, attributes or relationships.If each ontology were deployed in a self-contained application, overlapping conceptualizations would not be problematic. However, in the context of the semantic web, being self-contained is counterproductive, and overlapping conceptualizations can lead to semantic ambiguity across ontologies.

The vision of the semantic web is one in which both the information resources on the web and the "things" to which they refer, whether in cyberspace or real space, are defined and described in relation to one another such that a software agent can interpret and act on the relationships between "things." These descriptions and definitions are provided in machine-readable records that rely on the elements defined in one or more ontologies. These ontologies, in turn, contain assertions about entities, attributes and relationships, which accounts, in part, for their usefulness. For example, an ontology may define the class *lion* as a subordinate or "kind" of the class *cat*, a semantically meaningful relationship that can be processed by a software agent.The real power of ontologies, however, is in using their logical structure to infer knowledge that is not explicitly stated. If an agent has the ontological knowledge that a *lion* is a kind of *cat* and then encounters an assertion that *Sid* is a *lion*, the agent can infer that *Sid* is a kind of *cat*. This simple example demonstrates the inferential power of ontologies on the

semantic web. Indeed, the semantic web is so reliant on the ability to make inferences that Allemang and Hendler (2011) refer to it as "an inference-based system" (p. 117).

The ability of software agents to make inferences depends on the semantics encoded in the ontology, and those semantics depend on the logical structure of the ontology itself. The inference that *Sid* is a kind of *cat* is only possible because of the hierarchical relationship between the classes *lion* and *cat* that is defined by the ontology. While the definition of an element in the ontology will generally include a human-readable description of that element, a software agent must rely on the semantics represented as structural relationships among elements. This structure of relationships constitutes a network that can be represented as "a semantic graph of concepts and relations" (Alani & Brewster, 2005, p. 52).

The extent to which an agent can make inferences across the heterogeneous resources that make up the semantic web depends upon the extent to which elements used in those resources can be linked. While *ex post facto* methods such as crosswalks can relate elements in one schema to those in another, semantic interoperability can be built into an ontology by co-opting elements from existing ontologies. This approach is recommended by Heath and Bizer (2011) and by Coyle (2012), who argues that "reuse is preferred to the creation of new, redundant terms" (p. 16). Co-opting can take the form of borrowing an element wholesale from another ontology or of subordinating one element to another in an existing ontology (e.g., defining a class as a subclass of a class defined in another ontology). Linking one ontology to another by creating relationships between elements connects the graphs of the two ontologies, allowing an agent to process the elements in one ontology within the larger context provided by the combined graph of both ontologies. A single ontology may link to any number of other ontologies, which may, in turn, link to still other ontologies, allowing a software agent to traverse the entire interconnected network in order to process assertions in the original resource and make semantic inferences. This linking between ontologies is the fundamental concept underlying interoperability and linked data and is the basis for proposing a set of metrics for the evaluation of semantic interoperability.

In any ontology, it is likely that some elements will be defined internally (i.e., without reference to another ontology), while others will be defined externally (i.e., by creating relationships with other ontologies). An initial approach to understanding the semantic interoperability of an ontology is thus simply to ask what percentage of the elements are co-opted from another ontology. Obviously, if an ontology does not have external links, it cannot be said to be interoperable and the 0% calculated by this measure would support that conclusion. Conversely, if all elements were co-opted from other ontologies, 100% of the elements would be linked externally and the ontology would be completely interoperable, a conclusion supported by the value of this metric. Following this logic, if 50% of an ontology's elements were externally linked, then half of the *elements* would be considered interoperable. However, it would not be

appropriate to characterize two ontologies with 50% co-opted elements as equally interoperable. Between 0% and 100%, it matters *which* elements are linked to external ontologies since the linked elements might be peripheral or might be concentrated in one corner of the structure (e.g., properties related only to date). For this reason, it is important to have a way to assess the importance of the linked elements in the context of the ontology. Given that an ontology can be represented as a semantic graph, the analytic tools developed for evaluating social networks offer a potentially viable means for determining the semantic relevance of linked elements.

Wasserman and Faust (1994) note that "one of the primary uses of graph theory in social network analysis is the identification of the 'most important actors'" (p. 169) in a network. A number of metrics have been proposed for determining the relative importance of any given actor in a network and for determining its significance in the network as a whole. Each of these measures, collectively referred to as *centralitymeasures*, approaches the notion of importance from a slightly different perspective, providing a different understanding of an actor's relative significance in the network. The network itself is composed of a group of actors or *nodes* and the connections or relationships that exist between these nodes. Some kinds of relationships are directional (e.g., the passing of information) while others are non-directional (e.g., co-authorship). The links between nodes may represent any kind of relationship and are referred to as *edges* in the case of non-directional relationships or as *arcs* in the case of directional relationships. While some centrality measures can be calculated on either directional or non-directional networks, others can only be calculated for networks with edges rather than arcs. However, it is possible to convert a directed graph to a non-directed graph in order to perform certain calculations.

Ontologies are inherently directed networks, but it makes more sense to analyze them as non-directed networks for two reasons: 1) In order to usefully compare measures, they should all be performed on the same representation of the ontology; and 2) given the emphasis here on semantics and inference, a symmetrical relationship makes more sense. If an ontology asserts that a *rolltop* is a kind of *desk*, an agent can equally infer that the class of things known as *desk* includes a subclass of things called *rolltop*. In other words, neither the semantic relationship itself nor the inferential processes need to be unidirectional. Therefore, the following discussion of centrality measures assumes a non-directed network.

This paper focuses on the use of measures of *degree centrality*, *eigenvector centrality* and *betweenness centrality* in evaluating semantic interoperability. *Degree centrality* is perhaps the simplest and most intuitive of the three centrality measures discussed here. The degree of any particular node is simply the number of edges that impinge on it. A node with more edges is obviously more well-connected and can therefore be said to be more central and thus more important to the network. An obvious drawback to this measure is that it is limited by the number of nodes in the

network: In a network with three nodes, the maximum possible degree is 2 when self-loops are prohibited, whereas, in a network with 100 nodes, any given node could be connected to 99 other nodes. Degree centrality is a normalized index of node degree that takes the size of the network into account. It is calculated as the ratio of edges that exist to all possible edges. The denominator in this calculation normalizes the measure and makes measures from networks of different sizes comparable.

A drawback of degree centrality is that it considers each node individually, ignoring the network as a whole. *Eigenvector centrality* takes into account the degrees of all of the nodes to which a particular node is connected and is thus a more refined version of degree centrality (Borgatti, 1995). It uses an algorithm that weights the score of each node according to how well-connected its connections are. In other words, nodes connected to important nodes become important themselves.

Finally, *betweenness centrality* is based on tracing the paths in a network and rating each node on how frequently it lies on the path between other nodes. The assumption is that nodes that are frequently encountered while traversing the network must be central and, therefore, important. In the context of semantic interoperability, betweenness reflects the inheritance that is fundamental to ontological relationships (e.g., subclass and subproperty relationships) and expands understanding of which elements are linked outside the ontology. Rather than considering only direct relationships with other ontologies, this measure considers any elements that are linked to these relationships or to elements so linked, and so on. In other words, if an element lies on a path that leads outside the ontology, it can be considered linked outside the ontology. Thus creating the property *author* as a subproperty of Dublin Core *creator* asserts that *author* is a kind of *creator*. If the property *ghostAuthor* is defined as a subproperty of *author*, an agent can follow the chain of relationships from *ghostAuthor* to *author* to *creator* and infer that *ghostAuthor* is a kind of *creator*, thus semantically linking *ghostAuthor* to *creator* and increasing the betweenness score of the co-opted element *creator*.

The scores of these metrics for external elements can help to determine how much impact the external elements have on the semantic structure of the ontology. External elements with higher degree scores are, by definition, connected to more elements in the ontology, and these connections are therefore more significant for semantic interoperability. External elements with higher eigenvector centrality scores are connected to more central elements: Even if such an element has a low degree score, it may still be adding semantic interoperability by providing a bridge to *hubs,* i.e., elements with high degrees. External elements that lie between many elements are likely to add semantic interoperability to more elements, including hierarchies within the structure. By comparing the values for these measures, a researcher can begin to understand how external elements are operating within an ontology. Low degree coupled with high betweenness may indicate a high level element with many subordinates. High degree and low betweenness, on the other hand, would be seen with

a class that serves as the range for a number of properties but is otherwise not connected to the rest of the classes in the ontology. A low degree with a relatively high eigenvector score and relatively low betweenness might characterize a property that is not highly connected to other properties, but which shares its domain and range with a number of other properties. In short, these measures can aid in interpreting the impact of an ontology's relationships to external elements.

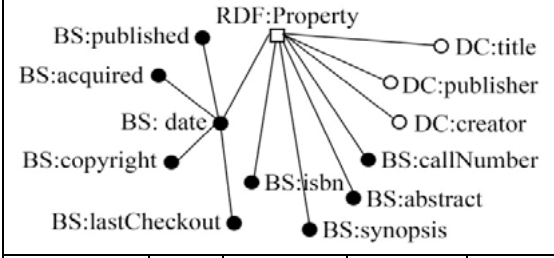
A single ontology may co-opt elements from many other ontologies, provided that each ontology is explicitly declared. Any measure of semantic interoperability must make a distinction between internal elements native to a particular ontology and external elements that have been co-opted from another ontology. If an ontology co-opts elements from only one other source, analysis will only consider the two groups of internal and external elements. If more than two ontologies are involved, the decision may be made to compare the impact of each ontology. There is one caveat in considering external elements, however. In order for the network to remain cohesive, which is important for some centrality measures, the graph must include element declarations such as rdf:type, which specify the class to which an element belongs, making all defined classes subclasses of rdfs:Class and all defined properties subsets of the class rdf:Property. Combined with other relationships, this makes an ontology into a single component while simultaneously presenting a more accurate representation of the ontology's semantics. While elements from the rdf and rdfs ontologies connect the network in semantically meaningful ways, an argument can be made to ignore them during analysis of semantic interoperability: rdf and rdfs elements are superordinate to most other elements in an ontology and are therefore central to the ontology, which will be reflected in any measure of centrality; but their importance is a foregone conclusion and does not add to an understanding of interoperability. In other words, because these elements provide the machinery by which elements are defined in an ontology, they are infrastructural and non-optional. Rather than making the ontology interoperable, these elements make the ontology possible.

Several simple examples illustrate how centrality measures can contribute to a measure of semantic interoperability. The Bibliographic Schema (BS) is a very simple ontology created for this purpose. It has nine internal elements, some of which are linked, and it co-opts three external elements from the Dublin Core Metadata Element Set (DC), none of which is linked to any other element in DC.Figure 1 shows the structure of the BS ontology and the values for degree, eigenvector and betweenness centrality for each element as well as a breakdown of these measures by internal and external elements. The rdf:Property element is included for the sake of comparison but is not considered in the analysis of internal and external relationships.

The BS:date element has the highest centrality for all measures and is the closest thing to a hub in this structure. The rest of the elements are fairly similar on all three measures, although the importance of the BS:date element does mean that its

subproperties gain a slight advantage in the eigenvector score. Because the DC properties are not directly linked to and do not interact with any of the internal elements, they are not expected to be important in terms of interoperability since, while reusing these elements is preferable to creating them anew, their semantics do not inform the BS ontology itself. Thus they net the same scores as internal elements that are not linked to any other element in the ontology. Looking at the ontology as a whole, 25% of the elements are external. All other things being equal, 25% of the nodes should have 25% of the value for the centrality measures. In this case, these nodes are less connected than would be expected although the difference is relatively small because of the low degree across the ontology. The external eigenvector is slightly higher than might be predicted because each of the DC elements is connected to the node with the highest degree in the ontology. Finally, the DC elements have no betweenness because they do not lie on any paths that include other elements. This analysis indicates that, while there is some semantic interoperability between these two ontologies, the DC elements are underperforming in terms of adding semantic value.

Figure 1. BS Schema Centrality Measures by Element and by Schema



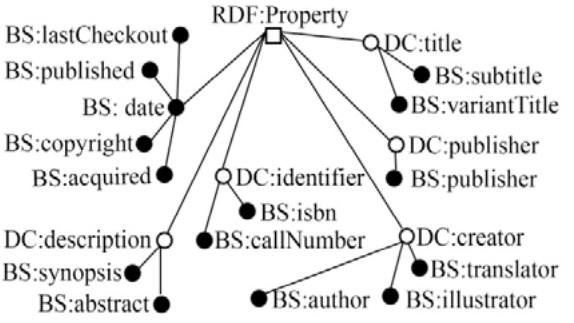
SCHEMA	*n*	DC	EC	BC
BS	9	1.08	1.82	38
DC	3	.25	.66	0
% INSIDE	.75	.81	.73	1
% OUTSIDE	.25	.19	.27	0

ELEMENT	DC	EC	BC
BS:abstract	.083	.22	0
BS:acquired	.083	.135	0
BS:callNumber	.083	.22	0
BS:copyright	.083	.135	0
BS:date	.417	.402	38
BS:isbn	.083	.22	0
BS:lastCheckout	.083	.135	0
BS:published	.083	.135	0
BS:synopsis	.083	.22	0
DC:creator	.083	.22	0
DC:publisher	.083	.22	0
DC:title	.083	.22	0
RDF:Property	.667	.653	56

Compare this with an improved version of BS (IBS) that links some elements externally by making IBS elements subproperties of DC properties, effectively making the DC properties an integral part of the IBS network. Ten IBS properties have been created as subproperties of five DC properties. Again ignoring the rdf:Property element, there are a total of 20 elements, 25% of which are external. With the external elements more integrated into the semantic structure, the DC elements should fare better on centrality measures, and this is the case both for individual elements and for the ontology as a whole. IBS:date still emerges at the top of the ranks, but the five DC elements are clearly more structurally important than the rest of the internal elements according to all three centrality measures. It is worth noting that the external elements that rank highest are those that provide a wider semantic context for the most elements.

Considering the revised ontology as a whole, the external elements are now netting a disproportionate amount of the total centrality scores even though the percentage of internal and external elements is the same as for the original BS/DC analysis.

Figure 2. IBS Schema Centrality Measures by Element and by Schema



SCHEMA	*n*	DC	EC	BC
IBS	15	0.95	1.839	70
DC	5	.75	1.328	184
% INSIDE	.75	.56	.58	.28
% OUTSIDE	.25	.44	.42	.72

ELEMENT	DC	EC	BC
IBS:abstract	.05	.09	0
IBS:acquired	.05	.13	0
IBS:author	.05	.11	0
IBS:callnumber	.05	.09	0
IBS:copyright	.05	.13	0
IBS:date	.25	.38	70
IBS:illustrator	.05	.11	0
IBS:isbn	.05	.09	0
IBS:lastCheckout	.05	.13	0
IBSpublished	.05	.13	0
IBS:publisher	.05	.08	0
IBS:subtitle	.05	.09	0
IBS:synopsis	.05	.09	0
IBS:translator	.05	.11	0
IBS:variantTitle	.05	.09	0
DC:creator	.2	.31	54
DC:description	.15	.26	37
DC:identifier	.15	.26	37
DC:publisher	.1	.23	19
DC:title	.15	.26	37
RDF:Property	.3	.59	164

These are simplified examples, but they do demonstrate that, when applied to a semantic network, centrality measures can reflect the importance to the entire representational structure of any concept or group of concepts. This is only the first step in finding a reliable metric for semantic interoperability and much remains to be done, including both understanding typical centrality patterns among various kinds of elements to determine if they should be weighted differently and determining how a measure should be normalized to make comparisons between ontologies accurate.

References

Alani, Harith & Brewster, Cristopher (2005). *Ontology ranking based on the analysis of concept structures*. In Paper presented at the Proceedings of the 3rd international conference on Knowledge capture, Banff, Alberta, Canada.

Allemang, Dean & Hendler, James (2011). *Semantic Web for the working ontologist: Effective modeling in RDFS and OWL* (2nd ed.). Waltham, MA: Morgan Kaufmann/Elsevier.

Berners-Lee, Tim, Hendler, James & Lassila, Ora (2001). The Semantic Web. *Scientific American,* 284(5): 34-43.

Borgatti, Stephen P. (1995). Centrality and AIDS. *Connections, 18*(1): 112-15.

Coyle, Karen (2012). *Linked data tools: Connecting on the web.* Chicago: ALA TechSource.

Easley, David & Kleinberg, Jon (2010). *Networks, crowds, and markets: Reasoning about a highly connected world*. New York: Cambridge University Press.

Francq, Pascal (2011). *Collaborative search and communities of interest: Trends in knowledge sharing and assessment.* Hershey, PA: Information Science Reference.

Giunchiglia, Fausto, Dutta, Biswanath & Maltese, Vincenzo (2014). From knowledge organization to knowledge representation. *Knowledge Organization, 41*(1): 44-56.

Gómez-Pérez, Assunción (2004). Ontology evaluation. In *Handbook on ontologies.* Berlin, Heidelberg: Springer Berlin Heidelberg. Pp 251-73.

Gruber, Thomas R. (1993). A translation approach to portable ontology specifications. *Knowledge Acquisition, 5*(2): 199-220.

Gruber, Thomas R. (1995). Toward principles for the design of ontologies used for knowledge sharing. *International Journal of Human-Computer Studies, 43*(5/6): 907-28.

Guarino, Nicola & Giaretta, Pierdaniele (1995). Ontologies and knowledge bases: Towards a terminological clarification. In *Towards very large knowledge bases: Knowledge building and knowledge sharing* . Amsterdam: IOS Press. Pp. 25-32.

Guerrini, Mauro & Possemato, Tiziana (2013). Linked data: A new alphabet for the semantic web. *Italian Journal of Library & Information Science, 4*(1): 67-90. doi: 10.4403/jlis.it-6305

Heath, Tom & Bizer, Christian (2011). *Linked data: Evolving the Web into a global data space*. San Rafael: Morgan & Claypool.

Herre, Henrich (2013). Formal ontology and the foundation of knowledge organization. *Knowledge Organization, 40*(5): 332-39.

Miller, Paul (2000). Interoperability. What is it and why should I want it? *Ariadne, 24.* [http://www.ariadne.ac.uk/issue24/interoperability]. Acessed on April 5 2016.

Ohly, H. Peter (2013). Challenges of knowledge organization and ISKO. *SRELS Journal of Information Management, 50*(6): 807-18.

Ouksel, Aris M., & Sheth, Amit (1999). Semantic interoperability in global information systems. *ACM SIGMOD Record, 28*(1): 5-12.

Staab, S., & Studer, R. (2009). Preface. In *Handbook on ontologies* (2nd ed.).Berlin: Springer.

Tolk, Andreas (2006). What comes after the semantic web - PADS implications for the dynamic web. In *Proceedings of the 20th Workshop on Principles of Advanced and Distributed Simulation*.Washington, DC: IEEE Computer Society. Pp. 55-62.

Vrandečić, D. (2009). Ontology evaluation. In *Handbook on ontologies* (2nd ed.). Berlin: Springer. Pp. 293-313.

Wasserman, Stanley & Faust, Katherine (1994). *Social network analysis: Methods and applications.* Cambridge: Cambridge University Press.

Yao, Haining, Orme, Anthony M., & Etzkorn, Letha (2005). Cohesion metrics for ontology design and application. *Journal of Computer Science, 1*(1): 107-13.

Laís Barbudo Carrasco and Silvana Aparecida Borsetti Gregório Vidotti

Handling Multilinguality in Heterogeneous Digital Cultural Heritage Systems trough CIDOC CRM Ontology

Abstract

Given the enormous flow of information available on the Internet in different languages and regarding the possibility of users to search independently their native language and retrieve relevant information, the issue of multilingual access and multilingual information retrieval has become of great significance. This paper covers the issue of handling multilinguality in heterogeneous digital cultural heritage systems trough CIDOC CRM ontology.

1 Introduction

Given the enormous flow of information available on the Internet in different languages and regarding the possibility of users to search independently their native language and retrieve relevant information, the issue of multilingual access and multilingual information retrieval has become of great significance.

> Language is the foundation of communication between people and is also part of their cultural heritage. For many, language has far-reaching emotive and cultural associations and values rooted in their literary, historical, philosophical and educational heritage. For this reason, the users' language should not be an obstacle to accessing the multicultural heritage available in cyberspace. The harmonious development of the information society is therefore only possible if the availability of multilingual and multicultural information is encouraged (Unesco, 2003).

World Wide Web should provide systems that are easy to navigate, with flexible tools, which help and orient the user in the search for information. In addition, by this the users can access a wide diversity of unrestricted information sources that give them the opportunity to select and discard, retrieving exactly those texts that are of interest. (Peters, 1996)

> Multilingualism refers to both a person's ability to use several languages and the co-existence of different language communities in one geographical area (Com, 2005).

The multilinguality provides political, economic, cultural, social integration among the countries and, in this way, develop their economy. Therefore, that knowledge access represents development. On the other hand, this knowledge should be multilingual to reach the higher number of citizens.

> A multilingual information society requires the deployment of standardized and interoperable language resources (dictionaries, terminology, text corpora, etc.) and applications for all languages, including the less widely used languages of the European Union (Com, 2005).

In the digital world, there is the predominance of the English idiom, and the most frequent second languages spoken are English, French, German, Spanish and Russian. Otherwise, people prefer to search on the Internet in their own tongue and they usually claim that they miss interesting information because the content was in a language they did not understand.

An important initiative for the multilingualism issue is the *EuropeanaConnect*, which 'supports the creation of a diverse and inclusive Europeana facilitating access to culture by all communities and individuals and representative of various cultures and language-groups'. It is important to emphasize that Europeana portal provides access to over 20 million digitized cultural heritage objects. (Europeana)

Cultural heritage domain uses different metadata to describe information resources. In order to integrate information from heterogeneous systems, ontologies as semantic technologies are already being used, e.g. in Europeana. In this context, ontologies are used as an important tool for achieving information integration, in other words, metadata can be semantically mapped and integrated into an ontology, which has the competence not only to conceptualize specific domains, but also to express their semantics.

> Managing heterogeneous data is a challenge for cultural heritage institutions, archives, libraries, and museums, which usually develop collections with heterogeneous types of material, described by different metadata schemas. […] The wide use of a number of cultural heritage metadata schemas imposes the development of interoperability techniques that facilitate unified access to cultural resources. One of the widely implemented techniques is the Ontology-Based Integration. Ontologies provide formal specifications of a domain's concepts and their interrelations and act as a mediated schema between heterogeneous sources. (Papatheodorou, 2012).

According to Gruber (1993) 'an ontology is a specification of a conceptualization'. More specially, the CIDOC CRM ontology is the specification of the Cultural Heritage conceptualization, which has been created as a tool for information integration. Metadata can be mapped into CIDOC CRM, hence, these mappings provide interoperability between heterogeneous systems.These mappings of the different metadata schemes into one conceptual reference space primarily support conceptual integration. It does not automatically create multilingual interfaces. They can be generated from such a common conceptual core, however, if one goes beyond current applications.

The multilingual access to content is used to increase and enhance the users' possibilities to access the cultural heritages systems in their native or preferred language. In summary, multilingual information access could allow users to search for information produced in different languages without having to make their search query (question) in each language. In this context, multilinguality in ontologies has become an impending need for institutions worldwide with valuable linguistic resources in different natural languages. Since most ontologies are developed in one language, obtaining multilingual ontologies implies to localizeor adapt them to a concrete language and culture community. (Gomez & Peters, 2008).

In this paper, CIDOC CRM ontology was used as a conceptual representation of cultural heritage domain to promote semantic integration between different metadata schemas, such as Encoded Archival Description (EAD), Machine-Readable Cataloging (MARC) and Lightweight Information Describing Objects (LIDO). The semantic mapping provides interoperability between the digital cultural heritage systems. In

other words, it integrates the ontological representation of the CIDOC CRM with earlier ways to represent metadata. According to the literature, there are many XML metadata mapping to the CIDOC CRM ontology efforts, since this ontology is considered one of the most appropriate models in integration architectures. In this paper, the ontology was translated and added, for each concept and property, di□erent labels in different languages, including synonymous.

2 State of art

By the fact that Digital Cultural Heritage Repositories - as Libraries, Archives and Museums - use different metadata standards to describe their information resources, the metadata harmonization from the cultural heritage field is a challenge, because the data models are more designed on the community requirements than on requirements of cross-community interoperability.

To examine the similarities and differences between the metadata standards of those three sectors the most prominent modern standards from these fields will be described next. It was selected MARC (Machine-Readable Cataloging), which is 'a standard for the representation and communication of bibliographic and related information in machine-readable form' (MARC). EAD (Encoded Archival Description), which project's goal was to create a data standard for describing archives.

> EAD stands for Encoded Archival Description, and is a non-proprietary de facto standard for the encoding of finding aids for use in a networked (online) environment. Finding aids are inventories, indexes, or guides that are created by archival and manuscript repositories to provide information about specific collections. While the finding aids may vary somewhat in style, their common purpose is to provide detailed description of the content and intellectual organization of collections of archival materials. EAD allows the standardization of collection information in finding aids within and across repositories (EAD).

For the field of museums, it was selected LIDO (Lightweight Information Describing Objects), which has an event-oriented approach compliant with the CIDOC CRM.

> The strength of LIDO lies in its ability to support the full range of descriptive information about museum objects. It can be used for all kinds of object, e.g. art, architecture, cultural history, history of technology, and natural history. (LIDO)

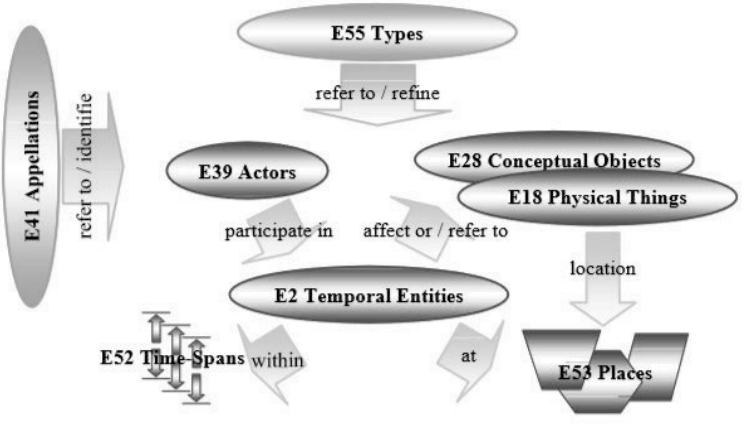
In order to integrate information from heterogeneous sources, ontologies as semantic technologies are already being used, e.g. in the Europeana.CIDOC Conceptual Reference Model (CRM) is a very prominent ontology used for such purposes. The CIDOC CRM arised from the CIDOC Documentation Standards Group in the International Committee for Documentation of the International Council of Museums and CIDOC CRM was accepted as the ISO 21127 in 2006. (CIDOC CRM)

> The CIDOC CRM is intended to promote a shared understanding of cultural heritage information by providing a common and extensible semantic framework that any cultural heritage information can be mapped to. […] In this way, it can provide the "semantic glue" needed to mediate between different sources of cultural heritage information, such as that published by museums, libraries and archives. (CIDOC CRM)

In order to provide information integration semantics mappings can be a solution for it, in this work, mappings derivated from the cultural heritage metadata standards, as MARC (Library), EAD (Archive), LIDO (Museum) into CIDOC CRM will be built. Hence, these mappings provide interoperability between those fields of the Cultural Heritage Universe.This mapping of the different metadata schemes into one conceptual reference space primarily supports *conceptual* integration. Therefore, CIDOC Conceptual Reference Model (CRM) is used as the mediated schema to integrate Cultural Heritage metadata sources.

The CRM scope can be defined as all the necessary information for the scientific documentation of cultural heritage collections, in order to enable a broad exchange of information of the area and the integration of heterogeneous sources.

Figure 1 – Main model entities.



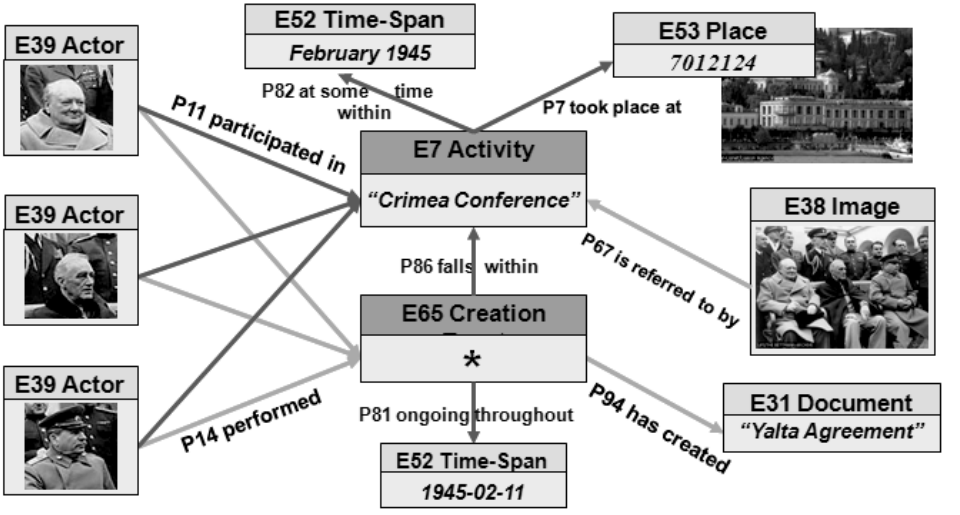
Source: Doerr, Ore & Stead (2007)

The CIDOC CRM contains classes and logical groups of properties. These groups have to do with the notions of participation, structure, location, assessment and identification, purpose, motivation, use, and so on. These properties have placed temporal entities, and with them, the events in a central location. In a technical context, the CIDOC CRM ontology can be used as a basis for data archiving, exchange and integration, being seen, in this way, as an important contribution to the creation of a global network of cultural heritage information. (Lima, 2008).

3 Results and discussion

Crimea Conference is an activity performed by three actors - Churchill, Roosevelt, Stalin - at a location represented by TGN code in a specific period of time. This episode is recorded in a photo (image) and reported in a written document.

Figure 2 – Crimea Conference



Source: Doerr & Stead (2008).

The following scheme describes such information by using entities and properties of CIDOC CRM ontology, which can be expressed into XML, as follows:

```
<?xml version = "1.0" encoding = "ISO-8859-1"?>
<?xml-stylesheet type="text/xsl" href="crm.xsl"?>
<CRMset>
<CRM_Entity> Crimea Conference
<in_class> E39: Actor @ pt E39: Ator
<is_identified_by> Churchill </is_identified_by>
<is_identified_by> Roosevelt </is_identified_by>
<is_identified_by> Stalin </is_identified_by>
</in_class>
<in_class>E53: Place @ pt E53: Local
<is_identified_by> TGN 7012124
<has_type> Crimea </has_type>
<has_type> Ukraine </has_type>
</is_identified_by>
</in_class>
<in_class>E7: Activity @ pt E7: Atividade
<is_identified_by> Crimea Conference
<has_type> Conference Region </has_type>
</is_identified_by>
<in_class>E52: Time-Span @ pt E52: Período de tempo
<has_note> February 1945 </has_note>
<has_note> 1945-02-11</has_note>
</in_class>
<in_class> E65: Conceptual creation @pt E65: Criação conceitual
<took_place_on> 1945-02-11 </took_place_on>
```

```
</in_class>
<in_class> E31: Document @ pt E31: Documento
<documents> Photo
        <has_note> Actors </has_type>
        <has_note> Event </has_type>
</documents>
<documents> Agreement Text </documents>
</in_class>
<in_class> E38: Image @ pt E38: Imagem
< is_identified_by> Photo </is_identified_by>
</in_class>
</CRM_Entity>
</CRMset>
```

CIDOC CRM ontology was used as a conceptual representation of cultural heritage domain to promote semantic integration between different metadata schemas, such as Encoded Archival Description (EAD), Machine Readable Cataloguing (MARC) and Lightweight Information Describing Objects (LIDO). Table 1 shows the harmonization between CIDOC and metadata.

Table 1 – Metadada mappings into CIDOC CRM

CIDOC CRM	MARC	EAD	LIDO
E39 Actor	100 Main entry – personal name (NR)	ead_author	Title
E53 Place	852 Location (R)	ead_origination	namePlaceSet Event Place
E7 Activity	111 Main entry – meeting name (NR)	ead_event	displayEdition
E52 Time-Span	518 Date/time and Place of an event note (R)	ead_date	Event Date
E31 Document	008 All materials	ead_frontmatter	Object
E38 Image	O08 Visual materials	ead_dao	objectWorkType
E65 Creation	508 Creation	ead_creation	Cultural context

When users want to extract information from a system, they need to ask a question on a standard computer. This means that they have to formulate a search strategy and then the computer should be able to find a result for the search. On the other hand, the most common way for people to get information is through questions in natural language. Thus, interfaces built in natural language facilitates the user's thought process and by being flexible, people need little training to use them in interaction with a computer system. The main use of natural language interfaces is that they support the view of the domain user and the system. In other words, they turn the user concepts into concepts effectively used by the system. Thus, natural language interfaces are systems that translate a sentence in natural language for a search engine.

In order to enhance the information retrieval in heterogeneous systems, we created a mapping of CIDOC CRM entities (Conceptual Reference Model) with the labels in the

search interface. In addition, to achieve a multilingual access, we selected entities of CIDOC CRM vocabulary, which uses English as a standard language, and connected with a vocabulary in Portuguese. Thus, the correspondence between labels potentiate multilingual access to content.

Table 2 – Mapping between CIDOC CRM ontology and its entities in a search interface

Search Interface		CIDOC CRM entities		Search Result
English	Portuguese	English	Portuguese	
Who	Quem	E21 Person E39 Actor	E21 Pessoa E39 Ator	Churchill Roosevelt Stalin
What	O quê	E7 Activity E31 Document	E7 Atividade E31 Documento	Crimea Conference Yalta Agreement
Where	Onde	E53 Place	E53 Lugar	7012124 (TGN)
When	Quando	E52 Time-Span	E52 Período de Tempo	February 1945 1945-02-11

When translating the CIDOC CRM vocabulary terminology into other languages, we would create a multilingual environment for information retrieval. For example, E50 Date entity can be translated into Portuguese as E50 Data, in German as E50 Datum, in French as E50 Date. In addition, E35 Title entity can be translated into Portuguese as E50 Título, in German as E50 Titel and in French as E50 Titre.

4 Final Considerations

Multilingualism in ontologies has become an imminent need for institutions around the world with valuable language resources.As most ontologies are developed in a specific language, obtaining multilingual ontologies implies locate them or adapt them to a specific language and culture community. (Gomez & Peters, 2008).

Multilingualism has become an important target to achieve because it means plenty of languages and in order to improve search results, search engines associated with a multilingual ontology leverage access and information retrieval. It is important to note that multilingualism also provides economic, social, political and cultural integration of countries. This means that the globalization process puts multilingualism challenge into an opportunity situation.

The CIDOC CRM vocabulary can be translated into many languages. Therefore, the search interface can use these labels translated to enhance multilingualism in the system. Here, it is important to mention that the search interface labels would not need to be translated one by one in the software, because the ontology would have a multilingual vocabulary. In addition, using a multilingual vocabulary of CIDOC CRM ontology, the search interface would apropriate translated labels into the search engine and could use them instead of translating label by label in the software system.

CIDOC CRM promotes multilingual information access and provides a higher information retrieval to the user's query. As language is not neutral, multilingual access and multilingual information retrieval is a topic for further reflection.

References

A New Framework Strategy for Multilingualism. (2005) In: *Communication from the Commission to the Council, the European Parliament, the European Economic and Social Committee and the Committee of the Regions*, 2005, Brussels. *Proceedings...* Brussels: COM, 2005.

Artur, O., Crofts, N.; Le Boeuf, P. (2002). Elag presentation ontologies. Semantic web and libraries. In: *Library Systems Seminar*. 2002, Rome.

CIDOC CRM. http://www.cidoc-crm.org/

Crofts, Nick et al. (Ed.). (2005) *Definition of the Cidoc Conceptual Reference Model*.Cidoc, Icom, 2005.

Crofts, Nick (2004) *Museum informatics*:the challenge of integration. 2004. PhD dissertation. Genebra: Univerity of Geneva.

Doerr, Martin (2003). The Cidoc CRM: an ontological approach to semantic interoperability of metadata. *AI Magazine*, 24 (3): 75-92.

Doerr, Martin, ORE, Christian-Emil & Stead, Stephen (2007). The CIDOC Conceptual Reference Model - A New Standard for Knowledge Sharing ER2007 Tutorial. In: *26th International Conference on Conceptual Modeling (ER 2007)*. 2007. Auckland: CRPIT, 2007.

Doerr, Martin & Stead, Stephen (2008) *The Cidoc CRM: standard for the integration of cultural information.*Glasgow, 2008.

EAD - http://www.loc.gov/ead/eadabout.html

Espinoza, Mauricio, Gomez-Perez, Asuncion &Mena, Eduardo (2008). Enriching an ontology with multilingual information. In *Proceedings of the 5th European Semantic Web Conference on The Semantic Web: Research and Applications*. Spain, 01-05 June, 2008. Pp 333-47.

European Commission (2011) –http://ec.europa.eu/languages/pdf/inventory_en.pdf.

EUROPEANA. http://www.europeana.eu/portal/

EUROPEANACONNECT. http://www.europeanaconnect.eu/

Gomez-Pérez, Assuncion & Peters, Wim (2010). *Modelling multilinguality in ontologies*. In *Proceedings of COLING*. 2010. Manchester: Coling. Pp 67–70.

Gruber, Thomas R.(1993). Towards principles of the design of ontologies used for knowledge sharing. In *International Journal Human-Computer Studies, Substantial Revision Of Paper Presented At The International Workshop On Formal Ontology*. 1993. Padova: Standford University.

ISO 639 – Code for the Representation of the Names of Languages, 1989. http://xml.coverpages.org/iso639a.html>.

Le Boeuf, Patrick (2006). Using an ontology-driven system to integrate museum information and library information. In *Symposium On Digital Semantic Content Across Cultures*.2006. Paris: Louvre.

LIDO. http://network.icom.museum/cidoc/working-groups/data-harvesting-and-interchange/what-is-lido/

Lima, João Alberto de Oliveira (2008). Modelo genérico de relacionamentos na organização da informação legislativa e jurídica**.***PhD Thesis.* Brasília: UnB.

MARC. http://www.loc.gov/marc/

Peters, Carol (1996). Guaranteeing multilinguality in the information society. In: *The Information Society In The Euro-Mediterranean Context* Conference. 1996.Antipolis: Ercim.

Papatheodorou, Christos (2012) *Ontology-based Integration of Cultural Heritage Metadata*.

Thesaurus of Geographic Names. http://www.getty.edu/research/tools/vocabularies/tgn.

UNESCO (2003). *Cultural and linguistic diversity in the information society*. Paris: Unesco.

Webert Júnio Araújo, Gercina Â. B. de Oliveira Lima and Ivo Pierozzi Júnior

Data-Driven Ontology Evaluation Based on Competency Questions: A Study in the Agricultural Domain

Abstract

Although ontologies have potential in knowledge representation, they must be tested for the guaranty of quality. The ontology evaluation is one way to ensure the quality of ontology. Therefore the guiding objective of this research is to evaluate a domain ontology using a data-driven ontology evaluation proposal. The research methodology is based on a proposal for a data-driven ontology evaluation developed by Brewster et al (2004). The results obtained in this research, so far, have been helpful because it was possible to achieve the proposed objective. We could implement a proposal for ontology evaluation based on a text corpus by adaptations to the context of research and to the ontology evaluated in this study.

1 Introduction

Climate changes and water shortages are real concerns that affect various spheres of our society. Brazil, for example, lived in 2015 one of its major water crisis, as water availability in the country is directly related to climate. Thus, research is needed in various fields of knowledge in order to understand climate changes and its impacts on water resources. The Knowledge Organization's contribution can relate to the modeling of Knowledge Organization Systems (KOS), which are systematic tools aimed at building abstract models of the real world through the representation of domain concepts. These instruments make it possible, among other things, a shared view on a specific area, and facilitate communication between people with different views through common vocabulary.

A KOS that has gained prominence in the Knowledge Organization field is ontology. An ontology can be defined as a shared model of reality, which is formally represented by classes, properties, relationships and axioms. Although ontologies have potential in knowledge representation, they must be tested for the guaranty of quality. The ontology evaluation is one way to ensure the quality of an ontology. Since the purpose of ontology evaluation is to verify the conformity of world model to the world modeled formally.

The ontology evaluation is a relevant subject for study in the context of Knowledge Organization since it allows to measure the efficiency of ontologies in knowledge representation. Moreover, the evaluation helps to identify possible failures in the model, improving the representation of the modeled field.

The guiding objective of this research is to evaluate a domain ontology using a data-driven ontology evaluation proposal. The ontology intends to evaluate in this work was developed by specialists from Embrapa Campinas, and it is named OntoAgroHidro.

2 Background

The OntoAgroHidro was built using Web Ontology Language 2 and the ontology editor Protégé (version 4). The main purpose of the OntoAgroHidro is to represent knowledge about the impacts of climate changes and agriculture on water resources. As pointed out by Bonacin, Nabucco and Pierozzi Junior (2015), it is a quite ambitious scenario to model, insofar as it covers the domain particularities of climate changes associated with agriculture and hydrology.

The development process of OntoAgroHidro involved the work of knowledge engineers and domain experts. To start the modeling process, they created two questions that represented the "macro scenario"; as follow: 1) What are the impacts of agriculture and climate changes on water resources?; 2) What are the impacts of water quantity and quality of water in agriculture? It then carried the modeling of the main concepts related to the "macro scenario". In addition, it was made the knowledge reuse of ontologies that addressed relevant themes about the modeling domain. They used the Cuashi [1] and SWEET [2] ontologies for knowledge reuse.

This research is characterized as exploratory, qualitative and applied, starting with the exploration of the relevant topics to the research in the Information Science and Computer Science literatures. The theoreticaland methodological foundation permeate concepts on topics such as organization and knowledge representation, definitions of term and concept, ontologies and competency questions.

In the context of this study, we make a distinction between the definitions of concept and term. We consider the term as a standard unit, which can consist of one or more words, forming an expression, and is considered as a lexical unit in a given domain, to denote a concept. With regard to the definition of concept, we consider it as something abstract that needs to be represented by a sign (which can be a term), so it can be used primarily within the knowledge organization systems.

The literature review found that there were several proposals for ontology evaluation. These proposals can be classified into four categories:

1) those based on comparing the ontology to a gold standard. In this kind of proposal the ontology is compared to another model built optimally, known as Gold Standard, which may be an ontology itself. e.g. Maedche and Staab (2002), Gangemi et al. (2006),
2) those based on using the ontology in an application or task. The ontology is evaluated based on their performance on a specific task, e. g. Porzel and Malaka (2004), Fernández, Cantador and Castells (2006),
3) those involving comparisons of the ontology against a data source (a collection of documents, for example) about the domain to be covered by the ontology, e.g. Brewster et al. (2004), Hlomani and Stacey (2013),
4) those where the evaluation is done by humans who try to assess how well the ontology meets a set of predefined criteria, standards, requirements, etc. e.g. Lozano-Tello and Gómez-Pérez (2004), Almeida (2009).

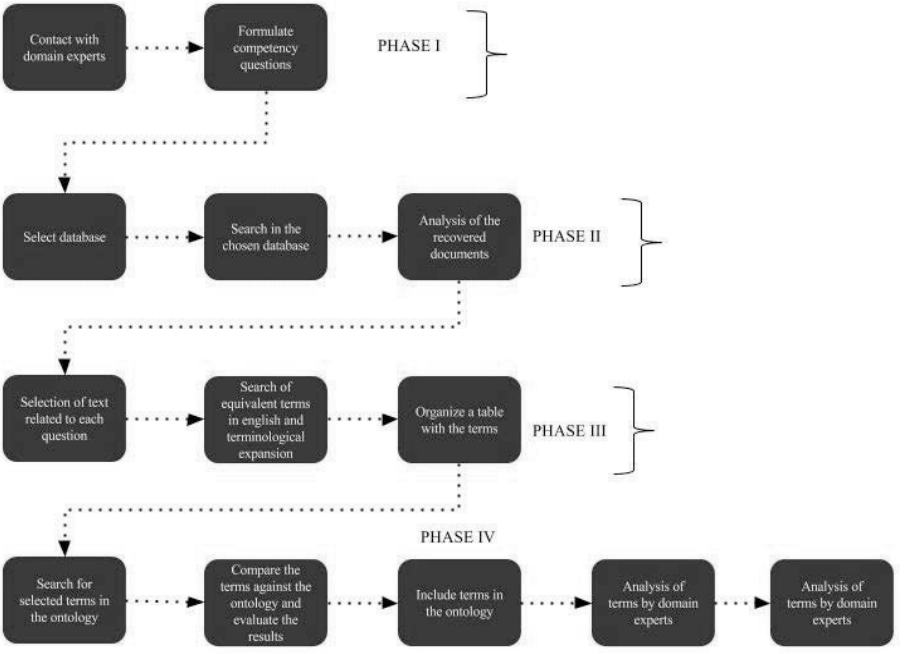
3 Methods

The research methodology is based on a proposal for a data-driven ontology evaluation developed by Brewster et al (2004), which consists of three steps: 1) identifying keywords / terms in the collection of texts; 2) query expansion; 3)mapping of identified terms in the ontology. The Brewster et al (2004) proposal was adapted to this research and has four steps: 1) definition of competency questions [3]; 2) selection of the collection of documents that will compose the data source for the ontology evaluation; 3) selection of terms related to the concepts of the competency questions; 4) evaluation and mapping the terms in the OntoAgroHidro.

Basically, the adaptations of Brewster et al (2004) proposal are related to the creation of the first two steps (step 1 - definition of competency questions, and step 2 - selection of the collection of documents that will compose the data source for the ontology evaluation) that do not exist in the Brewster et al (2004) proposal. Moreover, steps 1 and 2 of the Brewster et al (2004) proposal were synthesized in only one step (step 3- selection of terms related to the concepts of the competency questions) in this research proposal.

The evaluation of OntoAgroHidro was developed as follow. From three competency questions formulated by a domain expert, we extracted the key concepts from each competency question and used the concepts to search for documents in a domain database with the purpose of identifying documents related to each concept. Then we extracted from the documents, terms associated with the concepts from the competency questions and other terms from the AGROVOC and NAL thesaurus and created a list with all terms. Finally, we compared the ontology concepts against the list of terms with the purpose of identifying how much from the list the OntoAgroHidro represents and include some terms that the ontology didn't represent in its structure. Figure 1 shows a framework with each one of the phases performed.

Figure 1 - Phases for the ontology evaluation



4 Results and Discussion

Based on three competency questions which led to 11 concepts [4] we recovered 60 documents from a specific database in the agricultural domain (BDP@ [5]). We extracted 130 terms from the 60 documents and the AGROVOC and NAL thesaurus. The 130 terms were compared against the OntoAgroHidro.

The evaluation of the ontology against the corpus found out that the OntoAgroHidro represents 68% of the concepts related to competency questions. Detailing the result, it was discovered that the ontology represents 78% of the concepts related to the competency question number 1, 56% of competency question number 2 and 42% of the competency question number 3. Table 1 presents in more detail the data, which shows the number of concepts represented by OntoAgroHidro in each competency question.

Table 1 – OntoAgroHidro Representativeness

Questions	*N° of terms*	*N° of concepts* [6]	*N° of concepts represented by OntoAgroHidro*	*OntoAgroHidrorepresentativeness*
Question 1	66	50	39	78%
Question 2	40	25	14	56%
Question 3	24	7	3	42,85%
Total	130	82	56	68,29%

Table 2 discusses in more detail the representation of OntoAgroHidro for each concept originated from the competency questions. This allows a more specific knowledge of what are the concepts that are not well represented in OntoAgroHidro and points to the need for improvements in the representation so that the domain knowledge can be better represented by the ontology.

Table 2 – OntoAgroHidro representativeness discriminated for each concept of the competency questions

Questions	***Concept searched***	***Nº of terms***	***Nº of concepts***	***Nº of concepts represented by OntoAgroHidro***	***OntoAgroHidrorepresentativeness***
Question 1	Water quality	47	47	37	78,72%
	Organic crop	13	1	0	0%
	Bassin	6	2	1	50%
Question2	Water	18	15	11	73,33%
	Water reservoir	14	6	2	33,33%
	Population	5	3	0	0%
	Time	3	1	1	100%
Question3	Deforestation	10	1	0	0%
	Draw-well	5	1	0	0%
	Water quantity	4	1	1	100%
	River	5	4	2	50%

In a qualitative analysis, it was observed that the ontology is representative with regard to concepts related to the subdomain "water resources". This qualitative analysis was performed through the separation of concepts that were related to water resources. The idea of this analysis is directly related to the main objective of OntoAgroHidro development, which consists of representing knowledge about the impacts of climate changes and agriculture on *water resources*. So, we believe that the OntoAgroHidro must represent the concepts related to climate changes, agriculture and water resources in a very specific and complete manner.

Based on the results of this research, we believe that the gap that avoided the ontology to have a better representation of the concepts presented in the sample selected for this study was the absence of a more specific domain modeling. Some concepts in the field were not represented by OntoAgroHidro because it lacked specification, i.e. there was a broader concept, but the most specific concept was not represented in the ontology structure. This failure in representing some specific

concepts in OntoAgroHidro can be solved with the inclusion of new concepts in the ontology, which was exactly what we started doing in this research. In addition, the reuse of other Knowledge Organization Systems such as thesauri, can also contribute to the terminological and the conceptual enrichment of OntoAgroHidro and help to eliminate this lack of specific concepts of the domain.

5 Conclusions

The results obtained in this research, so far, have been helpful because it was possible to achieve the proposed objective. We could implement a proposal for ontology evaluation based on a text corpus by adaptations to the context of research and to the ontology evaluated in this study.

It is worth mentioning that in the context of this study we evaluated only the concepts of the OntoAgroHidro. Axioms and relationships were not evaluated because the main idea was to know how was the conceptual representation of the ontology. Besides, concepts are one of the main parts in a Knowledge Organizations System such ontology. In addition, the proposal on which it relied for the development of this research mainly focuses on the evaluation of the conceptual level.

The methodology presented in this work differs from most existing methods in the literature; it makes use of an approach that uses competency questions in combination with a data source. The results achieved are considered sufficient for identification of positives and negatives aspects of the OntoAgroHidro. Although it is believed that an expansion of the study scope can prove useful to detect other characteristics of the evaluated ontology and thus suggest improvements.

Notes

[1] SWEET (Semantic Web for Earth and Environmental Terminology): https://sweet.jpl.nasa.gov

[2] CUAHSI (Consortium of Universities for the Advancement of Hydrologic Science, Inc.): ontologia sobre hidrologia. http://his.cuahsi.org/ontologyfiles.html

[3] Question 1: Does organic crops cause some impact on water quality of our basins? Question 2: There will be water shortage in water reservoirs for supply to the population of the State of São Paulo, Brazil in 2016?

[4] The 11 concepts extracted from the competency questions are: organic crop, water quality, basin, water, water reservoir, population, time, deforestation, draw-well, water quantity and river.

[5] https://www.bdpa.cnptia.embrapa.br/consulta

[6] The number of concepts is based on an intellectual analysis of how many concepts the list of terms represents. Because one concept can be represented by several different terms.

References

Bonacin, Rodrigo, Nabuco, Olga Fernanda & Junior, Ivo Pierozzi (2015). Ontology models of the impacts of agriculture and climate changes on water resources: Scenarios on interoperability and information recovery. *Future Generation Computer Systems.*

Maedche, Alexander & Staab, Steffen (2002). Measuring similarity between ontologies. In *Proceedings of CIKM*. vol. 2473.

Gangemi, Aldo et al. (2006). A metaontology based framework for ontology evaluation and selection . In *Proceedings of 4th International Workshop on Evaluation of Ontologies for the Web*, Edinburgh, UK.

Porzel, Robert & Malaka, Rainer (2004). A task-based approach for ontology evaluation. In *ECAI Workshop Ont. Learning and Population.*

Fernández, Miriam, Cantador, Ivan & Castells, Pablo (2006). A tool for collaborative ontologyreuse and evaluation. In *Proceedings of 4th International Workshop on Evaluation of Ontologies for the Web*, Edinburgh, UK.

Brewster, Cristopher et al. (2004). Data-driven ontology evaluation. In*Proceedings of the 4th International Conference on Language Resources and Evaluation*, Lisbon, Portugal.

Hlomani, Hlomani & Stacey, A. D. (2013). Contributing evidence to data-driven ontology evaluation: Workflow ontologies perspective. In *Proceedings of the 5th International Conference on Knowledge Engineering and Ontology Development*, Vilamoura, Portugal.

Lozano-Tello,A., Gómez-Pérez,A. (2004). Ontometric: A method to choose the appropriate ontology. *J. Datab. Mgmt*., 15(2): 1–18.

Almeida, Mauricio B. A proposal to evaluate ontology content. *Applied Ontology*, Birmingham, 4(34): p. 245-265.

Dagobert Soergel and Olivia Helfer

A Metrics Ontology. An Intellectual Infrastructure for Defining, Managing, and Applying Metrics

Abstract

This paper presents the beginnings of a comprehensive ontology for organizing information about metrics and its potential application to defining and managing metrics in the CTSA (Clinical and Translational Science Award) project. The aim is to support an integrated database of all metrics used by CTSA components. The ontology is given as an entity-relationship conceptual data schema. Its completion should draw on metrics definition templates that can be found in many places.

0 Introduction and aims

This paper presents the beginnings of a comprehensive ontology for organizing information about metrics and its potential application to defining and managing metrics in the CTSA (Clinical and Translational Science Award) project.

Accountability is integral to the success of any program or endeavor. Accountability requires metrics that truly reflect the desired outcomes of a program. Too many metrics measure what can be easily measured rather than what is really important, leading to distortions of program execution: programs try to maximize performance on faulty measures, getting away from what really matters and giving accountability a bad name. So defining and documenting proper metrics is important for good management.

Metrics are based on variables. Variables are concepts or intellectual constructs. Any science or scholarly endeavor (from natural science to social science even to humanities) defines variables so it can state propositions that explain the world around us or the mind within us and that allow us to take intentional action. Just think about all the variables included in a model of climate change. In program management, variables characterize the environment in which the program works, processes the program carries out, and, most importantly, desired and actual outcomes. The program administrator wants to know whether or to what extent the actual outcomes match the desired outcomes and take action if they do not.

In order to serve their function, variables must be assigned values in the specific context under consideration. If that is not possible directly, one needs to define *indicator*s that shed some light on the variable and that can be assigned values. The process of assigning values is called *measurement,* and the result is often called a *metric*. The methods and processes of measurement must be carefully defined so as to measure what one wants to measure. Such a definition of how the values of a variable are to be measured is called *operationalization*. For some scholars, the conceptual definition of a variable and its operationalization are on different planes, for others the operationalization *is* the definition ("Intelligence *is* what the intelligence measures").

Variables or their indicators can be measured at any measurement scale – nominal, ordinal, interval, ratio.

In thinking about metrics, there are two major considerations

(1) *Substantive*: the meaning and role of the variable in the subject field.

(2) *Formal* (the focus of this paper)*:* The formal properties of the metric as metric (such as measurement scale, data collection, how the metric is computed) and considerations on how the metric is used in the management of a program (what decisions are based on the metric, who uses the metric in making decisions).

Our ontology of metric properties and use serves the following purposes:

(1) *Support better thinking about metrics*, tying metrics to objectives that matter.

(2) *Support organizing a database of metrics* that allows for
- storing all pertinent information about a metric and
- retrieving metrics and navigating through the metric space using multiple perspectives, including retrieval of metrics that are important when considering a given decision or metrics that need to be monitored.
- support teams developing metrics by storing and tracking their work in progress.

(3) *Support implementation of the metrics* by tying them to "on-the-ground" data that need to be collected and by improving the definitions used in collecting these data.

(4) *Support presenting causal models* (influence diagrams) in a given domain using hypothesized and/or empirically confirmed causal dependencies and influences among the metrics. This includes logic diagrams often used in planning a program.

2 Background: The problem in context

Our research context is the *Clinical and Translational Science Awards (CTSA)* program of the *National Center for Advancing Translational Sciences (NCATS)*, a center in the *US National Institutes of Health (NIH)*. The purpose of this program is to increase the effectiveness and efficiency of clinical research and to speed up the translation of research results into improvements of patient care, thereby increasing the return on NIH's enormous investment in clinical research. The program gives large, multi-year grants mostly to universities ("CTSA Hubs", now 60+) for the purpose of improving research infrastructure (so that researchers have less red tape and more and easier access to support services, such as consultation on research methods, recruitment of study participants, and use of equipment) and intensifying communication with hospitals and other health care providers to foster quick application of the results of clinical research.

Running a CTSA hub is a massive management problem. NCATS assists the hubs with implementing best management techniques through training in *Results-Based Accountability*

(RBA)(Friedman 2015) for all hubs in conjunction with a Common Metrics Initiative. In RBA management decisions are geared towards improving outcomes that are monitored by carefully chosen metrics, keeping in mind that many factors influence the outcomes.. Each CTSA hub defines its own metrics, but NCAT's common metrics initiative aims to establish a core set of metrics used by all.

Defining metrics requires much intellectual work and consensus building. Table 1 shows the definition for a metric for innovation that was elaborated by a team in the Common Metrics Initiative working hard through several meetings and many email exchanges.

Table 1. Example of a metric definition following the NCATS template. Condensed from the original

Template Element	Description
Operationalized Metric Title	Count of Innovations (*The Number of Innovations (by area) to Improve Translation)*
Common Function Group	Collective Impact of the Hub
Common Sub-function	Fostering the innovation of new technologies, methods, or approaches to improve translation.
Operational Specification	Definition: ***Innovation*** **is the discovery, development, demonstration, dissemination, adoption/implementation of a new product (good or service), or process, a new organizational structure or approach to research administration, or a new translation method in a CTSA hub**…. [Source] Six counts: Count of all innovations and a separate count for each of the following five areas **A** Research capacity building**B** Research methods**C** Diagnosis, prevention, treatment, medical practice **D** Translation methods**E** Other: please specify
Technical Description (include key definitions, timeframe, data scope)	b TimeframeRetrospectively: collect data and compute this metric for CY2015. Prospectively: collect data and compute this metric semiannually. c Data Scope. *The universe of innovations supported by the CTSA functions provided within a Hub (i.e., pilot awards, key personnel directly connected to the CTSA, support through a CTSA core).*
Data Sources, Method of Data Collection, Exclusion Criteria	a Data Source. Reports of innovations found in CTSA Hub databases, patent records, electronic records, other reporting documents, and/or survey of CTSA researchers and clinicians with e support from CTSA functions. b Method of Data Collection. Find and read documents, include a question(s) in a survey. Each innovation must be characterized with an area A – E; if several areas apply, select the primary area. c Inclusions and Exclusion Criteria .Include all innovations supported by CTSA functions in a defined period. Include an innovation when at least one stage (from Discovery, Development, Demonstration, Dissemination, and Adoption/Implementation) has been

Table 1. Example of a metric definition following the NCATS template. Condensed from the original	
Template Element	**Description**
	completed; do not include the same innovation again when another stage is completed (Refer to 10.3 Notes/Comments below)**.**
Unit of Analysis	Data will be collected within each hub at the innovation level and reported and aggregated at the CTSA hub level. This will include innovations occurring within the CTSA partner sites that comprise the Hub.
Scoring	This metric consists of simple counts (Refer to section 4 above).

There are many template definition templates (some CTSA hubs have their own). Table 2 gives some examples of template elements.

Table 2. Metric template elements		
Infotech	**Metripedia**	**Elements from various sources**
Metric # Name [Title] Purpose [Rationale] Creation Date Data Source(s) Collection frequency Metric Owner Customer(s) Process for collection/ analysis/ reporting Process for expeditious mitigation Review frequency Historical data lifespan Link to requirements document Metrics lifecycle end date	What Does It Say? What Is High Performance? How Easily Can It Be Compared? How To Collect It? How Is It Calculated? [Formula] How Frequently Reported? Dangers, Traps, & Pitfalls How Related To Other Measures? Public Domain Sources **Simply Excellence** (just new here) Example Accessed through Business specific issues Key learning points	Type [need typology of measures] Contextual data Unit of measure Scale of measurement Basic or Primitive vs. Computed or Derived Degree of subjectivity Ease of interpretation Degree of management control Convincing for stakeholders Convincing for public
https://www.infotech.com/research/it-metrics-definition-template?c=unlock1 http://metripedia.wikidot.com/template:metric-template http://www.simplyexcellence.co.uk/media/downloads/2779-Metrics%20DefinitionTemplate.doc		

The challenge is to derive from all these templates one comprehensive ontology which would form the intellectual basis for a database of metrics across, in the researchcontext of this paper, all CTSA hubs. The beginnings of the ontology are given in Section 3, and some notes on the implementation of such a database in Section 4.

3 Beginnings of a comprehensive metrics ontology

The format of our ontology is an entity-relationship conceptual data schema. There are no attributes (the introduction of which was a mistake), and relationships can have multiple arguments. Table 3a shows entity types (classes, universals) and Table 3b relationship types (f called properties in the RDF world). These tables should beself-explanatory.

The central entity type is CM – Construct or Measure. Since the distinction between variables and measures / metrics / indicators is not clear-cut, and since many relationship types apply to either, it is parsimonious to have just one entity type and assign to each individual CM one or more types, including the types *variable*, *metric, indicator*. RBA (Results-Based Accountability) makes an orthogonal distinction between variables and metrics that apply at the *population level* (called Results and indicators, respectively) and at the government/organization *program level* (where metrics are called performance measures). To give examples: result: people in a community have sufficient means to sustain an adequate standard of living; indicator: jobless rate; performance measure: percent of graduates of a job training program that find jobs within three months. These distinctions can also be handled by CM <*hasType*> CMType. There is a problem of handling the situation that the value of a metric such as *innovation count* depends on the geographic or organizational scope and on the time scope as well as on the decision what kinds of innovations to include.

The literature on metrics ontologies is sparse. Some documents focus on properties of metrics, for example, NIST, 2006, USAID, 2014, including units of measurement, for example Rijgersberg, 2013. There are many documents and websites on characteristics of good indicators, for example Brown, 2009. Most papers focus on variables and metrics in a specific domain, for example de los Angeles. & Olsina, n.d.; Kotenko et al. 2013; Singhal 2010; Smillie 2006, Fox, 2014; Okhmatovskaia, 2013

Table 3a. Entity types

Entity type	Sample values or comment
CM (construct or measure)	a broad entity type that includes outputs, outcomes, measures, metrics, or a group of CMs.)
CMLevel	Base (measured directly), Derived from Base, Composite index
DataCollectionMethod	This applies to base-level CMs
UnitOfAnalysis	IndividualStudy, Program, Service, LegalEntity
Interval	Monthly, Quarterly
Scale	Nominal, Ordinal, Interval, Ratio
Formula	A mathematical formula used for computing a metric
Direction	Up, Down
Power	Linear, Squared, Cubed

Agent	
FeasibilityRating	Ease and cost of tracking all data needed to compute the measure determined according to the best way to track the required data. 1 (hard, low feasibility) to 5 (easy, highly feasibility)
Significance Rating	
ImpactRating	
LegalEntity	A person or organization
DecisionType	For example, PersonnelDecision, FundingDecision
Decision	An individual decision
InformationArtifact (e.g., report)	
ProgramFunction	Program functions can be a hierarchy
SubjectDomain	This requires a taxonomy of scientific disciplines and transdisciplinary fields and other taxonomies, such as a taxonomy of diseases.
Study	
Grant	
Milestone	Each Milestone has a unique ID

Table 3b. Relationship types

CM	*<hasName>*	Text (String)	alternative name: *<hasTitle>*
CM	*<hasDefinition>*	Text (String)	
CM	*<hasType>*	CMType	CMType can express several properties of CM
CM	*<hasIssues>*	Text (String)	
CM	*<pertainsTo>* UnitOfAnalysis		
CM	*<hasIndicator>*	CM	
CM	*<operationalized*By>	CM	
CM	*<computedWith>*	Formula	
CM Direction, Power)	*<formulaIncludes>*	(CM,	direct or inverse.linear, square, …
CM	*<isUsefulFor>* (DecisionType,	LegalEntity)	Needs typology of DecisionTypes
CM	*<hasFeasibility>* *Feasibility*Rating		Scale 1 hard to 5 easy
CM	*<hasSignificance>* SignificanceRating		Scale 1 to 5
CM Decision)	<usedBy>	(LegalEntity,	
CM	*<determinedWithFrequency>*	Interval	
CM	*<includedIn>* InformationArtifact		
CM	<usesMeasure> (UnitOfAnalysis, DataCollectionMethod)		
CM	*<hasImpact>* ImpactRating		
CM	*<servesFunction>*	Function	
CM Value,	*<associatedWith>* PlannedOrActual)	(Milestone,	Value is the value of the measure that is to achievedby the milestone
Milestone	*<occursOn>*	Date	
Milestone	*<hasResponsible Party>*	LegalEntity	
CM	*<influences>*	CM	

4 Applying the metrics ontology to a database of CTSA metrics

Once completed, the metrics ontology will provide the basis for defining an SQL database structure that could store information about all these metrics without loss due do inadequate database structure, limiting access as required by the data owner. This could be a rich repository of data about metrics and thinking about metrics in the domain of the CTSA and effectiveness and impact of research in generally. It could be complemented by including information from reports of CTSA metrics working groups over time and from publications such as Dembe et al., 2014; Grether et al,. 2014; Hermann-Lingen et al, 2014; Rubio et al., 2015; and Generic Logic Model 2011. The VIVO Integrated Semantic Framework would be a useful source of variables for base data.

5 Outlook: Th broad applicability of the metrics ontology

While the metrics ontology was discussed in a specific context, it is general and can be applied in any subject area. It could even form the basis for a large repository of variables used in the sciences, particularly the behavioral and social sciences.

Acknowledgement: This work was supported in part by the National Center for Advancing Translational Sciences of the National Institutes of Health under Award Number UL1TR001412. The content is solely the responsibility of the authors and does not necessarily represent the official view of the NIH.

References

Brown, Denise (2009). Good Practice Guidelines for Indicator Development and Reporting. In *Third World Forum on 'Statistics, Knowledge and Policy*. 27-30 October 2009 Busan, KOREA. [https://www.oecd.org/site/progresskorea/43586563.pdf]

de los Angeles Martin, Maria & Olsina, Luis (nd). *Towards an ontology for software metrics and indicators as the foundation for a cataloging web system*. http://users.dsic.upv.es/~west/iwwost03/articles/Olsina%20IWWOST%202003.pdf

Dembe, Allard E. et al (2014). The translational research impact scale: Development, construct validity, and reliability testing. *Eval Health Prof,* 37(1): 50-70.

Fox, Mark S. (2014). *A foundation ontology for global city indicators*. Global Cities Institute Working Paper No. 3. [https://rd-alliance.org/sites/default/files/GCI_Ontology_v7.pdf]

Generic Logic Model for CTSA Community Engagement Research (2011) [rg/download/attachments/6357025/Rainwater+Logic+Model+Handouts+for_Group062111.pdf?version=1&modificationDate=1309384525000]

Grether, Megan et al. (2014). New metrics for translational research.

Hermann-Lingen, Cristopher et al. (2014). Evaluation of medical research performance – position paper of the Association of the Scientific Medical Societies in Germany (AWMF). *German Medical Science*, 12, np.

Kotenko, I. et al. (2013). The ontology of metrics for security evaluation and decision support in SIEM systems. In P*roc. ARES '13: Eighth Internat. Conf. on Availability, Reliability and Security*. Pp. 638-645. [http://tinyurl.com/z22jaek]

NIST. Info. Technology Laboratory, Software and Systems Division. SAMATE 2006. *Metrics and Measures*. [https://samate.nist.gov/index.php/Metrics_and_Measures.html]

Okhmatovskaia, Anya (2013). *Health Indicators Ontology* (last edited 2013-09) [http://surveillance.mcgill.ca/wiki/Health%20Indicators%20Ontology]

Rijgersberg, Hajo; van Assem, Mark; Top, (2013).Ontology of units of measure and related concepts. *Semantic Web* 4(1): 3-13. [www.semantic-web-journal.net/sites/default/files/swj177_7.pdf]

Rubio, D.M. et al. (2015).Developing common metrics for the clinical and translational science awards (CTSAs): Lessons learned. *Clin Transl Sci*, 8(5): 451-9.

Singhal, Anoop. Ontologies for Modeling Enterprise Level Security Metrics. In *CSIIRW '10*, April 21-23, Oak Ridge, Tennessee, United States.

Smillie, Robert. *Human Systems Performance (HSP) Assessment Capability Metrics Repository*. 2006-04-12. PowerPoint slides. Pacific Science & Engineering Group. (nd).

USAID 2014. *Metric reference sheet template. An additional help for ADS Chapter 597*. New ed. 2014-06-09. [www.usaid.gov/sites/default/files/documents/1868/597sae.pdf]

VIVO Integrated Semantic Framework [https://wiki.duraspace.org/display/VIVO/VIVO-ISF+Ontology]

Aarti Jivrajani, K. H. Apoorva and K. S. Raghavan

Ontology-based Retrieval System for Hospital Records

Abstract

Ontologies are being seen as important knowledge organization tools with a role to play in, organizing, representing and assigning meaning to data and information. This paper reports an experiment that was carried out to transform a database of patient records in a large hospital into an ontology. The paper also seeks to demonstrate how relevant open data (LOD) could be linked to the knowledge base to enhance its value and utility from an information retrieval point of view. Some of the issues that were encountered in the process of building the ontology are highlighted.

Introduction

The notion of *information structures* has received attention in different disciplines – library and information science, linguistics, computer science – to mention a few. In a general sense an information structure is essentially the way information is packaged keeping in mind its intended use and users. If we consider communication or transfer of required information relative to some specified need expressed as a query as defining the need for and purpose of structuring information, it is not difficult to realize why structuring information has been at the core of most information-related activities. Different kinds of information structures have been in use: structured databases, semi-structured texts, web pages, etc. Among others one of the reasons behind building structure into a document is to make it easier for the user to identify, retrieve and extract information / knowledge from the information structure. Knowledge organization systems (KOS) of all kinds and varieties developed by the LIS community are essentially information structures in this sense. Databases are information structures that store large volumes of data about a domain.Databases are used for data management in practically every domain. Technology has enabled quick access to more information and data than can be comprehended and managed effectively. Given this environment of *information overload* new techniques are being experimented with for handling and structuring information. Ontologies are being seen as important knowledge organization tools with a crucial role to play in capturing, organizing, representing and assigning meaning to information and data. However, despite the promise of enabling intelligent applications, ontology development has largely remained within the realm of research. A domain ontology could be seen as a logic-based conceptual data model of the domain; an important feature is the emphasis on accurate and semantically rich representation of concepts, properties and relations that exist in the domain resembling how they are actually seen and employed in the real world. Therefore, in comparison with other data models, ontologies are believed to bridge the semantic gap between the real world and its representation in an information system. Recognizing the value and utility of ontologies in data and information management, some experiments have been carried out and reported in the literature

(Kavi Mahesh, 2013; Kavi Mahesh and Pallavi Karanth, 2012). Ontology modeling deals primarily with describing in a declarative and reusable way the domain information. The use of ontology is believed to make it easier to share information within as well as between organizations. There is no one standard way of building a domain ontology. When a database and its associated schema are available for a specific domain it is helpful to conceive the design of the ontology from the available database schema through a process of generalization. In this paper we describe an experiment that was carried out to transform an existing database of patient records in a hospital into an ontology to support intelligent information retrieval.

Background

Most large hospitals create and maintain patient records. A patient record is a unique document and is essentially a collection of recorded data, information, signs, and images based on conjectures and facts related to health of the patient and the care that the patient is given. A patient record is intended to be a permanent documentation of a patient's health and is an invaluable source of information in healthcare and delivery as it assists healthcare personnel by contributing to improvements in making diagnoses and providing treatment. The value of a patient record is also to be seen in terms of its utility as a reference tool for crosschecking if patients with a problem similar to the one under examination had been treated in the hospital at some other point of time. Patient records, thus, could be valuable work tools for healthcare personnel not only for monitoring of patients, but also for improving the quality of prescriptions and reducing the consumption of certain medications to other patients by promoting the adoption of tested and clearly defined procedures. In practice, however, the useful information that is stored in these records too often does not get used as effectively as it should for want of effective systems for organization of the knowledge contained in these to support their search and retrieval. Healthcare and delivery is a multidisciplinary task involving different groups of healthcare professionals. Often these groups speak different languages and as a result the terminology of a patient record may reflect the practices of the different groups. Research has shown the utility and value of knowledge organization systems as switching mechanisms for effective knowledge sharing and communication between different groups working with a *boundary object* such as patient records. (Shepherd and Sampalli, 2012) Since it is now common practice for hospitals to maintain structured electronic records, it is important to explore how these electronic records could be organized.

Objectives of the Study

In this paper we look at structured patient records of a hospital to:

a) Explore the feasibility of transforming the information / data in the electronic patient records into an ontology to support intelligent retrieval;

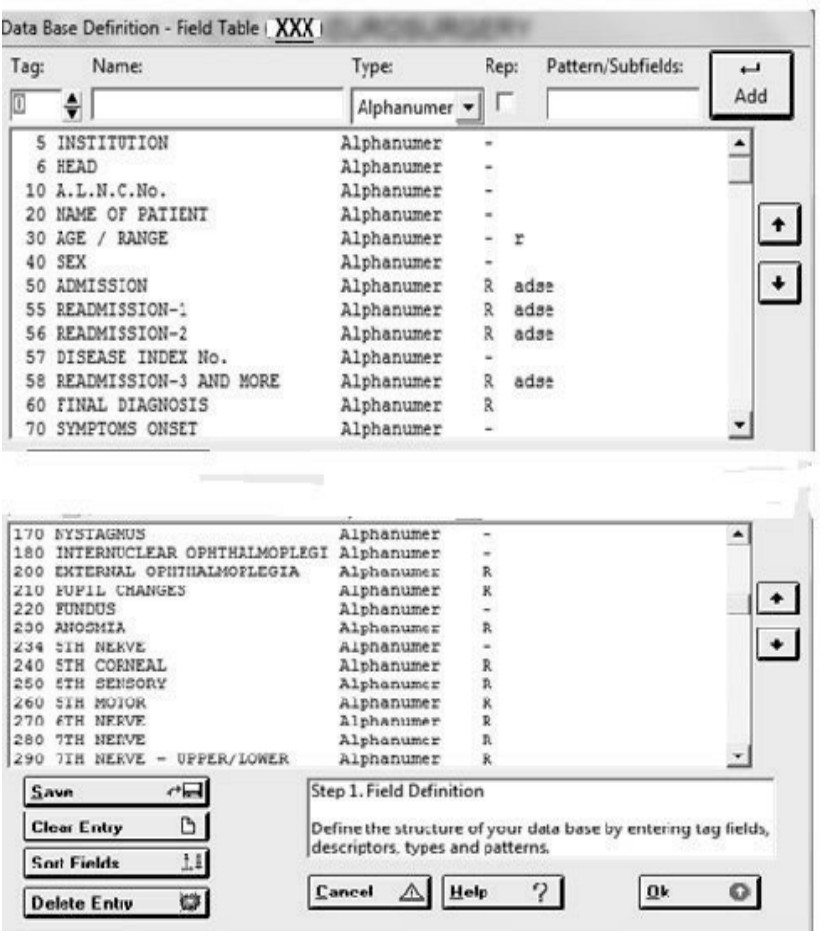
b) Explore how the evolving dynamic and flexible information environment could be exploited to semantically enrich knowledge structures.

Data and Methodology

This study is exploratory in nature. However, the results suggest that the approach can be extended by continued research; the present work serves as a baseline for future work.

Over 1000 electronic patient records maintained in the Neurology unit of a large hospital constituted the source data for the experiment reported in this paper. The source database maintained using the text retrieval software WINISIS, is a comprehensive database with over 100 data elements (several of them repeatable fields) for every patient admitted and treated at the hospital. The database has been in use at the hospital for several years and right from the inception the effort has been to gather, store and maintain as comprehensive information about every patient as possible. A rough idea of the kind of data collected and maintained (not exhaustive) by the hospital's unit can be obtained from the figure 1. The patient records database is quite extensively used in the concerned hospital and inputs – both creating new patient records as also updating of available records -- are done by a number of different healthcare personnel including physicians, surgeons, radiologists, physiotherapists, nurses, etc. As one of the objectives of this research was to explore the feasibility of transforming the data in the structured textual database into an ontology, a few of the data elements were identified for building the ontology. Since the data available in the database was quite detailed, it was not felt necessary, at least for the purposes of this experiment, to design a new data model for the proposed ontology. However, a few changes to the schema vocabulary were made. The selection of data fields for inclusion as classes in the ontology was made on the basis of the kind of questions that the knowledge base was expected to respond to. For the selected fields an ontology was defined using Protégé 4.3 and a knowledge base containing the knowledge originally recorded in the patient records maintained by the hospital was created. The ontology was populated with data from the WINISIS database. Since this was primarily an exploratory study, only a couple of dozen patient records were used for populating the ontology.

Figure 1: Illustrative List of Data Elements in the WINISIS database



The Ontology

Figure 2 presents the overall design of the ontology in terms of kinds of classes and properties (illustrative and not exhaustive) that were considered in developing the ontology.

Figure 2: Visualization of the Structure of Ontology

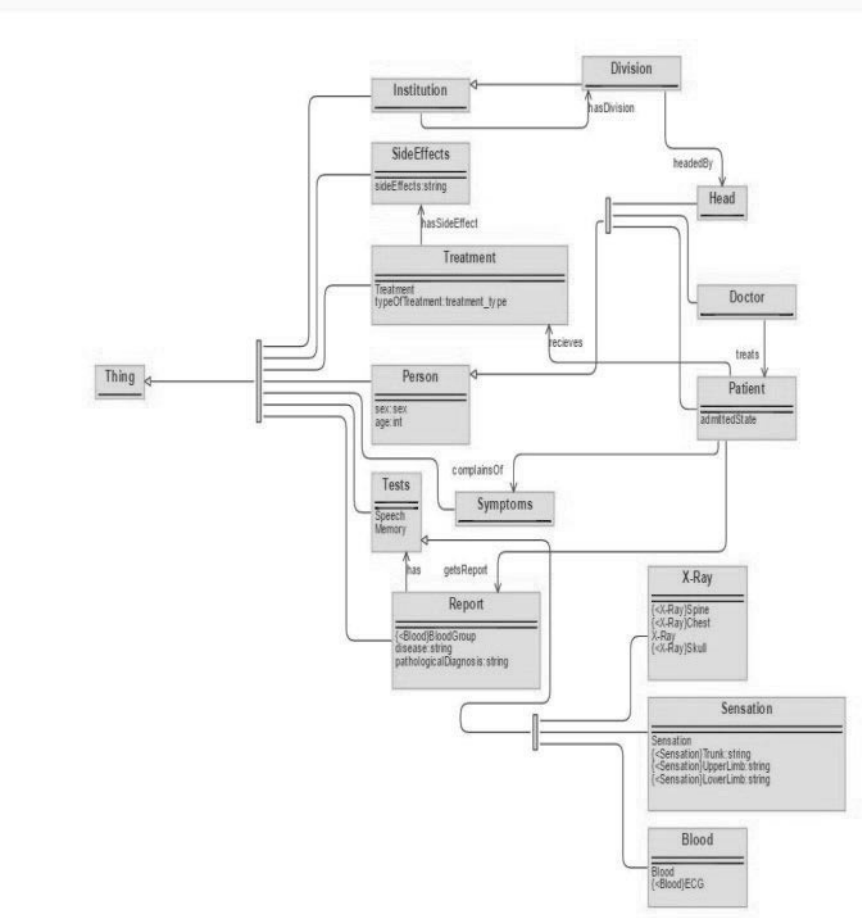
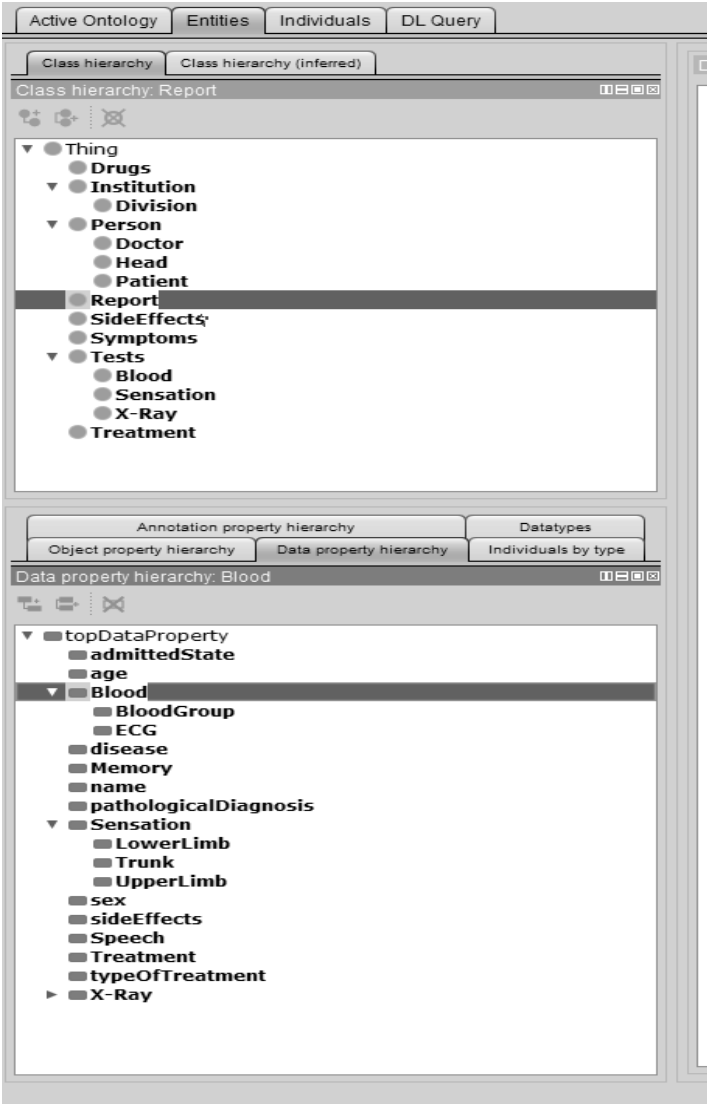


Figure 3 indicates some of the classes and properties that were defined for the ontology. This research sought to establish that an ontology has advantages over conventional databases even in an offline environment such as the one that is the setting for this study; in other words to demonstrate the ability to allow querying of the ontology which are not effectively handled by databases. The querying for knowledge from the knowledge base could concern general classes of concepts and properties between instances of these classes. For example, the present ontology could be queried for records or names of all '*patients*' of 'Dr.X' who underwent 'surgery' for *brain tumor*. The query could be more complex and specify restrictions on the class '*Patient*' in terms of '*gender*' and 'belonging to a certain age group'. For example, records of 'Diabetic Women patients aged between 50 and 60 having *Meningioma* and symptom of *speech disorder*.

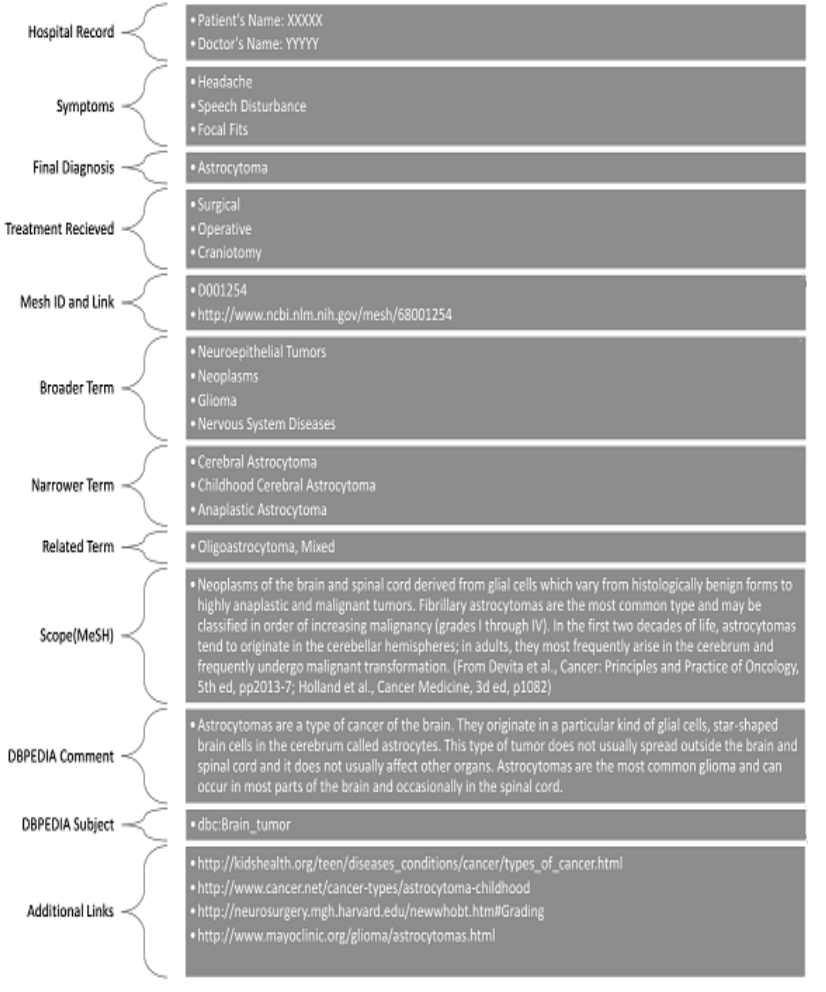
Figure 3: Some Classes and Properties in the Ontology



Another important objective of the study was to demonstrate the capability to exploit open linked data on the Web. Several discipline-specific as also multi-disciplinary open datasets are available on the Web. Linked Data builds on browsable links (URIs) spanning a seamless information space. Datasets on the Web using RDF and URIs constitute a global information graph that users can seamlessly browse. Just as an experiment to demonstrate what can be done in terms of enhancing the quality and nature of information retrieved by exploiting linked data, a couple of concepts in

the knowledge base were used to query dbpedia and from dbpedia the MeshID for the concepts were obtained which in turn was used to query Mesh endpoint. Querying such linked open datasets is relatively simple using a SPARQL Endpoint thus enabling users to query the knowledge base returning results. The result of a search for '*Astrocytoma*' with extracts from different datasets including open datasets is presented in figure 4. The semantic richness of the output is quite evident.

Figure 4: Search Output



Findings and Suggestions

The major findings of the study are:

a. Ontologies offer several advantages over conventional databases in information retrieval.
b. Ontologies can support complex queries as compared to the more conventional text retrieval systems;
c. Given the availability of several open data sets on the Web, it is important for information support systems to take advantage of facilities to link to open data sets.
d. A major issue that was faced in the present study while transforming patient records in the database to an ontology was the fact that there were in this will be the significant differences between the vocabularies employed by different healthcare personnel to refer to a certain medical situation, health problem, symptom, etc.
e. There were also differences between the vocabulary used by the healthcare personnel who created the patient records in the hospital that suggested by MeSH;

Acknowledgement: This work was supported in part by the World Bank / Government of India grant under TEQIP Programme to the Centre for Knowledge Analytics & Ontological Engineering at the PES Institute of Technology, Bangalore, India.

References

Mahesh, Kavi (2013). Ontology-based content management. In *ICDL 2013:Proceedings of the International Conference on Digital Libraries.* New Delhi*:* TERI

Mahesh, Kavi and Pallavi, Karanth (2012). A new knowledge organization scheme for the Web: Superlinks with semantic roles. In *Categories, contexts and relations in knowledge organization* edited by A. Neelameghan and K. S. Raghavan. Wurzburg: Ergon Verlag. Pp. 90-5

Shepherd, Michael and Sampalli, Tara (2012). Ontology as boundary object. In *Categories, contexts and relations in knowledge organization* edited by A. Neelameghan and K. S. Raghavan. Wurzburg: Ergon Verlag. Pp. 131-7

Ana M. Cunha, Carlos H. Marcondes, Joyce Messa, Monnique S. P. A. Esteves, Nilson Theobald Barbosa, Rosana Portugal and Tatiana de Almeida

Proposal of a General Classification Schema for Museum Objects

Abstract

The current schema used by the Rio de Janeiro`s Museum Network website to classify the museum objects in different museums' collections contains 16 categories that are no longer enough to encompass all the collections of museums about to adhere to the network. These new collections include scientific and intangible cultural heritage objects that needed to be fitted in categories of their own. In order to expand the classification schema, an Ontological approach was used, as well as the Aristothelic classification theory, to analyze and distinguish the different types of museum objects, define new categories and clarify the present ones, including them on the new broadened schema proposed, guaranteeing compatibility with museums already connected to the network. The categories suggested include a broader one, Museum objects, which contains Natural objects (subdivided in Inorganic and Organic objects) and Physical or conceptual products of human culture- Man-made objects (comprising Material culture objects or Artifacts and Conceptual products of human Culture – the first one containing all 16 pre-existing categories); and a new broad category for Cultural heritage objects. This proposal constructs a broader schema then the one in use, while encompassing it and allowing the insertion of any new categories that may appear in the future.

1 Introduction

The Web Museum Network of the state of Rio de Janeiro, Brazil, holds a website where users can search for records and images of museum objects available in different museum collections (http://www.museusdoestado.rj.gov.br/sisgam/). To support transversal searches over objects of the same type in different museum collections, the Web Museum Network uses a classification schema of broad object categories based on Ferrez and Bianchini (1987). This is a pioneering and established schema used in Brazilian museums, holding 16 categories: 1. Hunting and War, 2. Visual Arts, 3. Pecuniary objects (coins, etc), 4. Building, 5. Interior decoration objects, 6. Work, 7. Recreation, 8. Insignia, 9. Ceremonial objects, 10. Communication, 11. Transport, 12. Person al objects, 13. Penance and torture objects, 14. Measurement, recording, 15. Packing, 16. Samples, fragments. This schema is based on what a museum object is,its ontological nature. Although there are other specific facets by which museum objects may be classified as (type of material or technique, style, etc), these facets only do not apply to all existing categories of objects. The ontological facet is the most general and could be virtually applied to any museum collection to adhere to the Web Network. As the Web Network expands, by the adherence of several new museums holding scientific collections and intangible cultural heritage, it becomes necessary to expand the original schema to encompass the new categories of objects.

This paper addresses the following questions. How to update (reengineering) Ferrez and Bianchini’s (1987) museum object classification schema to include categories such as those needed to integrate and classify scientific collections, and intangible cultural

heritage objects comprising museum collections that adhere to the Web Network? What are the objects in Ferrez and Bianchini's museum object classification schema categories? Is there a category or categories that can subsume all or some of these categories? What is a museum object and in what general categories should it be subdivided?How to integrate the original the original Ferrez and Bianchini's museum object classification schema with such a general schema? How to integrate other museum thesaurus used in Brazil within the proposed general museum classification schema?

The objective of this paper is to propose a general museum objects classification schema that expands the original one proposed by Ferrez and Bianchini (1987) with the aim of support scientific collections and intangible cultural heritage objects, alongside the sixteen categories of the previous schema.

The paper is organized as follows: after this Introduction, section 2 presents methods and material used. Section 3 presents theoretical and methodological bases used. Section 4 presents the definitions collected and used as input to expand and complement the original classification schema. Section 5 presents and discussed the final classification schema. Section 6 presents the concluding remarks and future directions of research.

2 Methods

Literature and different museum classification schemas were used as sources for categories that could subsume the sixteen original ones and could hold both scientific and intangible cultural heritage collections. Definitions of museum object, artifact, natural objects, among others, were also collected. Definitions of intangible cultural heritage manifestations were collected from official documents concerning Brazilian policy for intangible cultural heritage (Cavalcanti & Fonseca, 2008). These definitions were used as inputs to define classes of entities.Ontological analysis (Guarino and Welthy, 2000, 2009), conceptual definitions (Dahlberg, 1981, 1983) and Aristothelic classification theory were then used as methodology to analyze and distinguish the different types of museum objects in order to classify them into a unique general schema.

3 Theoretical bases

Ontology as a philosophical inquiry is the science of what is, of all kinds of beings, their properties and relations in all domains of reality. It aims at answering questions such as: What is? What types of entities exist? What are their differences, what are their similarities? (Welthy & Smith, 2001, p. 2; Grenon & Smith, 2004, p. 138; Guarino, 1997, p. 1).

Within the scope of Semantic Web, ontological analysis have been used as a tool to formally model the knowledge of different domains and record it in artifacts such as computational ontologies, thus enabling computers to reason on this knowledge.

Specifically ontological analysis seeks to identify “formal distinctions between the elements in a domain, independently of their actual reality” (Guarino, 1997, p. 1).

Accordingly, ontological analysis tries to answer the following questions concerning all aspects of reality or a specific domain:What is it? What types of entities exist? What are their differences, what are their similarities? What are the properties that define an entity to be this specific entity? What is the difference of essential properties, accidental properties and observed-relative properties? What is implicit of is an assumption for something to exist? What makes something a whole, what makes it a part? What entities are independent, what are dependent, of what entities? When and under what conditions one entity begins to exist, evolves and ceases to exist? What entities precede the existence of other entities?

In order to answer these questions, ontological analysis uses theoretical and methodological tools, the meta-properties, which can be applied to the classes and relationships that comprise the taxonomic backbone of a knowledge organization system.These meta-properties are: Identity, Dependence, Essentiality and Integrality (Guarino & Welthy, 2000, 2009). Of these meta-properties, Identity is the most important as, according to the definitions of museum object collects, we are dealing with objects that maintain a persistent identity throughout all their existence. We can build the backbone of the proposed schema by assigning properties that assure identity to their instances; these properties are the different object types.

Another relevant methodological and theoretical contribution to ontological analysis is Searle’s (1995) theory of the process of social construction of reality, in which features of objects are socially attributed/added and became embedded within their essence.

Here ontological analysis was used to precisely identify what kind of entity is a museum object and within what types of entities could this class be subdivided, as we can see below:

> Applied to the problems of knowledge organization this means that any entity with which we are dealing ought to be understood and described ‘according to extension and intension’, in other words, the concepts existing in our minds, books, text, and discourses are more or less concealed and must be made explicit by adequate methods. (Dahlberg, 1992, 69).

To achieve this objective we seek for definitions of different museum objects. Dahlberg in his Referent-oriented, Analytical Concept Theory highlights the role of definitions in knowledge organization. She identifies three kinds of definitions, namely partitive, functional and generic. Generic definitions are building by declaring the “nearest genus” and the “differentiae” from this “genus”, an essential and unique characteristic holding for the “differentiae” but not for the “genus”. As a consequence an exclusive class may be defined by declaring the “nearest genus” and a property that do not hold for that “differentiae”.

Since Aristotle (Berg, 1982) definitions are strongly related to classification. As stated by Dahlberg (1981, 19):

> If the genus proximum is said to be an essential characteristic, then it is also the genus proximum of the genus proximum and so on until one reaches the ultimate category of a genus supremum and thus creating a hierarchy of genera proxima.

A "genus supremum" is what Dahlberg considers "Form-categorial relationships [which] help to distinguish and define concepts according to their form classes of being Objects, Properties, Activities, Dimensions (space, time, position)". (Dahlberg, 1992, 67).

Dahlberg (1981) suggests that, in order to construct concept systems, concept definitions within a domain must be collected, formalized or constructed and then used as inputs to systematization.

4 Definitions Collected

The meaning of a term within a knowledge domain is established by a definition statement. Definitions can explicit characteristics, functions, constitutive elements of a term, thus delimiting its semantic in this specific context (Campos, 2010). Accordingly, different sources such as literature, specialized dictionaries and thesaurus were consulted, seeking for formal or informal definition, aiming to clearly define the meaning of the categories proposed as an extension of the original schema.

- Museum object

> "the object-oriented methodology has recently received a good deal of attention among museologists. This approach met considerable support within the International Committee for Museology. The museum object is considered to be the basic unit of the museum working procedures" (Van Mensh, 1992, 67).
>
> "Museum objects are objects separated from their original (primary) context and transferred to a new, museum reality in order to document the reality from which they were separated." (Van Mensh, 1992, 104).
>
> "As documents museum objects (in the sense of primary museum material) are direct (authentic) witnesses of cultural and natural phenomena." (Van Mensh, 1992, 106).

Museum objects are "ontologically coincident with objects in general, but as to their semantic, they have a new function, i.e. the function of authentic witnesses, documents, and/or the testimony of natural and social facts" (Stransky 1985, 98).

Accordingly we can claim that museum objects have a dual nature, they are primary objects (natural or man-made) in addition to artifacts – descriptions of the primary object with the aim of adding a semantic function and enrich its role as documents andtestimony of natural and social facts. As documents the characteristics assigned, added or highlighted are dependent on the natural or social relevance of the specific object, a curator choice. Therefore, due to the different types of museum objects, some characteristics are assigned to all types of objects, other just to some types. The object facet is one of those characteristics that may be assigned to all types of objects.

- Artifacts

Borgo and colleagues (2009, 1) define

> […]technical artifacts are objects that exist by human intervention; and that technical artifacts are to be contrasted to natural entities. Yet the perspectives are different in the way they spell out these intuitions: the relevant human intervention may range from intentional selection to intentional production.

Hilpinen (2011) proposes a synthetic definition: “an artifact may be defined as an object that has been intentionally made or produced for a certain purpose”.

Within the artifact category fit, with minor changes, the original sixteen categories of Museum Collections Thesaurus, thus assuring the compatibility with museums that already uses the old schema.

- Natural X Man-made objects

Encyclopaedia Britannica's definition of Life [1] is a clue to distinguish between natural objects and man-made ones “Life, living matter and, as such, matter that shows certain attributes that include responsiveness, growth, metabolism, energy transformation, and reproduction”.

Baker’s claim shows the relevance and the extension of Artifacts Category among museum objects. According to this author

> Artifacts are objects intentionally made to serve a given purpose. The term ‘artifact’ applies to many different kinds of things − tools, documents, jewelry, scientific instruments, machines, furniture, and so on. Most generally, artifacts are contrasted with natural objects like rocks, trees, dogs, that are not made by human beings (or by higher primates). The category of artifact, as opposed to the category of natural object, includes sculptures, paintings, literary works and performances (Baker, 2004, 99).

Distinctions between natural objects and those made by man are also made by the CIDOC Conceptual Reference Model. Its hierarchy of classes makes an “*a priori distinction*” (Guarino, 1995, 5) from the class E 70 Thing and one of its subclass E71 Man-made Thing, which comprises “Everything that is not natural” (Oldman, & Labs, CRM, 2014, 9). The British Museum Materials Thesaurus has as its three Top terms (or Categories): “Organic”, “Inorganic” and “Processed Material” [2]. The Art and Architectural Thesaurus, Getty Foundation [3], makes a distinction between Man-made objects and Natural objects.

Different knowledge organization systems make a clear differentiation between objects and processes (CIDOC CRM, 2013, SUMO [4]), or what is called continuants and occurents (BFO [5]), endurants and perdurants (THE WONDERWEB LIBRARY OF FOUNDATIONAL ONTOLOGIES, 2003), SNAP and SPAN (Grenon and Smith, 2004). This differentiation concerns the modes of existence in time of entities. Objects are entities that maintain their identity during all their existence; processes happen during their existence. Processes are associated with, or depend on, objects. For example, the IALTA Conference at the end of World War II which decided the destiny of Europe, is a process. It has objects – actors -such as the prime minister of Soviet Union, Stalin, the prime minister of United Kingston, Churchill, and the president of United States, Roosevelt, as participants; and it occurred inside an object, a place, the city of Ialta, Crimea.

Once museum objects are separated from their original context, collected, guarded, preserved and exhibited with the intention of being testimonies of relevant natural and social phenomena during large periods of time without changing their properties, we can reasonably consider them as objects, or continuants, or endurants, or SNAP

entities. Processes, as the historical process of Ialta Conference, however, due to their inherent temporal characteristics, can only be “musealized” if they are registered as objects, for example, by taking and preserving a photo of the Conference, or its proceedings.

- Intangible cultural heritage

According to UNESCO [6]:

> Intangible or immaterial cultural heritage encompasses life expressions and traditions that communities, groups and people from all over the world inherit from their ancestors and pass their knowledge to their descendants. Besides sound and video recording, and archives, UNESCO considers that one of the most effective ways of preserving intangible heritage is to ensure that the bearers of this heritage can continue producing it and transmitting it”.

UNESCO also enumerates different expressions of: “Intangible cultural heritage: oral traditions, performing arts, rituals [7].

In recent decades UNESCO enlarged the meaning of the term ‘cultural heritage’ beyond traditional monuments and object collections, including also

> […] traditions or living expressions inherited from our ancestors and passed on to our descendants, such as oral traditions, performing arts, social practices, rituals, festive events, knowledge and practices concerning nature and the universe or the knowledge and skills to produce traditional crafts [8].

5 Results

The most general category that subsumes all the other should be “Museum Object” which classical definition is based on musealization as a cultural and value-added process that separates an object from its original physical, functional and cultural context with the aim of representing or recording that aspect of the reality.

Subsumed to this general category are two other: Physical or conceptual products of human culture and Natural objects. To this last category are subsumed Organic objects, those which have their origin in living beings, and Inorganic object. These two categories will support history/natural history sciences museums and herbariums, etc. Within the Physical or Conceptual products of human culture category are the categories Material Culture Objects or Artifacts. Examining the scope notes and the subclasses of the original sixteen categories of Museum Collections Thesaurus is proposed that they fall, with minor changes, within this last category, thus assuring the compatibility with museums that already use the old schema. The new category Intangible cultural heritage highlights in its definition the need to record these manifestations (Cavalcanti & Fonseca, 2008); indeed, recording and documentation are prerequisite to the musealization of these manifestations.

The resulting schema can be presentedas follows.

- Museum objects
  - Natural objects
    - § Inorganic objects (originally)
    - § Organic objects (originally)
  - Physical or conceptual products of human culture - Man-made objects
    - § Material culture objects or Artifacts
      - · Ferrez and Bianchini 16 Categories
    - § Conceptual products of human Culture
      - · Cultural heritage objects (records)

6 Conclusion

The resulting general schema inherits the object facet from Ferrez and Bianchini's schema as its systematization principle. Currently the object facet is a common fact to all museum object collections. This feature helps different thesaurus used by Brazilian museums to be integrated to the general schema by their respective object facets.

See, for example, the Tesauro de Cultura Material dos Índios no Brasil - Thesaurus of Brazilian Indigenous Material Culture (by its Artifact facet); the Thesaurus de Acervos Científicos em lingua portuguesa – Scientific Instruments Collections Thesaurus in Portuguese -(http://thesaurusonline.museus.ul.pt/hierarquica.aspx, by all of its Categories: scientific instruments, experiments and demonstration instruments, machines, reference objects, andutensil); and also the Tesauro de Folclore e Cultura Popular – Folclore and Popular Culture Thesaurus – (http://www.cnfcp.gov.br/interna.php?ID_Secao=30) by its Artifact Categorie. The proposal here presented do not aims at being exhaustive but just to propose a broad schema that encompasses the old one by Ferrez and Bianchini (1987) and also any others that may be used by the new museums adhering to the Web Network, including different types of museum objects.

Notes

[1] Avalilable at: http://global.britannica.com/topic/life. Access Apr 21 2016.
[2] Avalilable at: http://www.collectionstrust.org.uk/assets/thesaurus_bmm/matintro.htm
[3] Avalilable at:http://www.getty.edu/research/tools/vocabularies/aat
[4] Available at http://www.adampease.org/OP
[5] Available at http://ifomis.uni-saarland.de/bfo
[6] Avalilable at:http://www.unesco.org/new/pt/brasilia/culture/world-heritage/intangible-heritage
[7] Available at: http://www.unesco.org/new/en/culture/themes/illicit-trafficking-of-cultural-property/unesco-database-of-national-cultural-heritage-laws/frequently-asked-questions/definition-of-the-cultural-heritage
[8] Avalilable at: http://www.unesco.org/culture/ich/en/what-is-intangible-heritage-00003)

References

Baker, Lynne Ruder (2004). The Ontology of Artifacts.*Philosophical Explorations,* 7, 99-112. http://people.umass.edu/lrb/files/bak04ontM.pdf

Berg, Jan (1982). Aristotle's theory of definition. *ATTI del convegno internazionale distoria della logica,*19-30. http://ontology.buffalo.edu/bio/berg.pdf

Bloom, Paul (1996). Intention, history, and artifact concepts. *Cognition*, 60, 1-29.htttp://www.philosophy.dept.shef.ac.uk/hangseng/readinggroups/media/bloom.pdf

Borgo, Stefano, Franssen. Maarten, Garbacz., Pawel, Kitamura, Yoshinobu, Mizoguchi Riichiro & Vermaas, Pieter E. (2014). Technical artifacts: An integrated perspective. *Applied Ontology,* 9, 217-35. doi: 10.3233/978-1-60750-785-7-3

Borgo, Stefano & Vieu, Laure (2009). Artefacts in formal ontology. *Handbookof philosophy of technology and engineering sciences,* 273-308. ftp://ftp.irit.fr/IRIT/LILAC/BV-HBPT09.pdf

British Museum Materials Thesauros. (1997). doi: htttp://www.collectionconstruct.org.uk/assets/thesaurus_bmm/matintro.htm.

Campos, Maria Luiza de Almeida (2010). O papel das definições na pesquisa em ontologia. *Perspectivas em Ciência da Informação*, *15*(1), 220-38. https://dx.doi.org/10.1590/S1413-99362010000100013

Cavalcanti, Maria Laura Viveiros de Castro & Fonseca, Maria Cecilia Londres (2008). *Patrimônio imaterial no Brasil: legislação e políticas estaduais.* Brasília, Brazil: UNESCO. http://unesdoc.unesco.org/images/0018/001808/180884POR.pdf

Oldman, Dominic & Labs, C. R. M. (2014). The CIDOC Conceptual Reference Model (CIDOC-CRM): PRIMER. http://83.212.168.219/CIDOC-CRM/sites/default/files/CRMPrimer_v1.1.pdf

Definition of the CIDOC Conceptual Reference Model, version 5.1.12. (2013).

http://www.cidoc-crm.org/docs/cidoc_crm_version_5.1.2.pdf

Cleland, C. E., Chyba, & C. F. (2002). Origins of Life and Evolution of the Biosphere. *NASA's Astrobiology Magazine,* 32, 387-93.

Dahlberg, Ingetraut. (1981) Conceptual definitions for INTERCONCEPT. I*nternational Classification*, 8, 16-22.

Dahlberg, Ingetraut (1983). Terminological definitions: characteristics and demands. In: *Problémes de la définition etde la synonymie en terminologie*. (pp.13-51).Quebec, Canada: Girstern.

Ferrez, Helena Dodd & Bianchini, Maria Helena S. (1987).*THESAURUS para acervos museológicos.* Rio de Janeiro, Brazil: Fundação Nacional Pró-Memória.

Grenon, Pierre & Smith, Barry (2004). SNAP and SPAN: Towards dynamic spatial ontology. *Spatial cognition and computation.*. 4, 69-104. htttp://ontology.buffalo.edu/smith/articles/SNAP_SPAN.pdf

Guarino, Nicola & Welty, Cristopher (2000). Identity, unity, and individuality: Towards a formal toolkit for ontological analysis.*Proccedings of the 14th European Conferenceon Artificial Inteligence. 219-223.* Berlin, Germany: IOS Press.

htttp://pdf.aminer.org/000/165/249/identity_unity_and_individuality_towards_a_formal_toolkit_for_ontological.pdf

Guarino, Nicola & Welty, Cristopher A. (2009). An Overview of OntoClean. In: *Handbook on ontologies* (pp. 201-220).Berlin, Germany: Springer BerlinHeidelberg. http://www.loa.istc.cnr.it/Papers/GuarinoWeltyOntoCleanv3.pdf

Hegenberg, Leonidas. (1974). *Definições: termos teóricos e significado*. São Paulo, Brazil: Cultrix.

Hilpinen, R. (2011). Artifact. In: E. N. Zalta (Ed.), *The Stanford Encyclopedia of Phiilosophy*.http://stanford.edu/archives/win2011/entries/artifact

Van Mensch, Peter (1992). Toward a methodology of museology. *Unpublished Ph.D. dissertation*. Zabreb, Croatia: University of Zagreb. http://xa.yimg.com/kq/groups/23466284/1995686355/name/Towards

Miller, Seumas (2007). Artefacts and collective intentionality. *Techné: Research in Philosophy and Technology,* 11, 1-7. htttp://scholar.lib.vt.edu/ejournals/SPT/v9n2/miller.html

Searle, Jhon. (1995). The construction of social reality. New York, United States of America: The Free Press.

Thomasson, Amie. (2007). Artifacts and human concepts. In: S. Laurence and E. Margolis, (Eds), *Creations of the mind: Theories of artifacts and their representation* (pp.52- 73). Oxford, England: Oxford University Press.

THE WONDERWEB LIBRARY OF FOUNDATIONAL ONTOLOGIES. (2003). Preliminary Report. MASOLO, C., BORGO, S., GUARINO, N., OLTAMARI, A., SCHNEIDER, LISTR-CNR, Padova, Italy. [http://wonderweb.semanticweb.org/deliverables/documents/D18.pdf.]

Rick Szostak

Employing a Synthetic Approach to Subject Classification across Galleries, Libraries, Archives, and Museums

Abstract

There has been much interest in recent years in developing a subject classification system that could be utilized across galleries, libraries, archives, and museums (GLAM). This paper will argue that the solution lies in a synthetic approach to classification grounded in basic concepts: those for which there is broadly shared understanding across communities. The paper analyses the classification systems utilized in archives, galleries, and museums, shows how a synthetic approach using basic concepts would enhance user access while easing classificatory challenges. It then applies a synthetic approach to samples of archival documents, museum artifacts, and works of art. It is hoped that this exercise both establishes the feasibility of a synthetic approach and identifies strategies for pursuing this approach across GLAM.

Introduction

There has been considerable interest among both scholars and curators in recent years in developing a subject classification system that could be utilized across galleries, libraries, archives, and museums (GLAM). The main motivation is an increased appreciation that users often search across multiple components of GLAM, especially as these various institutions develop an online presence. And the task of developing an online presence forces the GLAM sector to confront issues of classification. Yet galleries, archives, and museums have displayed little interest in the subject classification schemes employed in libraries.

The argument of this paper is simple but powerful: The goal of achieving interoperability across GLAM is best achieved by a synthetic approach to classification grounded in basic concepts – those for which there is broadly shared understanding across communities. Such an approach is easy for classificationist, classifier, and user to master: Classifiers and users can both proceed directly from an object description to a sentence-like synthetic subject (they could be facilitated in doing so by a general thesaurus). Simplicity in application is critical given human resource constraints across GLAM. The inherent flexibility of a synthetic approach allows each of the GLAM enterprises to signal both the general and unique attributes of each object or document. Importantly this very approach is also well-suited to the needs of the Semantic Web, and should thus allow facilitate computer navigation across the GLAM sector.

This paper analyses the classification systems utilized in archives, galleries, and museums, showing how a synthetic approach using basic concepts would enhance user access while easing classificatory challenges. It then applies a synthetic approach to samples of archival documents, museum artifacts, and works of art. It is hoped that this exercise both establishes the feasibility of a synthetic approach and identifies strategies for pursuing this approach across GLAM. [See Szostak, Gnoli, and López-Huertas 2016 for justification of a synthetic approach to subject classification in libraries.]

General Observations

There is a considerable overlap in the items held by each element of the GLAM sector. An engraved silver bowl might appear in a museum or gallery. A rare book might be held in library or archive. A poster might be archived or held in a museum or perhaps a library. An art catalogue might be held in a gallery or library or (more rarely) archive. This significant overlap in coverage provides a powerful further motive for employing the same classification across GLAM. It is likewise worth noting that scholars of art may visit galleries for sculptures and paintings but will need to consult archives for theatre programs or architectural drawings and libraries for musical recordings.

The subject for many GLAM objects has three distinct elements. The one that receives the most attention here is purpose. This is what a subject heading for library materials ideally captures: what is a book or article attempting to communicate? Art scholars tend also to focus on the ideas or emotions an artwork communicates. Museum artifacts are generally described foremost in terms of their use. Archival documents, when identified by subject, are generally classified by purpose.

The second element is material. What is an object made of? Museums generally, and art galleries often, will note the materials of which particular artifacts are constructed (including types of paint). Archival and library materials may generally be of paper, but are increasingly in digital format. Other materials, such as cloth or papyrus, deserve to be indicated. Even poor quality paper subject to decay may merit note.

The third element is manufacture: how was an object made? For a minority of museum and gallery artifacts, this element is critical. Archival documents and books could also sometimes be distinguished in this way: handwritten, typeset by hand, computer generated, and so on.

A synthetic approach is useful for all three: (axe)(for)(war);(wooden)(shaft)(steel)(head); (mass)(produced). A synthetic approach also allows these to be combined into one longer subject entry. This approach would allow the classifier to stress the most important characteristics of a particular artifact – but necessarily then encouraging one or two of the three elements to be ignored. The latter approach may also better identify artifacts whose specialness lies in unusual combinations of the three elements: (golden)(axe).

Art Galleries

There is increased focus on thematic study in art scholarship and gallery exhibitions. Surveys show that art scholars would benefit from subject access to works of art. So also would other scholars, as well as the general public. There is thus much interest in subject classification of works of art, but very limited progress. Social tagging is often recommended, but consensus here will be far more likely if a controlled vocabulary is employed.

It is common, following Panofsky, to speak of three levels of subject description. One level simply describes the main elements (woman on horse). Another level gives specifics (name of woman). The third records cultural significance (e.g. Christian parable). Existing approaches are thought to cope poorly with these, but especially the third. Yet these past efforts do signal that the subjectivity of ascribing subjects to works of art does not prevent useful classification (Szostak 2014).

Some works of art may be about a single thing or relator. But most works are better described in terms of combinations of basic concepts: (girl)(smiling) or (vineyard)(at)(sunrise). And many/most works of art will express some causal relationship: (girl)(smiling)(because)(gift).Note that verb-like terms are often important in identifying the subject of a work of art.

Classification Schemes

Categories for Description of Works of Art, developed by the Getty Museum (https://www.getty.edu/research/publications/electronic_publications/cdwa/18subject.html), is the most developed, and likely the most utilized, classification guide for works of art. “Subject” is one of many metadata elements that it recommends. Indeed it states "Indexing the subject is core. All works of art and architecture have subject matter. Subject matter is critical to any researcher of art, both the scholar and the general public. The *SUBJECT MATTER* category may include an identification, description, and/or interpretation of what is depicted in and by a work or image."

Despite expanding at length regarding the importance of subject access, the CDWA expects a lengthy written description of each item rather than a subject heading using controlled vocabulary. It does ask for a general term (landscape, funerary art) and provides some 30 of these, but allows others. It encourages references to specific people, places, occupations, and so on, and recommends that debate regarding these be indicated. It refers the classifier to a variety of controlled vocabularies for “the proper names of the following: historical events; fictional characters, places, and events; religious or mythological characters or events; literary themes; iconographical themes. An authority with hierarchical structure, cross referencing, and synonymous names is recommended.”It thus recognizes the advantages of controlled vocabulary but stops short of insisting on a particular vocabulary (It lists dozens). [Getty’s Cataloguing Cultural Objects (http://www.vraweb.org/ccoweb/cco/) takes a broadly similar approach.] Still, in recognizing the complexity of the subject matter of works of art, these resources provide some limited support for the idea of a synthetic approach to subject classification grounded in basic concepts that might be both widely accepted and able to handle complexity.

The most widely used classification of the subject of works of art is ICONCLASS (2014). Szostak (2014) described ICONCLASS in some detail. Like any enumerated scheme, it limits the ability for synthesis. Though its creators have created many compound terms they cannot provide for all possible combinations that an artist might

pursue (and ICONCLASS has a notably Western bias in coverage). It proved fairly easy to translate all ICONCLASS terminology into the synthetic format of the Basic Concepts Classification (Szostak 2013; see the Appendix to this paper). Art works can thus be classified in terms of simple compounds without the necessity of mastering a detailed enumerative scheme. Yet the synthetic approach allows a much wider variety of compounds to be classified with recourse to shorter schedules.

Individual Works of Art

The National Gallery of Canada lists highlights under a handful of categories on its website (https://www.gallery.ca/en/see/collections/category_index.php). It is thus possible to survey the first few under each of these categories.

Under 'Canadian' art, the first is 'Lady with a dog,' easily rendered as (woman)(holding)(dog). Extra detail on the woman's attire might be provided. The next is 'A saint': (statue)(saint)(man). One could add (hand)(outstretched). The next, 'Virgin and child' is (statue)(standing)(Mary)(holding)(Jesus). The particular evokes the general mother/baby theme. The next several works have the same religious theme.

Indigenous artworks lend themselves fairly directly to synthetic description. 'Totem pole' is (Totem pole)(with)[key components such as 'raven' could be described]. 'Fort Simpson' is (Landscape)(Fort Simpson)(from)(distance). Fort Simpson evokes town. 'Seated man holding a fox by the leg' is (seated)(man)(holding)(fox)(by)(leg). The last bit, admittedly, may only rarely be searched. 'Man and woman' is (pair)(statues)(man)(and)(woman). The material they are made of might be the most important descriptor here. 'Kneeling hunter with seal' is (kneeling)(male)(hunter)(with)(seal).

The Asian highlights reflect Hindu or Buddhist religious beliefs. Various Hindu gods would have to be recognized in a classification (and the Buddhist boddhisatva). Their actions or position can then be described.'Buddha Shakyamuni' is properly (Buddha)(topknot)(denoting)(intelligence)(and)(drooping)(earlobes)(denoting)(royalty)(and)(oval)(halo)(and)(hands)(raised)(for)(protection)(and)(assurance).These are all classic elements of Buddhist art signifying perfection, and so perhaps the work could simply be classified as (Buddha)(signifying)(perfection).

Contemporary art provides the greatest challenges. 'I want you to feel the way I do … the dress' can be rendered as (wire)(mesh)(statue)(of)(dress)(signifies)(discomfort). 'I can hear you think' is (cast iron)(heads)(joined by)(wire). It is not obvious that anyone not looking for the precise work (or a very similar one) would search for these terms. 'Bridge at Remagen' is tricky as it involves two drawn hands holding a color photograph of a couple: (two)(drawn)(hands)(holding)(photograph). A painting of tattooed quotes from four authors is (painting)(tattoo)(quotes) [The four authors are listed in the work's title]. 'Hymne an die nacht 1' is(drawing)(inspired by)(prose cycle)(Hymnen an die nacht); (metaphor

of)(cemetery)(and)(death);(two)(sheets)(arranged)(vertically). For this work, there is a clear logic to distinguishing content, meaning, and form, all of which merit description.

There are no obvious challenges in the photography or prints and drawings categories, beyond identifying the particular subject of some works. The sculpture category opens with the same works as the Asian category.

A similar analysis of the top ten highlights from the United States National Gallery is provided in an Appendix to this paper. A key issue addressed is how to capture emotion synthetically.

Archives

Archives have traditionally categorized manuscripts by provenance. Historians have long opined that it would be useful if documents were classified by subject (see Gnoli 2014). As noted above, archivists should be able to apply a synthetic approach in much the same manner as library classificationists. A letter that addresses (buying)(wheat), a report on (import duties)(collected), minutes of a meeting about (planning)(royal)(coronation) can all be captured readily through synthetic construction. And thus the historian can potentially employ the very same search terms for locating both primary and secondary sources.

Archivists are increasingly seeing their role as providing access to archival materials rather than simply maintaining these (Theimer 2011). Cuts in archival staffing in recent decades increase the desirability of enhanced subject access, since users can no longer rely (as much) on detailed advice from archivists. Archivists have thus displayed some openness to social tagging, and increasingly embrace metadata standards such as EADS or DACS – neither of which pays much attention to subject classification (Theimer 2011). Progress is nevertheless limited by the idiosyncratic classifications in use at individual archives, and the limited familiarity of most archivists with information technology (Yaco 2011). Daniels and Yakel (2010) studied the users of online finding aids for archival materials and discovered that users had difficulty locating materials, were often unaware of the possibility of subject search, and were unaware of the controlled vocabulary employed.

Ribeiro (2014) shows that archives deviate from the provenance approach quite a bit, with an eye to facilitating historical research. But there is little systematic approach, and no theoretical basis. Different archives take quite different approaches. She concludes that in a world where users want to find information without visiting archives, and where users want to search across multiple databases, it no longer makes sense for libraries and archives to be classified according to different principles, and for archivists and librarians to be trained in completely different practices.

Whereas the library classifier will have access to titles and abstracts that hint at the subject of a work, archivists will often need to read a manuscript in order to identify the subject. And many manuscripts – and especially boxes of manuscripts – will have multiple subjects. One possibility here is crowd-sourcing: as historians or other users

read documents they could tag these with subject headings. Such tags will be most useful if they utilize controlled vocabulary, and this is only likely with a very accessible controlled vocabulary.

Pattuelli (2011) developed topic maps to aid teachers in finding resources in cultural heritage archives. Topic maps also take a synthetic approach to identifying documents. And the terminology employed by Pattuelli employs basic concepts.

Altermatt and Hilton (2012) describe the particular difficulties faced by archives with respect to ephemera: things intended to be discarded such as flyers, tickets, and posters. These tend to be donated in small amounts and/or by individuals other than the creators. Yet such ephemera often give a glimpse of especially working class culture unavailable in published works. The authors describe how a poster of an African American woman at work in 1937 tells us about race relations, attitudes toward woman's work, and the struggle for esteem of African-American women; likewise a 1909 union label advertisement shows how ideas of masculinity intertwined with craft unionism. The authors had participated in a two-year project to classify ephemera in a particular museum. Their work would have been easier, it would seem, if they could have employed subject chains such as (poster)(African-American)(woman)(working)(race)(relations)(esteem)or
(poster)(advocates)(craft)(unions)(illustrates)(masculinity)

Classification Schemes

There is, notably, an international effort toward developing standards that would maintain the traditional emphasis on provenance while allowing for increased subject access. This effort has progressed farthest with respect to government records.

The Queensland State Archives provide a detailed discussion on their website (http://www.archives.qld.gov.au/Recordkeeping/GRKDownloads/Documents/functional_vs_subject-based_classification.pdf).They recommend that government records – and indeed those of any large organization – be classified in terms of the 'function' of a particular record. Documents relating to travel authorization would be distinguished from those relating to career counselling or job evaluation. It would thus be possible to readily access documents on job evaluation (or other functions) from across government departments. [They note a further advantage of a functional approach:the legal requirements to maintain documents for a certain time period differ by function.]

Interestingly, they contrast this functional approach with what they describe as a subject approach emanating from library science. They note that classifying real estate records in terms of a particular property will result in very dissimilar documents – dog registrations and property sales – being grouped together. They indicate here and elsewhere that they are concerned for the most part with collocation of documents. But our concern in this paper is with access. And it seems obvious that users will often wish to search for combinations of what the archives term 'function' and 'subject,' but that we might well perceive as different aspects of subject: a user might be interested in

job descriptions within a certain department or dog registrations in a certain neighborhood. This a synthetic approach can achieve. The basic concepts to be linked would include the sort of classification of functions that the Queensland site provides, but also other subjects, such as typical government (or company) departments.

It should be stressed that the critique of 'subject classification' reflects a very narrow view of what this entails. In particular, users can be guided to records about dog registrations associated with a particular property only if documents are assigned compound subjects.

Archives Canada (http://www.collectionscanada.gc.ca/007/002/007002-2084-e.html) also advocates a movement toward functional classification of records, and compares this favorably to a subject approach. The approach here is similar to that in Queensland and elsewhere, but more detail is provided. The draft finance management classification has four steps and several sub-steps:

- plan: define requirements, assess, cost, report
- budget: forecast, allocate, monitor, adjust, report
- manage: account, reconcile, quality assurance, report
- measure: evaluate, analyze, adjust, report

The human resources model has different set of relators (compensate, deploy, monitor) but also many of the same (evaluate, analyze).

It is notable that all of these terms are basic relators. They could be linked both to the originating department and to the subjects of financial management or human resource management: (Department X)(evaluates)(program Y). Some of these relator terms occur more than once so care would need to be taken to clarify any differences in use (likely through compounding).

Particular Documents

The United States National Archives does provide limited subject access to documents. But different collections are classified with respect to different subjects. For example, http://research.archives.gov/description/6046816describes a set of customs records with a limited set of subjects: Accounts, Bounties, Imports, Inspections, Letters (Correspondence), Prohibition, Regulations, Ship captains, Statistics, and Tonnage. These describe for the most part the type of documents, not what they might contain.When you click on these various subjects there is a list of un-clickable broader and narrower terms.

Elsewhere, a lengthy list of 'subjects' relevant to agricultural extension in Missouri is provided at http://research.archives.gov/description/286143. Some of these are specific to Missouri (counties) or the agricultural service, others to agriculture (agriculture, agronomy, soils,farms, forests -- but nothing more specific), but others are general terms such as maps, photographs, leaflets, newsletters, memoranda, land use survey, dams, engineering, and erosion.

This subject access is certainly valuable. But allowing searches across collections would be very useful: the National Archives holds a bewildering array of collections. Users will struggle to identify appropriate collections to search. And then they need to master different subject classifications for each. Users with particular interests will want to search by combinations of concepts, a strategy that is not facilitated at present.

The Bentley Historical Library at the University of Michigan provides some 30 main subjects on its website (http://bentley.umich.edu/research/guides/index.php) and then for each lists the relevant archival resources. It lists subjects that users often request or that reflect collection strengths. This is an exceptional attempt to provide subject access throughout an entire archive. But individual collections within the archive are nevertheless treated with further subject headings. For example, the Tom Downs papers can be accessed through a lengthy set of subject headings (which follow the LCSH style). Though the Bentley is to be applauded for its efforts toward subject access, there is obvious scope for further extending subject search capability across its entire collection, and for utilizing similar subject terminology as other archives. And once it is appreciated that there are multiple subject entries to a particular document the advantage of allowing subjects to be readily combined in search is apparent.

Museums

Museums were addressed in Szostak (2016a). Again there is great interest in subject classification but no consensus on how to achieve this (Menard, Mas, and Alberts 2010, Zoller and DeMarsh 2013). Szostak (2016a) provides dozens of examples of synthetic classification of items from the Smithsonian and British Museums. Some of these involve several linked terms; still, a sentence-like synthetic structure provides far greater clarity than a mere listing of such terms. The Appendix to this paper provides classifications for a variety of ancient artifacts from the British museum. It also translates many terms in the archaeological classifications employed by the US National Parks Service. And it translates the Reference Model of the International Council of Museums (CIDOC 2013). The Appendix is available at: https://sites.google.com/a/ualberta.ca/rick-szostak/publications/synthetic-classification-of-museum-artifacts-using-basic-concepts.

Conclusion

The purpose of this paper was to provide empirical support for the conjecture that items across the GLAM sector can best be classified utilizing a synthetic approach grounded in basic concepts. Recent literature and classificatory efforts for archives, museums and galleries were reviewed. Perhaps most importantly, a synthetic classification was developed for a variety of works from an international selection of museums and galleries, as well as archival documents. These classifications hopefully establish the feasibility of providing very detailed classifications of museum, gallery, and archival items by compounding basic terms in concept chains of manageable

length. Museums, galleries, and archives could decide how detailed they wished to be in their classifications. GLAM staff can go fairly easily from an object description to a sentence-like synthetic classification; users can in turn search using sentence-like strings of nouns, verbs, and adjectives/adverbs (and a sentence-like structure achieves the best of both pre- and post-coordination; Szostak 2016b). Notably, the basic concepts employed came from a general scheme. This allows users to search for items related to any human or natural activity or thing with which they are interested.

References

Altermatt, Rebecca & Hilton, Adrien (2012). Hidden collections within hidden collections: Providing access to printed ephemera. *American Archivist* 75(1): 171-94

CIDOC: Conceptual Reference Model, Definition of the (2013).Produced by the ICOM/CIDOC Documentation Standards Group, Continued by the CIDOC CRM Special Interest Group, Version 5.1.2, October 2013.

Daniels, Morgan G. & Yakel, Elizabeth (2010) Seek and You May Find: Successful Search in Online Finding Aid Systems. *American Archivist* 73(2)535-68.

Gnoli, Claudio (2014).Boundaries and overlaps of disciplines in Bloch's methodology of historical knowledge. In *Knowledge Organization in the 21st Century: Between Historical Patterns and Future Prospects.* Proceedings of the 13th ISKO Conference, Krakow. Würzburg: Ergon Verlag.

ICONCLASS (2014).www.iconclass.nl.

Menard, Elaine, Mas, Sabine & Alberts, Inge. (2010). Faceted classification for museum artefacts: A methodology to support web site development of large cultural organizations. *Aslib Proceedings* 62 (4/5): 523-32.

Pattuelli, M. Cristina. (2011). Modeling a domain ontology for cultural heritage resources: A user-centered approach, *Journal of the American Society for Information Science & Technology*, 62(2): 314–342.

Ribeiro, Fernanda (2014). The use of classification in archives as a means of organization, representation, and retrieval of information. *Knowledge Organization* 41:4, 319-26.

Szostak, Rick (2013). *Basic Concepts Classification*. [http://www.economics.ualberta.ca/en/FacultyandStaff/~/media/economics/FacultyAndStaff/Szostak/Szostak-Basic-Concept-Classification2.pdf] Acessed on 30 April 2016.

Szostak, Rick. (2014). Classifying the Humanities. *Knowledge Organization,* 41(4):263-75.

Szostak, Rick.(2016a). Synthetic classification of museum artifacts using basic concepts.In *Proceedings, Museums and the Web Conference, Los Angeles, April, 2016.*

Szostak, Rick.(2016b). Poly-coordination. In *Proceedings, Canadian Association for Information Science.* [http://www.cais-acsi.ca/ojs/index.php/cais/issue/view/32]

Szostak, Rick, Gnoli, Claudio & López-Huertas, Maria (2016). *Interdisciplinary Knowledge Organization.* Berlin: Springer.

Theimer, Kate. (2011). What is the meaning of archives 2.0? *American Archivist* 74(1): 58-68.

Yaco, Sonia. (2008). It's complicated: Barriers to EAD implementation. *American Archivist* ,71(2): 456-75.

Zoller, Gabriela and Katie DeMarsh (2013) Museum Cataloging from a Library and Information Science Perspective. *Art Documentation*, 32:1.

Rodrigo de Santis and Claudio Gnoli

Expressing Dependence Relationships in the Integrative Levels Classification Using OWL

Abstract

This article presents the use of Web Ontology Language (OWL) to represent existential dependence relationships between phenomena in the Integrative Levels Classification (ILC). Existential dependence allows expressing that a higher level of reality depends on a level below it for its existence (for example, a forest depends on plants). Since most traditional knowledge organization systems (KOSs) reduce classes to a linear sequence, they are not able to represent this kind of non-linear relations. Computational formats like OWL are based on automatic processing and inference, bringing new capabilities of expressiveness that are explored in this work by some examples extracted from the Integrative Levels Classification schedules.

1 Introduction

In a connected society, the need for knowledge organization systems (KOSs) that emphasize interoperability of concepts instead of mere interoperability of data is an important challenge. Conceptual approaches to achieve this seemed unrealizable for decades, but are now becoming feasible due to the arising of new technologies, including those related to the Web.

This article presents the use of Web Ontology Language (OWL) to represent existential dependence relationships between phenomena in the Integrative Levels Classification (ILC) and discusses its implications and possibilities.

While such traditional KOSs as thesauri or taxonomies are based on the classical hierarchical (class / subclasses) and associative relationships ('see also' or 'related terms'), in a system based on integrative levels, new properties and different kinds of relationships can also be implemented (Gnoli, De Santis & Pusterla, 2015).

Briefly, in the theory of integrative levels, as formulated during the 1950s by philosophers like James Feibleman and Nicolai Hartmann, a higher level depends on the level below it for existence, but, at the same time, has a more complex organization with new emergent properties, which makes each level essentially a different thing. The relationship that allows expressing this kind of connection between levels is existential dependence (Gnoli, Bosch & Mazzocchi, 2007; Lowe, 2015).

The development of a KOS from the theory of integrative levels is an initiative that refers to the work of the British Classification Research Group (CRG). It is registered in the CRG bulletins regularly published between 1952 and 1968 and in individual works of some of its members, notably Douglas J. Foskett and Derek Austin (Foskett, 1978; Austin, 1971). A draft of a bibliographic classification scheme based on the theory of integrative levels developed by the CRG has been published in 1969, but could not be further developed at that time (Classification Research Group, 1969).

The growth of micro-computing since the 1980s began to allow for the development of new approaches to KOSs. Brian Vickery asserted in 1986 that new KOSs should be

designed to take into account not only retrieval, but also the possibility of automated reasoning (performed by the computer itself) leading to the redefinition of search strategies: from seeking and browsing to automated techniques (Vickery, 1986).

2 Integrative Levels Classification

The Integrative Levels Classification (ILC) project is an initiative that has been continuously developed since 2004, managed by an international team including researchers, librarians, computer scientists and philosophers, among which are the present authors. ILC is currently implemented in a web system that operates upon a MySQL relational database. This kind of technological construction brings significant progresses in the use of a classification scheme, including management of freely faceted combinations (Integrative Levels Classification, 2004; Slavic, 2008).

The ILC scheme consists in a single schedule listing all classes of phenomena, expressed in notation as lower-case letters. ILC main classes are listed in Table 1.

Table 1. ILC main classes

a	forms	*n*	populations
b	spacetime	*o*	instincts
c	branes	*p*	consciousness
d	energy	*q*	signs
e	atoms	*r*	languages
f	molecules	*s*	civil society
g	continuum bodies	*t*	governments
h	celestial objects	*u*	economies
i	weather	*v*	technologies
j	land	*w*	artifacts
k	genes	*x*	artworks
l	bacteria	*y*	knowledge
m	organisms	*z*	religion

Taking phenomena as main classes is an innovation as compared to most traditional bibliographic classifications, such as Dewey, UDC, Colon or Bliss, which are based on disciplines (Gnoli, 2016).

Each class of phenomena has subclasses expressed by further letters, just as in any other classification scheme, as exemplified in Table 2.

Table 2. Some ILC classes and subclasses

j land *jy* soils	*m* organisms *mp* plants *mq* animals	*n*ulations *ny* ecosystems *nyr* forests *nyu* deserts	*v* technologies *vh* horticulture *vo* husbandry

Additionally, it can be freely combined with different classes by a set of facets. For example*n7mq* ‘animals as parts of populations’, or *x8nyr* ‘artworks representing forests’. General facets are listed in Table 3 (their set has recently been updated as compared to ILC edition 1).

Table 3. ILC general facets

0 under aspect
1 at time
2 in place
3 by agent
4 suffering from disorder
5 with transformation
6 featuring property
7 with part
8 like form
9 of kind

Facets follow a standard citation order of fundamental categories similar to that recommended by the CRG (Type, Part, Property, Material, Process, Operation, Agent, Space, Time) except from introducing such original categories as Form, Disorder, and Aspect.

A class of phenomena can also have its own special facets, that is, facets that are typical of this particular class of phenomena (in ILC2, the second edition of this KOS currently under development, these are introduced by *9* followed by the appropriate category digits), such as volume as a facet of 3D geometrical shapes. Unlike general facets, these facets only have meaning when applied to their particular class (a language has no volume). Syntactically, special facets work in the same way as facets of disciplinary faceted classifications. General facets, on the other hand, work like phase relationships of disciplinary faceted classification, though being applied more commonly and extensively, or like role operators in such verbal indexing systems as Precis.

In this paper, however, we are particularly concerned with the representation of dependence relationships. This is another type of relationship that is complementary to types and facets and especially relevant in the theory of levels.

3 Existential dependence relationships

Existential dependence is the relationship holding between a level n and a previous level $m < n$. For example, *vh* ‘horticulture’ depends on *mp* ‘plants’ for its existence. In turn, *mp* ‘plants’ depend on *jy* ‘soils’ for existence. The sequence of main classes of phenomena (table 1) should indeed reflect the sequence of existential dependences.

However, several classes may depend on the same class (e.g. both *vh* ‘horticulture’ and *nyr* ‘forests’ depend on *mp* ‘plants’ for existence). Decision on which of them should be listed before others has to be informed by other dependence relationships

(e.g. horticulture also depends on civil society, while forests do not). Thus the network of dependences is more complex than a single list of levels.

Reduction of main classes to a linear sequence is needed for the management of classes in a systematic display, which is a basic function of any classification. The ability of managing and displaying the same relationships in different ways is not reachable in a traditional KOS, but can become feasible when considering new emergent technologies as is the case with the Web Ontology Language.

4 Web Ontology Language (OWL)

The Web Ontology Language (OWL) is a knowledge representation language built upon W3C XML standard for objects called the Resource Description Framework (RDF) and is part of the W3C's Semantic Web technology stack (WEB ONTOLOGY LANGUAGE, 2012).

OWL is a computational logic-based language such that knowledge expressed in OWL can be reasoned by computer programs either to verify the consistency of data or to make inferences – which consist in using automatic reasoning rules to make implicit knowledge explicit. OWL documents, usually called *ontologies* [1] are designed to provide interoperability and to be published in the World Wide Web.

An OWL document consists of class axioms, property axioms and facts about individuals [2]. In an OWL document, an axiom is a statement that might be either true or false given a certain state of conditions defined by other axioms and by processing rules.

A class in OWL must have a unique identifier (that forms an URI – Uniform Resource Identifier), usually something like "http://www.url.com/project#class_id". The value of *class_id* is the value that identifies a class. A class may also have one or more labels used for describing it for human reading, using natural language. OWL natively provides features for expressing hierarchy, equivalence, disjoint and union of classes [3].

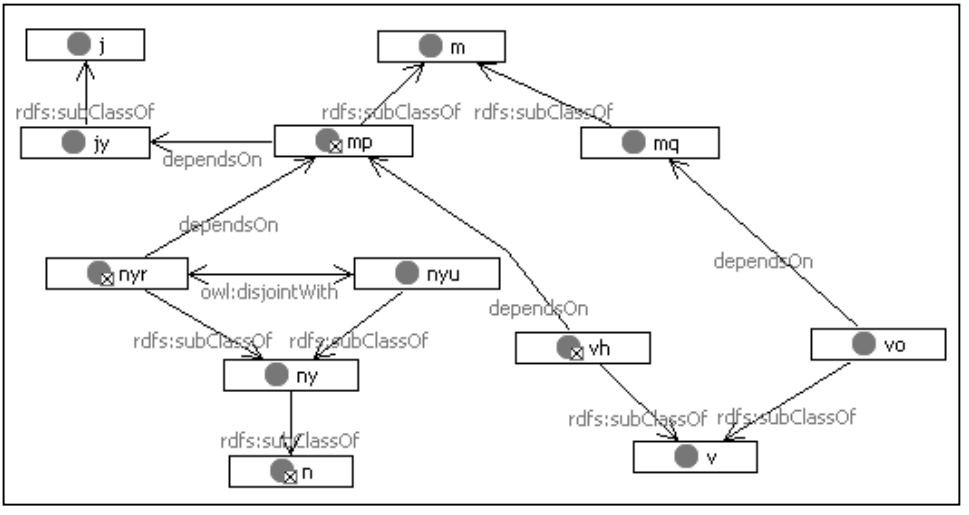
A property (which is, in fact, a special type of class) provides OWL with expressiveness and allows for specification of several kinds of relationships. Besides all native features of a class, a property also has the following predefined axioms: inverse, domain, range, functional, inverse functional, reflexive, irreflexive, symmetric, asymmetric and transitive.

Some examples of classes and properties are provided in the next section. At this point, it is important to emphasize that one of the major differences between OWL and traditional KOS formats is that OWL was conceived to be processed by machine, allowing for the dynamic inclusion of new properties and rules without the need of redefining the whole schema. This makes an OWL-based KOS flexible and extensible, and allows for the creation of user-defined properties, as is the case with existential dependence, which is the focus of this paper.

5 Expressing dependence relationships using OWL

The classes listed in Table 2, as well as some existential dependence relationships among them, have been written in OWL. The result is illustrated in Figure 1, that has been generated by the software TopBraid Composer, version 5.1.3.

Figure 1. Graphical representation of some ILC classes and their dependence relationships



The existential dependences shown here state that *vo* 'husbandry' depends on *mq* 'animals'; *vh* 'horticulture' depends on *mp* 'plants'; *nyr* 'forests' depend on *mp* 'plants'; and *mp* 'plants' depend on *jy* 'soils'.

Table 4 shows an excerpt of OWL code, including declaration of the transitive property *dependsOn* and the class *nyr*.

Table 4. Excerpt of OWL code: property *dependsOn* and class *nyr* 'forest'

```
<owl:AsymmetricProperty rdf:ID="dependsOn">
<rdfs:range rdf:resource=" http://www.iskoi.org/ilc/owl#Class"/>
<rdfs:domain rdf:resource="http://www.iskoi.org/ilc/owl#Class"/>
<rdfs:label rdf:datatype=" string">dependsOn</rdfs:label>
<rdf:type rdf:resource="http://www.w3.org/2002/07/owl#TransitiveProperty"/>
</owl:AsymmetricProperty>

<rdfs:Class rdf:ID ="nyr">
<owl:disjointWith rdf:resource="#nyu"/>
<dependsOn rdf:resource="#mp"/>
<rdfs:label rdf:datatype="string">forests</rdfs:label>
<rdfs:subClassOf rdf:resource="#ny"/>
</rdfs:Class>
```

The property *dependsOn* is declared as transitive because in integrative levels, when a level depends on a lower level, the following class hierarchy will also depend on it. In the example, as forests depend on the existence of plants, any subclass of forests, like for instance tropical rainforests, will also depend on plants. Transitiveness among levels is also achieved in OWL, but through new dependence relationships, as is the case with *mp* 'plants' *dependsOnjy* 'soil' that transitively makes *nyr* 'forests' *dependsOnjy* 'soil'.

Intuitively, dependence is an asymmetrical property, as the higher level will depend on the lower while the opposite will not usually be the case. In the previous example, if the property *dependsOn* was declared as symmetric, that would mean that plants also depend on forests for their existence.

The property *dependsOn* has two attributes: domain and range, corresponding respectively to values which may be dependent and values which may cause dependence. In the example, both are set to accept any ILC class.

The class *nyr* 'forests' is declared as a subclass of *ny* 'ecosystems' and as disjoint with *nyu* 'deserts'. This means that an ecosystem cannot be simultaneously a forest and a desert. This kind of consistence constraint is ensured by OWL, and is implemented in the major editing tools.

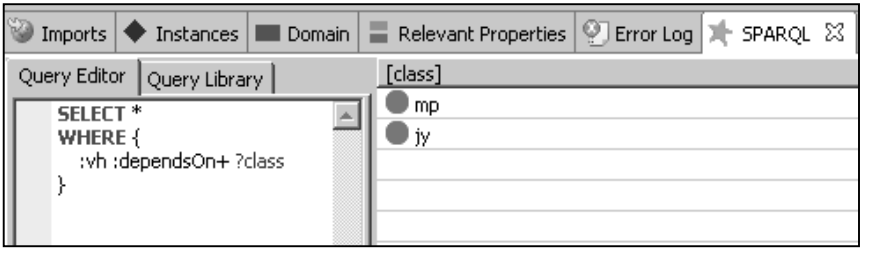
As can be deduced, expressing all kinds of properties axioms for all classes in such a large KOS as ILC may result in an endless task. For this reason, usually only main restrictions are set. In OWL, a restriction is a special kind of property and can be used as a 'negative relationship'. In the example, stating that *mp* 'plants' depend on *jy* 'soil'would fail when referring to some species of hydroponic or air plants that do not need soil for their existence. A restriction may be expressed for those particular cases through a restriction axiom operating on a relationship, as exemplified in table 5:

Table 5. Syntax of a restriction axiom on relationship *mpdependsOnjv*

```
<owl:Restriction>
<owl:onProperty rdf:ID= "dependsOn"/>
<owl:someValuesFrom rdf:resource="ilc:enumerate_allowed_classes" />
</owl:Restriction>
```

Aforementioned as one of the advantages of OWL, the inference mechanism allows expansion to several steps in the graph of a KOS through the principle of recursion. In the example from ILC, it is explicit that *vh* 'horticulture' depends on *mp* 'plants', but it is not declared that *vh* 'horticulture' depends on *jv* 'soil'. However, a SPARQL [4] query that searches for all dependences related to *vh* will be able to retrieve both results: *mp* and *jv*, as shown in the screenshot taken from TopBraid Composer software (Figure 2).

Figure 2. Query on ILC database using SPARQL language



The other branches of the given example follow the same principles and serve to illustrate that OWL is structurally extensible. More classes can be added without disturbing the existing schema. This is possible because relationships in OWL are implemented as properties over classes, not on each individual (neither on each subclass, except when this is strictly mandatory). This characteristic also makes possible for a user or a system to browse the KOS either superficially or deeply without needing to know the whole model.

6 Concluding remarks

The use of OWL for implementing a subset of ILC classes has confirmed the capability of this language to represent phenomenon classes and to manage relationships in ways different from the linear approach of traditional KOSs.

The task of expressing dependence relationships among classes has been achieved through activities of indexing and retrieving, and a main result obtained was a demonstration of automated inferences throughout the schema.

The potential relations of this work to Linked Data are also a remark that may lead to future works and possible applications of ILC project. Linked Data is an initiative conducted by W3C and proposed by the creator of the Web, Sir. Tim Berners-Lee. It refers to a set of best practices for publishing and connecting structured data on the Web using RDF to describe things in the world (Bizer, Heath & Berners-Lee, 2009). From this approach it is possible to aim for a general and open KOS, which manipulates shared knowledge, and that can be used in different ways and from several perspectives, such as the Web itself as opposed to local descriptions of closed systems.

Notes

[1] Concerning the discussion about adoption of the term ontology in a different sense from that originally considered in philosophy, in the present work,the authors decided to use “OWL document” when referring to the resulting artifact and consider ontology as defined by Roberto Poli: “ontology is not a catalogue of the world, a taxonomy, or a terminology. If anything, an ontology is the general framework within which catalogues, taxonomies, and terminologies may be given suitable organization” (Poli, 1996, p. 313).

[2] In this work, the emphasis is on classes and properties. Facts about individuals (instances) will not be explored here.

[3] A complete description of OWL syntax is available in thespecification website: [https://www.w3.org/TR/owl2-syntax/]

[4] SPARQL is a semantic query language able to retrieve and manipulate data stored in Resource Description Framework (RDF) format . For details, see: [https://www.w3.org/TR/rdf-sparql-query]

References

Austin, Derek (1971). Two steps forward. In: Palmer, Bernard I. *Itself an education*: six lectures on classification. Part two. The Library Association: London.

Bizer, Christian, Heath, Tom & Berners-Lee, Tim (2009). Linked data-the story so far. *Semantic Services, Interoperability and Web Applications*: Emerging Concepts, p. 205-27.

Classification Research Group (1969). Classification and information control. *Library Association Research Pamphlets*, n. 1. The Library Association: London.

Foskett,Douglas J. (1978). The theory of integrative levels and its relevance to the design of information systems, *Aslib proceedings*, 30 (6): 202-8.

Gnoli, Claudio, Bosch, Mela & Mazzocchi, Fulvio (2007). A new relationship for multidisciplinary knowledge organization systems: dependence. In *La interdisciplinariedad y la transdisciplinariedad en la organización del conoscimiento científico*: actas del VIII Congreso ISKO. Held at Universidad de León, April 18-20, 2007. Pp. 399-409.

Gnoli, Claudio, De Santis, Rodrigo & Pusterla, Laura (2015). Commerce, see also Rhetoric: cross-discipline relationships as authority data for enhanced retrieval. In *Classification & authority control: expanding resource discovery*: *proceedings of the International UDC Seminar*. October 29-30 , 2015. Würzburg: Ergon. Pp. 151-61.

Gnoli, Claudio (2016). Classifying Phenomena, part 1: Dimensions. *Knowledge Organization*, 43 (6), in print.

Integrative Levels Classification: research project ISKO Italia (2004). [http://www.iskoi.org/ilc/]

Lowe, Edward J. (2016) Ontological dependence. In: *Stanford encyclopedia of philosophy*. [http://plato.stanford.edu/entries/dependence-ontological].

Poli, Roberto (1996) Ontology for Knowledge Organization. In Green, Rebecca (ed.), *Knowledge organization and change*: *proceedings of the fourth international ISKO conference*. 1996. Frankfurt/Main: Indeks, pp. 313-9.

Vickery, Brian (1986). Knowledge representation: a brief review. *Journal of documentation*, 4 (3):145–59.

Web Ontology Language (OWL). OWL Working Group. (2012). [https://www.w3.org/2001/sw/wiki/OWL]

Lidiane Carvalho

The Knowledge Organization (KO) Studies in the Health Field: A Relational Perspective

Abstract

This research is about the construction and Organization Knowledge (KO) perspective in the health field The methodological aspect consider the search begins with the mapping of the scientific production using bibliometric techniques which led to an analysis corpus of 16 (sixteen publications) expressing the mobilization of 43 (forty-three) researchers in theoretical and empirical studies about KO in health fields. The analysis method uses the Social Network Analysis (SNA) and part of bibliometric data collected in the Web of Science. The distribution by countries regarding scientific publications stands out Canada with 31,25 %, the Brazil with 31,25 % and 12,50 % for the England and the United States. The discussion of the results shows the description of the actors with a greater degree of centrality and the epistemological objects of them.

Introdution

This work emerges from the production of scientific knowledge and analysis of the contribution to the fields of knowledge in a relational perspective and critical to the theoretical studies of KO. As an objective, we sought to investigate the production of knowledge in the field of knowledge organization in the health.

The representation and organization of knowledge in interdisciplinary/ transdisciplinary domains have in Huertas Lopez- Ramirez (2007, p.34) is vision attracted little attention among specialists in Information Science. The questions that guide the overall objective are: (1) What the knowledge structure in the fields of health research from the published literature about KO in the health? (2)What is the direction of the subjects addressed? (3) What are the prospects? (4) How is collaboration the among co-authors in KO / health interface?

The analysis measures of the social network, especially the degree of centrality measures show how an actor has connected to the scientific network and the human resources mobilized. The centrality measures of an actor also reveal their position and scientific and political influence in the field. In the scientific practices, the scientific authority ensures the power of the fundamental mechanisms.

Theoretical Approach

An important aspect of the approach to knowledge as an object in permanent construction of culture, is that cognitive structures are relevant in order to provide a social history and developing this view, Hjørland (2002, p. 464) is supported by the pioneering work on the domain classification. Studies of knowledge domains according to Nascimento and Marteleto, (2008, p.397) to suggest that scientific knowledge "is built by individuals who seek to exchange their experience, experienced individually with others causing the displacement information to the significance collective discursive communities "The discursive community according Hjørland (2002 p.423) is

a community where the orderly and defined communication process takes place. This communication is structured by a conceptual framework of knowledge.

Co-authoring in the scientific communication process, for example, is a collective production of subjects who 'seek to exchange their experience, experienced individually with others'. The bibliometrics studies, are recommended as a method of research areas of knowledge by Hjørland (2002, p. 431) because they express the connection between words, researchers, disciplines, geographies, and are considered explicit manifestations of patterns of interaction, communication and sociability.

How can we show a relational structure of knowledge through bibliometric indicators? The Social Networks Analysis (SNA) has provided a theoretical and conceptual framework to show the relationships between social actors in areas of knowledge, the study of its position to another actor.

The analysis measures of the social network, especially the degree of centrality measures (Table 1) show how an actor has connected to the scientific network and the human resources mobilized. The centrality measures of an actor also reveal their position and scientific and political influence in the field. In the scientific practices, the scientific authority ensures the power of the fundamental mechanisms of the field and for Bourdieu (1983, p.127) is susceptible to subjective influences of the actual structure of the scientific field, such as peer review, and other symbols of assessment and recognition of technical competence.

For example, Carvalho (2014) describe the production of knowledge of Brazilian geneticists in the scientific research. The mapping results was: the artificial intelligence and bioinformatics applied to the annotation of gene by using computational statistical methods exploit the protein function and likeness transfer sequence trial date on a large scale in order to extract subsets of variables collectively are able to distinguish different types of disease, grouping, and classification into subsets gene and the use of semantic Web and ontology, and documentary languages in the collaborative effort of building controlled vocabulary in the cataloging of the sequences (eg, projects, and HUGO Gene Ontology) and also the emerging paradigm of bio-information, and the idea of life the informational event.

This perspective KO/Health to has been approached and investigated by Carvalho (2014) and Marteleto and Carvalho (2015) in a relational and critical perspective from theoretical-methodological presuppositions of the sociology of knowledge from Pierre Bourdieu and the organization of knowledge domains proposed by Birger Hjørland. The authors suggest the paths built in the dialogue between this concept for the knowledge domain analysis.

Methodology

The search start with the mapping of the scientific production using bibliometric techniques which led to an analysis corpus of 16 (sixteen publications) expressing the

mobilization of 43 (forty-three) researchers in theoretical and empirical studies about KO in health fields.

The analysis method uses the Social Network Analysis (SNA) and part of bibliometric data collected in February 2016 in the Web of Science, according to the following procedure: Topic: (Knowledge Organization) AND Topic: (Health) AND Topic: ("Information Science") considering all the years from 1945 to the present.

To get the social network the author employed the VosWiever, a specific software for data viewing. For the calculation of centrality measures employed the Ucinet, specialized software for social network analysis (SNA). The discussion of the results shows the description of the actors with a greater degree of centrality and the epistemological objects of them.

Discussion

The should be noted that the co-authorships are expressed forms sharing of meanings and research subjects in the scientific field. The first survey results highlight the 16 publications produced by the group of 43 authors. These papers have been cited about 492 times (four hundred ninety-two times) and the average citation per article is 30.75 and the H -index of each researcher are 5, so five citations per author in the general average. The distribution by countries regarding scientific publications stands out Canada with 31,25 %, the Brazil with 31,25 % and 12,50 % for the England and the United States.

The recognition of scientific research can be measured by the number of times a paper is quote (M -index). A much-cited work can have been or not published in co-authorship. For example, the article was written by Alejandro Jadad and Ana Gagliardi (1998), researchers from the Department Epidemiology and Biostatistics at McMaster University in Canada have the most number citation. The work "Rating health information on the Internet - Navigating to knowledge or to Babel? Identifies the instruments to measure the web and the providing health information on the internet sites, and criteria used by them. The question: What the knowledge structure in the fields of health research from the published literature about KO in the health? For can a understand the relational perspective the scientific network that's imperative know the actors and your positions in the structure. The centrality measure of social network analysis (see Table 1) presents the number of contacts mobilized by an author, and the centrality this leadership in the scientific writing process.

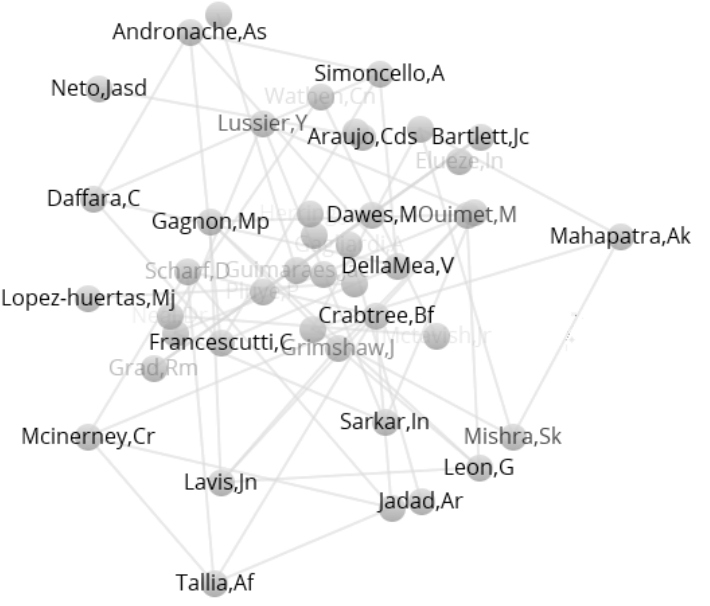
TABLE 1:KO IN HEALTH: SOCIAL NETWORK ANALYSIS (SNA)

Centrality mensures

ACTOR	**Degree**	**Betweenness**
1. Andronache,As	9,524	0
2. Crabtree,Bf	9,524	0
Daffara,C	9,524	0
DellaMea,V	9,524	0
Francescutti,C	9,524	0
Gagnon,Mp	9,524	0
Grimshaw,J	9,524	0
Lavis,Jn	9,524	0
Leon,G	9,524	0
Mcinerney,Cr	9,524	0
Orzano,Aj	9,524	0
Ouimet,M	9,524	0
Scharf,D	9,524	0
Simoncello,A	9,524	0
Tallia,Af	9,524	0
Bartlett,Jc	7,143	0
Dawes,M	7,143	0
Grad,Rm	7,143	0
Kapoor,L	7,143	0

Of the 43 authors located in the study, only five has published alone. In knowledge production structure in KO / Health, the dynamic the co-authored are with actors from areas interdisciplinary. The researchers are located in medical departments and departments bioinformatics in US and Canadian universities. The studies have concentrations in social analysis on access, use and appropriation of the knowledge.

GRAPH 1: NETWORK RESEARCHERS: KO IN HEALTH



The italian doctor Adrian Stefan Andronache and colleagues (2012, p.124) arefeatured in the investigation entitled "Semantic aspects of the International Classification of Functioning, Disability, and Health: towards sharing knowledge and unifying Information " and they describe the following features of the KO [...] the semantic interoperability of ICF: first, the representation of ICF using ontology tools; second, the alignment of upper-level ontologies; and third, the use of These tools to implement semantic mappings between ICF and other tools, such as disability assessment instruments, health classifications, and at least partially formalized terminologies. The other general aspect is of the betweenness centrality reports, with measure "zero", that is, the actors are not connected with agents outside the circle Which published.The groups are independent.

The other group with a high degree of centrality is the researchers John Orzano, Alfred Tallia and Benjamin Crabtree (2008, p.439) of the Medicine University of New Jersey with the publication " A knowledge management model: Implications for enhancing quality in health care". The article presents the vision the services in healthcare, and how the Information Science should understand the practical theory of the use of information. The authors suggest that knowledge management is key to the best practices in medicine because it is a profession based on knowledge. They propose

a model that is intended to inform the intervention protocols to improve the quality of care in health.

The researchers Philip Payne, Taylor Pressler, Indra Neil Sarkar and Yves Lussier (2013, p.1) from the Department of Biomedical Informatics and the University of Ohio in the United States, has to a high degree of centrality in the search network. The publication of the research entitled " People, organizational, and leadership factors impacting informatics support for clinical and translational Research " presents the map of specialists in the domain the Clinical and Translational Science (CTS) in the United States. In total were assigned 31 experts in CTS with competence in Biomedical Informatics (BMI), Computer Science (CS), Information Science (IS) and Information Technology (IT).

The researchers Pierre Pluye, Roland Grad, Martin Dawes Joan Bartlett (2007, p.39) from the Department of Family Medicine has published together “Seven Reasons Why health professionals search Clinical Information- Retrieval Technology (CIRT): toward an organizational model ". This research reports a case study with six Family doctors about the cognitive and organizational categories for decision making in clinical, health assessment and information sharing with the patient.

TABLE 2: KO IN HEALTH FIELDS

ACTORS	KO IN HEALTH FIELDS.
Group 1: Stefan Andronache, Adrian Stefan, Andrea Simoncello, e Vincenzo Della Mea, Carlo Daffara e Carlo Francescutti.	Group 1: Ontology; Information Science; Medical Informatics; Knowledge Bases.
Group 2: John Orzano, Claire McInerney, Alfred Tallia e Benjamin Crabtree.	Group 2: Organizational knowledge; learning organizations; sharing knowledge; decision-making; performance; firm; perspective; technology; framework; creation
Group 3: Philip Payne, Taylor Pressler, Indra Neil Sarkar e Yves Lussier.	Group 3: Biomedical informatics; cyberinfrastructure; challenges; Science.
Group 4: Pierre Pluye, Roland Grad, Martin Dawes Joan Bartlett.	Group 4: family-practice residency; primary-care; general-practitioners; medical informatics; seeking behavior; online evidence; patient-care; computer use; physicians; questionst
Group 5: Maria Lopez-Huertas de Torres e Isabel Ramirez.	Group 5: classification systems, women body
Group 6: Gregory Leon Mathieu Ouimet, John Lavis, Jeremy Grimshaw, Marie-Pierre Gagnon.	Group 6: Health care; Information science; Library science; Knowledge transfer; Research evidence.

The study of the Maria Lopez-Huertas de Torres e Isabel Ramirez (2007, p.34) is part of broader research aiming to analyze an interdisciplinary domain, in this case, Gender Studies, to identify its terminological behavior and its conceptual dynamics followed in therepresentation of concepts related to the health, image and body of

women, and their visualization in texts specialized in gender and in information retrieval systems, particularly in Gender thesauri.

Conclusion

This study mapped the degree of collaboration and refraction of knowledge areas under research KO, and its interface with the health and the existence of six groups in sociograms - that is, researchers has shared co-authorships with others researchers. This aspect is important because when many scientists/investigators have shared co-authorships, they shared epistemological aspects. The study has observed a triad - that is, only one publication that brought together three researchers. Other data refers to the publication of an article six dyads and fourth authors isolated.

The network analysis from the bibliometric indicators points to the following observations in relational aspects: The first, the collaboration between research groups on the subject KO / Health is still isolated groups around the world. The intermediation indicator characterized by statistical measure " betweenness " is “zeroed". The second aspect, groups in sociograms represent epistemological aspects and consensus about constant themes.

The main themes are: Ontology, Information Science, Medical Informatics, Knowledge Bases, Organizational knowledge, learning organizations, sharing knowledge, decision-making, performance, firm, perspective, technology, framework, creation, Biomedical Informatics, cyberinfrastructure, family-practice residency, primary care, general practitioners, medical informatics, seeking behavior, online evidence, patient-care, computer use, physicians, questions, Health care, Information Science, Knowledge transfer.

It also notes that although embryonic research in KO oriented to health objects is booming, this observation can be made by increasing number of papers and proceedings that have been published in recent years.

References

Andronache, Adrian Stefan, et al. (2012). Semantic aspects of the International Classification of Functioning, Disability and Health: towards sharing knowledge and unifying information. *American Journal of Physical Medicine & Rehabilitation* 92(13): 124-8.

Borgatti, Stephen (2005). Centrality and network flow. *Social Networks***,** 27(1): 55–71.

Bourdieu, Pierre (2004). *Os usos sociais da ciência: por uma sociologia clínica do campo científico*. São Paulo: Editora UNESP.

Carvalho, Lidiane dos S (2014).Informação e Genética Humana: O sequenciamento de uma cultura científica. *PhD Thesis.* Rio de Janeiro: Instituto Brasileiro de Informação em Ciência e Tecnologia, Universidade Federal do Rio de Janeiro/ Escola de Comunicação.

Del Moral, R.; Navarro, J. MariJuan, Pedro (2011). From Genomics to Scientomics: Expanding he bioinformatios paradigm. *Information***,** 2: 651-71.

Hjørland, Birger (2001). Towards a theory of aboutness, subject, topicality, theme, domain, field, content . . . and relevance**.** *Journal of the American Society for Information Science and Technology*, 52(9): 774 -8.

Hjørland, Birger (2002). Domain Analysis in information Science: eleven approaches – traditional as well as innovative. *Journal of Documentation*, 58: 422-62.

Jadad, Alejandro R. & Gagliardi, Anna (1998). Rating health information on the Internet: navigating to knowledge or to Babel?.*Jama*,279(8): 611-4.

López-Huertas, María J. & Ramírez, Isabel de Torres (2007). Gender terminology and indexing systems: The case of woman's body, image and visualization. Libri, 57(1): 34-44.

Marteleto, R. M. & Carvalho, L. S. (2015). Health as a Knowledge Domain and Social Field: Dialogues with Birger Hjørland and Pierre Bourdieu. *Knowledge Organization*, 42: 581-590.

Orzano, A. John et al. (2008). A knowledge management model: Implications for enhancing quality in health care. Journal of the American Society for Information Science and Technology, 59(3): 489-505.

Payne, Philip R. et al. (2013). People, organizational, and leadership factors impacting informatics support for clinical and translational research. BMC medical informatics and decision making, 13(1): 1.

Pluye, Pierre et al. (2007). Seven reasons why health professionals search clinical information-retrieval technology (CIRT): toward an organizational model.Journal of evaluation in clinical practice, 13(1): 39-49.

Hemalata Iyer

Alternative System of Medicine, Ayurveda: Challenges to Knowledge Organization and Representation

Abstract

The world today is much more open to embracing ideas that come from an indigenous culture.Yoga has become quite popular and is practiced all over the Western world. So have alternative systems of medicine originating in different part of the world. Ayurveda emerged independently of biomedicine in India and is an integral part of the healthcare system in the U.S. today.It is recognized as one of the four systems of complementary medicine (CAM) by National Center for Complementary and Alternative Medicine (NCCAM). It is a holistic system of medicine. Holistic health is a concept in medical practice that holds that all aspects of people's needs, psychological, physical and social, and mental should be taken into account and seen as a whole.The aim of this paper is to examine the challenges and issues involved in organizing and representing Ayurveda information. Part of the challenge in the ontological structure and the vocabulary to represent that structure. Diversity is also an issue.It is being practiced both as folk tradition, as well as an empirical, scientific system of medicine.Since it is a holistic approach to healing, it integrates medicine with culture, philosophy and religion and hence draws concepts from various disciplines.The technical terms are both in 'Sanskrit' as well as in English and often have multiple meanings.Due to these factors unique challenges have arisen with reference to methods of representing, organizing and communicating this information. This paper examines some key aspects such as professionalization, representation, terminology, social tagging, knowledge structures and semantic relations in the context of Ayurveda.

Introduction

The use of Complementary and Alternative Medicine (CAM) has increased considerably. This has been in spite of the tremendous advances in healthcare technology and its use by conventional medicine, also known as biomedicine or allopathy. Ayurveda is one of systems of CAM that emerged independently of biomedicine which has been transplanted from its native India to the United States.It is an integral part of the healthcare system today. The World Health Organization estimates that approximately 80% of the world's population relies on traditional systems of medicines for primary health care, defining these systems as those in which plants form the dominant component over other natural resources (Mukherjee & Wahile, 2006).Ayurveda is a system of alternative medicine that originated in India and is used prevalently in the United States and in several other parts of the world. In 2007 the National Health Interview Survey included a comprehensive survey of CAM use by Americans. The results indicated that more than 200,000 U.S. adults had used Ayurveda medicine in the previous year (NCCAM).

Ayurveda is a combination of two words *Ayu* and *Veda*, meaning the knowledge of life. It is a comprehensive natural holistic healthcare system which in its move to the west has adapted to the needs of the western consumers. It is not necessarily practiced in its classical form. Certain areas such as message therapy, rejuvenation, herbal supplements are popular. Understanding how it is being represented to consumers choosing to utilize it either in conjunction with or as a replacement for biomedicine is

important.Due to the increasing use of the Internet for health or medical information by the general public, websites were chosen as the method of representation to examine how these representations of Ayurveda on the Web might impact its use. (Iyer& Amber, 2013).

Objectives and Methodology

The objective of this study is to examine the web representation of Ayurveda and discuss issues of communication, professionalization, legitimization, knowledge structures and semantic relations appropriate to this domain.

The methodology includes examining the Ayurveda websites, the original indigenous texts and interviews of practitioners and Ayurveda physicians. Thirty US websites were used as a sample in this study. The following open directories on the Web were chosen for analysis.

(1) From the DMOZ directory, sites were chosen from those listed within Health: Alternative: Ayurveda: http://www.dmoz.org/Health/Alternative/.From this directory the sites listed in the categories "clinics and practitioners" and "schools" were selected.

(2) From the alternative medicine directory, sites from the Ayurveda: Clinics category were used: http://www.alternativemedicinedirectory.org/ayurveda-clinics.html.

The first step in the process was to search Google as it a popularly used search engine and an average user browsing the web for information on Ayurveda is likely to use it. The following search terms were used to search Google and top ranking sites/directories were randomly selected. The search term "Ayurveda open directory" resulted in a listing and the DMOZ directory was the top ranking website. The listing had a subheading for "Ayurveda": http://www.dmoz.org/Health/Alternative/. The search term "Ayurveda medicine directory" resulted in a listing in which Yahoo directory was first. The Yahoo directory had a subcategory for Ayurveda: Ayurveda - Alternative Medicine and Health Directory. Further this had another subcategory called "Ayurveda" with listing of sites: http://www.alternativemedicinedirectory.org/ayurveda-clinics.html. The rest of the websites consulted were chosen from a separate search rather than from directories. The search "Ayurveda clinics in USA" returned several sites and a random selection was made from the first two pages of results. While selecting sites from the above sources, only clinics and centers located in the US were chosen. Exclusively commercial stores were omitted. Each of the thirty randomly chosen sites was examined to determine the stated objective of the site, the Ayurveda themes represented, treatment, diseases, navigation, evidence of an effort to establish legitimacy, and references to India through language and symbols. In essence its organization and representation.

The search "Ayurveda clinics in USA" returned several sites and a random selection was made from the first two pages of results. While selecting sites from the above

sources, only clinics and centers located in the US were chosen. Exclusively commercial stores were omitted. Each of the thirty randomly chosen sites was examined to determine the stated objective of the site, the Ayurveda themes represented, treatment, diseases, navigation, evidence of an effort to establish legitimacy, and references to India through language and symbols. In essence its organization and representation.

Analysis and Discussion

The websites underlying philosophy/objectives can be seen from scope of topics they cover:wellness and health, prevention of diseases, inner and outer beauty, yoga, rejuvenation and detoxification, treatment in general and of specific diseases, the manufacture of medication, education and research and the transformation of consciousness by body, mind and soul integration. Website philosophies fall into three broad categories: health and wellness, education and research, and beauty. In general, most of the websites deal with health and wellness. The three common approaches represented are *Rasayana Chikitsa* (Rejuvenative Therapy), *Dinacharya* (Daily routines to be practiced), and *Rithucharya* (Seasonal regimem to be practiced). These are preventive measures. Some sites focus on specific illnesses such as cancer, hair loss, Parkinson's disease and cardiovascular diseases and treating these together with conventional medicine. They are not necessarily in competition with conventional medicine. The overall focus is on well-being rather than healing or curing. In contrast to conventional medicine the philosophy behind Ayurveda medicine advocates overall fitness by being proactive rather than reacting to illness.By placing equal emphasis on the body, mind and spirit, Ayurveda endeavors to restore the innate harmony of a person. The primary goal of this system of medicine is prevention and longevity rather than curing diseases.The websites portray this philosophy. One of the websites characterizes Ayurveda as "use of herbs, minerals, natural remedies, herbal products, life style modification, dietary modification, nutrition, meditation, purification and rejuvenation as methods of natural healing. We emphasize using natural products for living a healthy life and adopting a natural life style than depending on medicines."Tracing the history and transference of Ayurveda to United States in the late 1980's, it was linked with religion and mysticism. However as early as 1995 the image had started to change. (Reddy, 2004).It shifted from being intertwined with religion to an independent medical system. Globalization resulted in a cultural paradigm shift in Ayurveda and Yoga. It resulted in the dissolution of holistic Hindu body of knowledge. By representing Ayurveda as "products of ancient civilizations" and as "the oldest system of healing", it is portrayed as belonging to the human race as a whole thereby removing the cultural boundaries.

Efforts to legitimize the system is apparent in the websites. This happens when an indigenous system is transferred and practiced in a different environment and culture. Cultural symbols are used as one of the means of legitimizing the system. For instance,

the symbol of banyan tree, fire, the lotus flower, hands are the common symbols found on the websites.Symbols can be perceived as being socially determined and acceptable or as characterizing the culture. The banyan tree is the national tree of India. The roots sprout into more trunks and branches. Due to this attribute and its longevity, the banyan tree is considered immortal and sacred. The symbol of lotus flower possibly alludes to a purer state of mind as being conducive to good health. Though a lotus has its roots in marshy soil, but rises up and blooms in pristine purity. The depiction of hands in various gestures is drawn from the *mudras*, a Sanskrit word that refers to gestures that are associated with healing and channeling ones energy for better health. These symbols have the potential to locate Ayurveda within its cultural origin and thereby offer it the legitimacy of association with an ancient system of healing (Iyer & Amber, 2013).

Other ways of legitimizing occurs with reference to the herbs used by and sold by Ayurveda practitioners. The method of authentication revolves around established regulatory institutions such as the mention of conformity to ISO company 9001-2000 standards or certifications in organic and kosher practices or U.S. Food and Drug Association Good Manufacturing Practices (GMP). Websites might also provide the source of their herbs, claiming to have strict quality control standards in place. Some provide lengthy descriptions of the process while others claim that the herbs are grown naturally in the pristine Himalayan region without the use of chemical fertilizers and pesticides. These methods rely on disclosure. It empowers the consumers to take informed decisions. In addition, the images, videos and interviews with practitioners demonstrating the "Ayurvedic lifestyle" they live also serves to establish their legitimacy.

Knowledge Structures

The foundational concepts of Ayurveda have to do with three basic ideas: the body's constitution (*prakriti*), life forces (*doshas*) and universal interconnectedness (Valiathan, 2009). Man being a part of all that exists, the human body is attuned to the constituent elements of the universe or the macrocosm. Ayurveda envisions the human body as the microcosm and the universe as the macrocosm. *Panchabhutas* is the Sanskrit word that refers to the five constituent elements in nature: sky (*aakaash*), air (*vaayu*), fire (*agni*), water (*jala*), and earth (*bhoomi*). This underlying synergy with the elements that forms the basis of Ayurvedic therapeutics dictates the choice of food and drugs and the effect that they produce in the body. The idea of *samya*, or equilibrium, is also very important in Ayurveda.

Ayurveda information may be organized using its foundational principles known as *Siddhanta* in Sanskrit.Incorporating scope notes or annotations explaining the concept and the Sanskrit term is essential. Following are illustrative examples of structuring two siddhantas/principles using appropriate characteristics of division.

The three biological humors: The theory of omni substances is known as *Tridoshas*. Ayurveda is based on a unique fundamental principle, the *Tridosha* theory. An imbalance of the three humors is considered to be the root cause of diseases. The three humors, *Vata*, *Pitta*, and *Kapha* are aligned to the five constituent elements in nature.

< by *Tridosha*>

Vata,

(Annotation: Vata is a combination of air and ether)

Pitta

(Annotation: Pitta is combination of earth and fire)

Kapha

(Annotation: Kapha is a combination of ether and water)

Personality and human behavior approach: Another fundamental *Siddhanta* is based on Triguna or manas dosha theory. It rests in the three fundamental qualities or gunas that provide a platform for understanding personality as a dimension of human behavior and its impact on health and wellness.

< by *Triguna*/ Quality>

Sattva (beingness).

(Annotation: Manifests as joy, happiness, positive attitude, lightness, consciousness)

Rajas (activity)

(Annotation: Manifests as ambition, drive to achieve, passion and activity)

Tamas (darkness)

(Annotation: Characterized by negativity, resistance, apathy, lethargy)

Further, different kinds of foods are characterized and classifiedas related to the three qualities, *sattvic*, *rajasic* or *tamasic*and each of theseimpacthealth, wellness,and helppreventdiseases.For example vegetarian diet is considered to be sattvic and conducive to health. There are detailed treatises on type of diet, methods of cooking and recipes. Diet is also prescribed based on time, seasonal and climatic conditions.

Thus in addition to diagnosis, diseases and treatment, diet may be considered as a major category that can be classified using the three gunas/qualities as the characteristic of division. Other relevant attributes are also listed.Diet and Nutrition

<by *gunas*/qualities>

Sattvic diet

Rajasic diet

Tamasic diet

< by type of food>

< by diet for treating various diseases>

<by age>

(Diet appropriate to different age groups and stages in life)

<by gender>

(Diet for men and women; Women during pregnancy)

<by recipes>ant role
(Health recipes, ingredients, method of cooking, the kind of pots and pans to use)
< by time/ seasons/ climate>
(Diet appropriate for different seasons of the year, climatic conditions)

Terminology

Ayurveda uses complex Sanskrit terminology. The pre-coordinated terms may be simplified and wherever possible. Syntactic and semantic factoring of terms may be considered.In addition, some Sanskrit words can have more than one meaning (homonyms):For example Lakshana is a Sanskrit word derived by combining two words *lakshya* plus *kshana*, which means either symptom or indication. *Lakshana* also stands for attribute, quality and lastly for auspicious mark. Often context determines the connotation of terms. It is helpful to link the Sanskrit terminology to a glossary of Ayurveda terms. Ayurveda vocabulary reflects the holistic nature of this domain. The terms are inter-related and depict interactive processes. For instance,any two terms such as *prana* (life force), *manas* (mind) *prakriti* (individual constitution), *vishamatva* (imbalance), *samya* (balance), *dhatus* (tissues), *ritus* (seasons), *shodhana* (cleansing the body) are not compartmentalized but are inter-related. Hence along with the definition of the terms a description of the inter-related processes is needed.This can be embedded in the annotations/scopenotes for terms.

Another significant approach will be the user tags to complement the controlled vocabulary. Given the nature of this domain leveraging and utilizing the user tags assigned to Ayurvedic resources for improved access is particularly helpful. Towards this end, the results of the study (Iyer & Bungo, 2011) comparing the semantic relationship between the subject headings and user tags assigned to books on Alternative and Complementary medicine is relevant.The tag categories were created and these were then mapped onto the subject headings through a set of semantic relationships assisted by the UMLS Current Relations in the Semantic Network chart.The UMLS Current Relations framework identifies several types of relationships and associations under the general rubric of "associated _with."It lists different relationships. These subject headings covered a wide range of topics such as Complementary Therapies, Evidence-based Medicine, Massage Therapy, Self-help Groups, Diet, Spiritual Healing, and Medicine, and Ayurveda. Less than 1% of tags matched terminologically with the subject headings.Results indicated 46% semantic matches and 54% non-matches.Personal, Genre/Form, Location, Time Period and Belief Systems were the frequently occurring patterns among non-matches.Of the semantic matches, frequently occurring relationships were physical, functional, and conceptual relationships. Among the physical relationship schemas the sub categories, Part of; Ingredient of-; Branch of- occurred most frequently.

Physically_related_to (part_of): inherent part of a larger field. Eg. Alternative medicine is the Subject heading (SH) and alternative therapies is the Tag category

(TC); Holistic Health (SH) and mind (TC) [the parts of the field are all related to one another as well.]

Physically_related_to (branch_of): Discipline of knowledge and its subdisciplines.Eg. Science (SH) and medicine (TC); Holistic Medicine (SH) and alternative medicine (TC); Gynecology (SH) and medicine (TC)

Physically_related_to (ingredient_of): the core entities that are required in order for the larger concept to exist.Eg. Energy medicine (SH) and energy (TC).Energy is required for energy medicine to work, as energy medicine involves the manipulation of energy.

Among the functional relationship schemas the sub categories were (interacts with), (produces), (causes), (carries out), (practices), (occurs in) (processor).Much of the data was evenly distributed among these subsets.However, the subsets of (result of) and (manifestation of) had a much higher representation.

Functionally_related_to (manifestation of): An expression of an entity.Eg. Errors, Scientific (SH) and doubt (TC); Health Behavior (SH) and food (TC)

Functionally_related_to (result of): things that contribute to the end result.Eg. Mental healing (SH) and health (TC)

Conceptual relationships schemas: It included the subsets (evaluation of), (analyzes) (assesses_effect_of)), (property of), (method of), (conceptual_part_of) and (issue in).The most frequent sub-relationships were (property of) and (issue in).

Conceptually_related_to (property_of): attributes of an entity or a process.Eg. Meditation (SH and awareness (TC)

Conclusion

It is evident that there are several challenges that need to be addressed. The Colon Classification (CC) scheme is used in many of the Ayurveda libraries in India. It covers Ayurveda as a System within the Medicine schedule (L). L-B is the notation assigned to Ayurveda. The facets listed in the L schedule along, with the various common isolates can be used with the Basic subject L-B for classifying Ayurveda materials. (Ranganathan, 1987). Many libraries use the colon classification to classify their Ayurveda collection and for the rest of the resources the Dewey Decimal Classification is used. Given the specialized nature of Ayurveda, even a faceted classification such as CC does not seem to meet the needs of such collections. The Central Council of Ayurveda and Siddha Medicine library uses a home grown scheme. A depth classification schedule of CC is needed for Ayurveda system of medicine.The needs of libraries in the West for organizing Ayurveda collections is slightly different from the ones in the East. This is both with regard to the topics covered and the level of detail.

As for terminology, the vocabularies developed for information retrieval must address the inter-relatedness of the Ayurveda terms and incorporate description of the inter-related processes as well. Mapping the technical Sanskrit terms for diseases and medicinal plants with the appropriate International standard codes will help immensely.

Another area of research is ontologies for Ayurveda system and selected areas within that domain.Jayakrishna Nayak discusses the ontological challenges with regard to Ayurveda, compares it with conventional medicine and details the three phases of the ontological flowchart, the initial assumptions about nature (premise), the methods of gaining knowledge and final vocabulary. Ontology is a particular perspective of an object of existence and the vocabulary needed to share that perspective. (Nayak, 2012). Further research in KOS for Ayurveda is needed for improved organization and retrieval.

References

Iyer, Hemalata & Bungo, Lucy. (2011). An examination of semantic relationships between professionally assigned metadata and user-generated tags for popular literature in complementary and alternative medicine. *Information Research*, 16(3): 482.

Iyer, Hemalata & D'Ambrosio, Amber J. (2013). Ayurvedic medicine and the web: Representations and interpretations.*Annual Review of Cultural Heritage Informatics*, 1 (1).

Mukherjee, Pulok K. & Wahile, Atul. (2006). Integrated approaches towards drug development from Ayurveda and other Indian system of medicines. *Journal of Ethnopharmacology,* 103(1): 25-35.

National Center for Complementary and Alternative Medicine. NCCAM website [https://nccih.nih.gov/research/statistics/2007/camsurvey_fs1.htm].

Nayak, Jayakrishna. (2012). Ayurveda research: Ontological challenges. *Journal of Ayurveda and Integrative Medicine.* 3(1): 17–20.

Ranganathan, S.R. (1987). *Colon Classification*. (7th Ed.).M.A. Gopinath (Ed.). Sarada Ranganathan Endowment for Library Science, Bangalore.

Reddy, Sita. (2004). The politics and poetics of 'magazine medicine': New age Ayurveda in the print media. In *The Politics of healing: Histories of alternative North America medicine in twentieth-century*. New York: Routledge. Pp. 207-229.

Valiathan, M.S. (2009). The Ayurvedic view of life. *Current Science*, 96(9): 1186-1192.

Widad Mustafa El Hadi and Marcin Roszkowski

The Role of Digital Libraries as Virtual Research Environments for the Digital Humanities

Abstract

In this paper we will reflect on the state of the art of our knowledge of Research Infrastructures (RIs) for the Humanities and the role of digital libraries as Virtual Research Environments (VRE) for the Digital Humanities (DH). We will first give a brief note on the relationship between Research Infrastructures and digital libraries. We will explore next the concept of semantic annotations in digital libraries.Assuming the fact that the World Wide Web is a natural environment for digital libraries and research infrastructures, Digital Humanities-friendly Digital Libraries should be based on Web standards in order to be part of the Web and not only on the Web.

1 Context and Rationale

Digital Libraries are considered "as far more than simple digital surrogates of existing conventional libraries, they are considered as an important part of the Digital Humanities infrastructure", Svensson, (2010). They have in fact the potential to be complex *Virtual Research Environments* (VREs). This concept is defined by the UK Joint Information Systems Committee (JISC) Virtual Research Environments Program, as comprising: "A set of online tools and other network resources and technologies interoperating with each other to support or enhance the processes of a wide range of research practitioners within and across disciplinary and institutional boundaries. A key characteristic of a VRE is that it facilitates collaboration amongst researchers and research teams providing them with more effective means of collaboratively collecting, manipulating and managing data, as well as collaborative knowledge creation [1]. In our paper we to try to show how Digital Libraries could play the role of Virtual Research Environments (VREs) and what are the basic requirements. The major role for libraries in terms of DH research projects is to provide high quality information resources both on the level of relevant content and appropriate level of information representation. High quality of metadata for digital collections is the basis for efficient information retrieval and further processing but the application of computing to cultural heritage is paving the way for new opportunities for studying and engaging with an ever-increasing volume of heterogeneous cultural heritage artifacts, along with a wide range of publication genres on them and computational objects that derive from them (Fay & Nyhan, 2015, 118).

2 Research Cyber-infrastructures

Today, Research Infrastructures [2] moved to what is called "*Research Cyber-infrastructures*". This concept is defined by Geoffrey Rockwell in his blog [3] as: "*Anything that is needed to connect more than one person, project, or entity is infrastructure. Anything used exclusively by a project is not."* This type of

infrastructure is essential for setting the place of humanities within the digital realm. In the same way Alison Babeu (2011) presents the components of cyberinfrastructure as follows: the network, discipline-specific software, data collections, tools, expertise/best practices, and standards. At its core, cyberinfrastructure is made up of extensive and reusable digital collections, but each of the categories mentioned above is also vital to the success of a cyberinfrastructure. Categories of Research Infrastructures relevant for Humanities are defined by the European Commission Framework Program Moulin et al. (2001). Within this program a set of categories or types of research infrastructure has been proposed. Research libraries and Research archives are considered as Transversal RIs. We will therefore focus on digital libraries and their role as potential Virtual Research Environments for the Digital Humanities.

2.1 Research Cyber-infrastructures for the Humanities

Historically, humanities researchers have long been familiar with Research Infrastructures (RIs) and the objects that populate them such as archives, museums, galleries and libraries where collections of physical objects such as archaeological fragments; paintings or sculptures; inscriptions manuscripts books and journals were kept. An infrastructure is thus considered as the technical and operational framework that allows researchers to collaborate and share data and results, (Moulin et al 2011). Many definitions of RIs have been formulated over the years. Regarding Humanities, it should be stressed that there are special dynamics and aspects that must be considered while defining this type of RIs.

Three elements are essential for Virtual Research Environments: i) *Textmining*: data mining tools and their accompanying visualizations, which facilitate pattern finding across a wide range of data, can play a role in this process 2) *Interfaces*: Interface design researchers have worked on systems intended to help users access digital images, work with electronic text files, and apply data mining algorithms to a variety of problems, both in the sciences and in the humanities; 3) *Semantic annotation*(SA) of texts within textual VREs: Adding meaningful structures to document resources. It is particularly useful for making computers communicate with each other more effectively. Semantic annotation can be one of the elements fostering systems and resources interoperation and better highlight the nature of digital humanities research as collaborative and interdisciplinary, crossing the traditional academic borders between computer Scientists, engineers, library and Information professionals as well as humanities Scholars.

2.2 The need of annotations as a form of information representation in Virtual Research Environments for the Digital Humanities

The concept of annotation in VRE is related to the process of content analysis and indexing, personal information management and also has social and collaborative layer of interpretation. Morbidoni at al. (Morbidoni, Grassi, Nucci, Fonda, & Ledda, 2011)

argue that Digital Humanities community have understood that Digital Libraries (DL) should no longer be simple ‘expositions’ of digital objects but rather provide interaction with users, enabling them to contribute with new knowledge, e.g. by annotating and tagging digital artefacts. In this crowdsourcing paradigm annotations are seen as virtual modifications of data objects by patrons (Nuernberg, Furuta, Leggett, Marshall, & Shipman, 1995) and user-generated metadata that can be stored, searched and organized. The added value of annotations in VRE is built on engaging the community, recording new types of information about digital assets (often subjective and discourse-pragmatic information), increasing of library's flexibility and responsiveness, codifying professional judgement of research community and offering new ways for information exploration (see alsoArko, Ginger, Kastens, & Weatherley, 2006).

In many studies the need for annotation features in DL especially in the domain of humanities become crucial (eg. Maristella Agosti, Ferro, Frommholz, & Thiel, 2004; Maristella Agosti & Ferro, 2003; Barbera, Meschini, Morbidoni, & Tomasi, 2013). The idea of annotations is strongly related to the concept of interpretation which is one of the basic tasks that scholars perform (Maristella Agosti et al., 2004, p. 246). In this perspective annotations support user in expressing information about digital assets and accessing his own collection of annotations and support research communities in sharing these type of recorded information. Agosti et al. (Agosti et al., 2004) distinguish three layers of annotations that can coexist in the same document:i) a private layer of annotations accessible only by the annotations author himself, ii) a collective layer of annotations, shared by a team of people, and finally iii) a public layer of annotations, accessible to all the users of the digital library; in this way user communities can benefit from different views of the information resources managed by the digital library.

3 Semantic Annotations – the model

The term *annotation* in the context of DL means both the process of annotating digital assets and the result of that process. In the latter perspective it can be a comment,a tag, a bookmark or some kind of structured or “formally meaningful” metadata. The concept of annotations in networked environment is strongly related to social annotation systems such as Connotea [4] or Delicious [5] which fostered the idea of user participation in creating and responding to online resources an resonates with he philosophy of Web 2.0. Tagging and bookmarking of web resources became an extremely popular feature in many web applications and services and also were the subject of interest for software developers for digital libraries and repositories. With the advent of the Semantic Web technologies and Linked Data solutions, expressing annotations in machine-understandable formatsbecome possible.

The difference between semantic annotation (SA) and simple annotation lies in the fact that the former describes the data using a common, established way while the latter

using keywords or other *ad hoc* serialization which impedes further processing of the metadata. (Konstantinou & Spanos, 2015, p. 7). Oren et al. (Oren, Möller, Scerri, Handschuh, & Sintek, 2006) present conceptual model for annotations in networked environment. According to them, we can distinguish four elements of annotation which are expressed in their definition: *"'An annotation A is a tuple (a_s, a_p, a_o, a_c), where a_s is the subject of the annotation (the annotated data) a_o is the object of the annotation (the annotating data) a_p is the predicate (the annotation relation) that defines the type of relationship between a_s and a_o, and a_c is the context in which the annotation is made.*."

By endorsing this conceptual approach it is possible to distinguish three main types of annotations:

1) informal annotations,
2) formal annotations, that have formally defined constituents and are thus machine-readable, and
3) ontological annotations, that have formally defined constituents and use only ontological terms that are socially accepted and understood.

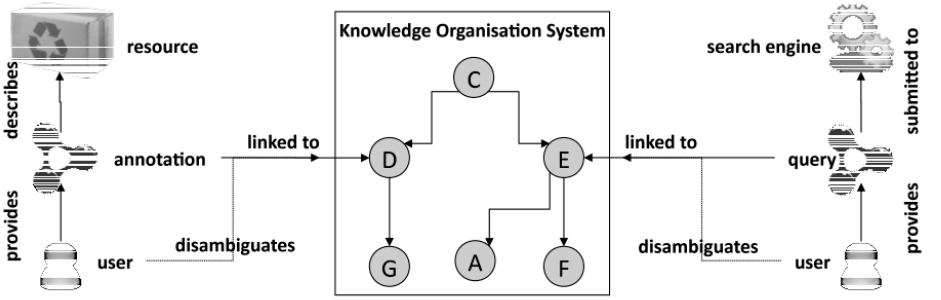
The most interesting types of annotation in the context of Digital Humanities-friendly digital library are formal and ontological annotations which according to Oren et al. (Oren, Möller, Scerri, Handschuh, & Sintek, 2006) can be defined as follows: *"Formal annotation: A formal annotation A_f is an annotation A, where the subject a_s is a URI, the predicate a_p is a URI, the object a_o is a URI or a formal and the context a_c is a URI. Ontological annotation: A ontological annotation A_s is a formal annotation A_f, where the predicate a_p and the context a_c are a_n (arbitrarily complex) ontological term, and the object a_o conforms to an ontological definition of a_p."*

Semantic annotation as a process can be also investigated regarding the level of automation. Annotations can be manual (performed by one or more people), semi-automatic (based on automatic suggestions), or fully automatic. (Oren et al., 2006)

3.1 Knowledge Organization Systems and Semantic Annotations

The application of Knowledge Organization Systems (KOS) in the semantic annotations model is in most cases related to the expression of annotation object $a_o$by referencing to relevant concept in KOS by the means of its unique and constant identifier (URI). This scheme (KOS) can be implemented in VRE or operates as external and remotely dereferenced namespace (e.g. Library of Congress Subject Headings [6]). The key idea is that terms in user annotations and queries can be explicitly linked to the elements of the underlying KOS and, therefore, their meaning can be disambiguated (Andrews, Zaihrayeu, & Pane, 2012) and the annotation object can be machine-computable (Fig. 1).

Figure 2. KOS-based Annotation and Search (Andrews et al., 2012)



Andrews et al. (Andrews, Zaihrayeu, & Pane, 2012) proposed three types of annotation models which differ in the way they are based on a KOS. In the first model annotation objects are not linked to KOS elements which may generate problems related to variation, polysemy, synonymy, and the specificity gap. The second model is based on authority files (AF) and the third one on thesauri. In the AF annotation model the key idea is to disambiguate concepts (term lists) and named entities (authority files) by using their preferred names from controlled vocabularies. Implementation of thesaurus in annotation model can solve the specificity gap problem because query terms can be mapped to the annotation elements through the parent-child relations of the thesaurus and, therefore, resources which are more specific in meaning than the user query can also be retrieved. This is rather narrow approach to KOS implementation in SA because it takes into account only term lists and thesauri. It seems that more suitable approach to this problem would be reusing the networked knowledge organization systems typology by G. Hodge (Hodge, 2000). Then we could investigate the implementation of term lists (AF, dictionaries, glossaries, gazetteers) classifications and categories (subject headings, classification schemes, taxonomies, and categorization schemes) or relationship lists (thesauri, semantic networks, ontologies) in SA.

The most advanced approach to implementing a KOS in SA is ontology-based semantic annotation. In this scenario some formally specified knowledge base is being used for semantic annotations (manual or automatic). In this approach not only annotation object can be formally referenced but also other annotation constituents. On the conceptual level ontology consist of terminological box (TBox) and assertion box (ABox). The TBox represents concepts organization and stores a set of universally quantified assertions (inclusion assertions) stating general properties of concepts and roles whereas ABox comprises assertions on individual objects (instance assertions) (Giacomo & Lenzerini, 1996). Domain or application ontologies seem to be the best solution for this approach because they describe concepts depending on a particular domain and/or are relevant to goals of specific applications and user needs. This means

that ontology-based SA offers formal references to subject and object of annotation (classes and individuals) and also for annotation predicates (object properties – roles, relationships).

3.3 Semantic Annotations in Virtual Research Environments for the Digital Humanities

Semantic annotation is one of the core elements in collaborative and interdisciplinary nature of Digital Humanities. In the context of Virtual Research Environments we can point at least two scenarios for application of SA – content analysis and indexing. Kruk and McDaniel (Kruk & McDaniel, 2009) argue that it is expected that semantic digital libraries can provide both automated and user-based annotations.

One of the examples of ready-to-use technology for SA in DH is Annotator [7] which is an open-source JavaScript library to easily add annotation functionality to any webpage. Annotations can have comments, tags, links, users, and more. It is being used in many DH related projects on the Web. One of them is *Infinite Ulysses* [8] which is research and reading community focused on interpretation of Ulysses by James Joyce. Members of this community have access to full version of transcribed work of Ulysses where each chapter, page, paragraph, sentence or word can be bookmarked and annotated. Readers create annotations by highlighting piece of text and adding comments. They have created over 1,000 annotations on Ulysses and tagged them with over 200 unique terms to make them filterable by theme, reader needs. Each annotation has its unique URI and can be dereferenced. There is a rating system for annotations where users can respond positive or negative to others' comments. This approach falls into the manual and semi-formal category ofSA. Only annotation subject has its own web identifier but the content its expressed in natural language.

One of the major projects that can be adopted for DH projects is hypothes.is [9]. This is an open-source web application based on Annotator. Hypothes.is, which was founded in 2011 in San Francisco, California, and is supported by philanthropic grants, has a bold mission: "To enable conversations over the world's knowledge" (Perkel, 2015). In 2015 Hypothes.is together with many publishers, libraries and research communities founded coalition on annotating scholarship fostering adding annotations over their content. Hypothes.is works as a browser plugin which allows to annotate any web content or users can add comments directly through its webpage https://via.hypothes.is/.Each time when plugin is activated user can add annotation to web content and sees comments created by other hpothes.is members. Annotations are stored on a dedicated Hypothes.is server and in 2015 there were around 250,000 annotation created by about 10,000 users (Perkel, 2015). Each annotation has its own URI and can be shared across the Web. Users can also reply to other annotations by making the comments. Hypothes.is allows for making annotations public and private

and for working in groups with invited users. The latter feature is interesting for potential application in education and for small research groups.

More advanced approach to SA presents Pundit annotator [10]. The software architecture is quite similar to Hypothes.is but Pundit introduces formal and ontological annotations. This platform allows for highlighting and commenting web resources and introduces subject-predicate-object model based on knowledge base. Annotations are organized in virtual notebooks. In this scenario the annotation subject (e.g. text fragment) is formally identified with the URI. Pundit offers eight default predicates: *identifies, is related to, describes, has author, has type, cites, quotes, replies to*, which represent the type of relationship or role in annotation act. The annotation object is represented with the URI of the relevant concept from Dbpedia [11] knowledge base. Annotations are created manually, simple by filling the subject-predicate-object boxes where the Pundit mechanism suggest relevant annotation predicate and object. Therefore manual semantic annotations are being expressed in machine-readable format.

Pundit has been implemented in DH project – Wittgenstein Source, which provides free access to Wittgenstein primary source editions and its being developed by the Wittgenstein Archives at the University of Bergen [12]. In this case we have ontological approach to semantic annotation. The application ontology developed for this project is called “Wittgenstein ontology” and it was intended primarily for the browsing of Wittgenstein's writings and their internal and external relations, including bibliographic metadata such as relations between Nachlass sources and "works", references to persons and works of others, datings of the single remarks, and also text genetic paths [13]. This ontology formally specifies entities and relationships between them relevant to the content and interpretation of Wittgenstein works. Within this virtual research environments user can make informal annotations by highlighting and commenting digital assets, add tags and use controlled vocabulary for concepts and named entity (e.g. from Wittgenstein Ontology, Word Net, DBpedia, Europeana), connect two different text by the means of annotation with appropriate predicate and use triple paths for creating formal semantic annotations. In the latter scenario users create knowledge graph by the means of formal RDF-based statements what makes an essential collaborative input into Linked Open Data Cloud.

Automatic approach to semantic annotations is related to the concept of named-entity recognition (NER) and named-entity linking (NEL). For textual content there are several widespread commercial services that automatically perform a light type of this form of annotation with a constantly improving degree of relevance [14] (Barbera et al., 2013). Named-entity recognition approaches can be divided in two families - those using text similarity and those using graph based methods. The first one takes into account string similarity between the mention in the text and concept or entity label in the knowledge base and according to the used algorithms return best candidate for

matching. Graph-based approaches rely on formalised knowledge described in graph form that is built from a Knowledge Base (KB). Frontini et al. (Frontini, Brando, & Ganascia, 2015) argue that the key idea of this approach is that for all ambiguous words in the context, senses that belong to the same semantic space should be selected, and that in this way two ambiguous words can mutually disambiguate each other.

The first approach to NER is exemplified by DBpedia Spotlight [15] which performs NER by referencing to DBpedia entity labels. This application extracts from text named-entity and matches them to DBpedia Knowledge Base. Annotated text includes URIs to matched entities from DBpedia. Frontini et al. (Frontini, Brando, & Ganascia, 2015) argue that this method is known to be very efficient, but it can only provide linking towards resources such as DBpedia, whose entries come with a description in the form of unstructured text. The other problem is that DBpedia is general knowledge base.

The example of graph-based approach to NER is the SemLib Project: Semantic Annotations for Digital Libraries (Morbidoni et al., 2011). This prototype application is resulted in Pundit software development and has been founded underEuropean Union's Seventh Framework Programme.

The promising solution for NER as SA can be reusing public Application Programming Interfaces (API) to external general and domain oriented knowledge bases within Virtual Research Environments. For example the Alchemy API product [16] allows not only for NER but also returns formally specified sentiment analysis for textual documents. The Library Linked Data initiative resulted in many projects where both authority and bibliographic data were converted to machine-computable datasets according to Linked Data methods and Web standards. One of the solutions for NER is VIAF - Virtual International Authority File [17]. This is a joint project OCLC with national libraries and agencies across the world with the main goal to aggregate authority data from national authority files and make them available on the Web. VIAF datasets can be access via web services and are also ready for downloading.

The implementation of KOS and NAF in semantic annotation requires the formal representation of the former.This includes the "semantization" of KOS using Semantic Web standards like Simple Knowledge Organization System (SKOS) or Linked Data approach. Therefore, formally represented/specified knowledge organization systems will allow for automatic approach to SA, e.g. named-entity recognition.

4 Conclusion and perspectives

Semantic annotation is one of the core elements in collaborative and interdisciplinary nature of Digital Humanities. In the context of Virtual Research Environments different scenarios for applyingSA – content analysis and indexing- are possible. Digital libraries using SA are expected to provide both automated and user-based annotations.

We argue that sustainability approach to semantic annotations can be understood as follows:

- SA should be based on formal model (eg. Open Annotations Data Model – Sanderson, Ciccarese & Van de Sompel, 2013),
- SA should be identifiable on the Web – the application of persistent identifiers,
- SA should be computable and shareable on the Web – the application of Web standards (eg. RDF-based solutions),
- SA should use ontological approach (concept oriented) – the application of formal knowledge bases (eg. domain ontologies) and KOS, NAF.
- Formal representation of knowledge organization systems and authority files and publishing them on the Web as a Web services is an important input of our community for development of the Semantic Web. Re-using controlled vocabularies for the purpose of semantic annotation creates additional domain of application for library tools and standards beyond the OPAC.

Notes

[1] http://www.jisc.ac.uk/publications/publications/pub_vreroadmap.aspx
[2] Digital RIs were developed earlier in the hard sciences than in the Humanities and currently receive a larger proportion of funding.
[3] Rockwell http://blogs.ischool.utexas.edu/f2011dh/tag/cyberinfrastructure
[4] http://www.connotea.org
[5] http://delicious.com
[6] http://id.loc.gov/authorities/subjects
[7] http://annotatorjs.org
[8] http://www.infiniteulysses.com
[9] http://hypothes.is
[10] http://thepund.it
[11] DBpedia (http://wiki.dbpedia.org/) is a crowd-sourced community effort to extract structured information from Wikipedia and make this information available on the Web. The DBpedia Knowledge Base consists facts from Wikipedia converted to ontological statements with Resource Description Framework and by the means of DBpedia ontology and additional formal specification of metadata schemes. Each entity from DBpedia Knowledge Base has its own URI that can be referenced on the Web, e.g. using SPARQL query language.
[12] http://www.wittgensteinsource.org
[13] http://wab.uib.no/wab_philospace.page
[14] e.g. Opencalais, Zemanta, AlchemyAPI
[15] http://dbpedia-spotlight.github.io/demo
[16] http://www.alchemyapi.com
[17] http://viaf.org

References

Agosti, Maristela & Ferro, Nicola (2003). Annotations: Enriching a Digital Library. In *Research and Advanced Technology for Digital Libraries*. Springer Berlin Heidelberg. Pp 88–100. doi:10.1007/978-3-540-45175-4_10

Agosti, Maristela, Ferro, Nicola, Frommholz, Ingo & Thiel, Ulrich (2004). Annotations in Digital Libraries and Collaboratories – Facets, Models and Usage. In *Research and Advanced Technology for Digital Libraries*. Springer Berlin Heidelberg. Pp 244–55. doi:10.1007/978-3-540-30230-8_23

Andrews, Pierre, Zaihrayeu, Ilya, & Pane, Juan (2012). Classification of Semantic Annotation Systems. *Semantic Web*, *3*(3): 223–248. doi:10.3233/SW-2011-0056

Arko, Robert A., Ginger, Kathryn M., Kastens, Kim A., & Weatherley, John (2006). Using annotations to add value to a digital library for education. *D-Lib Magazine*, 12: 21–35. doi:10.1045/may2006-arko

Babeu "Rome Wasn't Digitized in a Day": Building a Cyberinfrastructure for Digital Classicists, Babeu, 2011, http://www.clir.org

Barbera, Michele, Meschini, Federico, Morbidoni, Christian & Tomasi, Francesca (2013). Annotating Digital Libraries and Electronic Editions in a Collaborative and Semantic Perspective. In *Digital Libraries and Archives. 8th Italian Research Conference, IRCDL 2012, B*ari, Italy, February 9-10, 2012,. Springer Berlin Heidelberg. Pp. 45–56. doi:10.1007/978-3-642-35834-0_7

Barret, Elydia (2014), *Quel rôle pour les bibliothèques dans les humanités numériques ?*,Mémoire Master 2, Sous la direction d'Isabelle Westeel Directrice du SCD de l'Université Lille 3

Berry, David M. (2011). The computational Turn: Thinking the Digital Humanities. In *Culture Machine*, 12. [www.culturemachine.net]

Borgman, Christine (2000). The invisible library: Paradox of the global information infrastructure. *Library Trends*, 57(4): 652-75.

Frontini, Francesca, Brando, Carmen, & Ganascia, Jean-Gabriel (2015). Semantic Web based Named Entity Linking for digital humanities and heritage texts. In *Semantic Web for Scientific Heritage 2015. Proceedings of the First International Workshop Semantic Web for Scientific Heritage at the 12th ESWC 2015 Conference*. CEUR. Pp 77-88. [http://ceur-ws.org/Vol-1364/paper9.pdf]

Giacomo, Giuseppe de & Lenzerini, Maurizio (1996). TBox and ABox Reasoning in Expressive Description Logics. In*Proceedings of the Fifth International Conference on Principles of Knowledge Representation and Reasoning (KR'96)*. doi:10.1.1.22.8293

Hodge, Gail (2000). *Systems of Knowledge Organization for Digital Libraries: Beyond Traditional Authority Files*. Digital Library Federation. [http://www.clir.org/pubs/reports/reports/pub91/pub91.pdf]

Kruk, Sebastian Ryszard & McDaniel, Bill (2009). Goals of Semantic Digital Libraries. In *Semantic Digital Libraries*. Berlin, Heidelberg: Springer Berlin Heidelberg. Pp. 71–6. doi:10.1007/978-3-540-85434-0

Morbidoni, Christian, Grassi, Marco, Nucci, Michele, Fonda, Simone & Ledda, Giovanni (2011). Introducing the semlib project: Semantic web tools for digital libraries. In *CEUR Workshop Proceedings*, *801*. Pp 97–108.

Moulin, Claudine, Nyhan, Julianne, Ciula, Arianna *et al.*(2011). *Research Infrastructures in the Humanities*, European Science Foundation, Strasbourg.

Mustafa El Hadi, Widad (2015) (forthcoming). Digital Libraries & Digital Humanities: Some Reflections on their Synergy. Keynote address. In: *Proceedings of the First International Conference on Information Management and Libraries (ICIML).* Held at University of the Punjab at Lahore, Pakistan, 10-13 November 2015.

Oren, Eyal, Möller, Knud Hinnerk, Scerri, Simon, Handschuh, Siegfried, & Sintek, Michael (2006). *What are semantic annotations?* [http://www.siegfried-handschuh.net/pub/2006/whatissemannot2006.pdf]

Perkel, Jeffrey M. (2015). Annotating the scholarly web. *Nature*, 528(7580): 153–4. doi:10.1038/528153a

Svensson, Patrick (2010). The Landscape of Digital Humanities. *Digital Humanities Quarterly*, 4(1).

Tasovac, Toma (2015). Not All Texts Were Born Equal: Toward DH-friendly Digital Libraries. In *European One-Day Seminar*, Held at University of Lille 3, Lille, 27 May 2015.

Warwick, Claire, Terras, Melissa, Galina, Isabel, Huntington, Paul and Pappa, Nikoleta (2008). Library and information resources and users of digital resources in the humanities. *Program*, 42(1): 5 - 27

Wusteman, Judith (2008). Virtual Research Environments: What is the librarian Role? *Journal of Librarianship & Information Science.*

Johanna W. Smit, Clarissa M. dos Santos Schmidt, Lilian M. Bezerra, Marli M. de Souza Vargas and Ana Silvia Pires

Functional Classification of Archival Records: Some Questions and a Case Study with Records Produced by the University of São Paulo, Brazil

Abstract

Functional classification in archives is still not very common in practice. This paper presents a discussion on archival science theoretical framework, and subsequently, a case study on practical issues to classify current records based on a functional classification scheme.

Introduction

Although the literature on archival science highlights the importance of records classification as an essential activity of records management (Sousa, 2014a), issues regarding classificatory function are still overlooked, in our opinion, lacking theoretical development (Foscarini, 2010, p. 42) and do not discuss issues perceived as central by some researchers.

Initially, it is necessary to distinguish bibliographic classification from archival classification, considering both essential challenges of information organization, but distinct in their methods as they emerge from different informational perspectives. Briefly defining a complex controversial activity, while bibliographic classification aims to group documents that address the same subjects or information subjects, archival classification is not driven by subject matters, but by groups of records originated from the same activities, which part of the development of the functions fulfilled by public or private organizations. We shall restrict our paper to archival theories and purposes, which are not always addressed as an information organization issues, but considered by us as such. Even if "until the late twentieh century the archival discipline did not recognize information as its object of study, studying only records and archives" (Guimarães, Tognoli, 2015, p. 566), we have to consider that the *function* assigned to information is a very special one, since the records – and the information they carry – signify something that goes beyond information solely, but represent traces of the organizational functioning (Menne-Haritz, 2004). We state, thus, that archival systems are information systems and consequently can be discussed in the realm of KO, but that the information embedded in archival records has a special meaning – and this meaning has to be taken into account.

Functional classification

As stated before, the meaning of information from the archival perspective, aims at something beyond the records, but still plays a very important role to these records, since they are the traces of the organizational transactions: records show, in addition to

the information they carry, how transactions were organized, who ruled them, who executed them, on which basis, who counter-argumented them, what was decided and by whom, etc. Of course, records can be seen from different perspectives, but the archival perpective leads us to analyze the records and their information in a close and obligatory (mandatory) relationship with the organizational transactions they literaly record. It becomes, thus, relevant to understand the context in which every record was produced, and also to preserve the information about this condition and the original contents of the records because they are considered true evidence of the transactions. We stand far from a subject organization and indexing, but close to a more administrative view of records production and organization.

In order to underpin our argument, we need to highlight the main concept of archival theory, the *provenance principle*, which has been developed since the 19th century. The principle of provenance is central in the archival science literature and is defined as "the relationship between records and the organizations or individuals that created accumulated and/or maintained and used them in the conduct of personal or corporate activity" (International Council on Archives, 2000, p.15). Related to the provenance principle, we bring the concept of *archival bond*, as defined by InterPARES Project: "The relationship that links each record, incrementally, to the previous and subsequent ones and to all those which participate in the same activity. It is originary (i.e., it comes into existence when a record is made or received and set aside), necessary (i.e., it exists for every record), and determined (i.e., it is characterized by the purpose of the record)" (InterPARES Project, 2001). "[…] Therefore, the archival bond determines the meaning of the record" (Duranti, 1997, p. 217). The concept of archival bond completes the provenance concept, highlighting the fact that records carry contextual information, but "context is by definition outside the record, even if it conditions its meaning and, in time, its interpretatiuon, while *the archival bond is an essential part of the record*, which would not exist without it" (Duranti, 1997, p. 217).

Considering both concepts (provenance and archival bond) together, one can realize the importance of two groups of information that, together, represent the conditions in which records can be, and have to be, understood: a) information about the organizational context in which the records are produced or received, i.e., the identification of the administrative source (locus) that produced or received the record; and b) the place in which these records can be understood within the global context of the activities fulfilled by the organization. These two groups of information explain the two most discussed criteria for records organization in archives: the structural classification prioritizes the organizational (structural) context whereas the functional classification prioritizes the place of the activities in the global organizational functioning.

The opposition between structural and functional classification schemes has been, to some extent, avoided by the archival science community, which adopted the functional

classification, regarded as a more efficient response, but overshadows the discussion about the relationship between the two classification criteria, taking the risk of ignoring the provenance principle when functional classification is carried out to its extreme. We will resume this point later on.

Before developing our argument, two points have to be clarified [1]. 1) We focus our discussion about classification from the moment the records are produced, i. e., the moment when current records are created and the concepts of fund and series are still virtual (Bearman, 1999), but assume that the reason for determining the record's production does not suffer any change througout its life cycle; and 2) in our view, classification applies both to digital records and to traditional ones. If records organization was important in the past, with digital records it is certainly more important, since these records are not tangible and shelved (Bailey, 2007, Heredia Herrera, 2013, p.144).

As pointed out by Foscarini (2009), after a careful analysis of functional classification projects developed in five central banks' archives, functional classification is seen as far more difficult by the employees who have the mission of classification. Functional classification is neither "automatic", nor "intuitive" (Bak, 2010), it presuposes the identification of the activity represented by the records, and this identification is not always simple. Structural classification, in opposition, does not require an intellectual identification of activities, but the identification of the production *locus*, usually shown in the record's heading or signature. Functional classification presuposes knowledge of the big picture.

Returning to the functional classification scheme, it states that the development of the organizational mission can be organized – intelectually – by the identification of the functions accomplished and how these functions can be subdivided in activites, performed through actions or transactions.The concept of function and the criteria for its subdivision have different versions. A *function* can be understood as "all of the activities aimed to accomplish one purpose, considered abstractly" (InterPARES Project, 2001). "[…] The function does not produce records, but determines them" (Heredia Herrera, 2011, p.116). Schellenberg specifies: "records, as a rule, should be classified according to function. They are a result of function, they are used in relation to function, they should therefore be classified according to function" (2002, p.95). The other side of the coin is the *activities*, "a class of actions that are taken in accomplishing a specific function" (Schellenberg, 2002, p. 84). The activities are performed via transactions, "the smallest unit of business activity in an organization" (NAA, 1996, apud Foscarini, 2012, p. 22). Functions, activities and transactions summarize Schellenberg's F-A-T model. Conceived as a hierarchy, between the functions and the activities, several levels are sometimes cited. It is not our purpose to develop the discussion about these levels, considering they are all logical divisions of the functions and that these divisions can be more "structural" or more "functional".

Heredia Herrera states “we identifiy and classify functions and procedures before records. The classification of records comes from their assignment to the procedures of the refered functions/activities” (Heredia Herrera, 2013, p. 162).

What is classified?

Trying to get closer to our case study, another fact attracted our attention when reviewing the literature about functional classification. The literature discusses the conceptual framework for functional classification and details how to construct a functional classification scheme. We can sumarize the propositions in the scope of ISO 15489, which does not mention the structural classification scheme at all, but cites vocabulary control as an important support. The same caution regarding words choice was pointed out by Sierra Escobar (2004) and Heredia Herrera (2013). However, once the functional classification scheme is developed, the strategies to use it empirically are often not discussed, but may lead to challenging questions.

If archival literature agrees about the distinctions between functions, activities and transactions, the same agreement does not seem to occur when the discussion lies about what is classified, the central node of records organization.

We acknowledge the importance of functions and their practical development through activities. Activities, in turn, are performed through different record types, i. e., that one activity can be performed through different record series [2]. We adopt the distinction, according to which a series is a “sequence of items of the same record type” (Camargo, Bellotto, 1996, p.69), while record species is “the configuration asssumed by a record in accordance to its disposition and the nature of the information it contains” (Camargo, Bellotto, 1996, p.34). According to this Brazilian tradition, contracts, for example, are considered species, but an employment contract names a record type, and thus denominates a series. When Barbadillo Alonso (2010, p.96-97) claims that the key for an archival system lies in the adoption of the concept of series and that the acceptance of the fact that items are organized in series, the question is not how we classifiy the records, but how we classifiy the series; our response to that is how do we classifiy activities, since we consider the series as subdividions of activities.

We classifiy activities and detail them in series with potentially different disposal dates. Our propositon on this point assumes the following hierarchy for business activities:

- Function
- (functional sub-divisions, if applies)
- Activities
    - Record series – not concerned with classification, but a result of the identification of different record types related to the same activity, but possibly having different disposal dates.

Even if we are well aware of the difficulties represented by the functional view, its importance pleads for its adoption – we hope to clarify this point in the conclusions.

Case study

The University of São Paulo (USP), founded in 1934, is a public university, distributed in 11 campuses, where 6,090 professors and 17,199 technicians develop teaching and research for 59,081 undergraduate students and 30,039 graduate students (Master's and PhD degrees) [3]. In addition to teaching departments, USP counts 4 universitary hospitals (2 general hospitals, 1 odontological and 1 veterinary hospital), farms, libraries, museums, an orchestra and different cultural services provided for the community – these figures illustrate the university's size.

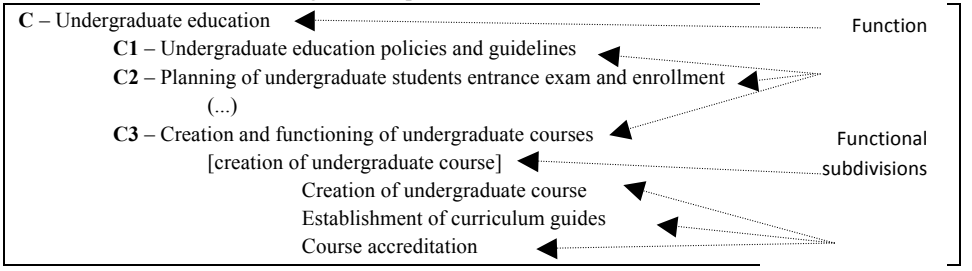
In the begining, the deanship had a record service that fulfilled the necessities of naming and accessing documents, aided by a digital system to control their flow. There were no disposal rules, which caused problems when recquired records were not found or were already destroyed. The initial centralized organizational structure has been replaced by a decentralized structure, and different transactions began to migrate from the traditional system to digital systems. After a two-year effort to produce the first records management tools, the Sistema de Arquivos da USP (SAUSP) was created in 1997. The oficial act was completed with 3 documents: a disposal list, a classification scheme and a glossary of record types produced or received at the university. Since the disposal list followed a structural logic, and the classification scheme followed a functional one, the development of a digital system to give support to the different actions involved in records management was impossible, unless we normalized both logics. In line with the literature, we began with transforming the disposal list (from structure to function) and, since time has passed, an update of the classification scheme.

Fig.1. USP's functional classification scheme – the main functions

A - General administration and university organization	**M** - Technological resources management
B - Cultural and extramural activities	**N** - Finance management
C - Undergraduate education	**P** - Human resource management
D - Graduate education	**Q** - Research
E - Pre-school, basic, intermediate and technical education	**R** - Health care services
F - Information resources management	**S** - Veterinary health care services
G - Communal areas and equipment management	**T** - Students support and counselling services
J - Patrimony and materials management	**U** - Sports and physical activities services and management
L - Agriculture and livestock resources management	**V** - Publishing and media services

An example of the hierarchy follows.

Fig.2. Example of function C subdivision



Steps to classify "real" records

Based on a functional classification scheme, we propose and tested some rules to classifiy records at their production moment, since literature does not mention these practical issues very often. The difference between simple and complex series [4] has to be taken into account, as well as how to classifiy records that are related to different functions (cross-functional records) [5]. In short:

1. Simple series – classify them into the most specific activity;
2. Complex series - classify them into the most specific activity;
3. Cross-functional series - classify them into the most specific original activity and classify into the most specific activity of arrival, but maintain the original classification recorded, in order not to lose its track and its original production's reason.

Conclusions

A relational digital system has to be created, in which some metadata are necessary to correctly represent the structural data that identifiy a record and distinguishes it from others. Other metadata will supply information about the recorded activity and, since the activities are organized in a hierarchy, the broader function in which the records preserve their meaning. Functional classification, specifically, empowers distinct goals, the first of which characterizes the well-known information retrieval, by naming the access points that organize records retrieval (Schellenberg, 2002). This goal is often mentioned and effectively represents a major question in archives. A second goal, less mentioned, is also very important, because it preserves the link between records and shows how different records cooperate in the global functioning. This goal is, of course, central to archival theory and was well developed by Foscarini (2009, p.54). A third goal is scarcely mentioned, but also important: providing the necessary basis to develop disposal lists and complete the description of records, series and funds in the historical archives realm (Sousa, 2014b, Foscarini, 2009, Heredia Herrera, 2013). The first goal emphasizes information retrieval, the second goal ensures the preservation of

the archival bond, and the third goal supports the production of other archival management tools.

In archives, each record can be compared to a sound, but the music is produced by functional classification.

Notes

[1] As these points are not the focus of this paper, they will not be brought to discussion.

[2] We are aware that this understanding of record series is not consensual. Since the concept of series has different definitions, well known archivists propose the classification of series (e. g. Heredia Herrera, 2013, p. 152, p.169). Following the serie's definition we adopt, the classification schemes stop at the activities level, identified as divisions of functions.

[3] Data from 2014 statistical annuary [https://uspdigital.usp.br/anuario/AnuarioControle]

[4] Simple series are composed of records of the same type, opposite to complex series, which join different record types to fulfill one transaction, e.g. an employment contract (Jimenez González, 2003).

[5] Cross-functional series are series that are produced, for example, in an activity related to one function, e.g. evidence of a missing equipment (function G in USP's classification scheme) and ends up as a refund action, supposing the robber was identified (function N in USP's classification scheme).

References

Bailey, Steve (2007). Taking the road less travelled by: the future if the archive and records management profession in the digital age. *Journal of the Society of Archivists*, 28(2):117-24.

Bak, Greg (2010). La classificación de documentos electónicos: documentando relaciones entre documentos. *Tabula*, 13: 59-77.

Barbadillo Alonso, Javier (2010). Classificaciones y relaciones funcionales de los documentos de archivo. *Tabula*, 13: 95-112.

BearmaN, David (1999). Documenting documentation. *Archivaria*, 34: 33-49.

Camargo, Ana Maria de Alemida & Bellotto, Heloisa Liberalli (1996). *Dicionário de terminologia arquivística.* São Paulo: Associação dos Arquivistas Brasileiros – Núcleo Regional de São Paulo/Secretaria de Estado da Cultura.

Duranti, Luciana (1997). The archival bond. *Archives and Museum Informatics*, 11: 213-8.

Foscarini, Fiorella (2009). *Function-based records classification systems*: an exploratory study of records management practices in central banks. PhD Thesis – University of British Columbia, Vancouver.

Foscarini, Fiorella (2010). La classificación de documentos basada en funciones: comparación de la teoria y la práctica. *Tabula*, 13: 41-57.

Foscarini, Fiorella (2012). Undestranding functions: an organizational culture perspective. *Records Management Journal*, 22(1): 20-36.

Guimarães, José Augusto Chaves & Tognoli, Natália (2015). Provenance as a domain analysis approach in archival knowledge organization. *Knowledge Organization*, 42(8): 562-9.

Heredia Herrera, Antonia (2011). *Lenguaje y vocabulario archivísticos*: algo más que un diccionario. Sevilla: Junta de Andalucia – Consejeria de Cultura.

Heredia Herrera, Antonia (2013). *Manual de archivística básica*: gestión y sistemas. Puebla: Benemérita Universidad Autónoma de Puebla.

International Organization For Standardization. (2001). *Information and documentation: records management.* ISO 15489. Geneva.

International Council On Archives. (2000). *ISAD(G)*: General international standard archival description. 2ª ed. Madrid: Subdireccion General de Archivos Estatales.

INTERPARES Project. (2001). *Glossary.* [http://www.interpares.org/book/interpares_book_q_gloss.pdf.] Acessed on 5 May 2016.

Jimenez Gonzalez, Gladis (2003). *Ordenación documental*. Bogotá: Archivo General de la Nación de Colombia.

Menne-Haritz, Angelika (2004). *Business processes*: an archival science approach to collaborative decision making, records, and knowledge management. Dordrecht: Kluwer Academic Publishers.

Schellenberg, Theodore Roosevelt (2002). *Arquivos modernos*: princípios e técnicas. 2ª ed. Rio de Janeiro: Editora FGV.

Sierra Escobar, Luis Fernando (2004). Como identificar y denominar una serie documental: propuesta metodológica. *Biblios*, 5(20): 49-61.

Sousa, Renato Tarciso Barbosa de. (2014a). A representação da classificação e indexação automática de documentos de arquivo. In *ENANCIB, XV*. Held at Federal University of Minas Gerais. Pp798-811. [http://enancib2014.eci.ufmg.br/documentos/anais/prefcio_ANAISFINAL.pdf.]

Sousa, Renato Tarciso Barbosa de. (2014b). Alguns apontamentos sobre classificação de documentos de arquivo. *Brazilian Journal of Information Research Trends*, 8(1/2): 1-24. [http://www2.marilia.unesp.br/revistas/index.php/bjis/issue/view/289]

K. S. Raghavan, I. K. Ravichandra Rao and K. N. Bhargav

Knowledge Organization in a Multi-disciplinary Domain: Case Study of Forensic Science

Abstract

Knowledge organization in Multidisciplinary domains is a challenging task. It is questionable if either the top-down approach adopted by the traditional KOS or the network approach based on citation links individually prove adequate for effective organization and retrieval of information in multidisciplinary domains. This paper presents a case study of Forensic Science based on an analysis of literature in the domain indexed in the *Web of Science*. The multidisciplinary nature of the domain is brought out. The changes in the contours of the domain over a 15 year period are mapped and contrasted with the map of the domain in traditional KOS. An experiment to compare the degree of overlap between search outputs (document sets) retrieved using keyword searches and those retrieved using citation-based searches indicates that a combination of multiple approaches is useful in such domains. The inadequacy of any single approach as also the complementary nature of traditional KOS and citation networks for effective retrieval of information are demonstrated.

Introduction

Knowledge organization systems (KOS) have, among others, two major objectives:

(a) To support information retrieval in response to actual or perceived needs of users of information;

(b) To facilitate a logical order of information objects on the shelves of a library or their metadata records in a bibliographic file to support browsing and knowledge discovery.

Most KOS, including special KOS are discipline-oriented. Langridge (1992) and Hirst (1977) refer to disciplines as '*Forms of knowledge*'. Library classification schemes such as Dewey, UDC, Colon, and Bibliographic Classification are all examples of traditional KOS structured on the basis of the notion of *'forms of knowledge'*. All these essentially adopt the *Top-down approach* and seek to structure and map a discipline in a hierarchical fashion. Even faceted systems are hierarchical in their approach for handling and structuring concepts within each facet. Special classification schemes and micro-thesauri are also primarily discipline-oriented and, for pragmatic reasons, cover facets of closely allied disciplines also. MeSH is a good example of a micro-thesaurus focused on a single discipline. There are obvious advantages in adopting a discipline-based approach. Most KOS have proceeded on the assumption that a reasonably stable broad map of the structure and composition of a discipline is feasible if formulated based on consensus among the members of the research community. This is understandable when seen in terms of the kind of stability that the notion of *forms of knowledge* lends in providing a base for knowledge organization. However, with research becoming increasingly multidisciplinary in nature, a discipline-based approach to knowledge organization, by itself, may not be adequate to meet the information organization and retrieval needs, particularly in

multidisciplinary domains. An examination of the history of knowledge organization indicates that the need for creating classes of information objects to complement and supplement the kinds of groupings created by traditional KOS has for long been recognized; some solutions have been suggested and even implemented in information products, bibliographic databases and library catalogues to enhance retrieval. The *see* and *see also* references in library catalogues and bibliographies, Dewey's *Relativ Index*, Ranganathan's *Chain* Index, and what Jack Mills (Mills, 2004. p. 545) has referred to as '*collocation of distributed relatives*' are all aimed at grouping what is scattered by the KOS (a function similar to what a back of the book index accomplishes). The subject approach to information organization expanded during the 20th Century. Even as conventional KOS were being revised to cope with new developments in the universe of subjects, new KOS such as those based on facet analysis, KOS to serve as switching languages (e.g., BSO), Thesauri, Thesaurofacet, Classarus, etc were developed. Parallel to these there were attempts to exploit the capabilities of technology for indexing and information retrieval leading to a wide range of techniques based on *keywords*. Techniques were developed to improve precision performance of such systems by computing the degree of *relevance* of every document retrieved and generating a ranked output based on computed relevance. In the 1950s Eugene Garfield came out with a different approach based on citations to classing and grouping of documents to support information retrieval (Garfield, 1955) and implemented the idea in the *Science Citation Index*. Citation-based grouping of documents does not have as its basis the *Aboutness* of a document, a feature that characterizes all traditional KOS based on *forms of knowledge*.The bibliometric approach is believed to enhance retrieval by suggesting useful links to related documents (and thus between their topics) which are not provided by traditional KOS.

There are also some who argue that the systems developed by librarians are *over sophisticated, at least in the display forms in which they presently exist* to be of much use to the end user in information retrieval. (Taylor, 1968). He says: "*.... subject naming systems however appear to be more concerned with the description of physical objects (books, papers) than assistance to the user in defining his subject. This is an important and critical differentiation for present systems are object-oriented (static) rather than inquiry-oriented (dynamic)*". Today there is a considerable amount of empirical data available on information seeking behaviour of end users. Whether such empirical studies help identify the factors impacting users' information seeking behaviour so as to contribute to the theoretical base of knowledge organization andinformation retrieval needs to be examined.

The broad approaches referred to in the preceding paragraphs could be categorized as:

(a) Pragmatic approach adopted by library classification schemes and other vocabulary control mechanisms such as thesauri, etc that view the published

literature by members of the research community in a knowledge domain as defining the contours of the domain;

(b) Bibliometric approach adopting the view that networks of documents based on citations by members of the research community collectively represent the domain;

(c) The Co-word analysis approach adopting the view that clusters of keywords and / or descriptors adequately map a knowledge domain.

The Present Study

Knowledge organization to support information retrieval in multidisciplinary areas could be a challenge. Understanding a domain is a necessary prerequisite for knowledge organization. This paper looks at issues related to knowledge organization in a multi-disciplinary domain, viz., Forensic Science and how they impact retrieval. According to Paul Hirst, disciplines are closely associated with their knowledge base (Hirst, 1974). He identifies seven *forms of knowledge* and suggests that all *pure disciplines* relate to one of these. He also refers to *practical disciplines* derived from one of these *forms* that are aimed at solving practical problems. Looked at from this point of view, *Forensic Science* would qualify as a *practical discipline* based largely on the *form of knowledge* of Biological Sciences. However, literature on *Forensic Science* also includes a substantial number of papers falling within the *form of knowledge* of Social Sciences as its applications are mainly in Criminology and Law. The social importance of *Forensic Science* has led to a situation in which the knowledge domain has emerged as a multi-disciplinary area with inputs from and applications in several disciplines. One of the reasons that prompted the choice of the subject was the fact that there is no one index to forensic sciences literature. *Wikipedia* defines Forensic science as the application of science to criminal and civil laws. These sciences refer to a group of sub-disciplines that apply their principles and methods to legal questions of a criminal or civil nature. As a domain it integrates with many areas of science and social sciences. The Wikipedia also gives an idea of the sub-disciplines of Forensic Science which are drawn from Physiological sciences, Social sciences, Forensic criminalistics, Digital forensics and a few other areas**.**

The objective of this paper is to look at issues related to knowledge organization in forensic science and their impact on retrieval. It is hypothesized that traditional discipline-oriented KOS are not adequate for knowledge organization in multi-disciplinary domains. In order to test this, the study examines the structure and composition of Forensic science and how it has evolved and transformed over the last 15 years with a view to identify implications for knowledge organization and information retrieval. More specifically the study aims to:

(a) Identify the keyword clusters defining the contours of the domain;

(b) Identify differences between the contours of the domain (e.g., conventional KOS and that obtained through the present empirical study);

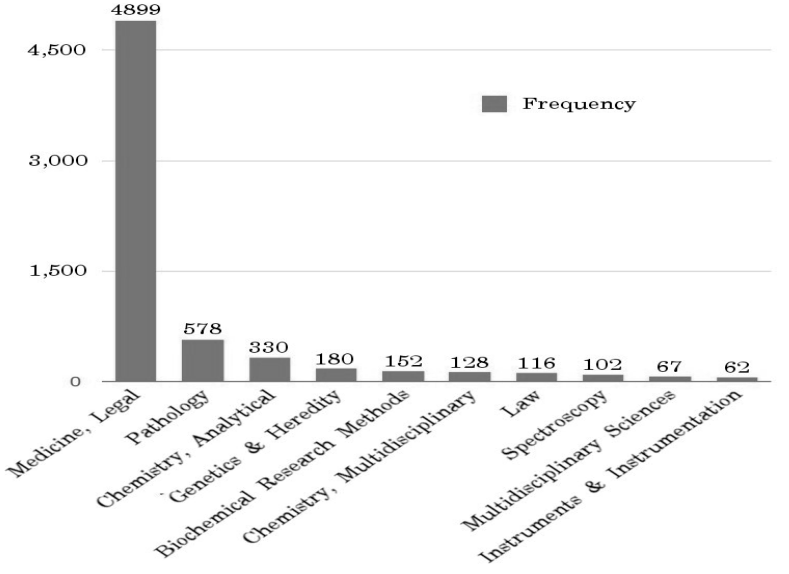
(c) Compare the search output based on search and retrieval using keywords and that using citation-based search.

Searches were carried out in the *Web of Science* (WoS) for documents with the word '*Forensic*' or '*Forensic Science*' or '*Forensic Research*' in document title / keywords / abstract. The database was searched for all journal articles in the English language on the subject published between 2000 and 2015. The data was analyzed and the following paragraphs present the analysis and the findings.

Analysis and Results

A total of 6359 records were retrieved from WoS. For every item retrieved all the data elements available in the database were downloaded for further analysis. For the purposes of this paper a few data fields were considered; viz., the *keywords* assigned by the authors of the paper (to each paper), the *web categories* to which the item was assigned by *Web of Science*, and the list of references appended to each document. The multidisciplinary nature of the domain is clearly indicated by the distribution of literature (Figure 1).

Figure 1: Distribution by WoS Categories



Maps of the top keywords for the years 2000, 2005, 2010 and 2015 indicated changes in the contours of the domain over time. While DNA / DNA Typing were the predominant research themes in 2000 and 2005, Forensic Anthropology had emerged as a major area of interest since 2010. The facets that collectively form the contours of the domain based on the most significant keywords are shown in the figure 2.

A matrix showing co-occurrence of author keywords assigned to the documents was prepared. This matrix was used as the input for identifying the strength of association between keywords using the Euclidean Distance Model. The association / proximity and distances between keywords were visualized using Multi-Dimensional Scaling (Figure 3). Concepts occurring in the same quadrant in close proximity to one another are closely related to each other. For example, the figure suggests that the concepts *sex determination; sudden death; suicide;* and *VWA* occur in the same quadrant and are located in close proximity to one another. It may be, therefore, inferred that the terms are closely related to each other. A cursory examination of the LCSH was made to examine the descriptors in the vocabulary that collectively define the domain of Forensic science. It was, however, found that many of the frequently occurring keywords in the domain did not figure among RTs to Forensic Science or to other descriptors covering the domain. It is proposed here that, among others, such visualization techniques could be used as the basis for generating RT links in thesauri and similar controlled vocabularies so as to more comprehensively cover the related terms as also to facilitate navigation while searching.

Figure 2: Composition of Forensic Science

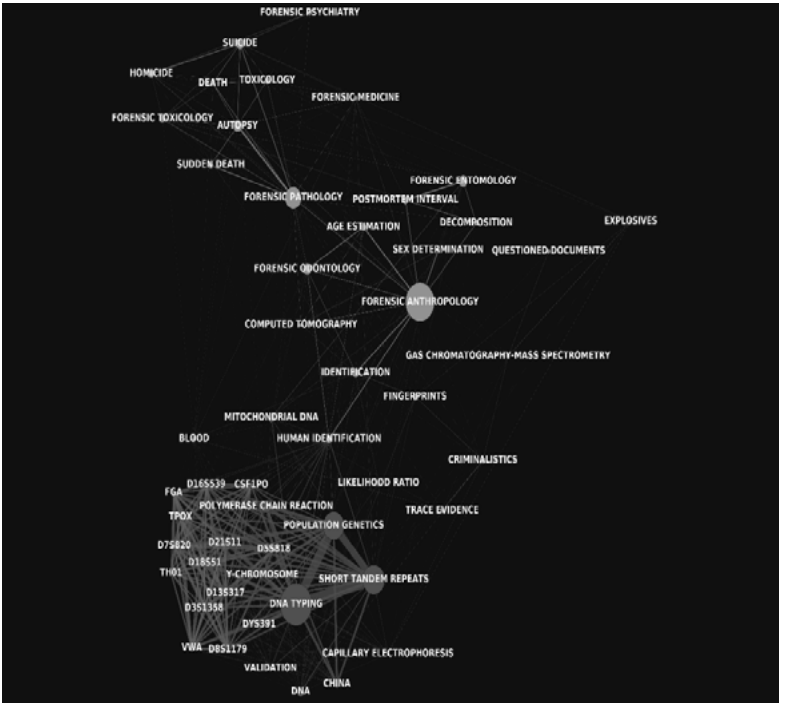
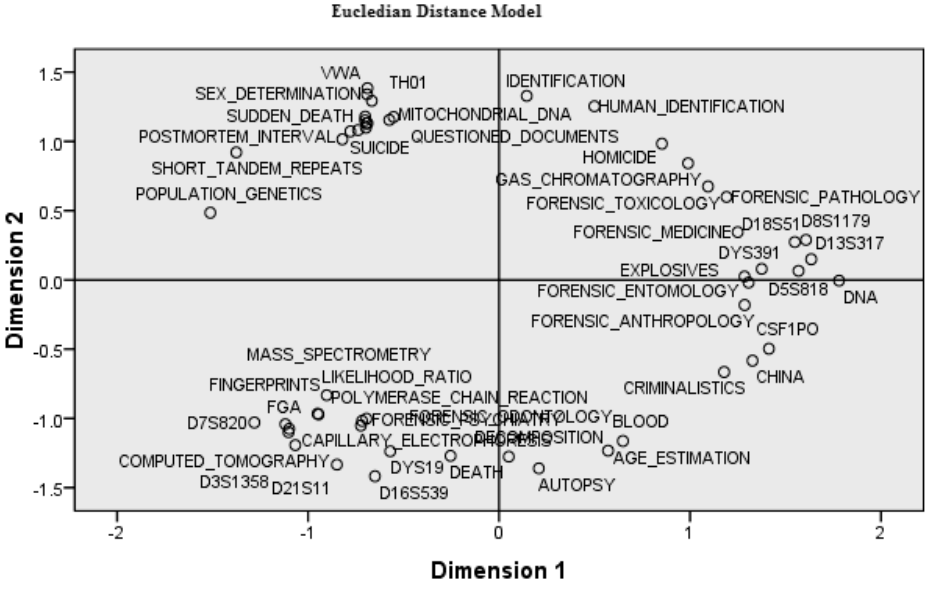


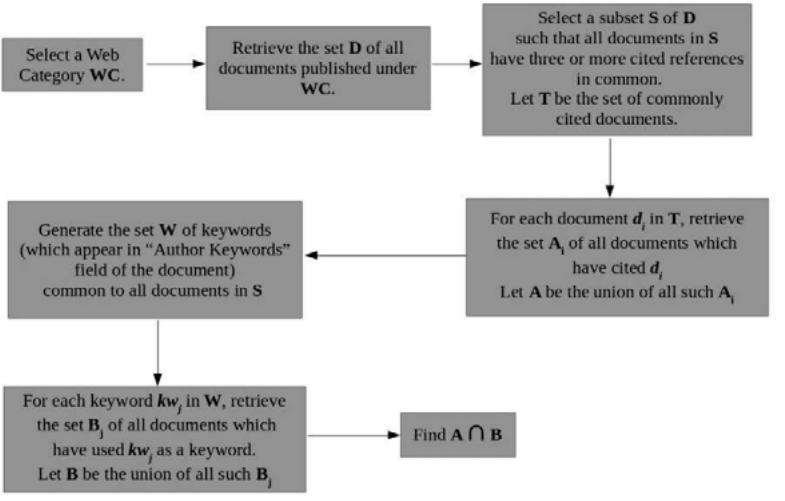
Figure 3: MDS Plot of Co-occurring Key Words



Knowledge Organization and Information Retrieval

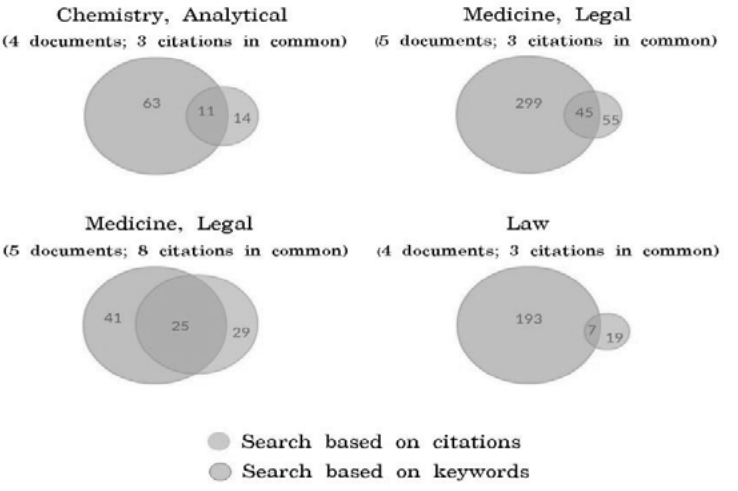
A major objective of this study was to examine the impact of KOS on information retrieval. A small scale experiment was carried out to see the impact, if any, of the different approaches to knowledge organization on retrieval effectiveness in the multi-disciplinary domain. The experiment sought to compare the document sets retrieved using keyword searches and corresponding cited documents-based searches. The idea was to see the extent or degree of overlap between the two sets of documents and the proportion of unique documents retrieved by each of the two different modes of searching. An idea of the methodology adopted for the test can be had from the flow diagram (Figure 4). A few searches were carried out in the database of documents that

Figure 4: Flow Diagram



was the initial output from the WoS database (6000+ documents). The sets of documents retrieved based on citations and based on keywords were compared to see the differences, if any, in the document sets retrieved. The results are presented in figure 4. In general the size of document classes retrieved using keyword-based searches was much larger than the document set retrieved using citations. Interestingly the degree of overlap was of a very small order as there were very few documents at the intersection of the document classes retrieved based on the two approaches. This clearly suggests that the two approaches to KO in multidisciplinary domains complement each other rather than functioning as substitutes for each other. Another interesting observation was that a union of search results based on common references to five or more source documents yielded an output that had much in common with the document set retrieved by keyword search. In the figure, the two sets of documents retrieved in the broad area of *Legal Medicine* indicate the differences. While it is difficult to generalize based on this study, this is an area that could be further explored.

Figure 5: Retrieval sets based on Keyword(s) and Citation(s)



Knowledge Organization in a Multidisciplinary Domain

Traditional KOS are based on the notions of disciplines, sub-disciplines and facets of these discipline / sub-disciplines and adopt a top-down hierarchical structure. However, for a multidisciplinary domain such as Forensic Science a top-down approach may not necessarily work. The citation indexes sought to overcome of the limitations of the traditional KOS by supporting cross-disciplinary searches for relevant documents. As research is becoming increasingly multi-disciplinary, it is becoming important to explore the adequacy of such approaches for effective information retrieval. The study reported in this paper brings out the need for taking care of different perspectives and viewpoints as no one approach by itself provides a comprehensive and adequate view of the contours of a knowledge domain / search request which is necessary for knowledge organization and information retrieval. Both the rational approach, widely employed by designers of controlled vocabularies, and the bibliometric approach to create clusters of documents based on references / word co-occurrences, view a domain as a subject of discourse whose extensions and intensions are represented by the research literature generated by its scholar community – the rational approach considering the facets constituting the themes of the research literature output while the bibliometric approach considers the network of research literature linked via references as defining the contours of the domain. Research on information-seeking behaviour is still in its early stages to provide an adequate basis for mapping the contours of a domain. However, data generated by empirical studies of users' information seeking behaviour as also folksonomy are potentially important for

mapping the contours of a domain as they complement and supplement the other two approaches. The results of the study and the experiment reported in this paper clearly validate the suggestion that knowledge organization in multi-disciplinary domains requires adoption of different perspectives and viewpoints and effective retrieval can be realized only by a combination of different approaches to knowledge organization and information search.

Acknowledgement: This work was supported in part by the World Bank / Government of India grant under TEQIP Programme to the Centre for Knowledge Analytics & Ontological Engineering at the PES Institute of Technology, Bangalore, India.

References

Garfield, Eugene (1955). Citation indexes for science. *Science* (*AAAS*), 122 (3159): 108–11.

Hirst, Paul (1974). *Knowledge and the curriculum*. London: Routledge & Kegan Paul.

Langridge, Derek (1992). Bliss, disciplines and the new age. *Bliss Classification Bulletin*, 34: 8-13.

Mills, Jack (2004). Faceted Classification and Logical Division in Information Retrieval.*Library Trends*, 52(3), Winter: 541–70.

Small, H. G. (1980). Co-citation context analysis and the structure of paradigms. *Journal of Documentation*,36(3): 183-96.

Smiraglia, Richard P. (2012). Epistemology of Domain Analysis. In *Cultural Frames of Knowledge*. Wurzbürg: Ergon Verlag. Pp 111-24.

Taylor, Robert (1968). Question negotiation and information seeking in libraries [https://faculty.washington.edu/jwj/lis521/Taylor1968.pdf]

Aderibigbe Stephen Ojo, Akhigbe Bernard Ijesunor, Afolabi Babajide Samuel and Adagunodo Emmanuel Rotimi

Towards a Sustained Collaborative Knowledge Sharing: The F2F Interactive Sharing Paradigm

Abstract

This paper addresses the possibility of using the Friend-to-Friend (F2F) collaborative platform to propose an interactive knowledge sharing paradigm towards sustained collaborative knowledge sharing in a connected society. Although social media network remains a veritable platform and tool that was created for knowledge sharing; the platform still suffers security and privacy exposure since users does not have control over their data. Relying on the theoretical support of the philosophy of social trust, the theoretics of F2F collaborative capability, the distributed hash table, and predecessor replication techniques were leveraged to propose a Trust-Aware Model (T-AM). This current paper approaches things differently from what obtains in literature regarding trust by towing the perspective of (i) trust that connote the feelings of vulnerability and (ii) how peers are likely to behave over time as exemplified in a strain-test situation. This context of trust makes the T-AM a novel contribution to the epistemological dimension of explicit and tacit knowledge sharing and the social dimension of knowledge organization. Interestingly, the research that resulted in the T-AM is still ongoing. In future an archetype of the T-AM will be implemented. Galuba's architecture was used in this paper to exemplify the workability of the T-AM. Thus, this paper is descriptive and not prescriptive, thus highlighting the need to validate the claims in this paper, which is a limitation in this work.

1 Introduction

Cultural diversity is an asset and should be utilised to facilitate innovation and global creativity (Lauring, 2007). For innovative thinking, this diversity can help people proffer different solution to identical problems (Ardichvili, 2006). The advantages of cultural diversity can be leveraged through Sustained Collaborative Knowledge Sharing, which is necessary in a connected society (Taylor, 2007). Nevertheless, the current cultural, scientific, and technological way of organizing and sharing knowledge still face a major challenge. The challenge has to do with what appropriate knowledge organization and sharing - the focus of this paper - paradigm to use? For explicit knowledge sharing, the existing ways of sharing them are ubiquitous and easy to interface with reality. It can be captured, codified and transmitted easily since its sharing practices are more communal in the working environment (Ali *et al.,* 2015). But, for tacit knowledge - the opposite of explicit knowledge - which are in the form of experiences, skills, know-how, or know-whom; a systematic and pragmatic way of disseminating them is needful. There are communities of practice that use the knowledge from experiences and skills (Hildreth *et al.,* 2000); hence a trustworthy and secured way of sharing such knowledge is required. This is because the knowledge is tacit and it is the innovative and creative potentials of people that can be harnessed.

In today's dynamic world, which is a knowledge based economy the flow of knowledge is very prompt, and the receiving and sharing of knowledge is from so many different sources (Ali *et al.,* 2015). This can be challenging for the portrayal and

visible performance role of the strength of a knowledge sharing paradigm. One of the ways to avoid this (we believe) is to look beyond the adoption and use of the informal communal ways, because the issue of trust among receiving and sharing peers comes in. We believe that even with the formal way of knowledge sharing with its successful collaboration (Leonard and Sensiper, 1998; Li, 2008; Hassandoust and Kazerouni, 2009); everyone would want to share (particularly tacit) knowledge with only those they trust. This is currently a huge challenge, which has been fuelled by the development of new technologies, such as e-collaboration and other communication technologies. The evolution of new forms of interaction and collaboration; the advent of the Internet and the World Wide Web that have engendered myriads of teams (communities) to be able to share knowledge and work remotely on projects (Hassandoust and Kazerouni, 2009); and the advent of the social network media that has provided an unprecedented medium for global communication and knowledge sharing (Vijayakumaran and Vinod, 2014; Aderibigbe *et al.,* 2015), have all also exacerbated the desire for users to want to collaborate and share knowledge with only trusted peers.

The context of this paper is the Social Network Media (SNM). It was adopted because of its wider space for both Explicit and Tacit Knowledge (E&TK) sharing and collaboration. Though the SNM offers attractive features, they also suffer some limitations. They rely on central authority thereby placing even personal information in the hand of the service providers (Aderibigbe *et al.,* 2015). This centralised nature of user data repositories, which can be used inadvertently for data mining and targeted advertising by the service providers (Buchegger *et al.,* 2009) further compound the foregoing concern. These data (some of which are E&TK) are valuable information, which ought to be at the exclusive reserve of the user. Users should have the right to determine what happens to their data. Based on these arguments, the need to protect user data is imperative in E&TK sharing expecially in the SNM environment. Firstly, users should know what is possible with the lots of data that accrue from their activities on the network. Secondly, this usual practice of having exclusive right to user data by social network owners need to be reconsidered. This is because the practice puts users' confidentiality at the risk of infringement (Steel and Vascellaro, 2010; Greenwald and MacAskill, 2013). Thus, the need for users in the SNM platform to have some measure of control over their data - E&TK is overarching.

The aim of this paper is to propose a trust-aware collaborative model that can manage trust between collaborating parties and put some level of control over data in the hands of SNM users. The paper is further structured as follows: 2.0 is the section committed to literature; 3.0 and 4.0 contains the theoretical foundation and the methodology employed (in the paper) respectively; 5.0 covers the proposed F2F system architecture; and the papers conclusion is presented in section 6.0.

2 Literature

A number of authors have attempted to use the theoretics of trust as it bothers on online collaboration. For example, the work of Clarke *etal.* (2010) proposed an F2F technique called the Darknet. It allowed users to communicate directly only with their own trusted peers, but with the proviso that peers' identity is only revealed to the peers they already trust. While everybody has to trust somebody in a network, there was no central party whom everybody must trust. An F2F computing platform exists where attempt has been made to transfer control of users' data from the service providers hosting the data to the actual users (Galuba, 2008). This work by Galuba (2008) highlighted the fact that F2F can be useful for creating platforms that will give users much more control over their data than it is currently with the SNM. The work also showed the possibility of putting in place a reliable level of privacy and security that is not possible in any centralized knowledge sharing architecture (Galuba, 2008). A push-pull-clone model for trust-based collaborative editing with contract deployment capacity over F2F network has also been proposed by Truong *et al*. (2011). In the model, contracts were specified when data are shared between friends and a log auditing protocol is used to detect user's misbehaviour. If the contract(s) is violated after receiving data by a peer, a penalty is meted out to the peer. For, Forster *et al*. (2012), a collaborative business process model was developed by extending the Cheetah experimental platform designed for investigating how a business model can be collaboratively created.

Other attempts at contributing a collaborative model for knowledge sharing include that of Varlamis*et al.* (2013), Chen and Pan (2014) and Avesani (2004). Varlamis*et al.* (2013) used social network metrics to formulate a trust-aware collaborative model for generating personalized user recommendations. The work suggested that the impact of various measures in recognizing trustworthy actors in a social network can be recommended to specific users. In Chen and Pan (2014) a social collaborative user model based on social network data that detects contact and collaborative relationship between collaborators from both contact and work information was developed. The model was also able to reveal social relationship between collaborators with a focus on recommending accurate collaborators. Earlier on, Massa and Avesani (2004) had developed a trust-aware solution that was motivated from the argument that trust awareness can be useful to solve traditional problems in recommender systems. However, this current paper approaches trust differently from others in literature by towing the perspective of (i) trust that connote the feelings of vulnerability and (ii) how peers are likely to behave over time as exemplified in a strain-test situation. This context of trust makes the proposed model in this paper a novel model and contribution to the epistemological dimension of E&TK sharing and the social dimension of knowledge organization.

3 The Theoretical Foundation

This paper drew from the theory of trust and used the distributed hash table and predecessor replication techniques to postulate a robust trust aware collaborative model that supports E&TK sharing. The transfer of a commendable level of control over data to users motivated this paper. We debate therefore the fact that organizing knowledge for a sustained world entails the belief that trust is a language that must be understood (particularly) in a world that is culturally diversified. This is because trust and knowledge sharing are correlated and both can mitigate the negative effects of diversity in knowledge sharing communities (Pinjani and Palvia, 2013). This highlights the plausibility to address the issue of trust in this paper. Thus, it was essential in this circumstance to consider mutual trust and knowledge sharing. There is therefore the need to ensure that the trustworthiness of any collaborating peers that are participating in the network is available.

Philosophically, trust is contingent upon the issue of reliability and effectiveness. It would be necessary to model processes that connote the feelings of vulnerability and the expectations of how a partner peer is likely to behave across time (Kramer and Carnevale, 2001). The strain-test situation provides a clue to want to model a typical trust relevant situation. For example, trusting a person that one has not had any collaboration with can be risky. This is because of the feeling of being vulnerable. It is also possible that a trusted friend can equally change and become dubious at some point. Interestingly, trust is not immutable. It changes from time to time (Truong *et al.,* 2011). When a trustor takes a risk in a trustee that leads to a positive outcome, the trustor's perception(s) of the trustee will be enhanced. Likewise, the perceptions of a trustee will decline when at another time attempting to trust others lead to unfavourable conclusions (Mayer *et al.,* 1995). This implies that engaging in trusting behaviour will affect trust directly. Base on this conception; the trustworthiness of a peer can be modelled as a function of self-experienced information (i.e. interaction-derived or first-hand information) and the ratings of second-hand information. This is consistent with the practice in Liang and Shi (2008), and Netrvalova and Safarik (2009, 2012). Evidently, drawing on the foregoing scenario; the recommendation from trusted friends and the continuous computation of their trustworthiness is needful. This is the only way to know who to trust and share E&TK with. This should be taking seriously. Based on this theoretical frame, the F2F interactive sharing paradigmatic methodology is thus presented.

4 Methodology

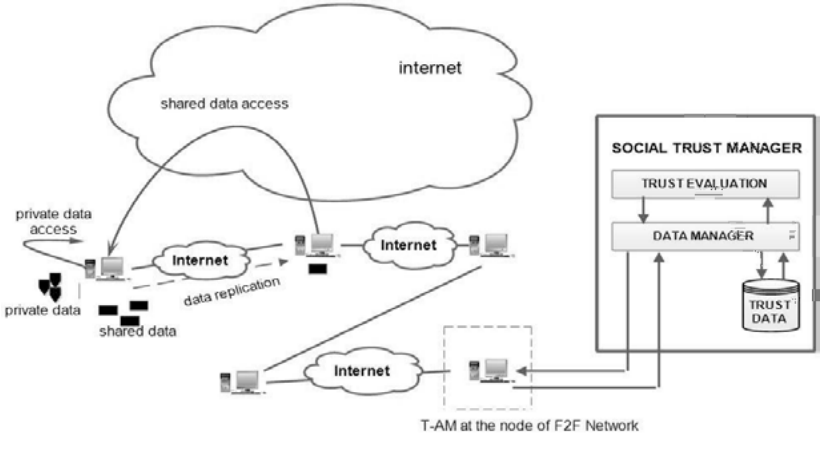
The formulation of a trust-aware model for implementation in an F2F network (based on our experience) requires some algorithms to perform specific tasks in ensuring that trust data are available when needed. These tasks included a lookup activity and trust computation. Additionally, since peer-to-peer networks are often susceptible to churn, it was therefore necessary to ensure that trust data would be

available for trust computation. The create the intended trust-aware model a Distributed Hash Table (DHT) and the Symmetric Replication Technique (SRT) is proposed. To be able to use the DHT and the SRT, the concept of trustworthiness would be reduced to personal trust, reputational trust and recommendation trust. This section is further partitioned into the subsections that discourses the trust-aware model as it pertains to the lookup feature, and the replication of data.

4.1 Trust aware model

The architecture in Figure 1 show the Trust Aware Model (T-AM) as it will be deployed on a typical F2F network. Deployed on the F2F network, the T-AM will perform the function of peer update with respect to a peer's collaborating level of trustworthiness to continue to collaborate. Each node on the F2F network will need to maintain a local trust data (as shown in Figure 1 as Trust Data) of all the nodes it has interacted with. As a result, an effective lookup service feature would be introduced using the Chord technique of the DHT. The lookup feature would be needed to handle challenges that bothers on issues of churn, scalability, decentralisation, availability and flexibility of naming. DHT is simple, has provable correctness and performance as a peer-to-peer lookup protocol. These and its usefulness to scale thousands of nodes and handle rapid arrivals and failures motivated its choice. Following the practice in Manju and Govindaraj (2014), the DHT would also be used to provide decentralised and effective access to trust information. Based on this, access to the location of trust data for the computation of personal, reputation and recommendation trust should be possible. These three trusts exist in literature; but they have not been conceptualized and modelled to perform the function of determining the level of peer trust as novelly proposed in this paper.

Figure 1: An F2F architecture showing the trust aware model deployed as a social trust manager



4.2 Data replication

The T-AM is intended to invoke the social trust manager to initiate the lookup service feature to acquire information on a specific node including its IP address, service port and location as a peer on the F2F network. To ensure trust data is available during lookup (see Figure 1), it must be replicated to appropriate nodes using appropriate data replication strategy. We propose to achieve this using the symmetric replication algorithm, which shall implement the replication strategy. The main reason for the replication is to maintain several copies of the same data at different sites. In this wise increased data availability, improved performance and load balancing between peers in the network, end-to-end fault tolerance, and increased security will be ensured (Ghodsi, 2005). As a result of this data replication, whether for a join or a leave; what is required is that the joining or leaving node should exchange data with its successor prior to joining or leaving.

5 The Friend-to-Friend System Architecture

The architecture presented in Galuba (2008) has remained profound. It has continued to serve as one of the *de-facto* standards in F2F collaboration. This research adopts and adapted it to show the novelty of the proposed T-AM. The T-AM with its intended working parts as components are as presented in one piece as shown in Figure 1. The architecture in Figure 1 is meant to show where the T-AM will be deployed to work, expecially at each of the nodes in the F2F computing environment. As shown in Figure 1, at each of the node level of the social network computing environment, the T-AM will be deployed to manage F2F collaboration. This abstraction is presented and shown Figure 1. This abstracted situation will be the same in all the nodes, when implemented. This is meant to ensure that trust is managed at every point in time. As a result, at the trust evaluation level (see Figure 1), the aforementioned three types of social trust will be handled. At the data management level, all the forms of trust data updating based on trust computation as earlier presented will be taking care of. Finally, the trust data will be the user data (see Figure 1).

6 Conclusion

It is interesting to highlight the fact that sections 3.0 and 4.0 were discussed in an instructive narrative. The aim is to highlight the fact that this paper is descriptive, and it is intended to whet the appetite of knowledge organization research community concerning the place of trust in E&TK sharing. Therefore, this paper is not prescriptive, which is one of its limitations. This paper as a result, proposes the T-AM as a model with trust aware computational capability. It is certain that as users gain control of their data the issue of trust will not be compromised. This way, the continuous security of nodes from unsuspecting malicious nodes will be guaranteed. As an ongoing work in progress model, the T-AM is targeted at contributing to the development of a

resourceful approach to knowledge sharing in the SNM context based on an F2F collaborative situation.

The T-AM as it is, is a model that promises to be useful in the development of reputational systems that requires a trust aware scheme to ensure secured E&TK sharing particularly in a diversified cultural context. No doubt, users will be able to have control over their data. This will be because of the control mechanism that is built into the T-AM at the level of each node. The T-AM is designed to operate at three levels: The trust evaluation level; the data manager level; and the trust data level as shown in Figure 1.

Cognizance of the contributions made so far, it is important to conclude that the effective management of trust in an F2F network can bring about a robust E&TK collaborative sharing. This is consistent with the postulation earlier presented by Pinjani and Palvia (2013). Evidently, the T-AM has been presented simplistically, and we find (based on our experience in the formulation of the model) that sustained collaborative E&TK sharing is possible. The use of the F2F interactive sharing paradigm to make this possible showed that the F2F paradigm of sharing is an invaluable platform for E&TK sharing in a connected society with diversified cultural background. However, the claims made so far still need to be verified and validated in real life situation. This is currently ongoing, and the postulations put forward so far are being implemented as a prototype model for deployment in a real social network environment. It will be interesting to consider in future the workability of the T-AM to provide the required support for collaborating parties to have control over their data, while trust among peers is not compromised.

References

Aderibigbe, Stephen Ojo, Afolabi, Babajide Samuel, Akhigbe, Bernard Ijesunor and Adagunodo, Emmanuel Rotini (2015).A Friend-to-Friend Approach for Secured Knowledge Sharing. In *Proceedings of the 10th International Conference of ISKO-France.* Held at Collège Doctoral Européen, November 5-6, 2015. Strasbourg, France.

Ali, S.M., Saleem, U., and Sikandar, S.M. (2015). Knowledge Sharing Prominence and Role in the 21 Century Organizations. *Researcher,* 7(1).

Ardichvili, Alexandre, Maurer, Martin, Li, Wei, Wentling, Tim and Stuedemann, Reed (2006). Cultural influences on knowledge sharing through online communities of practice. *Journal of Knowledge Management*, *10*(1): 94-107.

Buchegger, Sonja, Schiloberg, Doris, Vu, Le-Hung, and Datta, Anwitaman (2009). PeerSoN: P2P Social Networking: Early Experiences and Insights. In *Proceedings of the Second ACM EuroSys Workshop on Social Network Systems.* April 01-03, 2009. Pp. 46-52.

Chen, Xiang and Pan, Yao-Hui (2014). The Study of Open Source Software Collaborative User Model Based On Social Network and Tag Similarity. *Journal of Electronic Commerce Research*, 15(1).

Clarke, Ian, Sandberg, Oscar, Toseland, Matthew, and Verendel, Vilhelm (2010). *Private Communication through a Network of Trusted Connections: The Dark Freenet*. [https://freenetproject.org/assets/papers/freenet-0.7.5-paper.pdf].

Forster, Simon, Pinggera, Jakob and Weber, Barbara (2012). Collaborative Business Process Modelling. *EMISA*, 206: 81-94.

Galuba, Wojciech (2008). *Friend-to-Friend computing: Building the social web at the internet edges*(No. LSIR-REPORT-2009-003). [http://lsirpeople.epfl.ch/ galuba/papers/f2f.pdf]

Ghodsi, Ali, Alima, Luc Onana, and Haridi, Seifi (2005). Symmetric Replication for Structured Peer-to-Peer Systems. In *Proceedings of DBISP2P*. Pp. 74–85.

Greenwald, Glenn and MacAskill, Ewen (2013). *NSA Prism Program Taps into User Data of Apple, Google and Others*. [http://www.guardian.co.uk/world /2013/ Jun/06/ us-tech-giants-nsa-data]

Kramer, Roderick M., and Carnevale, Peter J. (2001). Trust and Intergroup Negotiation. In *Blackwell Handbook of Social Psychology: Intergroup Processes*, Malden, MA: Blackwell Publishers. Pp. 431-50.

Hildreth, Paul, Kimble, Chris and Wright, Peter (2000). Communities of Practice in the Distributed International Environment. *Journal of Knowledge Management,* 4(1): 27-38.

Lauring, Jakob (2007). Obstacles to Innovative Interaction: Communication Management in Culturally Diverse Organizations. *Journal of Intercultural Communication*, 15.

Liang, Zhengqiang and Shi, Weisong (2008). Analysis of Ratings on Trust Inference in Open Environments. *Performance Evaluation*, *65*(2): 99-128.

Manju, Jhon and Govindaraj, F. (2014). A Survey of Trust Management in Peer-to-Peer Systems. *International Journal of Computing and Technology,* 1(2): 2248-6090.

Massa, Paolo and Avesani, Paolo (2004). Trust-aware collaborative filtering for recommender systems. In *On the Move to Meaningful Internet Systems 2004: CoopIS, DOA, and ODBASE*, Springer Berlin Heidelberg. Pp. 492-508.

Mayer, Roger C., Davis, James H. and Schoorman, F. David (1995). An Integrative Model of Organizational Trust. *The Academy of Management Review*, 20(3):709-34.

Netrvalova, A. and Safarik, J. (2009). Interpersonal Trust Model. In *Proceedings of MATHMOD* Vienna. Pp. 530-7.

Netrvalova, A., and Safarik, J. (2011). Trust Model for Social Network. In *Proceedings of the 25th European Simulation and Modelling Conference*. October 24-26, 2011. Guimaraes: Portugal. Pp. 102-07.

Panahi, Sirous, Watson, Jason and Partridge, Helen (2013). Towards Tacit Knowledge Sharing Over Social Web Tools. *Journal of Knowledge Management*, 17(3): 1-17.

Pinjani, Praveen and Palvia, Prashanti (2013). Trust and Knowledge Sharing in Diverse Global Virtual Teams. *Information and Management*, 50(4): 144-53.

Steel, Emily and Vascellaro, Jessica E. (2010). *Facebook, Myspace Confront Privacy Loophole.* The wall Street Journal, May Edition.

Taylor, Hazel (2007). Tacit Knowledge: Conceptualizations and operationalizations. *International Journal of Knowledge Management,* 3: 60-73.

Truong, Hien Thi Thu, Bouguelia, Mohamed-Rafik, Ignat, Claudia Lavinia and Molli, Pascal (2011). *Collaborative Editing with Contract over Friend-to-Friend Networks*. [http:// home - pages.laas.fr/mkilliji/APVP2011/Site/Programme_files/Article_4.pdf]

Varlamis, Iraklis, Eirinaki, Magdalini and Louta, Malamati (2013). Application of social network metrics to a trust-aware collaborative model for generating personalized user recommendations. In *The Influence of Technology on Social Network Analysis and Mining*, Springer Vienna. Pp. 49-74.

Vijayakumaran, Nair K, and Vinod, Chandra S.S. (2014). *Informatics*. New Delhi, India: PHI learning Private Limited.

Wenger, Etiene (2011). *Communities of practice: A brief introduction*. [https://scholarsbank.uoregon.edu/xmlui/bitstream/handle/1794/11736/A%20brief%20introduction%20to%20CoP.pdf?sequence=1&isAllowed=y]

Wenger, Etiene, McDermott, Richard A., and Snyder, William (2002). *Cultivating Communities of Practice: A Guide to Managing Knowledge*. USA: Harvard Business Press.

Diogo Alves Cândido Pereira, Edson Marchetti da Silva and Renato Rocha Souza

Use of Lucene Framework to Retrieve Documents through Multiword Expressions as Search Descriptors

Abstract

This work used the frameworks Lucene e Tika as basis for development a prototype of software, an Information Retrieval System (IRS), capable to give support for all experiments realized. In this sense were indexed 1256 papers of 5 congresses. The goal is answer the following question: to extract search descriptors in an automatic way from a reference document, represented by a set of Multiword Expressions, can to improve the precision and reduce the recall of answers, if it compared to a traditional search made by key words from an ad hoc way by an end user? As results, are presented data that reveal improvements of the precision of answer. This makes feasible the use this strategy to digital libraries system, which deal with similar situation, to search relevants documents from a *corpus* of thesis, dissertations and papers. As a future works new experiments are recommended.

Introduction

According to Teixeira and Schiel (1997), with the fast developing of internet and the exponential information growth available on the Web, became necessary to develop search tools that help users to find what they need precisely. Nowadays the main search tools, such as Google, Yahoo and Bing work through a query interface that the user informs keywords and receives as answers, references ordered by relevance of thousands and even millions of links to web pages where these terms were found. Obviously, if on a hand this is good, because many references were retrieved, on the other hand, there is an overflow of contents, which the user has a hard work to interpret and to filter what he really wants for his interest. Furthermore, with the tendency of increase new information sources, day after day, it enhances the need to retrieve information more accurately. In this sense, and delimiting this problem to the context of digital libraries, this paper aims to conduct experiments in order to answer the following question: it to extract search descriptors in an automatic way from a reference document, represented by a set of Multiword Expressions, can to improve the precision and reduce the recall of answers, if it compared to a traditional search made by key words from an ad hoc way by an end user?

In that sense, this works aims to verify if in the context of digital libraries composed by thesis, dissertations and papers (considering a collection ofdocuments from aspecific domain), toperform searchesfroma reference documentcanproducea more accurateresponseto the end user. So, the idea behind this process is to extract multiword expressions from a reference document represented by of a set of *n*-grams used as descriptors in an automated search performed through of many concurrency searches, one for each *n*-gram. Then the score relevance is calculated by the sum of relevancies obtained from each *n*-gram that exists in the set of *n*-grams. Therefore, the

more bigrams were found by comparing the reference document of search with every document the *corpus*, greater will be its relevance to the set of answers.

To better describe the experiments the work is structured into the following sections which are presented in the following contents: Section 2 – theoretical concepts about Multiword Expressions, Section 3 - methodology, Section 4 – main findings, Section 5 - conclusions; Section 6 - Recommendations for future works.

2 Theorical concepts

The use of bigrams is justified, because according Sarmento (2006), the text is not just a bag of words. Themeaning of a textis produced by theorder in which thewords areplaced on it. So, the co-occurrence study of words brings important information about the meaning of the text. This can indicates that the words are connected, directly by compositionality or affinity, or indirectly by similarity. In this way, when an user tries to search using as reference a document which is relevant to him, the aim of the implemented software is to produce an automated search that retrieves documents, not because only one keyword was found, but because many bigrams were found at the same time.

Manning and Marneffe (2012) described *n*-grams are a set of two of more words extracted from a document, by heuristic methods, without necessarily to exist a semantic meaning between the *n*-grams for a set of *n* words also can be called as:

- Bigram: it is made of two words, such as: “Minas Gerais”, “São Paulo”, “beyond that”, etc;
- Trigram: a set of three words, such as: “UEFA Champions League”, “Information Retrieval System”, etc.

Independently of quantity of words that the *n*-grams are compound, each one is converted in a set of bigrams. The process of to extract*n*-grams is made like that: first are extracted the stop words, then it is checked the co-occurrence of words that appear together more then a threshold of times all over the text. In this work, three, was used as threshold. After that, in the situation that the quadgram “Federal Center of Technological Education” was found, it will be transformed in a set of three bigrams, like: “federal center”, “center technological” and “technological education”. Because the word “of” is considered as a stop word.

According to Ramisch, Villavicencio e Boitet (2009), these set of words can be identified as composed nouns, composed names, adverbial adjuncts, etc. In another study, Ramisch, Villavicencio e Boitet (2010 apud BIBER et al., 1999) claim that, the Multiexpression Words are important in Natural Language Processing and consequently in Information Retrieval (IR), which it is a sub-area of that, because they can correspond from 30% to 45% of English speaking and 21% of academic language.

According to Baeza-Yates, Neto (1999) IR can be defined as a science which studies methods and algorithms capable of to fetch information in document repositories in an efficient way. Toperform our experiments was developed a software prototype, which

it can considered as an Information Retrieval System (IRS). it was used the Lucene and Tika frameworks, both developed by Apache Software Foundation. Lucene is a framework that implements diverse techniques to create an IRS and Tika is used to extract a content of plain text from binary files such as ".doc",".pdf" and others proprietary formats, and to identify the idiom of document. The second feature cited it is a pre-requirement to the stemming process and the stop words remotion, techniques which are dependant of the idiom. Both frameworks are widely used as technological tools to the production of many different experiments in this area.

Thereby, in this work was implemented an IRS that were used the main IR techniques, as the indexing well as the searching. Also was implemented the concepts related to the extraction of Multiexpression Words, being used as descriptors of a reference document.

The first stage of an IR process consists in a set of five steps that operate in a text which is in a pre-defined *corpus*. In this way, it can be according to Baeza-Yates and Neto (1999):

1. Lexical analysis of the text with the aim of threating digits, punctuations, hyphens and capital and lower case letters;
2. Stop words remotion to filter words which are not relevant to the quality on the operations of indexing and searching;
3. To perform stemming operation for removing suffixes from words, which made many terms the of lexicon to be reduced in one only term;
4. Selection the termswords/stems (set of words) that will be used in the index;
5. The use of a Thesaurus which allows to expand the result of a query with terms related to it.

3 Methodology

To answer the main question, it was developed an Information Retrieval System (IRS) prototype capable of index and search a collection of documents to perform the experiments. It was used the Lucene and Tika frameworks, both developed by Apache Software Foundation. Lucene is a framework that implements diverse techniques to create an IRS and Tika used to extract a content of text from ".doc" and ".pdf" format files and to identify the document language as a pre-requirement to remove stop words, which are dependant of the idiom. Both frameworks are widely used as technological tools to the production of many different experiments in this area.

Thereby, this work also implemented the concepts related to the extraction of Multiword Expressions, being used as descriptors of a reference document.

To perform this experiment it was used a *corpus* of 1256 scientific documents, which contains articles extracted from five congresses: IARIA, ENANCIB, CBSoft, ACM, Contecsi and Perspectiva, Journal of UFMG. The indexed documents are related to Information Science, Information Systems, Software Engineering, Multimedia Systems and Computer Networks. 49% of these documents are in the Portuguese and

48% are in the English and 3% are in others idiom. The papers written in languages, other than Portuguese and English, were considered as outliers, therefore they were removed of the *corpus*, that now it have 1218 documents. The indexing process followed these steps:

1. Indexing of the *corpus*
   - processes each document into the *corpus* located in folders and sub-folders;
   - convert the documents to plain text;
   - automatically identify the idiom;
   - perform the lexical analysis;
   - remove the stop words;
   - index the current document;
   - store the index in the hard disk.
2. Identification of the bigrams without relevance
   - processes each document into the *corpus* located in folders and sub-folders;
   - converts the documents to plain text;
   - identifies automaticallythe idiom and chooses only an idiom to process all the of documents selected at the same time;
   - identifies *n*-grams for each document and counts the occurrence frequency global, generating a list with all bigrams with frequency below a defined threshold. In this case, the value was ten occurrences;
   - creates of the supervised form a list with the irrelevant bigrams.
3. Realization of the searches from *n*-grams extracted automatically from a reference document:
   - performs the searchby the prototype developed using 20 reference documents which there are not in the *corpus*, in order to find their similar ones, in two ways:
     - considering the complete list of bigrams extracted from the reference document;
     - do the same task but excluding of the search the bigrams that are in the false positive list.

The idea is to compare how much the search for descriptors extracted automatically as a reference document can be improved, if compare searches with and without considered the irrelevant bigrams list. In this sense, it consider as false bigrams as idiomatic expressions that have little relevance to the search because they are related to the phrasing such as “beyond that”, “even so” etc.

To be in the retrieval list, it is considered only documents with the coefficient relevance calculated, using the OKAPI-BM25 technique, above at least 60% of the greater coefficient value one found for all the documents of the answer.

4 Main Findings

For accomplish the first step of the methdology were worked four scenarios to evaluate of the indexing performance:

1. Using stemming and removing stop words;
2. Using stemming and without removing stop words;
3. Without using stemming and removing stop words;
4. Without using stemming and without removing stop words.

The aim of this stage was to check how the use of these techniques can reduce the of lexicon size. The results showed that when you use both techniques using stemming and removing stop words the reduction is the greater, around of 32% of the lexicon. Beyond this, it produces a significant reduction of memory consuming without affecting the searching process. In addition, the time to index the *corpus* is bigger when you use the stemming technique, on the other hand, the number of terms stored in the index is smaller. Because, many words when have its suffix removed, they become a same lexicon, in this way the frequence of co-ocurrence of both words are added.

Then, for continue those experiments was chosen create the index using the third scenario proposed, because the use of stemming reduces the word to a radical, and this situation could be bring noise for bigrams extraction.

After the index structure of the *corpus* be ready, thenext step, was to work at search process. First of all, it is necessary to extract the bigrams from a reference document to be used as descriptors in the searching process. Beyond that, the bigrams extracted are those that co-occurs in a frequency higher than a pre-defined threshold. In this case, we used greater or equal than three. However, the bigrams generated there are some that could be considered as false positives, which can be idiomatic expressions that do not have no semantic meaning to the searching process, because they are related to the linguistic style, such as: “além disso”, “desse modo” (in Portuguese) or “in this sense", "on the other hand" (in English). In this way, the next step of this work is to produce a supervised analysis of the extracted bigrams from all the documents belonging to the *corpus* to create a list of irrelevant bigrams to be used as a filter in the searching process. This list will be used to filter bigrams during the extraction of them from the reference document to eliminate those false positives during the searching process. In this way, the idea that was behind of this experiment is the use of the false positive list could produce a significant improvement on the answers accuracy.

In the sequence of the experiment, for time limitation, was realized only the analisys of the documents written in Portuguese. In that sense, It was performed the bigrams extraction for all the documents in Portuguese of the *corpus*. In that case, It to be consider a bigram was defined a threshold minimal for the frequency greater or equal ten. As result, 8060 bigrams were extracted. After that, was realized a supervisioned analisys aiming to separate the bigrams that could be considered as a false positives. It were considered as false positives 2030 which were put in a list. The Table 1 shows

some exemples for bigrams that were considered as relevants (that have discrimination for a search) and also the irrelevants.

Table 1 – Examples of the bigrams relevant and irrelevants

Relevants		Irrelevants (false positives)	
Term	Frequency	Term	Frequency
ciencia informação	6750	informacao rio	3721
inclusao social	3804	resultados obtidos	495
gestao conhecimento	1361	presente estudo	239
seguranca informação	660	ser considerado	181
inovacao tecnológica	254	fundamentaçao teorica	141

The table 2 shows three lines with numbers. The first indicates the number of the document used as a reference document. The second line shows the number of documents found as relevants, when it consider all bigrams extracted from the reference document. The third one shows the number of documents found as relevants, if the false positives bigrams contained in thelist were eliminated of the descriptors list used in the search. For all the cases (they were realized fourty searches), the threshold used as cut point of the coefficient was 60% of the greatest coefficient found for all the returned documents.

Table 2 – number of retrieved documents

Search Documents																			
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
4	5	7	7	9	10	10	11	11	15	18	20	23	23	23	26	26	28	33	45
3	3	9	7	10	6	11	9	10	14	23	15	21	14	19	19	19	18	16	26

In average the number of documents extracted were: in the first case, 17.7 and in the second case 13.6. That is, when we use the false positives list occurring the remotion of irrelevants bigrams, which leads to an improvement of precision, because occurring the reduction of answers. That is, when we use the false positives list occurring the remotion of irrelevants bigrams, which leads to an improvement of precision, because occurring the reduction of answers. However, if for the greatest majority of documents occurring the reduction of documents returned, on the other hand for the documents 3, 5, 7 and 11, when the search was executed removing the false positives bigrams, the set of answer was greater. It means that more documents were found. But, after analyzed the content of the documents by human read, it was note that nevertheless, for all cases the answersbecame better because actually false positive bigrams inserts noises for the calculation of relevance coefficient. As a conclusion, we can find that the recall becomes better in almost 19% as the case that more relevants documents were returned. On the other hand, when the number of answer was smaller, It means that minus irrelevant documents were found, in this sense, the precision become better 39%.

5 Conclusions and future works

The Lucene framework study could show many tools that support developing solutions to implement an Information Retrieval System. Also the use of Tika helped to build an interface to convert documents to a format which Lucene could work properly when you are performing indexing and searching tasks. The list of irrelevant bigrams to eliminate of the search task some of the descriptors extracted automatically of the reference document could result in an improvement of 21.3% in average comparing when you are not using this list.

6 Recommendationsfor future Works

The results obtained by the work is promising, because it showed better responses than those obtained for key words provided by an end user. Also the results can be improves, if after of extracted the bigrams we use a filter for removing the false positives that were extracted by the algorithm that identify bigrams only for the frequency without to use any semantic knowledge. So, this work bring improvements obtained from previous works produzed by Silva & Souza (2013). As future works wesuggest made the same experiment using the documents in English idiom, with the goal of compare the results. Also, the another possibility it is to use a list of false positives created by colaborative way, where people could during the use of the IRS of search take your contribution for add new bigrams in the list.

References

Baeza-Yates, Ricardo & Neto, Berthir Ribeiro (1999). Modern Information Retrieval. 1. ed. [S.l.]: Addison Wesley.

Barcala, Mario, Alonso, Miguel &Vilares, Manuel (2002). Tokenization and proper noun recognition for information retrieval. *Proceedings of the 13th International Workshop on Database and Expert Systems Applications*, 10(9): 246-50.

Hatcher, Erik et al (2009). Lucene in Action. [S.l.]: Manning Publications.

Manning, Cristopher D., Marneffe, Marie-Catherine & Green, Spence (2012). Parsing models for identifying multiword expressions. *Computational Linguistics*, 39(1).

Manning, Cristopher D.; Raghavan, Prabhakar & Schutze, Hinrich (2009). An Introduction to Information Retrieval. [S.l.]: Cambridge University Press.

Manning, Cristopher D. & Schutze, Hinrich (2003). Foundations of Statistical Natural Language Processing. [S.l.]: The MIT Press.

Milosavjevic,Branco,Boberic Danijela & Surla, Dusan (2010). Retrieval of bibliographic records using Apache Lucene. *The Electronic Library*, 28 (4): 525-39.

Porter, M. F (1980). An algorithm for su-x stripping. 14(3): 130-7.

Ramisch, C.; Villavicencio, Aline & Boitet, C. (2009) Statistically-driven alignment-based multiword expression identication for technical domains. In *Proceedings of the 2009Workshop on Multiword Expressions*. Pp. 1-8.

Sarmento, Luis (2006). Simpósio doutoral linguateca. [http://www.linguateca.pt/documentos/SimposioDoutoral2005.html]

Silva, Edson Marchetti & Souza, Renato Rocha (2013). Comparando três diferentes técnicas para recuperar documentos utilizando expressões multipalavras. In *10th International Conference onInformation Systems and Technology Management*.

Silva, Edson Marcheti; Souza, Renato Rocha (2013). Sistema de recuperação da informação por busca comparada, que utiliza como descritores expressões multipalavras obtidas através de uma técnica que avalia a estrutura do documento. In *International Conference on Information Systems and Technology Management.*

Teixeira, Cenivalda Miranda de Souza & Schiel, Ulrich (1997). A internet e seus impactos nos processos de recuperação da informação. *Ciência da Informação*, 26(1).

Villavicencio, Aline, Ramisch, Carlos & Machado, André (2010). Identificação de expressões multipalavra em domínios específicos. *Linguamática*, 2(1): 15-34.

Zitting, Jukka L. & Mattmann, Chris A (2011). Tika In Action. [S.l.]: Manning Publications.

Kavi Mahesh and Pallavi Karanth

Organizing Knowledge to Facilitate Analytics

Abstract

Knowledge is being created, shared and used increasingly through on-line media such as electronic mail, Web pages, social networking platforms, e-books, e-learning content, videos and other multimedia channels. These media continue to be organized minimally with some hierarchy of categories but largely through unrestricted or inadequately managed tags and inverted indices. While such organization supports search and browsing to a large extent, it is not sufficient for the emerging needs of two new functionalities that are already in universal demand: *analytics* and *visualization* of information. This article explores the relation between knowledge organization and analytics, including visual analytics. It attempts to bring out the requirements for performing analytics on knowledge structures and proposes a solution framework for organizing knowledge to suit those requirements.

1 Introduction

As on-line information grows in both size and availability, analytics and visualization together offer the promise of overcoming information overload by aggregating and quantifying information and presenting the results to users both visually and interactively.Analytics and visualization are already being applied across domains and applications as varied as medicine, finance, news, defense intelligence, forensics and humanities to name a few. Analytics is also being applied to all types of content including structured data in databases, Web pages, text and documents, video and multimedia and leaning content (Karanth and Mahesh, 2015).

Analytics involves the operations of counting, aggregating and computing of statistical measures about various kinds of knowledge structures or pieces of information present in a repository. It answers quantitative questions about knowledge: what is present, how much of it, how is one piece related to another and the relative importance or centrality of each piece. For example, even if the only organization scheme present in a system is a set of tags where each piece of content is tagged with one or more tags, analytics can tell users how many pieces of knowledge are present for a given tag, which are the most popular tags, how recently a certain tag was used, which two tags are assigned together most often, who uses which tag the most, and so on. Analytics can go beyond such capabilities and also predict future trends although it is not clear at this time how predictive analytics may work with knowledge structures.

A variety of contexts (Smiraglia and Lee, 2012) as well as structures (Iyer, 2012) are of importance in knowledge organization. In particular, when tags are organized hierarchically using a thesaurus, taxonomy or ontology, additional analytical operations can be defined. For example, counts and other aggregate measures can be *rolled up* or *drilled down* the hierarchies. That is, the count at a particular level in the hierarchy is the sum of(or, to be more precise, the total number of unique items in) all the nodes at the next lower level. Further, one can compute various structural measures such as the *fan-out* at a node in the hierarchy, that is, the number of branches or links or immediate

children nodes at the node; the average or maximum height of a branch of the hierarchy; or the degree of balance of the hierarchy across its branches.

Using an ontology to organize knowledge offers several benefits since it introduces precise definitions of concepts in terms of its relations to and differences from other concepts. Each entity as well as all its attributes and relations among other entities present can be defined semantically and the relations themselves organized in a subsumption hierarchy, using a standard language such as the Web Ontology Language (OWL). Category or class hierarchies can exploit inheritance of attributes and relations. Semantic constraints can be imposed to prevent meaningless or out-of-bound values. A clear separation can be maintained between conceptual definitions in the so called TBox (Terminological Box) and actual data about instances in the ABox (Assertional Box). The TBox is the ontology proper and is typically in the form of a tree or lattice structure, often with complex concept definitions. The ABox is the data proper and is best viewed as a graph whose nodes are the entities and edges the relations, a graph being a more general organization structure than relational tables used typically in databases.

2 Related Work

Analytics and knowledge discovery have been attempted with ontology in the domain of defense intelligence (Sheth, 2005). However, in this domain, analytics is carried out by a human analyst while the ontology is used primarily to make the analyst's queries efficient and to provide semantic metadata to the analyst. The focus of the work is not automatic analytics using quantitative measures of aggregation on knowledge structures.

Semantic networks have been used to help users analyze citations and publications in a domain (Plikus, Zhang and Chuong, 2006). Here too the use of the semantic network is to assist the human analyst and does not impact the analytical operations directly. Ontology has been used to facilitate analytics by carrying out spatial reasoning in the specialized domain of geospatial data (Arpinar, et al., 2004). Ontology has also been shown to help in visualization of Big Data (Rysavy, Bromley and Daggett, 2014).

3 Knowledge Analytics

Analytic operations on an ontology-based knowledge organization can be termed *knowledge analytics* - an area that has just begun to develop. Major questions that need to be addressed in knowledge analytics include (i) the impact of inheritance of attributes and relations in conceptual hierarchies, especially when inheritance is constrained by semantically or is nonmonotonic, (ii) effect of semantic constraints on roll-up and drill-down aggregation measures, and (iii) semantic filtering of edges in the graph for computing graph centrality measurements. Further, detection of communities and other such graph-analytic operations can also be defined on semantic grounds.

In the rest of this article, we propose a fundamental knowledge analytic operation of semantic filtering in agraph-structured organization of knowledge using an ontology. After applying the semantic filtering operation, a variety of analytic operations can be applied to obtain measures that are meaningful and relevant to the domain of interest. For example, if knowledge about people is organized as a social network, some of the relationships among the people may be business or professional ones while others may be social within which kind, some may be family relations while others are friendships and yet others are neighborhood relations, and so on. Graph analytics applied to such a social network without semantic filtering will not be meaningful. For example, graph *centrality* or *betweenness* may be misleading if business and social relations are mixed. The semantic filtering operation enables us to neatly dissect the knowledge by using relation (or property) hierarchies in the associated ontology.It may be noted that the resulting slices of knowledge, although somewhat similar to views in a database, are defined semantically using the ontology, not directly in terms of which columns or rows in a table to consider. Figure 1 shows a sample social graph (from the Indian epic Ramayana). The result of semantic filtering on family relations only is shown in Figure 2. This operation of semantic filtering has been enabled by the relationship hierarchy in the ontology shown in Figure 3. It can be seen that the *degree centrality* of the node "Rama" (defined as the number of relations or edges incident at that node) was 8 before filtering and became only 4 after semantic filtering on the sub-hierarchy of relations called FamilyRelation in the ontology. A degree centrality of 8 would have been highly misleading as an analytic measure if the domain of interest was focused only on family relations.Further, it may be noted that semantic filtering removed certain nodes altogether: places like "Lanka" and "Ayodhya" and persons like "Hanuman" who had no family relation to the rest of the network.

On top of this, knowledge organization may need to support multiple versions of data or time-varying snapshots of information (akin to time-series data). Knowledge analytic operations should take these into account by providing version or snapshot specific measures thereby avoiding the mistake of counting the same entity multiple times across versions or snapshots. In more challenging situations, the conceptual organization in the ontology itself may change over time or across multiple versions resulting in a multi-view ontology [Mahesh and Karanth, 2014] posing further challenges to knowledge analytics.

Fig 1. Sample Social Graph with Various Social, Political and Other Relations
(only relations other than family relations labeled)

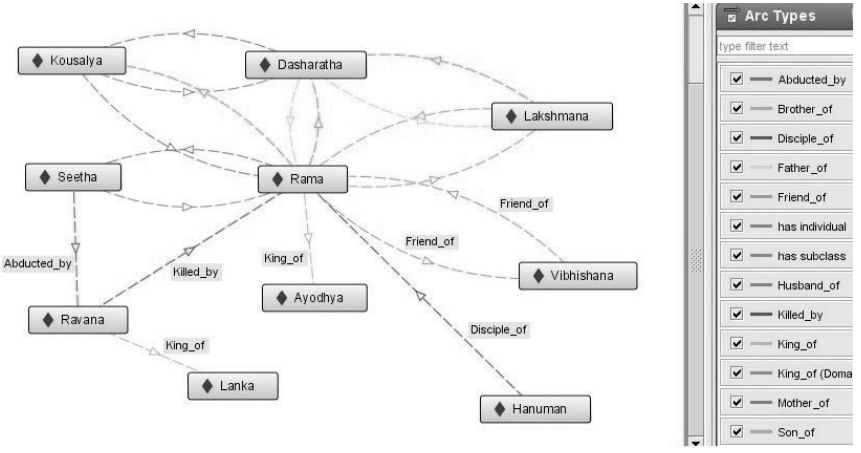
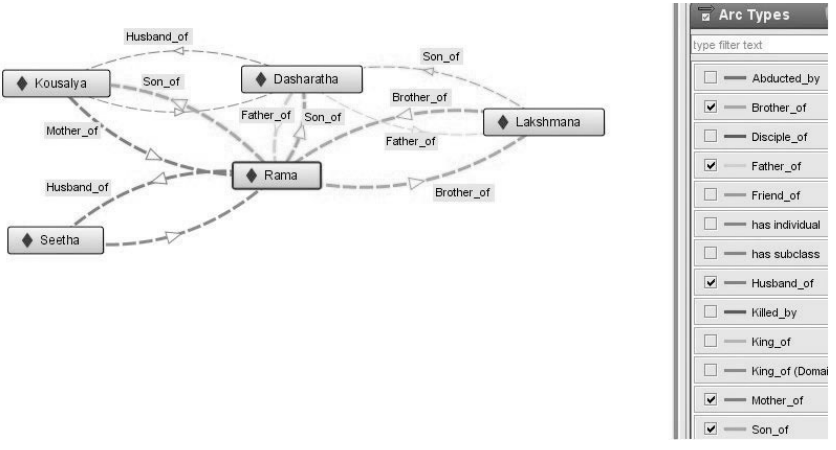
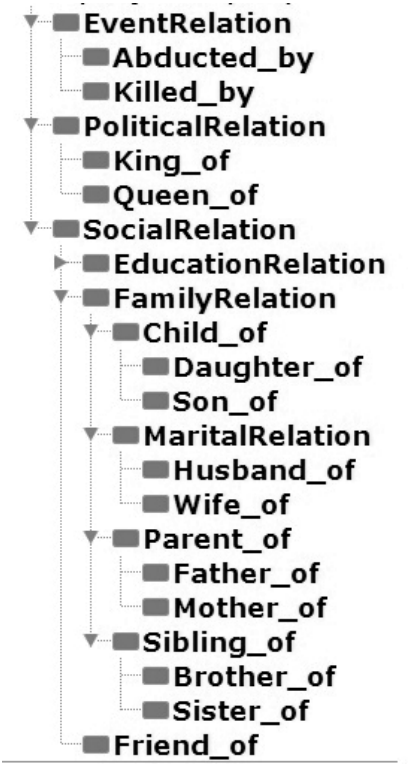


Fig 2. Result of Semantic Filtering on FamilyRelations



Alternatively, if knowledge is organized using faceted classification structures (Ranganathan, 1951), the facets can be seen as different dimensions of a multi-dimensional matrix. Suitable analytic operations based on matrix algebra can be applied to derive meaningful measures and insights from the knowledge organization. Such operators too need to apply suitable semantic filtering in terms of combinations of facets. It may be noted however, that in the general case, a graph-based organization where the nodes and relations in the graph are defined semantically in an ontology affords a greater variety of knowledge analytic operations with a consequent greater potential for users to derive insights from the resulting analytics and visualizations.

Fig 3. Relation Hierarchy in Ontology Showing Types of FamilyRelation under SocialRelation



Thus, *knowledge analytics,* that is, the effective application of analytics to knowledge structures, requires a careful consideration and integration of both knowledge organization schemes and mathematical operations involved in analytics. In particular, we propose that knowledge organization designed to make analytics effective must meet the following requirements:

(i) ability to uniquely identify each entity. For example, this overcomes the usual challenges faced in scientometrics in distinguishing between several authors with the same name. In spite of recent attempts to assign publisher neutral unique identities to authors, other metadata such as affiliation continue to be in unstructured forms without unique identities.

(ii) all types of knowledge to be organized in hierarchies of classes or categories having well-defined attributes and relations. Knowledge organization must be based on a well-defined ontology for domain.

(iii) have a controlled vocabulary for tagging in meta-data to avoid duplicate and ambiguous tags. In other words, the vocabulary used must not lead to the notorious problems of processing natural language.

(iv) unstructured data should only complement structured data. For example, an abstract may be present for an article but a set of keywords is essential and preferably be from a controlled vocabulary.
(v) attributes and relations must themselves be carefully organized in hierarchies to enable effective semantic filtering.
(vi) support multiple versions as well as time-varying snapshots.
(vii) have a clear separation between structured and unstructured parts to enable quantitative operations on the structured parts.

4 Discussion

Our work aims to paint a complete picture of how operations of knowledge analytics can be facilitated through suitable knowledge organization schemes by further developing models of the implications of semantic filtering along with structural and semantic constraints on a full range of analytical measures: counts, frequencies, statistical distributions, graph centrality measures, geometric measures such as graph diameter, coverage measures and sub-graph relationships.

It appears that a formal knowledge organization scheme such as a well designed ontology is better suited for knowledge analytics than an informal scheme or organization such as an informal taxonomy or folksonomy (Gruber, 2007). The latter may be good enough for search and navigation and perhaps even visualization but may not be accurate enough to support quantitative measures of knowledge analytics.

In applications such as bibliometrics and scientometrics, where the information is traditionally well structured, there is still a need for an ontology-based semantic organization to support some of the advanced operations of knowledge analytics. For example, finding authors of research papers who are most central or most influential in a domain of knowledge or automatically identifying communities of collaborative research requires a proper ontology of not just authors and their topics of interest, but also their locations, affiliated organizations, specialized equipment or methods used in their research, professional relationships of studentship, co-investigation, reviewer, editor and so on.

Bibliometrics is being used increasingly for research performance evaluation although it falls short of precise knowledge about the research domain in question. Citation analyses are based on the assumption that a linear relationship exists between the quantum and quality of contribution to the body of knowledge and citation counts.In reality, citation patterns vary greatly among different scientific disciplines, publication types, authors, types of research and its long term significance (Wallin,2005). To address such problems, it helps to have a rich graph representation of the relevant information about publications, citations and authors where every relationship in the domain is well defined in a domain ontology to give better insights through knowledge analytic operations.

To summarize, knowledge analytics is an emerging field and along with visualization techniques offers better ways of managing knowledge by enabling quantitative analytic operations on the knowledge structures in a repository. A key factor in the successful application of knowledge analytics is a well-designed knowledge organization scheme, with ontology-based and other such semantics-heavy organizations being the more promising schemes of knowledge organization for analytics. With the foundation of proper knowledge organization, knowledge analytics promises a bright future with a variety of powerful applications presenting quantitative insights and visualizations to users.

Acknowledgments: This work is supported in part by the World Bank/Government of India research grant under the TEQIP programme (subcomponent 1.2.1) to the Centre for Knowledge Analytics and Ontological Engineering at PES University, Bangalore.

References

Arpinar, I. Budaket al. (2004). Geospatial Ontology Development and Semantic Analytics. In *Handbook of Geographic Information Science*, Blackwell Publishing.

Gruber, Thomas (2007). Ontology of folksonomy: A mash-up of apples and oranges. *International Journal on Semantic Web and Information Systems* (IJSWIS), 3(1): 1-11.

Iyer, Hemalata (2012). Classificatory Structures: Concepts, Relations and Representation. Textbooks for Knowledge Organization , Vol 2, Ergon Verlag.

Karanth, Pallaviand Mahesh, Kavi (2015). From Data to Knowledge Analytics, *Information Studies*, 21(4):261-274.

Mahesh. Kavi and Karanth, Pallavi (2014). Supporting multiple points of view in knowledge organization. *Information Management*, 50(6):831-842.

Plikus, Maksim V., Zhang, Zina and Chuong, Cheng-Ming (2006). PubFocus: Semantic MEDLINE/PubMed citations analytics through integration of controlled biomedical dictionaries and ranking algorithm. *BMC Bioinformatics*, 2(7) October: 424.

Ranganathan, Shiyali Ramamrita (1951). *Philosophy of library classification*. Copenhagen: Munksgaard.

Rysavy, Steven J., Bromley, Dennis and Daggett, Valerie (2014). DIVE: A Graph-Based Visual-Analytics Framework for Big Data. *IEEE Computer Graphics and Applications*, 34(2):26-37.

Sheth, Amit (2005). From Semantic Search and Integration to Analytics, *Proc. Dagstuhl Seminar*, Dagstuhl, Germany.

Smiraglia, Richard P. and Lee, Hur-Lee, eds. (2012). *Cultural Frames of Knowledge*, Ergon Verlag.

Wallin, Johan A. (2005). Bibliometric Methods: Pitfalls and Possibilities. *Basic & Clinical Pharmacology & Toxicology*, 97: 261–275.

Maja Žumer and Marcia Lei Zeng

The New FRBR-LRM Model: Some Accents

Abstract

The FRBR - LRM *(Library Reference Model)* consolidated three previously developed conceptual models of the FR family, aiming at enabling consistent implementations of improved bibliographic information systems in libraries of the world. This paper reviews the background of the FRBR-LRM model and introduces the entities and relationships. It gives special attentions to the *Work* to *Res* relationship and *Work* to *Nomen* relationship that can be traced to the *Functional Requirements for Subject Authority Data* (FRSAD) model.

1 Background

The *Functional Requirements for Bibliographic Records* (FRBR) model was published by IFLA in 1998. Two additional models, *Functional Requirements for Authority Data* (FRAD) and *Functional Requirements for Subject Authority Data* (FRSAD) further developed the name authority and subject authority parts, thus forming the so-called "FR family". The three models were developed over an interval of more than twenty years by different working groups and, consequently, different modelling decisions were made for several common issues. It was therefore necessary to consolidate the FR family into a single coherent model to clarify the understanding of the overall model of the bibliographic universe and remove barriers to its adoption and implementation.

FRBR Review Group, the IFLA body responsible for the development of the model, constituted the Consolidation Editorial Group (CEG) charged with the consolidation process. In 2016 the draft of *FRBR Library Reference Model* (FRBR-LRM) was released for the world-wide review, which concluded on May 2nd. The comments are now analysed and the edited text will be prepared for final adoption. This paper is based on the draft version [1], published on the IFLA website for the review.

2 User tasks, Entities, and Relationships in the new FRBR-LRM Model

The FRBR-LRM is a high-level conceptual reference model of the bibliographic universe developed within an entity-relationship modelling framework. The model aims to show the general principles behind the structure of bibliographic information, without prescribing how that data might be stored in any particular system or application. It is therefore NOT a data model. The FRBR-LRM takes its scope from the user tasks when interacting with the bibliographic universe, in particular *find, identify, select, obtain, explore,* which are defined from the point of view of the end-user's needs and context. The model considers bibliographic information pertinent to all types of resources generally of interest to libraries, however, it seeks to be applicable in a generic way to all types of resources, or to all relevant entities. In consequence, any elements that are viewed as specialized or are specific only to certain types of resources, are not explicitly represented in the model. The model is comprehensive at

the conceptual level, but only indicative in terms of the attributes and relationships. The conceptual model as declared in the FRBR-LRM is a high-level abstraction and as such is intended as a guide or basis on which to elaborate cataloguing rules and implement bibliographic systems.

The so-called WEMI part of FRBR remains essentially the same, with slightly moderated definitions:

- *Work*, the intellectual or artistic content of a distinct creation.
- *Expression,* a distinct constellation of signs conveying intellectual or artistic content.
- *Manifestation,* a set of all carriers that are assumed to share the same characteristics as to intellectual or artistic content and aspects of physical form. That set is defined by both the overall content and the production plan for its carrier or carriers.
- *Item,* a physical object carrying signs resulting from a production process and intended to convey intellectual or artistic content.

What used to be group 2 entities in FRBR is now replaced by *Agent* (an entity capable of exercising responsibility relationships relating to *works*, *expressions*, *manifestations* or *items*) and its two sub-classes: *Person* (an individual human being) and *Collective agent* (a gathering or organization of persons bearinga particular name and acting as a unit). *Agents* are connected to WEMI entities by responsibility relationships. *Place* and *Time-span* are also introduced as entities.

A new entity class, *Res,* is introduced as superclass of all entities and is defined as "Any entity in the universe of discourse". Consequently, the (many to many) subject relationship is now *Work* has as subject *Res.*

Since all entity types are subclasses of *Res,* some important attributes and relationships are declared for *Res* and valid for all, such as relationships:

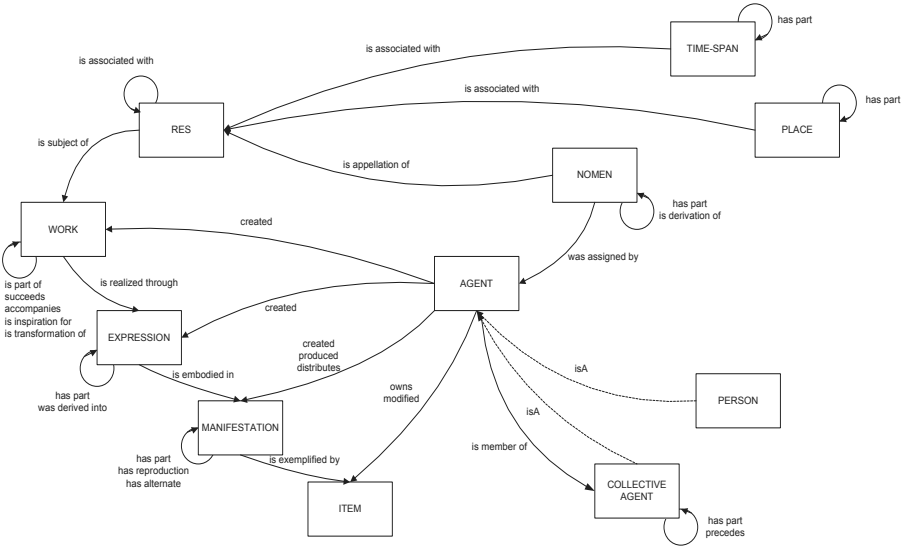
Res has appellation *Nomen*

Res is associated with *Place*

Res is associated with *Time-span*

The attributes and relationships are declared on a general level, thus allowing the implementers to introduce more specific relationships in a consistent and coherent way. For example, the relationships between instances of Work, i.e., *work* to *work,* are: "is part of", "precedes", "accompanies/complements", "is inspiration for", "is transformation of". Both the attributes and relationships of *Work,* or any other entity, may be further refined to serve specific needs of an implementation.

Figure 1:Overview of FRBR-LRM

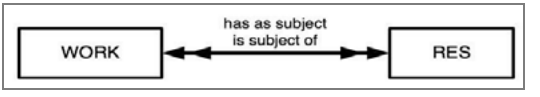


Source: http://www.ifla.org/files/assets/cataloguing/frbr-lrm/frbr-lrm_20160225.pdf(The isA relationship for all entities to *Res* is not shown. Only one way relationships are depicted, all without cardinality)

3 About the *Work* to *Res* relationship

As presented above, the new entity class, *Res,* (superclass of all entities, defined as "Any entity in the universe of discourse") has the (many to many) subject relationship: *Work* has as subject *Res*. [Figure 2]

Figure 2: Work-to-Res relationship in FRBR-LRM



Source: http://www.ifla.org/files/assets/cataloguing/frbr-lrm/frbr-lrm_20160225.pdf

This can be traced back to the FRSAD model. The IFLA Working Group on the Functional Requirements for Subject Authority Records (FRSAR) was formed in 2005 to address subject authority data and to investigate the direct and indirect uses of subject authority data by a wide range of users. It expanded on the 'has subject' relationship already defined in FRBR where *Concept*, *Object*, *Event*, and *Place* are four entities of the "third group" and are considered "additional things that can be the 'subject of' *Work*. FRSAD created and defined a *Thema* entity to cover any entity that is used as a subject of a *work* [Figure 3].

In FRSAD, the entity type *Thema* was defined as "any entity used as a subject of a *work";* A *thema*, together with other *themas*, represents the totality of what a particular *work* is about and/or any of the more atomic aspects of that totality.[Figure 4]

Figure 3: FRSAD's relation to FRBR after defining the *Thema* entity and its relationship with *Work*

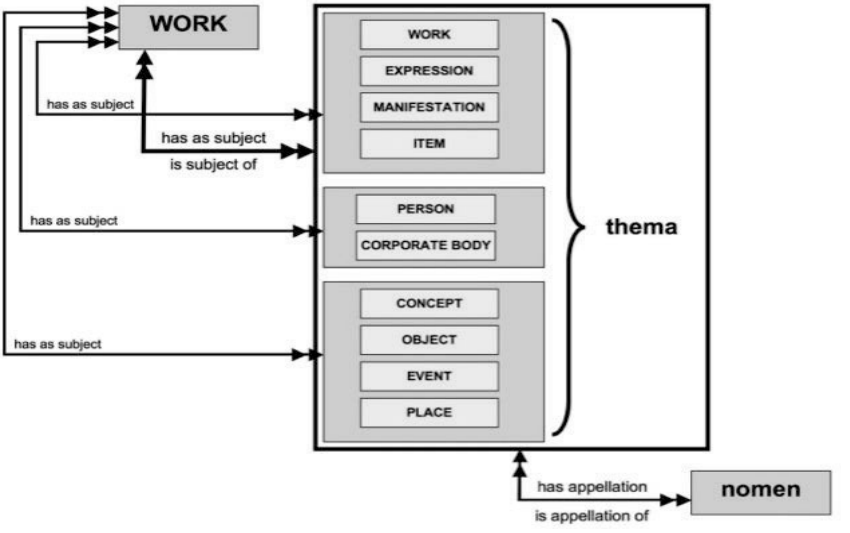
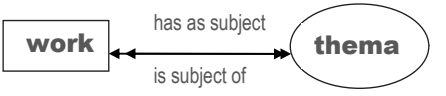


Figure 4: Work-Thema relationships in FRSAD

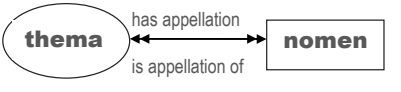


In the current FRBR-LRM model, *thema* is further generalised to all and any *Res*, not limited by their USE ("as a subject of a work") in the bibliographic universe.

4 Focusing on the *Res* to *Nomen* Relationship

In FRSAD, another entity, *Nomen,* is defined as "any sign or sequence of signs (alphanumeric characters, symbols, sound, etc.) by which a *thema* is known, referred to or addressed as". [Figure 5]

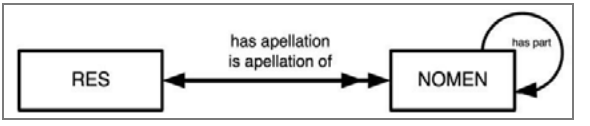
Figure 5: Appellation relationship (Thema-Nomen) in FRSAD



This relationship is now the *Res* to *Nomen* relationship in the new LRM model, but the framework remains: *Nomen* covers any form of appellation from the most common alphanumeric to sound and any visual representation.For example, "Infinity", "neskončnost", and "∞" are all *nomens* for the notion of unlimited in mathematics.

"Euro", "evro", "EUR", and "€" are all *nomens* for the European currency. *Nomen* is therefore what we use to communicate the meaning of any *Res*. In all cases they also include the sound (spoken) versions. In general, the relationship between *Res* and *Nomen* is many-to-many. Anything may have many different appellations and in some cases a *Nomen* may apply to several *Res* (homonyms). In the controlled environment of bibliographic information systems, though, we avoid the latter situation by disambiguating *Nomens* in order to allow only one *Res* for each *Nomen.* The many-to-many relationships is therefore declared as one-to-many. [Figure 6]

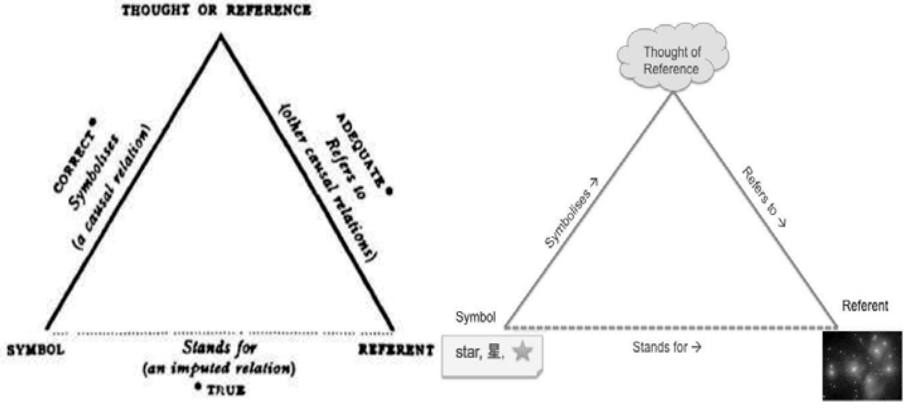
Figure 6: Appellation relationship in FRBR-LRM



Source: http://www.ifla.org/files/assets/cataloguing/frbr-lrm/frbr-lrm_20160225.pdf

The notion of the split between the thing itself (*res*) and the label(s) (*nomens*) we use to refer to it has been described in literature. Ogden and Richards (1923) semiotic triangleof meaning illustrated the relationships among: (1) language, (2) thought content, and (3) referent. Ogden's model was adopted by researchers in library and information science as the basis for building subject authority systems (Dahlberg, 1992; Campbell et al, 1998). [Figure 7]

Figure 7: Left: Ogden, C. K., and Richards, I. A. (1923). *The Meaning of Meaning: A Study of the Influence of Language Upon Thought and of the Science of Symbolism*.London: Routledge & Kegan Paul.p.11.

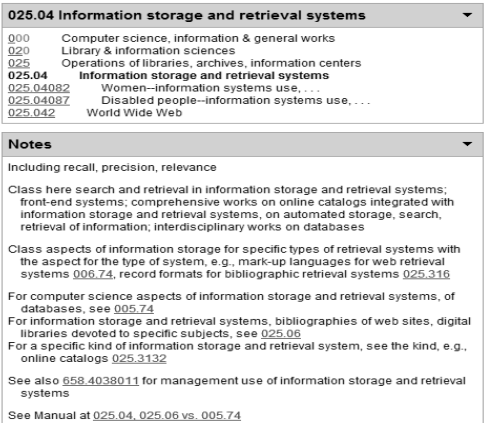


In laying out a layer of sign, icon, and index, Charles Peirce (1932)'s semiotics describes the interaction between a "representamen" (the form the sign takes) and an "interpretant" (the sense made of the sign), as well as an object (to which the sign refers). Beyond these, we also understand that this is not limited to the naming process,

i.e., a list of labels, each corresponding to the thing that it names. As augured by Ferdinand de Saussure (1983), and structuralism more generally, there is an arbitrary relation between the signifier (a sign) and the signified (meaning the concept corresponding to the signifier). The meaning of a signified comes from the differential relations between signs, or the place of a sign in a whole structure of interrelated signifying units.

In the bibliographic universe, it is important to separate what we usually call *concepts* (or topics or subjects) from what they are *known by, referred to,* or *addressed as*. Such relationships are usually incorporated into the structures of subject authorities. For example, in Dewey Decimal Classification (DDC), class "025.04 Information storage and retrieval systems" in Figure 8 is a *thema*. One *nomen* for the class is 025.04. The caption in its full hierarchical context serves as an alternative *nomen* for the class: Computer science, information & general works/Library & information sciences/Operations of libraries, archives, information centers/Information storage and retrieval systems.

Figure 8: DDC class 025.04



Source: *WebDewey*

The notion of *Nomen* as a separate entity seems to be difficult to understand – several comments during the world-wide review included the proposal of to model the appellations as merely attributes of the entities they name. This may be due to the fact that in our mind the thing itself and its label(s) are very closely coupled. What most do not realize is that such a 'simplified' model does not allow including attributes (e.g. language, script) of the *nomen* and neither the relationships (e.g. derivation). As a consequence a 'simplified' model that treats appellations as literals does not enable authority control.

5 Conclusion

FRBR-LRM confirms the importance of defining *Res*, defined as "Any entity in the universe of discourse" and the relationship with *Work*: [*Work* has as subject *Res*]. The modeling of *nomen* as an entity is also a significant decision, which supports authority control. The general appellation relationship [*Res* has appellation *Nomen*] is consistent with the conceptual models and best practices used by the Semantic Web. The FRBR-LRM is currently a draft. After the completion of the review period it will be amended and discussed within IFLA bodies (FRBR Review Group and relevant sections). It is expected to be approved by the end of 2016. This will finally enable consistent implementations of better, improved bibliographic information systems.

Note

[1] http://www.ifla.org/files/assets/cataloguing/frbr-lrm/frbr-lrm_20160225.pdf

References:

Campbell, Keith E., Oliver, Diane E., Spackman, Kent A. and Shortliffe, Edward H. (1998). Representing Thoughts, Words, and Things in the UMLS. *Journal of the American Medical Informatics Association* , 5 (5): 421–31.

Dahlberg, Ingetraut (1992). Knowledge Organization and Terminology: Philosophical and Linguistic Bases. *International Classification*19 (2): 65–71.

(FRAD) *Functional requirements for authority data : a conceptual model*. München : K.G. Saur, 2009. (IFLA series on bibliographic control ; vol. 34). Available at: http://www.ifla.org/files/assets/cataloguing/frad/frad_2013.pdf.

(FRBR) *Functional requirements for bibliographic records : final report.* München:K.G. Saur, 2009. (IFLA series on bibliographic control ; vol. 34). Available at: http://www.ifla.org/files/assets/cataloguing/frad/frad_2013.pdf.

(FRSAD) *Functional requirements for subject authority data (FRSAD) : a conceptual model*. München : K.G. Saur, 2009. (IFLA series on bibliographic control ; vol. 34). Available at: http://www.ifla.org/files/assets/cataloguing/frad/frad_2013.pdf.

Ogden, C. K., and Richards, I. A. (1923). *The Meaning of Meaning: A Study of the Influence of Language upon Thought and of the Science of Symbolism*. London: Routledge & Kegan Paul.

Peirce, Charles Sanders (1932). *Collected Papers of Charles Sanders Peirce. Volume II: Elements of Logic*. Edited by Charles Hartshorne and Paul Weiss. Cambridge, Massachussets: The Belknap Press of Harvard University Press.

Saussure, Ferdinand D. (2013, 1983). *Course in General Linguistics*. Translated by *R.* Harris.London: Duckworth.

Wiesław Babik

Information Logistics: Usability in Knowledge Organization

Abstract

The object of this paper concerns the information logistics relating to knowledge organization. It is assumed that the traditional logistics isrequired as one of the effective information and knowledge organization tools.The results of the previous considerations constitute one of the elements of the current research project on "Information Logistics and its Use in the Organization of Knowledge" whose selected outcomes will also be presented in this article. Specific examples will be used here to present the achievements of traditional logistics that can be used for the purpose of knowledge organization. The paper will indicate that information logistics allows for a wider and deeper view of knowledge organization and identification of the elements which have been omitted, or recognized to be insignificant. It also opens to knowledge organization a new research perspective and consequently a possibility of a further development in the consideration of knowledge organization.

Introduction

In general terms, logistics is understood as a process of planning, implementation, and controlling of effective and efficient commodity and service flow and storage, as well as of the related information, from the point of origin to the point of consumption, to meet the customers' needs, in accordance with their requirements and wishes. In the situation when information and knowledge are perceived as commodity, which is also a product, effective information flow control becomes a logistic issue. It is the task of logistics in this case to organize and coordinate the information and knowledge flow processes in logistic channels and chains. Such an approach is congruent with the understanding of logistics as coordination of a set of specific-class processes (information and knowledge-generating processes in the present case), carried out in the information environment identified within one or more frames of reference.

The considerations presented here will constitute a report on the current results of the research project intended to examine the usability and implementation options of the achievements of traditional knowledge organization logistics, in the creation of its structures and its implementation in information retrieval systems. The research project relies on the critical analysis of literature content.

Information vs. Knowledge

Information itself is a key element for the organization of both information and knowledge, as well as information logistics. There are hundreds of various definitions of information. We can reduce them to information as a property of message consisting in the reduction of indefiniteness or uncertainty as to the status of something, or a further development of the situation that the message is related to.

Information, being a notice transmitted by the informant and received by the informed, regarding something that the informed was not aware of, is always passed through people or other means serving as media. Such a definition of information

causes that the issue of information transmission is essentially analogous to the issue of transportation of goods, being the object of logistics. However, we should be aware that there are essential differences between the goods and information even if the latter is treated as a commodity. When trying to include logistics in the sphere of information transmission, it is indispensable to expand the concept of information in order to take into account the form of signs and their structural properties that constitute the basis of information organization.

Since information is a component of logistic processes, we should be aware of those information features which are essential in such an approach. They are the following: independence from the observer (objectivism); synergy; diversity; perpetuation of the resource; duplicability, transferability in space and time; processability without information destruction (exhaustion); ambiguity, or various meanings for various information users (with subjective evaluation of information); and the description of an object only in respect of a single feature (Stefanowicz, 1997, p. 25).

According to Marek Hetmański,

> […] information is an abstract and general term whose scope is determined not by one constitutive feature but rather by a set of several features, including consecutive ones. The content of that term is lexically rich because the definitions of the particular features are adopted from various disciplines and diverse types of knowledge. Information is also ambiguous more for the reason of numerous changeable uses than the very etymological meaning (Hetmański, 2015, p. 24).

"Knowledge" is an equally ambiguous a term. In a narrow meaning, knowledge is the entirety of credible information about reality, together with the capability to use knowledge. In a broad meaning, knowledge is a vast collection of information, views, belief etc. to which cognitive and/or practical value is assigned. Information is transformed by knowledge when information is ordered, systematized, prioritized, confronted, evaluated, or subjected to criticism. Therefore, knowledge is a type of processed information. Broadly understood knowledge is the entirety of contents enrooted in human mind as a result of accumulation of experience and learning.

It is the condition of the organization of information and knowledge to subject the existing resources to structuring and reasonable ordering. Information organization is carried out by the creation of the systems of information elements assumed within a given representation, being determined by the methods and tools of recording the information collection, the relations between information components, and the operations conducted on information. Information organization is decided in information retrieval systems primarily by the information retrieval language and the method of retrieval set organization. We organize information because we need to search for information. Information organization requires e.g. selection of an elementary unit of description and, consequently, the structure of information attributes, which could play the roles of access points to organized information and knowledge. A network environment prefers content, not any object which embodies content.

Organization of Information and Knowledge: Terms and Tasks

Knowledge organization is a field which deals with ordering of information about specific objects of reality and ensures effective search for information. The task of that field is to improve the systems, methods, and procedures for information resource ordering in all areas. According to Barbara Sosińska-Kalata "knowledge organization is the name of a separate field of interdisciplinary research, associated with information and knowledge ordering to ensure effective access to information for those who need specific information and specific knowledge for specific purposes. That field includes all research concerning conceptual aspects of knowledge representation in information systems, in particular: theory of terms, models of terms and their representations, theory and methodology of information classification, conceptual and terminological systems, methods and systems of information indexation and retrieval, models, methods and techniques of knowledge representation, or designs of the so-called knowledge organization systems." [Sosińska-Kalata, 2008, p. 104]. In a broad interdisciplinary understanding of the term, knowledge organization is a disciplinary subdivision of knowledge and social institutions (e.g. universities), languages and terminologies, terminological systems and scientific theories, scholarly writings and types of such literature.

According to the Danish theoretician of the information science, Birger Hjørland, knowledge organization understood narrowly (as limited to the area of library science and information science) means organization of information in the form of bibliographic records, including bibliographic citation indexation, full texts, and the Internet. (Hjørland, 2003, p. 87). The information science intends that such bibliographic records should be constructed in the way to make optimum searches possible. The most comprehensive knowledge organization definition was offered by the German scholar, Ingetraud Dahlberg (Dahlberg, 2014).

When treating knowledge organization as a process, people distinguish such technological stages of knowledge organization as: indexation and classification, documentation (abstracting and summarizing), information retrieval, search for information based on citations, complete texts hypertexts, and the Internet [Hjørland 2003]. Present-day knowledge organization systems include the following: lists of terms, model card catalogues, glossaries, vocabularies, lists of geographic names, classifications and categorizations, subject headings, classification schemes, taxonomies, categorization schemes, relation lists, thesauruses, semantic networks, and ontologies. The linguistic tools of knowledge organization entail information retrieval language, including the bibliographic citation language and metadata languages. Such tools are used for identification, analysis, and evaluation of the flows of information streams, mainly bibliographic information streams, by citations, co-citations and self-citations.

According to B. Hjørland the core subjects in KO are classification systems and concept systems. They are the two main aspects of KO: (1): knowledge organization processes and (2) knowledge organization systems (KOS). Knowledge organization processes are the processes of cataloging, subject analysis, indexing and classification by humans or computers. Knowledge organization systems (KOS) are the MARC format, classification systems, lists of subject headings, thesauri, ontologies and other systems of metadata [Hjørland, 2008]. Knowledge organization systems constitute important elements of knowledge organization. The application of knowledge organization systems to organize access to information and knowledge resources is associated with the creation of conceptual maps, representing fragments of reality described in a given resource.

A new look at knowledge organization is offered by cognitivism. In that approach, knowledge organization is oriented on the user and is user based. That allows us to answer difficult questions of philosophical nature, including those relating to knowledge on which the designs of *user-friendly* knowledge organization systems are based and the accepted within anthropological, psychological, and sociological paradigms.

Knowledge organization is associated with digitization to ensure broad access to information and knowledge resources and provide search capabilities for complete texts of documents, including those stored in digital forms. That process creates an additional interpretational context. As long as metadata become the search result, it is hard to talk about knowledge. However, when content appears at output, knowledge can become the search result as well. In the network paradigm of knowledge organization, the goal is moved from information provision to "knowledge possession", or from information to knowledge.

Bibliometric maps provide other modern tools of knowledge organization. Those are the key-word maps of a certain discipline of knowledge designed on the basis of an analysis of key word occurrence in the literature of that discipline. The maps of the researchers conducting studies within the given field are designed on the basis of co-citations of their works. Such maps show the relationships occurring among the research problems treated by the researchers. The maps belong to the ISO standard adopted for the description of knowledge structures and their integration with particular sources (or resources) of knowledge. Initially, the maps were developed as the foundations for information retrieval ontology. Presently, they constitute rich semantic models, adjusted to support the retrieval process. The maps use graphs to present formal text characteristics, with respective bibliographic relations, displaying institutional and personal connections between the text authors, as well as the terms contained in their texts. An example of such a map is given by Joseph D. Novak and Alberto J. Canes (Novak, Cañes, 2005). Other types of knowledge organization systems are mentioned by Marcia Lei Zeng (Zeng, 2008).

Information Logistics as a Tool for Managing Information-Stream Flows

Information logistics is presently treated as one of the most important tools for rational actions, i.e. rational handling of information processes, information structures, information systems and networks, and knowledge organization systems. Those are the logistic components of information and knowledge organization that influence the shape of information-feed stream flows.

Information logistics can be understood similarly to general logistics as the process of planning, implementation, control of effective and efficient flow, and storage of information and information services, from the point of generation to the point of consumption to meet the needs of customers (information receivers), in accordance with their requirements.

The need of developing information service process, oriented on the needs of a potential information receiver, is strong associated with information logistics, which is different from traditional logistics of functional information, because the latter is unable to meet present-day needs. Smart information logistics is based on a new quality of information supply processes. The new quality of information supply is assured by decentralization of information supply, which is the basis of the operation of dynamic networks that create virtual enterprises. That is an expression of the creative function of information logistics.

The task of information logistics is to coordinate the information-feed system of a given entity and particular information receivers (Haftor, Kajtazi, 2009). Owing to the determination of how, when, and where necessary information can be delivered, it is possible to bring closer the information supply offer (e.g. to provide information about new publications or new library acquisitions, or the need to return or prolong keeping the borrowed books, or the need to extend the library card validity etc., or informing readers about new acquisitions) to the needs of objective information supply and information services to the information receiver. Information logistics is therefore involved in the information distribution optimization. Thus, we can formulate a general quality criterion for information supply: information should be adjusted to the receiver's needs. Consequently, we should avoid redundancy. The form of offered information should be designed in the way which is the most suitable for information receivers. The design should encourage a far-reaching repetition of such a form of transmitted information, ensuring better comprehension by receivers. An information logistics system should be resistant to information manipulation or distortion. Information logistics should be close to a just-in-time system in its strategy: the right supply to the right receiver at the right time. We should be aware that information needs appear *ex ante* for the user, although the logistic systems create a possibility to develop a detailed analysis of information *ex post*. It would be necessary to reverse that sequence, so that information becomes first an essential element of the offering directed at the receiver and reaching the receiver.

The objectives of information logistics are intended to free information systems of irrational information relationships, and excessive details and redundancy, as well as to design the contents of information packages and the information supply times for all the structural elements of a given organizational unit and potential information users.

The significance of information logistics results not only from the continuous hyper-competition on the globalized information market and the alarmingly expanding asymmetry of market information, but also from the need of the implementation of sustainable development. For that reason, logistics is perceived as an opportunity to reduce the costs of information circulation, with a versatile support of the process by information technologies.

The information processes intended to satisfy the information needs require preparation of information for its consumption. The connections existing between the information generation and consumption or reception processes are equivalent to logistic processes. That concerns the processes of information gathering and processing and making it available. However, the systems in which such processes occur are also called logistic systems. Information systems are thus equivalent of logistic systems.

Presently, we observe the switch from functional perception or resource management to process perception or the perception oriented on flows in information logistics, and that is well corresponding to the differentiation of the existing information processes. That paradigm is expressed by the logistic strategy. The performance of an assumed logistic strategy is carried out in the form of an optimized logistic chain or process in which we see a constant and unmatched ideal of the integration of all information flows.

Various structures are used in information logistics, such as classifications, which allow to order large information collections, typologies, folksonomies, tag clouds etc. Logistics entails design and development of technical, organizational, and information systems, including the development of information and planning structures, efficient implementation, and mutual information process coordination.

Information Logistics and Knowledge Organization

Information logistics may be used in two ways (can be seen):

- as “managing and controlling information handling processes optimally with respect to time (flow time and capacity), storage, distribution and presentation in such a way that it contributes to company results in concurrence with the costs of capturing (creation, searching, maintenance etc)” (https//en.wikipedia.org/wiki/Information_logistics). Information logistics utilizes logistic principles to optimize information handling.
- as a concept using information technology to optimize logistics.

The goal of information logistics is to deliver the right product, consisting of the right information, in the right format, at the right place at the right time for the right people at the right price and all of this is customer demand driven.

The methods for achieving the goal are: the analysis of information demand; intelligent information storage; the optimization of the flow of information; securing technical and organizational flexibility; integrated information and billing solutions (Uckelmann, 2012).

Information logistics is closely related to information management, information operations and information technology. It deals with the flow of information between human and computer. The contemporary information logistics focuses on the organization, planning, control and implementation of the flow of information. Information means in a general sense everything that adds knowledge and thus reduce ignorance or lack of precision.

The organizational model of information should integrate traditional functions of logistics: supply, production, distribution, transportation, storage, and operation not to separate mutually dependent elements unnecessarily and make information-feed process management easy. The integration of logistic conceptions is a condition of theirimplementation in information and knowledge organization leading to the integration of information collected in a system. Computer databases are presently the most popular forms of information and knowledge organization. The quality of information obtained from an information system depends finally on the method of information processing and presentation.

Information needs play essential roles in logistics. Meeting those needs is necessary to perform all management functions. And that is possible only in the situation of continuous assurance of access to current, detailed, and reliable information.

Logistic management of information consists in a comprehensive management of all activities during the information stream flow, from origin to reception. That concerns control and coordination and control of information-feed stream flows, using logistic chains to make the flows effective. Rational information-stream flow management contributes to the improvement of the quality of serving information users by meeting of their needs, in accordance with their expectations [Dworecki, Berny, 2005].

Our research confirmed that information processes are hampered by the overwhelming amounts of information which is irrelevant or unhelpful to users. When rationalizing information-stream flows to information users by the managinginformation and knowledge organization, information logistics can become very useful as it is one of effective information and knowledge organization tools along the way of information flow to the user, and consequently, information logistics will contribute to cleaning information retrieval systems of irrational information connections and excessive detail (e.g. by elimination of unnecessary redundancy). The role of information logistics may be reduced in that case to an organizational tool for

effective integration of the information flow in information retrieval systems and coordination of feeding the system with information. Information logistics is beneficial for the optimization of the distribution of information which serves knowledge structure building.

In information and knowledge organization, information logistics may play two essential functions: a creative and an integrating one. Substantively, we can deal with a proper coordination of information streams. The use of the achievements of traditional commodity and service logistics in knowledge organization can significantly contribute to the substitution of energy by information. The amount of physical effort, which is indispensable for the fulfilment of information tasks, in fact depends on currently held information and knowledge (their type, quality, and structure), as well as the skills of using them. The more information one holds the less effort will one usually require to attain specific objectives. In the case of shortage or inadequacy of information and knowledge, one has to consume more energy to perform a task. Consequently, information allows one to save energy, and just the opposite: one spends more energy when missing information. The scope of the ability to substitute energy and information under the same processes but in various information and retrieval systems is not identical. Energy cannot be replaced with information for ever, since energy is also necessary at least for information keeping and processing in a system. A new quality of information service can also be achieved owing to decentralization of information supply from the existing databases and the bases which constitute the foundations of dynamic networks which create a virtual information environment. That is a vivid expression of the creative function of information logistics. The use of information logistics allows us to provide a new definition of information processes and that can enrich knowledge organization.

Conclusion

The theory of knowledge organization allows us to order various methods of knowledge organization, applied in practice in literature and digital communication. The methods concern the bibliographic relationships, thematic relationships, contacts between scholars, information and knowledge flow methods, the origin of ideas and their development, citations, co-citations, and self-citations, the structure of literature and the birth of new research disciplines, researcher networks (invisible colleges), thematic connections, including inter- and multi-disciplinary ones, the necessity of multi-aspectual approach to research problems, the most frequent citations and the most often cited researchers, new trends in research and science, the flows of new conceptions and ideas, including information and knowledge streams, in connection with the statistics applied for the purpose of the evaluation and ranking of research journals, together with their research papers.

Knowledge organization can be usually reduced to the use of the methods and tools in design and application of information retrieval languages, including those used for

the evaluation and analysis of research literature, various fields, and electronic messages.

The use of logistics in the knowledge organization for control and coordination of information and knowledge-generating processes may lead to the integration and creation of new added value. The logistic functions extended to the sphere of information processes also entail purely organizational aspects belonging to the information and knowledge organization. Those can lead in particular to the optimization of information resource streams and flows. Integration of those processes may ensure the desired quality of information service for users, with such information and knowledge organization which allows for obtaining an optimum level of the use of information and obtain new knowledge as added value.

References

Dahlberg, Ingetraut (2014). *Wissensorganisation*. Entwicklung, Aufgabe, Anwendung, Zukunft. Würzburg.

Dworecki, Stanisław E., Berny, Jan S. (2005). Logistyka racjonalnego działania (Zarządzanie strumieniami przepływów) [*Logistics of Rational Operation (Flow Stream Management)*]. Radom.

Haftor, Darek, Kajtazi, Miranda (2009). *What is Information Logistics*. [http://lnu.diva-portal.org/smash/get/diva2:344139/FULLTEXT01]

Hetmański, Marek (2015). Świat informacji [*Information World*]. Warszawa.

Hjørland, Birger (2003). Fundamentals of Knowledge Organization. *Knowledge Organization*, 2: 87-110.

Hjørland, Birger (2008). What is Knowledge Organization (KO)? *Knowledge Organization*, 2/3: 86-101.

Novak, Joseph D., Cañas Alberto J. *The Theory Underlying Concept Maps and How to Construct and Use Them.* [http://cmap.ihmc.us/docs/theory-of-concept-maps]

Sosińska-Kalata, Barbara (2008). Ewolucja modeli organizacji wiedzy w systemach organizacyjnych [*The Evolution of Knowledge Organization Models in Organizational Systems*]. In: *Książka. Biblioteka. Informacja. W kręgu kultury i edukacji*. Ed. E.B. Zybert&D. Grabowska. Warszawa. Pp. 103-16.

Stefanowicz, Bogdan (1997). Informacyjne systemy zarządzania [*Informational Management Systems*]. Warszawa.

Uckelmann, Dieter (2012). *Quantifying the Value of RFID and the EPC global Architecture Framework in Logistics*. Berlin.

Zeng, Marcia Lei (2008). Knowledge Organization Systems (KOS). *Knowledge Organization*, 2/3: 160-82.

Peter Ohly

Dimensions of Globality: A Bibliometric Analysis

Abstract

We analyzed over 12,000 publications on global issues ('globality') from the years 1945-2008, using literature references in a German social science database. Methodologically this study applies network analysis on document descriptors in combination with main document categories, what permits the visualization of thematic changes in the light of some stable concepts. During the whole period a partially interrupted continuous increase in the number of publications is shown. The breakdown in certain time periods reveals some shift in the understanding of global issues, though there is little specific orientation. An analysis of the authors gives their activity and cooperation over time. There seem to be no general research centers, but rather smaller groups of critical scientists.

1 Introduction

The concept 'globalization' is indistinct (Teune 1990, SEP 2014) and has no uniform definition. The references on literature 1945 to 2008 with respect to the concept 'globality' (global themes) were examined by a bibliometric analysis in the context of a state-of-the-art-report (Mayer 2011). The intention was to get ideas what different kinds of 'globalization' are treated, and how they develop over time, but as well, the structure of its context, could be detected by network analysis. Further, publication characteristics and main players in this field were of interest. Comparative analyses (different time, different schools) afford consistent vocabulary, what was given more or less by the standards of the data base manufacturer, but further methodological considerations had to be taken into account. As the above mentioned state-of-the-art report compiled other overview chapters on special globalization aspects, we can use them for some validity considerations, here.

Bibliometric studies intent to give an objective description of a scientific area (Nacke 1979; van Raan 1997). They are based on process-produced data that accrue in science without questioning or explicit collection (Bick/Müller 1984). It must be noted that this is a document analysis that means the material has been collected for an entirely different purpose and therefore contains only limited knowledge for the new analytical purpose. Of course not the knowledge or science discipline is measured, but only the publication production in a limited sense (Stock 2000). It is not considered whether a certain topic has been largely by repeated or published unchanged by an author, nor its work context or its reputation has been taken into account (cf. Gerhards 2002). Such issues are the subject of scientometric fundamental research, that is not intended here. Rather, we intend to compile indicators for the visibility of this field of knowledge in a generally accessible database (see Mandl 2006; Weingart 2003, chap. III, Wissenschaft als Kommunikationssystem, 31-39) and we will show traces of publication behavior on a highly aggregated level. Often bibliometrics is seen as a tool for extracting concepts from library stocks, e.g. as an aid to develop thesauri or

retrieval front ends (Schwanhäuser/Kind 1974; Hjørland 2013). In our study the denomination of the subjects (i.e. concepts in a wider sense) is already given by the data base manufacturer but the relations between these are found by bibliometric analysis. Thus it gives evidence of their usage in the data base, including some pre-defined structures of the thesaurus applied by the data base manufacturer. The empirical generation of main topics and concepts clusters used in literature on ‘globality’ and their relation corresponds with the claim for postmodern ‘domain analysis’, which describes the informational focus and needs of special discourse communities (Hjørland 2002; Smiraglia 2015). The community here is built by the authors in the field of ‘globality’, but it is represented in the view of social science information specialists (knowledge organizers, indexers etc,). Methodological our analysis can be seen as a sociological study in this field, as groups actors and their actual artifacts are examined (c.f. Merton 1985)

2 The information base

As already stated above the basis for the quantitative analysis on the development of publications on global topics was a literature database, namely SOLIS (Social Science Literature Information System) [1]. The starting point for the analysis was a search with keywords of the field of ‘globality' (what is understood here as a more general concept than ‘globalization’) in publications in the German language area within the publication years 1945-2008 [2] - not including misleading aspects, as 'world war ' or 'personality'. Accordingly, relevant terms such as ‘globalization’, ‘multinational’ or ‘global economy’ had to occur in the title or the key terms. This yielded a total of 12,374 references across all years (see Fig. 1 in Ohly 2011). A narrower period was based on the years 1999 to 2008 with 6,940 bibliographic references to reflect current trends. Occasionally, however, the periods 1947-68, 1969-78, 1979-88, 1989-93 and 1994-98 were used for comparisons. Not always the special analyses are based exactly on 12,374, resp. 6,940 publications. The reason is that for some questions proper information was not available in all publication references (e.g. author name).

With regard to all analyses, it should be emphasized that the documentary basic used for this analysis may not reflect all or only pertinent publications on the subject (e.g. due to: the selectivity of the database, refinement to the most significant search terms, incorrect assignment to descriptors). Therefore, comparisons should be made only in relation to this documentary body and then however with caution, since a number of uncontrolled factors, such as occasional additions of certain data providers or changing indexing rules, might have occured. In no case, the overall quality of certain study objects (database, authors, journals, etc.) should be evaluated as such, as for other topics there might quite different results exist. Note also that there are not equivalent publication types, such as grey literature (papers) or contributions in compilations (chapters). For this reason, no absolute evaluation of the analytical results is intended, but rather objective empirical figures are presented, which can describe the state and

development in the studied publication period and otherwise should be interpreted in suitable questions, and preferably together with other subject specific material.

Since 1972, with respect to the publication year (see Ohly 2011, Fig. 1) an increase of the amount of publications on 'globality' can be reported in the overall trend. This corresponds to one part to the general trend that more and more publications came on the market, and that the database widened its scope over the years. On the other hand, it can be shown that, at the beginning of the seventies, the literature increases remarkably when the Bretton Woods system, the system of fixed currencies between western states, was converted to floating currencies. After a small local peak in 1985, a political literature segment of the data base was abandoned and entailed the temporary decrease of globality literature (from 2.7% to 1.3% in relation to the entire literature of these years in the data base). On the other hand compiled publications and the adoption of the descriptor 'globalization' in the thesaurus in 1996 lifted the sensibility for global themes. But in 1997 the Asian financial crisis seems to have caused again a short setback in the attractiveness of this theme. Especially towards the end of the examined period, the number of documents is not definitive, since even after the search (here: in July 2009) input for the data base will be produced for the last publication years. As well not all bibliographic fields are filled in the last period. On the other hand a later search (in September 2015) with the only term 'globalization' showed as well a steady increase till 2010 but then as well a tremendous decline (from 20,136 for the period 2006-2010 to 10,921 for the next five year period). Maybe, meanwhile, this buzz word has lost some of its attraction.

3 Individual and collective actors

We analyzed how frequent individual authors are involved in the 12,374 publications (see Ohly 2011, Tab. 1). Again, in this 'ranking' of authors has to be noted that the frequencies only refer to this specific search in the respective database and that the given information results only from the bibliographic descriptions. Individual forms of publications (e.g. books vs. gray literature) are not differentiated. While publications with single authors can rather be considered as an independent achievement in this area, joint publications tend to demonstrate collective achievements. A comparison of the total number of publications with those by individual publication shows different ranking positions (74% of all bibliographic references here have only one author). *Zürn* ranks with all his publications in second position, however, as individual author he lies several rank positions behind *Witthauer* who is with 30 own publications at the first place. Vice versa *Witthauer* is at the fifth position including the publications with others. Like in other publication analyses the counting of publications shows that only few authors publish comparatively much, and vice versa many authors publish comparatively few (exponentially declining Zipf curve).

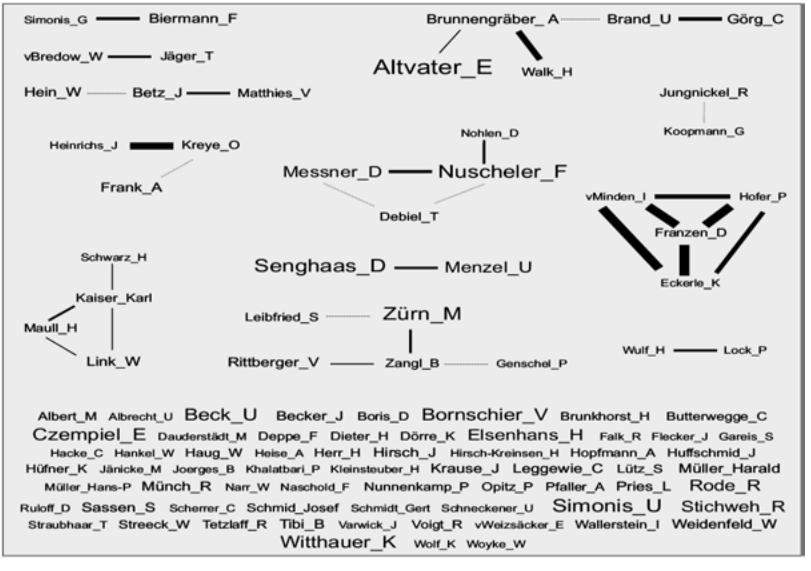
A publication distribution over the years is an indication of how continuously the respective author has been engaged in global themes (see *Tab. 1*). The years 1989-2008

were divided into four 5-year periods. Thus can be shown, which authors publish only recently in this field (e.g. *Zürn* and *Stichweh*), who publishes consistently on this topic (e.g. *Senghaas*, *Bornschier*, *Elsenhans*) and who may already have turned away from this subject (e.g. *Witthauer*). Also, there are individual authors whose 'leading position' primarily goes back to an extraordinary period of activity, but they are at the same time continuously publishing on this topic (e.g. *Simonis*) [3].

Table 1: Frequency of publications 1947 to 2008 (descending by frequency)

Author	Publication Period							
	1947-68	1969-78	1979-88	1989-93	1994-98	1999-03	2004-08	total
Altvater, Elmar		1	7	4	*10*	9	7	**38**
Zürn, Michael				1	10	9	*14*	**34**
Simonis, Udo E.			1	5	14	9	4	**33**
Senghaas, Dieter		3	*17*	5	2	2	3	**32**
Nuscheler, Franz			6	2	7	6	*10*	**31**
Witthauer, Kurt	*16*	12	2					**30**
Beck, Ulrich			1		8	*10*	*10*	**29**
Bornschier, V.		3	*15*	5	2	1	1	**27**
Czempiel, Ernst			3	1	8	*10*	5	**27**
Elsenhans, H.		1	*10*	5	4	2	3	**26**
Menzel, Ulrich		1	*8*	7	4	4	2	**26**
Messner, Dirk				3	5	7	*10*	**25**
Rode, Reinhard			1	1	4	*13*	4	**23**
Stichweh, R.					5	*10*	8	**23**
Müller, Harald			1	2	3	4	*12*	**22**

Figure 1: Co-authorship 1947-2008. Remarks: The ranking of the line width corresponds to the frequency ofco-authorship (between 2 and 10) and the size of the author name corresponds to his publicationfrequency



As already apparent, not all listed authors publish alone, e.g. they are working in a broader research context. The map of co-authorships (see *Fig. 1*) indicates, how far (in absolute frequencies), the same pairs or groups are repeatedly involved in publications. Conversely, the names not linked to others (at the bottom of the figure) indicate that they publish often, but either alone or always with changing co-authors. Some minor components can be detected: Around *Altvater, Brunnengräber and Görg* (left critics of globalization), around *Zürn and Rittberger* (de-nationalization approach), around *Nuscheler and Messner* (development policy approach), around *Link*, around *Franzen*, and around *Hein*. In addition, there are a few other minor author teams, like *Simonis – Biermann* (ecological approach) or *Senghaas – Menzel* (idealistic approach). These components are not mutually connected by further co-authorships (although they can quite often turn up to be 'side by side' in compiled editions) [4]. Although we examine an interdisciplinary topic and individual sciences are hardly nameable, it can be stated that those political scientists are mostly represented who have a more critical-theoretical orientation. Perhaps this explains why so many single authors are represented, as in this area less large-scale research is required.

4 Thematic areas of publications

It was analyzed to which extent certain global issues occur in the literature as hints for research trends. First, we referred to the thematic areas (more or less concepts in a narrower of indexing [5] of the references in the database [6]. We applied this classification only on the second level, with 23 different sub-themes. This addresses to deliberately rough concepts, which were used in *Fig. 2a* to connect the various keywords to wider semantic contents [7].

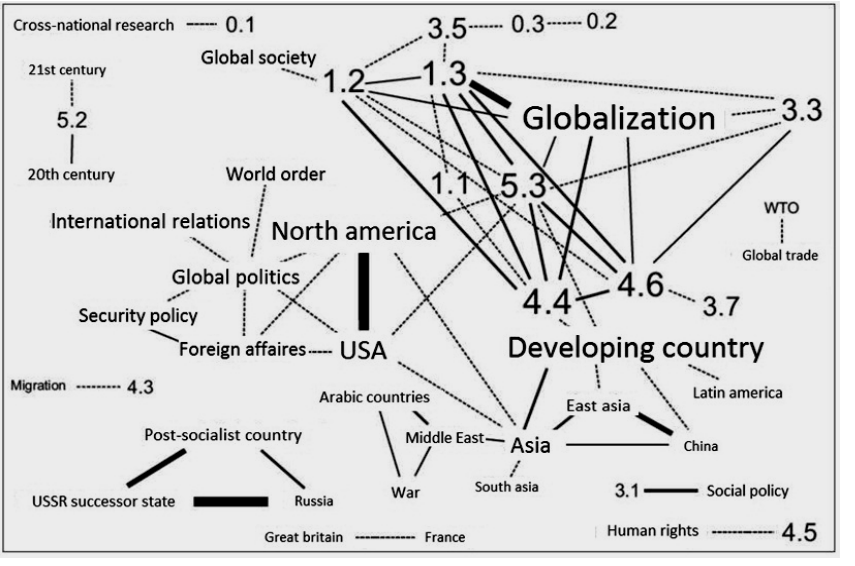
We analyzed the occurrence of the classes with respect to publication periods (see Ohly 2011, Tab. 5). Because of the wide scope of the selected areas only moderate changes are visible over time. The period 1947-1978 is less representative as the SOLIS database was created in 1979 and the publication years 1945-1977 were assembled only for a narrower scope [8]. Generally, it can be seen that throughout the period 1947-2008 the subject areas *4.4 (Politics)*, *4.6 (Economy)*, *5.3 (Regional studies)* are strongly represented. In all years, are only little represented life-quality related topics, such as *1.4 (Social indicators)*, *5.4 (Professions)*, *3.2 (Occupation)*, *4.1 (Education)*, *3.6 (Leisure)*, and *3.4 (Medicine)*. Only in the course of time, *4.5 (Law)* and *1.3 (Social change)* are significantly becoming stronger. *4.4 (Politics)* already has in 1969-1978 [9] a high level, but again it increases up to 85.8%, in the last period 2004-2008. *Scientific disciplines (0.3)*, as well as *Historical studies (5.2)* decrease over time, at least in the 90s. *Social data (1.4)*, i.e. social indicators, are in fact not applied or discussed (1.1%).

Relations of the themes with each other are represented in a conceptual map (see *Fig. 2a*). Now, for further clarification of the semantic content of the thematic areas (represented as *notations*), *keywords* are also included if they meet the frequency

criteria as well. Indirect semantic proximity of concepts is given by the crossed nodes, while as well the bivariate relationship strength is given [10]. Accordingly, chains such as *Foreign Affairs – Security policy – Global politics – International relations* are only associated with each other by one path, and cliques [11] as *Middle East – Arab countries – War* are directly and indirectly connected to one another.

Absolutely as well as with respect to the relationships in the period 1999-2008 thematic areas are most strongly represented, whereas the keywords are much more specificly applied within those areas. However, the descriptor *Globalization*, belonging to the field *1.3 (Social Change)*, has as well strong relationships with (keywords from) other areas, namely *4.4 (Politics)*, *4.6 (Economy)*, *1.2 (Social systems)*, *5.3 (Regional studies)*. Even, a keyword is not necessarily linked to his subject area, for example if its area occurs frequently due to other keywords (e.g., *3.3 (Labor)* vs. *Unemployment* (see Ohly 2011, Fig. 5); *4.4 (Politics)* vs. *Foreign affairs*). Evidently, as *Methodology (0.1)* the *Cross-national research* is represented, however, not specifically with a narrower topic, such as *Globalization*.

Figure 2a: Subjects relations as semantic network from 1999 to 2008. Remarks: The ranking of the line width corresponds to the Jaccard association degree (0.2 to 0.9) and the term size to the literature frequency of this topic

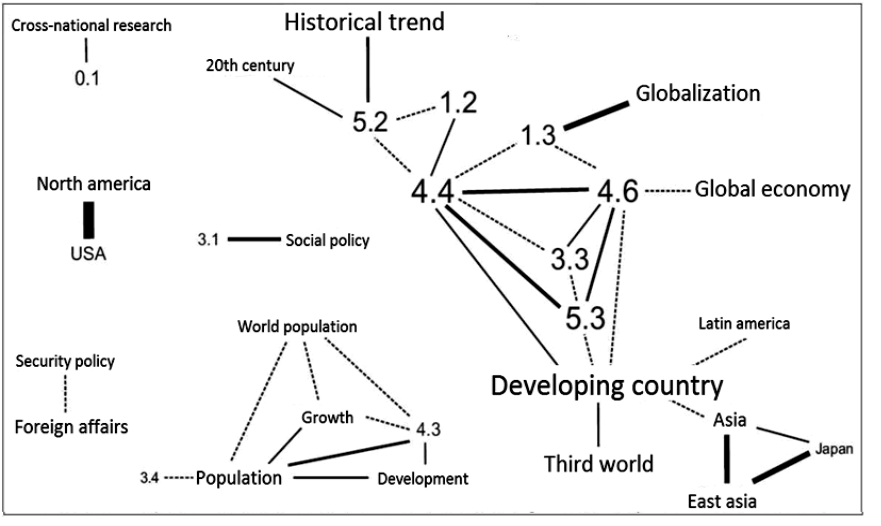


A main concept complex is becoming apparent around the subject areas *5.3 (Regional studies / Regions)*, which directly or indirectly (via others) is connected with *4.4 (Politics)*, *4.6 (Economy)*, *1.3 (Social change)*, *1.2 (Social systems)*, *3.3 (Labor)*, *3.7 (Environment)*, and *5.2 (Historical studies)*. *4.4 (Politics)*, *4.6 (Economy)*, *1.3 (Social change)*, *1.2 (Social systems)*, *3.3 (Labor)*, *3.7 (Environment)* are as well in a

closer connection, which can be considered as a confirmation of the interdisciplinary nature of global themes and the essential aspects of globalization. Another complex is formed in the neighborhood of *5.3 (Regional studies)*, relating to *Developing countri*es, *Asia* and the *Middle East*.

Looking at the period 1947-1998, the young term *Globalization* (only in use since 1996) is poorly connected with other themes there (see *Fig. 2b*). Only in its subject area *1.3 (Social change)*, given by the thesaurus structure, so rather theoretical, it is dealt with. In 1947-1998 the term *Global economy* is used in a similar relational context like *Globalization* later. In this period *Middle East* (even as a hotspot) is not dealt with, in this view. With regard to *Population* the (projected) *Growth* is predominant.

Figure 2b: Subjects relations as semantic network from 1947 to 1998



5 Conclusion

We examined 12,000 publications on global topics from the years 1945-2008 on the basis of a German Social Science database. Even when examining particular themes in connection with ‘globalization‘ and the older topic ‘global economy‘, there is little specific orientation in the whole period. But being seen in the light of different periods, there is not only a continuous increase but also a slight changing of the focus. By comparing these results with other studies in the field we have a means to interpret it, e.g. historical events might influence the literature outcome. Accordingly, there is a trend from economical considerations to security problems. In the later period, population is less seen under the perspective of growth but of migration. Economy now encompasses as well environmental issues. And human rights are as well coming into consideration of globalization research.

This analysis gives some insight into the phenomenon of globalization but also shows the possibilities and limitations of bibliometric studies. Often the history of the data base must be taken into account to locate artifacts. In this report, some validity is given by external findings published in a state-of-the-art-report (Ohly 2011a). As upcoming demands and research themes there are stated quality expectations, as democracy, life quality, world peace, that finds some representation in concept map of the later period. The *Atlas of Globalization* expresses this in a correspondence of *Gini coefficient* (Equal income distribution) and *Human developmentindex* (Le monde diplomatique 2006).

Notes

[1] The SOLIS database has been produced by Leibniz Institute for the Social Sciences GESIS-IZ. It provides information on specialized Social Science literature in German language including journal articles, contributions in compilations, monographs and gray literature.

[2] Actually publications in SOLIS start only since 1947, therefore 1947 is provided for as starting period.

[3] By the way, Ulrich Beck has yet published in the period 1989-1993 to 'risk' and 'modernity', however not explicitly in the context of 'globalization'.

[4] A graph for the period 1999-2008 reflected fewer network components, as connections from the previous periods were missing (see Ohly 2011).

[5] These keywords or descriptor terms are selected from an authoritative list.

[6] The classification of these areas of keywords was taken from the authoritative list by the database provider.

[7] Several keywords of a literature reference may belong to the same area but are then counted only once.

[8] With specific reference to Bette et al. 1980.

[9] Until the year 1986, there was input from the political science database PolDok FU Berlin, what may explain the local peak in the subject area 4.4 (Politics) for the years 1979-1988.

[10] Here a Jaccard coefficient of at least 0.2 is considered, what is already a significant association. The Jaccard coefficient sets the association (but not the shared non-occurrence) in relation to the frequency of the two concepts, where the shared occurrence is counted only once (see, e.g. Sodeur 1974: 107-11).

[11] In a so-called 'clique' all points are interconnected by a maximum number of shortest paths.

References

Le monde diplomatique (2006). *Atlas der Globalisierung*: Die neuen Daten und Fakten zur Lage der Welt: Menschliche Entwicklung. Berlin: taz.

Bette, Karl-Heinrich; Herfurth, Matthias; Lüschen, Günther (1980). *Bibliographie zur deutschen Soziologie*, 1945-1977. Göttingen: Schwartz.

Bick, Wolfgang; Müller, Paul J. (1984). Sozialwissenschaftliche Datenkunde für prozess-produzierte Daten: Entstehungsbedingungen und Indikatorenqualität. In: *Sozialforschung und Verwaltungsdaten. Historisch-Sozialwissenschaftliche Forschungen 17.* Stuttgart: Klett-Cotta. Pp 123-59.

Cronin, Blaise (1984). *The citation process*. London

Garfield, Eugene (1995). Citation indexes for science. *Science*, 122: 108-11.

Gerhards, Jürgen (2002). Reputation in der deutschen Soziologie – zwei getrennte Welten. *Soziologie*, 2002(2): 19-33.

Hjørland, Birger (2002). Domain analysis in information science: Eleven approaches: Traditional as well as innovative. *Journal of Documentation,* 58(4): 422-60.

Hjørland, Birger. (2013). Citation analysis: A social and dynamic approach to knowledge organization. *Information Processing & Management,* 49(6): 1313–25. [http://www.sciencedirect.com/science/article/pii/S0306457313000733].

Mandl, Thomas (2006). *Die automatische Bewertung der Qualität von Wissensprodukten*. Habilitationsschrift Universität Hildesheim, FB Informations- und Kommunikationswissenschaften.

Mayer, T. et al. (2011). *Globalisierung im Fokus von Politik, Wirtschaft, Gesellschaft*. Wiesbaden: VS.

Merton, Robert K. (1985). Entwicklung und Wandel von Forschungsinteressen. *Aufsätze zur Wissenschaftssoziologie.* Suhrkamp Verlag, Frankfurt am Main.

Nacke, Otto (1979). Informetrie: Ein neuer Name für eine neue Disziplin. *Nachrichten für Dokumentation,* 30 (6): 219-26.

Ohly, H. Peter (2011). Globale Themen im Spiegel der deutschen sozialwissenschaftlichen Literatur: Eine szientometrisehe Analyse. In *Globalisierung im Fokus von Politik, Wirtschaft,Gesellschaft*. Vs Verlag, p. 419-43.

Ohly, H. Peter (2011a). *Globalisierung: Woher? Was? Wohin? Der Versucheines Fazits*. In *Globalisierung im Fokus von Politik, Wirtschaft,Gesellschaft*. Vs Verlag, p. 445-452.

Patzke, Alexander (1998). *Grenzenlose Gesellschaft – Grenzen gesellschaftlichen Handelns. Reihe Gesellschaft im FOKUS der Sozialwissenschaften*. Bonn: IZ Sozialwissenschaften.

Schirm, Stefan A. (Ed.) (2006). *Globalisierung. Forschungsstand und Perspektiven. Internationale Politische Ökonomie Bd. 4*. Baden-Baden: Nomos.

Schwanhäußer, Gerhard & Kind, Friedbert (Eds) (1974): *Thesaurus Hochschulforschung, Hochschulbau. Ein automatisch erstellter Thesaurus.* München-Pullach: Verlag Dokumentation.

SEP (Stanford Encyclopedia of Philosophy) (2014). Globalization. [http://plato.stanford.edu/entries/globalization/]

Smiraglia, Richard (2015). *Domain analysis for knowledge organization*: tools for Ontology Extraction. Chandos Publishin.

Sodeur, Wolfgang (1974). *Empirische Verfahren zur Klassifikation. Studienskripten zur Soziologie 42*. Teubner: Stuttgart.

Stock, Wolfgang G. (2000): Was ist eine Publikation? Zum Problem der Einheitenbildung in der Wissenschaftsforschung. In *Wissenschaft und Digitale Bibliothek*: *wissenschaftsforschung jahrbuch*. 1998. Berlin: Gesellschaft für Wissenschaftsforschung. Pp 239-82.

Teune, H. (1998). *The concept of globalization*: Paper prepared for the 14th ISA Word Congress, Montreal.

Van Raan, Anthony F. J. (1997). Scientometrics: State-of-the-Art. *Scientometrics* 38(1): 205-18.

Weingart, Peter (2003). *Wissenschaftssoziologie*. Transcript Verlag: Bielefeld.

Helen de Castro Silva Casarin and Nayara Bernardo de Mattos

Child's Information Behavior in the Domain of Information Science: An Analysis through the Scopus Database

Abstract

This communication is part of a larger study about child´s information behavior that constitutes a discourse community. The present relate aims to verify how the Information Science domain has researched the information behavior of children, based on the bibliographical review carried out in Scopus during the last 10 years. The results presented here include identify the most productive authors; the most used keywords (and the issues studied); in which countries the articles were published in and which were the most cited articles, the publication frequency by year and the journal in that the articles were published.It was observed that most of the papers were published in the United States and Canada. It was also found that two papers stand out as the most cited, both dealing with the search for information in children´s daily lifeand they address methodological aspects of the research, which shows the interest of the researchers on this topic.

Introduction

Information behavior is a traditional issue studied in the field of information science in order to know how users deal with it in their day by day and play some role that is practiced by the people in several contexts (Case, 2012). Hjørland (2002, 2004) considers that users`s studies "[…] may represent an important approach to domain analysis in Information Science if they are informed by proper theory" (p.432). This studies can provide inputs about specific information needs, preferences and practices of different communities according to this speciality. According to Smiraglia (2014), information behavior is one "aspect of social behavior that arises from within a domain".

This communication is part of a larger study aboutchild´s information behavior that constitutes a discourse community. Although there is a predominance of studies on adults and professionals in certain areas (Case, 2012), in recent years some authors, including Chelton and Cool (2004), are concerned to understand the information behavior of children and adolescents. The results of the researches on this subject can be applied in the improvement of practices and services of information systems. Beak (2014), for example, believes that the knowledge organization has devoted little attention and consideration to this group, to the extent that this can be considered as a marginalized group of users.

The present relate aims to verify how the Information Science domain has researched the information behavior of children, based on the bibliographical review carried out in Scopus during the last 10 years. The results presented here include identify the most productive authors; the most used keywords (and the issues studied); in which countries the articles were published in and which were the most cited articles, the publication frequency by year and the journal in that the articles were published.

Knowledge Organization, Domain Analysis and the Information Behavior

Hjørland and Albrechtsen (1995) point the Domain Analysis as an important method for the characterization and evaluation of the science, in which it allows identifying the conditions under which the scientific knowledge is constructed and socialized. According to Guimarães (2014), the domain analysis contributes to the epistemological configuration of different areas, therefore, allowing their consolidation.

According to Smiraglia (2014) the domain analysis is the heart of Knowledge Organization and the research in this field today concern to knowledge of individual groups, e.g., children, that is approached here.

The Information Behavior consists of all human behavior that is related to the sources and channels of information, including the active or the passive search, and the use of it. Wilson (1999, p. 251) points out that the information searching behavior, which is a component of the information behavior, arises as a result of a need for specific information identified by the subject in his daily life. In order to fulfill this need, the individual uses formal or informal sources of information. At the time the individual requests some particular information to the system or to the sources the obtained results may be satisfactory or not. When this request presents a satisfactory result, the individual can then make use of it. But, when the result does not become satisfactory, the individual will resume the search process.

Case (2012) makes an extensive review of the different user groups that have been studied in researches on information behavior. Among these, we highlight those that consider the demographic aspects for choosing the research subjects. Age has been the focus of many studies and according to Case (2012), particularly those that focus on the extremes of life: the children and the elderly. This communication favors the studies that present children as subjects.

Beak (2015) points out that the knowledge organization systems are not developed considering children's characteristics. The author points out that the knowledge organization systems are usually developed for adult audiences, causing children to have to adapt to them because systems, generally, do not consider the process of cognitive development of children. This condition makes the information search difficult for this group and has ethical implications for the librarian's work. Beak (2015) emphasizes that there is a need to understand the information behavior of children, so that systems of knowledge organization that fully meet their information needs can be developed. Silva and Silva (2008) also emphasize the need for studies in information organization area to contemplate the characteristics of the children's audience. The authors point out that children develop at different rates and same age children's reading skills sometimes also varies. Thus, reading levels should be considered in the information organization instruments.

Child's Information Behavior

Todd (2003) proposes three lines of research on information behavior of children and adolescents. The first line of research refers to investigations that seek to identify the contribution of school libraries to the students' learning. The second line of research deals with the search for information on the Internet, that is, the way children and adolescents do their search for information on the Internet. The third and final line of research addresses the search for information to everyday use by children and adolescents. These daily searches relate to their life interests, such as drug use, health and professions, for instance.

Methodology

In order to carry out the data collection of the studies conducted on Information Behavior of children in the domain of Information Science, it was used the Scopus database because it is the largest scientific database peer reviewed. We selected articles from the last ten years, from 2006 to 2015, that addressed the "information behavior of children". The search was conducted with the following terms: "Information Behavior" AND children in the title, abstract and keyword. 44 documents were retrieved. The results was limited to articles and conference papers.

The exam of keywords, abstract and author affiliation show that some documents wasn´t of information science domain or because it wasn´t about information behavior of children but about their parents, etc. The final list included 21 documents (15 articles) and (six conference papers).

The articles were examined and the following data were extracted: authors´s name; keywords; author's country of origin; number of citation of each article, year of publication and the journal title of publications.

Results

At first, we identified the authors that have been devoted to the topic.It was identified 35 different authors in the 21 analyzed articles. It was noted that most of the works were published in co-authorship, given that fourteen articles have two or more authors, whereas that five articles has two author; for three and four authors there were four occurrences and one had five authors.

Regarding thefrequency publication of each author, it was found that Behesti, J. is the most productive author within the considered period, with three papers; 13 authors published two articles and 21 only one article in within the period analysed.

The keywords indicated by the authors in seven of the 21 articles examined were also analyzed. The general keywords such as Information Behavior, research, and Information Science and the others related to age and geographical location were excluded and we get the following presented in table 01.

It can be noticed that the main topics approached in the articles are focused on steps and/or specifics aspects of information behavior, like information seeking, sharing, use,

creating, collaborating etc. Technology is another issue approached in the set of articles The information organization has been covered in the some articles on the subject, particularly related to the design of the interface of catalogs and information retrieval. However, the information organization instruments, such as classification or indexing, were not founded in the articles addressed.

Table 01: Keywords of analysed documents

Keywords	Frequency
Information seeking	3
Information use	3
Collaborative information behavior	2
Information creation	2
Information literacy	2
Affective factors	2
Artefacts	1
Artificial intelligence	1
Association reactions	1
Bonded Design	1
Browsing	1
Catalog interface design	1
Children with special needs	1
Cognitive information behavior	1
Collaborative information	1
Collaborative searches	1
Conceptual frameworks	1
Contextual modeling	1
Curricula	1
Design representations	1
Design/methodology/approach	1
Digital information literary	1
Digital inquiry	1
Digital libraries	1
Disabilities	1
Disruptive technologies	1

It was verified the countries where the paper was published. As the table 01 shows the articles were published mainly in the USA, with eight articles, followed by Canada, with five articles, and in the United Kingdom, with three articles; Finland and Ireland with the fewest publications: one paper each.

Table 02 - Countries in which the articles were published

Countries	Number of papers
United States	8
Canada	5
United Kingdom	3
Finland	1
Ireland	1
Total	14

It was verified the countries where the paper was published and in which country there were more publications. As the table 03 shows the articles were published mainly in the USA, with eight articles, followed by Canada, with five articles, and in the United Kingdom, with three articles; Finland and Ireland with the fewest publications: one paper each.

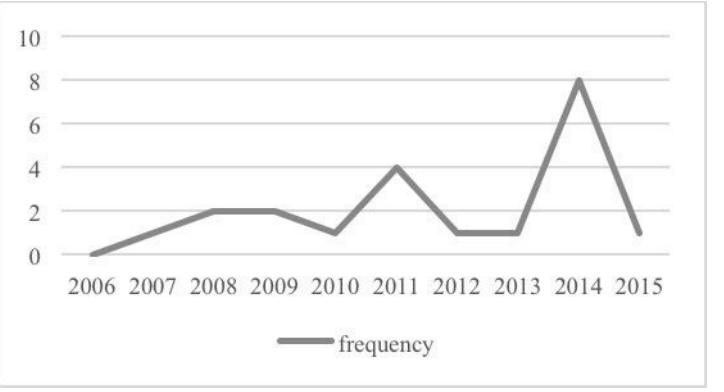
Table 03 : Frequency of publication about the theme by source (2005-2016)

Source title	Frequency
Library and Information Science	05
Proceedings of the Asist Annual Meeting	04
Aslib Proceedings New Information Perspectives	02
ACM International Conference Proceeding Series	01
Communications in Computer and Information Science	01
Information Research	01
Journal of Educational Media and Library Science	01
Journal of the American Society for Information Science and Technology	01
Journal of the Association for Information Science and Technology	01
Lecture Notes in Computer Science Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics	01
Library Quarterly	01
Library and Information Science Research	01
Proceedings of the 5th Information Interaction in Context Symposium IIIX 2014	01

As the table 03 shows, the 21 articles are distributed to 13 journals/proceedings, whithin Library and Information Science, with five articles, Proceedings of the Asist Annual Meeting, with four papers and Aslib Proceedings New Information Perspectives with two papers. Three of this sources are open access.

It was observed the publication year of the articles. Although the frequency of publication was inconstant amount the period of analysis (2006-2015), it is possible to note at the Graphic that the frequency of publication is growing.

Graphic 01: Frequency of publication of the analyzes articles by year (2006-2015)



It was analyzed in what journal the authors are dissemination their research. As showed in the table 02, were founded a set of six journal titles that published the theme in Scopus (2006-2015), the mainly was "Library and information science" with six articles; followed by JASIST with two articles; "Proceedings of the ASIST Annual Meeting" and "Aslib Proceedings: New Information Perspectives" with also two articles.

Finally, we checked the number of citations each article. Almost eigth of the 21 analyzed articles received citations in Scopus as the Table 4 illustrate. Two articles received the major of citations (32 e 20 citations) and coincidentally authors of this articles was the same: Meyers, E. M.; Fisher, K. E. ; Marcoux, E. This two articlesreports a study on adolescents and it focuses on information behavior in the everyday life approach.

Table 4 - Frequency of article citations

Articles cited	Frequency
Meyers, E. M.; Fisher, K. E. ; Marcoux, E. (2009)	32
Meyers, E. M. ; Fisher, K. E.; Marcoux, E.(2007)	20
Shenton, A.K.; 3Hay-Gibson, N. V.(2011a)	06
Beheshti, J.; Alghamdi, M.J.; Cole, C. et al. (2015)	02
Spink, A.; Heinström, J. (2011)	02
Trace, C. B. (2008)	02
Shenton, A. K.; Hay-Gibson, N. V.(2011b)	01
Lu, Y.-L. (2011)	01

Conclusion

This study points some research trends on the information behavior of children during the last ten years, identifying the most productive authors, the most cited papers, and the countries that are more devoted to the topic among those who have published in journals indexed by Scopus. It was noticed that the theme was not frequent in the

journals of the area included in Scopus in the last ten years, but we realize a tendency of growing.

It was observed that most of the papers were published in the United States and Canada. It was also found that two papers stand out as the most cited (Meyers, E. M., Fisher, K. E., Marcoux, E. (2009) and Meyers, E. M., Fisher, K. E., Marcoux, E. (2007), both dealing with the search for information in children´s daily lifeand they address methodological aspects of the research, which shows the interest of the researchers on this topic.

It is expected that this paper will assist those interested on the subject to identify recent trends of research and major reference centers for research on this subject.

References

Beak, Jihee (2015). Where is Children's Voice in KO? *Knowledge Organization*, 42(5) December: 284-89.

Case, Donald O. (2012). *Looking for Information*. 3 ed.Bingley: Emerald.

Chelton, Mary K.; Cool, Colleen (2004). *Youth Information Seeking Behavior II: Context, Theories, Models, and Issues.* Scarecrow Press.

Hjørland, Birger (2014). Theories of knowledge organization – theories of knowledge. In. *Meeting of the German. ISKO* [http://www.academia.edu/3466074/Theories_of_knowledge_organization__theories_of_knowledge]

Hjørland, Birger & Albrechtsen, H. (1995). Toward a new horizon in Information Science: domain-analysis. *Journal of the American Society for Information Science*, 46(6): 400-25.

Silva, Helen Castro.& Silva, A.P. C. 2008. Aplicações dos níveis de leitura para a mediação da leitura com crianças e para a organização da informação. *In Ensino e Pesquisa em Biblioteconomia no Brasil: a emergência de um novo olhar*. Fundepe; Cultura Acadêmica, Marília. Pp 74-86.

Smiraglia, Richard P. (2011). Domain coherence within knowledge organization: people, interacting theoretically, across geopolitical and cultural boundaries. In*MC Exploring interactions of people, places and information*. Fredericton, N.B. Canada: University of New Brunswick/St. Thomas University. [www.cais-acsi.ca/proceedings/2011/73_Smiraglia.pdf].

Guimarães, José Augusto Chaves (2014). Análise de domínio como perspectiva metodológica em organização da informação. *Ciência da informação* 41(1):13-21

Todd, R. J. (2003). Adolescents of the information age: patterns of information seeking and use, and implications for information professionals**.** *School Libraries Worldwide*, 9(2): 27-46.

Wilson, T. D. 1999. Models in information behavior research. *Journal of Documentation*,55(3) June: 249-70.

Analyzed Texts

Agarwal, Naresh Kumar (2014). Use of Touch Devices by Toddlers or Preschoolers: Observations and Findings from a Single-Case Study. *Library and Information Science*. 10

Barriage, Sarah C. (2014). Parental Perceptions of Young Children's Information Behavior Related to Free-Time Activities. *Proceedings of the ASIST Annual Meeting*, 51(1).

Beheshti, Jamshid et al. (2014). Designing an Intervention Tool for Students with Students. *Library and Information Science*, 10.

Beheshti, Jamshid et al. (2015). Tracking Middle School Students' Information Behavior Via Kuhlthau's ISP Model:Temporality. *Journal of the Association for Information Science and Technology* , 66 (5): 943-60.

Bilal, Dania et al. (2012). Learning to Discover: Youth Information Literacy in the "i" Digital Age. *Proceedings of the ASIST Annual Meeting* 49 (1).

Bilal, Dania et al. (2010). Children and Young People with Disabilities: Breaking New Ground and Bridging Information Worlds. *Proceedings of the ASIST Annual Meeting* 47.

Creel, Stacy (2014). Interface Design: The Impact of Images and Catalog Organization on the Information Retrieval of Children Ages Five to Eight while Subject Browsing. *Library and Information Science*, 10.

Dinet, Jerome, Simonnot, Brigitte and Vivian. Robin (2008). La Recherche Collaborative d!Information Sur Internet : Impact Du Lien Affectif Entre Les Jeunes Collaborateurs.

Lai, Ling Ling and Tsai, Chih-Hsin Tsai (2009). Studying Homeschoolers' Information Behavior on Network Forum: Taking the Discussion Board for Chinese Classic Literature as an Example. *Journal of Educational Media and Library Science*, 46 (3): 403-36.

Lu, Ya-Ling (2011). Everyday Hassles and Related Information Behaviour among Youth: A Case Study in Taiwan. *Information Research*, 16 (1).

Mehra, Bharat (2014). Perspectives of Rural Librarians about the Information Behaviors of Children with Special Needs in the Southern and Central Appalachian Region: An Exploratory Study to Develop User-Centered Services. *Library and Information Science*, 10.

Meyers, Eric M., Fisher, Karen E. and Marcoux, Elizabeth (2009). Making Sense of an Information World: The Everydaylife Information Behavior of Preteens. *Library Quarterly*, 79 (3): 301-41.

Meyers, E. M., Fisher, K. E. and Marcoux, E. (2007). Studying the Everyday Information Behavior of Tweens: Notes from the Field. *Library and Information Science Research*, 29 (3): 310-31.

Nicol, Emma (2014). Using Artefacts to Investigate Children's Information Seeking Experiences.

Shenton, Andrew K. andHay-Gibson, Naomi V. (2011). Modelling the Information Behaviour of Children and Young People: More Inspiration from Beyond LIS. *Aslib Proceedings: New Information Perspectives*, 63 (5): 499-516.

Shenton, Andrew K. andHay-Gibson, Naomi V. (2011). Modelling the Information-Seeking Behaviour of Children and Young People: Inspiration from Beyond LIS. *Aslib Proceedings: New Information Perspectives,* 63 (1): 57-75.

Spink, Amanda andHeinström, Jannica (2011). Information Behaviour Development in Early Childhood. *Library and Information Science*, 1.

Terra, Ana Lucia and Sá, Salvina (2013). Strategies to Assess Web Resources Credibility: Results of a Case Study in Primary and Secondary Schools from Portugal. *Communications in Computer and Information Science*, 397.

Trace, Ciaran B. (2008). Resistance and the Underlife: Informal Written Literacies and their Relationship to Human Information Behavior. *Journal of the American Society for Information Science and Technology* 59 (10): 1540-54.

Vanderschantz, Nicholas, Hinze, Annika and Cunningham, Sally (2014). Sometimes the Internet Reads the Question Wrong: Children's Search Strategies & Difficulties. *Proceedings of the ASIST Annual Meeting*, 51(1).

Vanderschantz, Nicholas, Timpany, Claire, Hinze, Anikka and Cunningham, Sally (2014). Family Visits to Libraries and Bookshops: Observations and Implications for Digital Libraries. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Vol. 8839.

Laura Ridenour

Practical Applications of Citation Analysis to Examine Interdisciplinary Knowledge

Abstract

This study examines the overlap of documents in the interdisciplinary area of network theory research, further exploring a dataset examined by Ridenour (2016). Citation analysis is employed to examine the social exchange of authors as demonstrated through paper citation and publication of works in multiple venues of interest to many communities. This study uses citation analysis to detect boundary-spanning authors, or authors who publish in multiple venues that are classified in separate high-level Web of Science (WoS) Essential Science Indicator (ESI) categories. The notion of boundary-spanning is important to knowledge organization, as interdisciplinary research is becoming increasingly common and ways to analyze and classify it are required to make such knowledge discoverable. A total of 75 papers were found to be cited by the two ESI categories examined, and twelve authors contributed to literature published by journals in both these categories.

1 Background

Boundary objects are points of mutual interest where compromise is required in order for members of multiple communities to interact about or through these objects (Bowker and Star 1999). Boundaries must be clearly identified in order for "boundary work," or any work involving the fringes of science, to progress (Szostak 2004 173). In examining boundary objects, Star and Griesemer (1989) emphasized the need to examine both the flow of concepts through their corresponding networks and their social worlds. An author's social world can be seen as rooted in their parent discipline, while an author's willingness to engage in publication outside of their own discipline can be seen as an act of boundary crossing (Pierce 1999). Thus, the application of the concept of the boundary object is well-suited to examining the classification of interdisciplinary works and the boundary-spanning authors of these works, especially in interdisciplinary fields where authors may publish in multiple venues.

Determining subject matter is critical to establishing classification, as classification must be agreed upon. The subject matter of a document is subjective in nature, and the interpretation of it differs between individuals (Hjørland 1992). Researchers in a field will tend to cite documents they determine to be relevant to their topic, creating bodies of literature that are considered pertinent to a field, disregarding other potentially relevant sources. Habermas (1998) created a theoretical framework which could be applied to this phenomenon, in which the acceptability of a speech act is pragmatic. Thus, an individual who wishes to have the truth of their ideas accepted by a given audience will craft their speech acts in a way that they are more likely to be accepted by the audience that the speaker is addressing. In the case of scientific literature, a researcher will choose terminology and cite sources that are likely to be deemed valid by their community. This illustrates Wilson's (1993) view that communication breakdowns occur between specialties, or communities, and not individual researchers.

Furthermore, information outside of a specialty is not considered to be as pertinent when used in support of a hypothesis inside of a field. Zhang and Jacob (2013) proposed a threefold theoretical framework when investigating boundary objects. Their framework addresses physical, epistemological, and virtual dimensions of boundaries; interdisciplinary communication must cross both epistemological and virtual boundaries in order to be effective. The classification of documents for information retrieval creates such virtual boundaries.

Coding of documents for information retrieval systems is done by trained indexers. Highly experienced indexers are assumed to be more likely to code documents similarly than inexperienced indexers (Olson and Wolfram 2008). Indexing is a process of describing a document according to what its aboutness is determined to be. Subject analysis should reflect the epistemological stance of a document, and what the document could mean to multiple communities (Hjørland 1997). Through indexing, such as is done in bibliographic databases, multiple dimensions of a document can be represented for retrieval. However, most knowledge organization systems are built to accommodate the epistemological stance of one discipline, neglecting ways to accommodate interdisciplinary areas of research and potentially resulting in issues of sustainability when knowledge changes (Gnoli 2008).

2 Methodology

In order to examine the interdisciplinary area of network theory research, the following questions research questions were posed:

1. Do these two areas share information exchange demonstrated by co-authorship?
   a. Which authors found in the dataset have published in journals found in both ESI categories?
   b. Which subject categories co-occur in boundary spanning authors' publications?
2. Are documents in the dataset co-cited by both analyzed categories?
   a. How many documents are co-cited in the dataset?

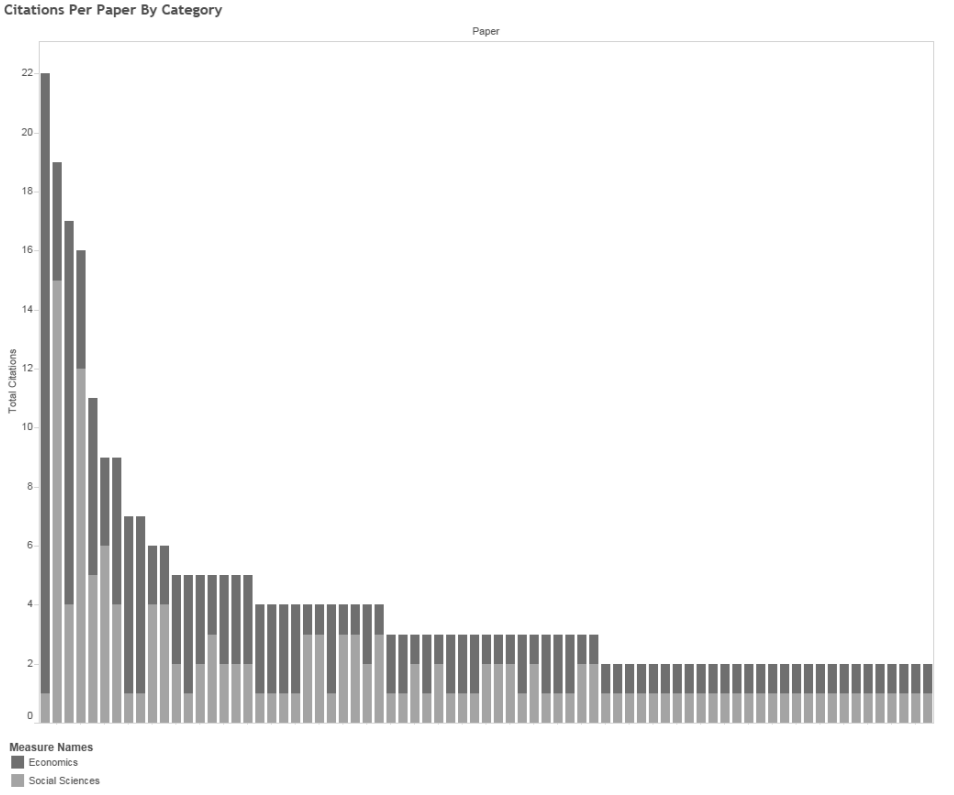
The Web of Science (WoS) was queried for "network theory," and the search was limited to articles and conference papers published in English. The ESI file documenting the top-level WoS classification of journals was used as the basis of clear boundaries for the purposes of this analysis (Thomson Reuters 2013). WoS results were merged into a single file and processed using Sci2 (Sci2 Team 2009). The resulting WoS file and the ESI file were processed to align ESI categories with corresponding journal information in a database. The Map of Science was used to determine proximity of disciplines, as it was created to empirically map connections between disciplines and articles and their likelihood to shift based off of five years of data (Klavans and Boyack 2007). As such, "Social Sciences, General" (SSG) and "Economics & Business" (E&B) Subject Categories were selected to conduct the analysis.

Two .csv files were created from the data, one for each ESI category, and were processed for co-authorship and paper citation in Sci2. This prevents the data from being treated as one file during the analysis by the software, as categories added in for analysis were stripped out when rendered by the algorithms applied to the file. Author names were disambiguated in Sci2, abbreviating all names and entries to the same abbreviated form. Overlap in unique author names and paper titles were calculated using pivot tables in Excel. Counts were extracted, and verified against the original Web of Science data files. Local citation counts were used for both the documents and authors accounted for in the dataset.

3 Results

From the analyses conducted, the number of works cited by each field, the number of commonly cited works, the numbers of authors, and the number of authors contributing to both fields were determined. The ESI category of Economics & Business cited a global total of1567 papers, while Social Sciences cited 1810 papers. A total of 75 papers contained within the local dataset were cited by documents in both ESI categories (Figure 1); however, the overall most cited document was cited twenty-one times by papers in Economics, and only once by a paper in Social Sciences. These 75 papers were cited a total of 3759 times; 1805 total times in Economics and Business, and 1954 times in Social Sciences.

Figure 1: Citations per Paper by ESI Category



Twelve authors wrote thirty-one papers that were cited in both categories in the dataset (Table 1). Five-hundred ninety-five total authors published in Economics, and 440 total authors published in Social Sciences. These counts were not adjusted for the twelve authors who published papers in journals classified in both categories of the dataset.

Table 1: Authors published in both ESI categories

Author	Social Sciences	Economics
Cooren, F	2	3
Peltonen, T	1	2
Callon, M	2	1
Gao, P	2	1
Poell, RF	2	1
Finch, JH	2	1
Denis, JL	1	1
Lee, N	1	1
Miller, NJ	1	1
Monteiro, E	1	1
Muniz, ASG	1	1
Van Der Krogt, FJ	1	1

Most authors in the dataset did not author more than one paper, following Lotka's Law in that very few authors contribute a great number of papers and many authors contribute only one paper (Table 2). This holds true for the authors represented in Table 1.

Table 2: Number of Papers Authored by Each Author'

Num. papers authored	Total authors
5	1
4	2
3	10
2	66
1	944

Of the twelve authors who published in both categories, eleven coauthored articles with others; F. J. Van der Krogt and R. F. Poell shared co-authorship on their papers represented in the dataset, sharing co-authorship with two other authors in Economics, and one in SSG; P. Gao was the only consistently sole author in this group; however, N. Lee, J. H. Finch, T. Peltonen, E. Montiero, and A. S. G. Muniz published both as sole authors and in collaboration with others.

Table 3: Boundary Spanning Authors and Expanded Subject Categories

Name	Social Sciences	Economics
Cooren, F	Communication; Social Sciences Other Topics; Sociology	Business & Economics; SS Other Topics
Peltonen, T	Education & Educational Research	Business & Economics
Callon, M	Business & Economics, Sociology	Business & Economics
Gao, P	Information Science & Library Science	Engineering; Business & Economics; Operations Research & Management Science
Poell, RF; Van Der Krogt FJ	Education & Educational Research	Business & Economics
Denis, JL	Social Sciences Other Topics	Business & Economics; Social Sciences Other Topics
Finch, JH	History & Philosophy of Science; Social Sciences Other Topics	Business & Economics
Lee, N	Sociology	Business & Economics
Miller, NJ	Business & Economics; Social Sciences Other Topics	Business & Economics
Monteiro, E	Information Science & Library Science	Business & Economics
Muniz, ASG	Business & Economics; Environmental Sciences & Ecology; Geography	Business & Economics

Topics published on by these authors varied, but overlap between high-level Subject Categories occurs in both E&B and SSG, as well as between lower level categories. SCs within data extracted from WoS occasionally contained multiple SCs, which provide a greater understanding of the breadth of each field.

4 Discussion

This analysis accounts for the "who" and "what" that can be determined within a WoS dataset to establish mutual discourse as it occurs between two domains investigating a shared phenomenon. The analysis of the dataset examined local citation counts more closely than global, as the shared discourse spanning boundaries contained within an interdisciplinary community helps identify which authors, and in turn which journals, engage in more interdisciplinary acts of publication and citation on the same topic.

WoS assigns Research Areas, Subject Categories, and Essential Science Indicators to various levels of publications. Subject categories were selected for this analysis because they allow for high-level categorical division of journals. SC and ESI are periodically updated and re-released, resulting in possible changes in journal categorical assignment over time. The analysis conducted in this paper utilized a late 2013 version of the file. Though journals are only assigned one subject category in the ESI, multiple assignment of SCs occurs within the database itself.

Limitations to the analysis conducted within this study include sampling, the version of the Web of Science database available, and available space. The sampling of the original Social Sciences ESI data hampered the co-citation analysis: by a peculiarity of the sampling and the way in which Sci2 processes papers for citation. Expanding the paper citation data to include all papers authored in the dataset and abandoning the matched portion of the analysis (295 papers in each category) would address this issue and provide a better understanding of who is contributing to this field in both categories. Expanding the analysis in this way would not affect any analysis of topicality, as a randomly selected, matched corpus sizes are better to analyze for word frequency distributions.

Further expanding the analysis of the epistemic boundaries shared across the entire dataset without attempting to work from a matched corpus would be another approach, and would work better when examining citations as opposed to corpora. Not limiting the analysis to two categories, but examining all ESI categories and comparing co-occurrences of SC assignment would yield a richer analysis.

The Web of Science penalizes authors for self-citations within their own publications, which is reflected in their database. This artifact of the database created a negative local citation count after processing by Sci2. Citations counted as negative were changed to positive for the purpose of analysis based on the assumption that a citation to a document is still a citation, and if a paper is referenced in the dataset it should be reflected in the overall paper counts. Other citation databases, such as Scopus and Google Scholar, may account for self-citation differently and consequently impact the number of times cited for documents.

5 Future Directions

This type of analysis provides a means to examine the dimensionality of conceptual overlap in classificatory systems by integrating the analysis of the social habits of authors (paper citation and publication habits) in domain analysis. Integrating this methodology with co-word and other topical analyses would further enrich the potential of domain analysis of interdisciplinary domains. Triangulation of boundary objects through examining "bursts" in word frequency changes could demonstrate shifts in topicality; though this could be done with the current dataset, a different approach to sampling would be required in order to account for the temporal nature of the dataset. A year-by-year co-citation analysis, such as that of Culnan (1986) could be modified to be applied to multiple categories and show themes in citation patterns across disciplines. Additionally, expanding the analysis of epistemic dimensions of boundaries to more than the two categories analyzed would provide a richer picture of an interdisciplinary topic and its methods, perspectives, and applications.

6 Conclusion

Boundaries in interdisciplinary communication are both epistemological and virtual, and understanding the dimensions of shared boundaries is important for knowledge organization. Analyzing overlapping subjects provides a greater understanding of these boundaries (Table 2). Methodology examined in this paper has potential to determine the epistemological origins of cited works, as well as the authors acting as boundary spanners between multiple disciplines. Additional data would likely increase the overlap in both co-authorship and shared paper citations, as the original dataset contained more papers in the Social Sciences than in Economics. Pinpointing current areas of interdisciplinarity through citation analysis and co-authorship could increase the acceptability of citing extra-disciplinary sources, as well provide a foundation for improving interdisciplinary communication through better classification of interdisciplinary material.

References

Bowker, Geoffrey C. and Susan Leigh Star. (1999). *Sorting Things out: Classification and Its Consequences*. Cambridge, Massachusetts: MIT Press.

Culnan, Mary J. (1986). The Intellectual Development of Management Information Systems, 1972–1982: A Co-Citation Analysis. *Management Science* 32 (2): 156–72.

Gnoli, Claudio (2008). Ten Long-Term Research Questions in Knowledge Organization. *Knowledge Organization*, 35: 137–49.

Habermas, Jürgen (1998). Actions, Speech Acts, Linguistically Mediated Interactions, and the Lifeworld. In *On the Pragmatics of Communication*. Cambridge, Massachusetts: MIT Press. Pp 216-55.

Hjørland, Birger (1992). The concept of 'subject' in information science. *Journal of Documentation,* 48: 172–200.

Hjørland, Birger (1997). Information Seeking and Subject Representation: An Activity-Theoretical Approach to Information Science. In *New Directions in Information Management 34*. Westport, Connecticut: Greenwood Press.

Klavans, Richard and Kevin W. Boyack (2007). Maps of Science: Forecasting Large Trends in Science. In *Places & Spaces: Mapping Science*. [http://scimaps.org].

Olson, Hope A. and Dietmar Wolfram (2008). Syntagmatic Relationships and Indexing Consistency on a Larger Scale. *Journal of Documentation,* 64: 602–15.

Pierce, Sydney J. (1999). Boundary Crossing in Research Literatures as a Measure of Interdisciplinary Information Transfer. *Journal of the American Society for Information Science,* 50: 271–9.

Ridenour, Laura (2016). Boundary Objects: Measuring Gaps and Overlap Between Research Areas. *Knowledge Organization,* 43: 44-55.

Sci2 Team (2009). Science of Science (Sci2) Tool. *Indiana University and SciTech Strategies*. [http://sci2.cns.iu.edu].

Star, Susan Leigh and Griesemer, James R. (1989). Institutional Ecology, 'Translations' and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39. *Social Studies of Science*: 387–420.

Szostak, Rick (2004). Classifying Science: Phenomena, Data, Theory, Method, Practice. *Information Science and Knowledge Management*. Norwell, MA: Springer.

Wilson, Patrick (1993).Communication Efficiency in Research and Development. *Journal of the American Society for Information Science*, 44: 376–82.

Zhang, Guo and Jacob, Elin K. (2013). Understanding Boundaries: Physical, Epistemological and Virtual Dimensions. *Information Research-an International Electronic Journal*, 18: C21.

Wan-Chen Lee

Challenges and Considerations of Adapting Foreign Classification Standards

Abstract

This paper explores the challenges and considerations of adapting foreign classification standards to local contexts by tracing the emergence of the Nippon Decimal Classification (NDC) and the New Classification Scheme for Chinese Libraries (CCL). NDC and CCL were established by adapting the Dewey Decimal Classification (DDC) and other classification standards. Through content analysis of the first edition of the NDC and the CCL, we have drawn three main findings: *contextual identity*, *fit to purpose*, and *search for foundations*. The findings add to the discussion about the relationships between culture and classification, as well as standardization and localization of classification standards.

Introduction

Classification schemes like the *Dewey Decimal Classification* (DDC) are widely used. The advantages of standardization and the cost of developing and maintaining a new classification scheme make adapting such standards a common choice for classifying local collections. However, such classification standards may not meet all local needs. This paper explores the challenges and considerations of adapting foreign classification standards by tracing the emergence of the *Nippon Decimal Classification* (NDC) and *the New Classification Scheme of Chinese Libraries* (CCL). NDC and CCL are the main classification standards in Japan and Taiwan respectively. Both schemes were established by adapting foreign classification standards, drawing primarily from DDC (see Table 1). Through content analysis of the 1st edition of NDC and CCL, we describe the context and motivation for the tandem creation of these classification schemes in 1929.

Table 1: Comparison table of classes in CCL, DDC, EC, and NDC (Liu, 1929; Mori, 1929)

Expansive Classification	NDC (1st edition)	DDC (12th edition)	CCL (1st edition)
A General works	000 General works	000 General works	000 General works
B Philosophy and religion	100 Philosophy and religion	100 Philosophy	100 Philosophy
E Historical sciences	200 Historical sciences	200 Religion	200 Religion
H Social sciences	300 Social sciences	300 Social sciences	300 Natural sciences
L Science and arts	400 Natural sciences	400 Philosophy	400 Applied sciences
	500 Technology	500 Pure science	500 Social sciences
	600 Productive arts	600 Useful arts	600 History and geography
	700 Fine arts	700 Fine arts	--
X Language	800 Language	800 Literature	800 Language
Y Literature	900 Literature	900 History	900 Arts

Previous research exists that addresses the adaptation of foreign classification schemes. Mcllwaine (1997) reviews the development of the *Universal Decimal Classification* (UDC), including how it was derived from DDC. Idrees (2012) interviews librarians applying DDC, LCC, or UDC to their rich Islamic collections. This research shows that classification standards have limited space for Islamic knowledge. This requires local expansions of the schemes. Batty (1976) looks at the worldwide adaption of DDC, pointing to challenges of localization. These include modifying cultural sensitive classes (e.g. religion, history), and reorganizing and relocating notations. Another line of research highlights the differences of classification schemes in different cultures. Lee (2012) examines the *Seven Epitomes* (*Qilue*), the first Chinese library catalog, and describes its epistemic foundations, which can be contrasted with counterparts based in other cultures. Chang (2012) compares the literature class in the DDC and the literature class in CCL.

Background

Reviewing the history of the NDC and CCL, we find that the editors of NDC and CCL share the concern of wanting to import from the U.S. and the conflicting goal of serving local library communities' needs. Theydescribe the deficiencies of existing schemes, and identify culturally sensitive classes and the need of updating the knowledge represented in the schemes. The elements of the DDC that were altered in the NDC and CCL mark what was essential for meeting the needs of Japan and China. These considerations and challenges showcase the tension between global and local knowledge organization, and make evident cultural influences on classification schemes. These cases contribute to our understanding of factors that surface in the design and adaption of global standards.

NDC was created by Kiyoshi Mori and published in 1929. His design addressed a number of goals. First, a national bibliographic classification standard was needed to support inter-library collaboration. In addition, he prioritized materials related to Japan and applied a universal scheme for books in different languages [1]. To achieve his goals, Mori extracted elements from several schemes. He adapted the structure and notation of DDC, and the class sequence of Cutter's Expansive Classification (EC). He also referenced Chinese classifications for the approach of adapting DDC notations (Mori, 1929).

The CCL was created by Kwoh-Chuin Liu and published in 1929, when Chinese classificationists realized the deficiency of existing Chinese classifications, and started proposing approaches to organize new knowledge and materials in different languages. A main goal of CCL was to use one scheme for books published in all eras. It was based on the collection of the University of Nanking Library, and made reference to Chinese classifications and Western classifications, drawing primarily from the DDC (Liu, 1929).

Methods

To investigate the challenges and considerations of adapting foreign classification standards, we conducted content analysis on the 1st editions of the NDC and the CCL, including the forewords, prefaces, and introductions of the two schemes. The forewords, prefaces, and introductions include the editors' motivations and descriptions of the context for developing the schemes. We analyzed these documents through open coding, and present prominent themes (Berg, 1989; Lindlof & Taylor, 2011). To see different perspectives and to avoid hearing only a single voice embedded in a single document, we also used secondary literature describing the context of classification development in 1929 Japan and China. Through source triangulation, we gain a better understanding of the context in which NDC and CCL were first developed.

Findings

We have three main findings which manifest as themes common to the genesis of both the NDC and CCL. The first finding relates to *contextual identity*. The contexts of developing the NDC and CCL are very similar. The lack of national classification standard, and the existing schemes being unsatisfactory led to the development of both schemes. The second finding relates to *fit to purpose*. Both the NDC and CCL construct culturally specific orders, so some things stay and some things are removed when they adapt foreign schemes. The NDC addresses the needs of shelving books in different languages together and prioritizing Japan related books. The CCL emphasizes using one scheme for both old and new books, and the flexibility of updating new knowledge to the scheme. The third finding is related to the *search for foundations.* The editors of the NDC and CCL went through thorough consideration of other schemes before developing these two schemes.

Contextual identity: lack of a classification standard and unsatisfactory schemes

Liu (1929) and Mori (1929) describe the library communities of China and Japan respectively. Some common themes surface. In both cases, there was no national classification standard. Liu and Mori recognized the negative impacts of the lack of classification standard on users, librarians, the development of libraries, and library collaboration. Standardization was regarded as a common yet difficult goal.

Existing schemes were considered unsatisfactory in both cases. Mori considers Japanese classification schemes published before 1929 to be superficial imitations of DDC. They are also not easy to use. Since most of the existing classification schemes lacked indexes, catalogers using the same scheme might assign classes inconsistently. Liu criticizes *Si kuquan shu*, a Chinese classification widely used at the time. According to Liu, *Si kuquan shu* has limited scope. In addition, Liu disagrees with its characteristics of division, which are generally format, and then discipline. He proposes division by discipline, and then by format, and adding divisions by geography, time, and language when appropriate. Jiang (1957), in his book reviewing the historical

development of Chinese classification schemes, adds to Liu's criticism of *Si ku quan shu*. Jiang thinks the failings of *Si ku quan shu* are: that it is influenced by Confucianism, it fails to prioritize philosophical literature, it arbitrarily merges categories with smaller bodies of literature, and it is inconsistent in its characteristics of division. For instance, in the History category, books are divided by both format (e.g. official history written in annals-biography style, edicts, and annals) and genre (e.g. biography, geography, and historical review). We can find books about history dispersed in different classes due to inconsistent principles of division.

Another aspect of contextual identity that surfaces is a desire to serve the academy. Liu noticed that over time, some disciplines expand or deepen research scopes, or change research methods. Liu predicted that with the new knowledge introduced from the West, the rapidly evolving professoriate would prefer specific classification, and would generate needs that older classification schemes like *Si kuquan shu* cannot satisfy. He called for reform of classification schemes to meet the needs of contemporary academy.

Fit to purpose: identifying goals

While developing a national classification standard was a common goal of both China and Japan around 1929, the CCL and NDC focus on similar but distinct goals. Mori's motivations for developing the NDC were the pursuit of a scheme that fully realizes the spirit and functions of the DDC, and (1) enables shelving books in different languages and formats together, and (2) prioritizes Japan-related books. In the case of the CCL, Liu's motivation was establishing a single classification scheme for books published in all eras. He wanted to change the practice of using different classification schemes for old books and new books, because the distinction of the old and the new is not clearly defined.

Jiang (1957) reviews the development of Chinese bibliographic classifications from the first Chinese catalog: the*Seven Epitomes* (*Qilue*) to schemes developed around his time. He categorizes Chinese classificationists' attitudes about the DDC into 3 periods, 4 systems, and 7 groups (see Table 2), and matches some Chinese classification schemes with his categories. The CCL is placed in the Expand and Modify DDC Group. Fan (2011) examines and agrees with Jiangwhen reviewing the development of Chinese classifications in the 20th century. While this categorization was created for Chinese classificationists' attitudes, could it describe Mori's attitude and the NDC? Since Mori (1929, p.12-13) references two Chinese classifications: Doo's *Universal Classification Scheme* (1925) and Wang's *Zhong wai tu shu tong yi fen lei fa* (1928) for common emphasis on shelving books in different languages together, we can argue that the NDC can be placed in the same category with these schemes, which is the Universal System. The distinctness of the goals addressed in the CCL and NDC are highlighted through Jiang's categorization.

Table 2: Jiang's (1957) categorization of Chinese classificationists' attitudes about the DDC.

Periods	Systems and Groups
Chaotic period	Tradition group: modifies *Si ku quan shu* and continues using it
	Revolution group: breaks the structure of *Si ku quan shu* and builds a new system
	Compromise group (Parallel of the New and the Old): uses *Si ku quan shu* for old books and Western classification schemes for new books
Importing Western Classification Schemes Period	Expand and Add to DDC group
Creative Period	Mixing the New and the Old System -- Imitate DDC group
	Mixing the New and the Old System -- Expand and Modify DDC group -- *CCL*
	Mixing the New and the Old System -- Adopt [the structure and decimal notations of] DDC group
	Universal System: Shelving books in different languages together -- *NDC*

Fit to purpose: adaptation

Despite having distinct goals, Mori and Liu have a common attitude about the DDC, which is adaptation, i.e., using DDC with localized modifications. In the preface, Mori addresses the necessity of having a Japan-oriented classification standard, and points out inevitable issues of fully adopting Western classification schemes.

Japan's situation does not allow full adoption of foreign systems. Even if Dewey's D.C. is a global standard, with full adoption, some difficulties in religion, law, literature, and history classes would arise. In addition, it would be difficult to meet the need of prioritizing Japan-related literature. (Mori, 1929, p.14)

From Mori's comments on Doo's imitation and modification of DDC, we can see that he supported adapting the DDC with modifications. He considers adding Roman characters before DDC notation as an approach to gain flexibility of class sequence and arrangement without destructing the functions of the scheme and the index. Mori (1929) explicitly addresses his concerns about using the DDC without modifications. He lists specific classes he considers especially problematic were he to adopt the DDC, including religion, law, and history (p.14). As a result, Mori adapts the classes of Cutter's EC into the NDC (see Table 1), including modifying classes related to Japan using mnemonic devices. For instance, materials for children are distinguished by adding the letter "J" in front of class numbers. Besides the modifications targeting particular classes, Mori prioritizes local needs. He asks catalogers to classify from a practical perspective by taking users, local context, local concern, and the nature and mission of the library into account (p.21-22).

Liu also emphasizes the necessity of adapting the DDC.

Considering the differences of research scopes, methods, and questions between China and the West, it is difficult to adopt [with no modification] Western

classification schemes. Forcing ourselves to imitate the West is like cutting the feet to fit the shoes. (Liu, 1929, p.1)

Liu not only addresses the inappropriateness of fully adopting Western classification schemes with no localized modification, but he also crafts a culturally appropriate sequence of classes (see Table 1 above). In the introduction of the CCL, Liu walks readers through his rationale of class sequence by referencing traditional Chinese classifications and elaborating his interpretation of the relationships between classes. Another modification needed by Chinese classification schemes is a standardized system of arranging Chinese characters that supports book arrangement. According to Wan (1929), who examined the issues of arranging and retrieving Chinese characters, there were 40 systems of arranging Chinese characters at the time. The complexity of the Chinese language leads to various arrangements (e.g. arrangement by sound, and form: strokes, parts of a character, etc.). This adds a roadblock to the development of a Chinese classification standard. Liu (1929) mentions his debate between using a combination of Roman alphabet, numbers, and Zhuyin, a system of phonetic notation of Chinese, and using number exclusively as notation (p.5).

Search for foundations: the foundations of NDC and CCL

Another theme surfaces from tracing the classification schemes Mori and Liu reference. Through examining the source materials of the NDC and CCL, we see references that reflect the rationale for designing the schemes. Liu established the CCL based on the collection of the University of Nanking Library. Classes are either drawn from traditional Chinese classification schemes and catalogs like *Han shu yi wen zhi*, *Shu mu da wen*, and *Si ku quan shu*, or adapted from Western classification schemes, namely Library of Congress Classification, DDC, EC, and Brown's Subject Classification, to represent new disciplines introduced from the West. Liu also gives credit to some special classifications like Huang's *Jin shi shu mu* (1926), and other Chinese classification schemes developed around the same period, including *Doo's Universal Classification* (1925).

Mori adapts the structure, notation system, and the mnemonic devices of the DDC. For instance, NDC's notation for Japan is 1. Thus, 210 is history of Japan. The class sequence is based on the EC (see Table 1), because Mori (1929) regards the class sequence of the EC to be the most theoretical of the western schemes (p.15). Having the common goal of shelving books in different languages together while prioritizing local materials, Mori adapts Wang's (1928) approach of adding notation in front of DDC numbers. At the time of Mori, books in the same class were arranged by accession order. Mori mentions that Miki (1928) proposed an approach imitating Cutter's author numbers, and thinks it will become a standard. Besides listing the standards referenced, Mori explains why he is inspired by Chinese classifications. He sees common challenges both Japanese and Chinese libraries face in developing

classification standard. This, in turn, explains the number of common themes between the NDC and CCL.

Conclusion

This study presents challenges and considerations of adapting foreign classification standards and provides libraries considering adapting classification standards a list of potential issues. It also points to further research questions. Do our concepts of *contextual identity*, *fit to purpose*, and *search for foundations* appear in other contexts? How might consideration of *contextual identity*, *fit to purpose*, and *search for foundations* be used in the design of culturally sensitive classification standards? How are the challenges and considerations of adapting foreign classification standards identified in 1929 similar to or different from current situations? Exploring these questions can help us identify whether these findings are common themes in similar cases in different geographic regions and time periods. Future research might include an extended comparison of classification schemes adapted from the DDC to discover whether other possible cultural considerations have arisen (c.f., Batty, 1976; Idrees, 2012; Mcllwaine, 1997). In addition, a more comprehensive picture of the challenges of adapting foreign classification standards could be presented by investigating perspectives from other cases similar to the NDC and CCL. These findings broaden our understanding of the relationship between culture and classification, as well as standardization and localization of classification standards.

Note

[1] Before NDC was widely used, most Japanese classification schemes used one classification table for Japanese and Chinese books, and another table for books written in other languages. The two sets of books were shelved separately. One of the main reasons of differentiation was binding. Mori considered it problematic to divide collections by language and binding. He developed NDC to address issues concerning books written in two or more languages, and Japanese books that are about foreign literature (Mori, 1929, p.13-14).

References

Ban, Gu, Yan, Shigu & Yao, Zhenzong (1963). *Han shu yi wen zhi* [Treatise on Literature in the Book of Han]. Xianggang: Taiping shu ju.

Batty, David (1976). Dewey abroad: The international use of the Dewey Decimal Classification. *The Quarterly Journal of the Library of Congress*, 33(4), 300–10.

Berg, Bruce L. (1989). An introduction to content analysis. In *Qualitative Research Methods for the Social Sciences*. Boston: Allyn and Bacon. Pp.105-27.

Brown, James D. (1914). *Subject classification: With tables, indexes, etc., for the subdivision of subjects* (2nd ed. rev.). London: Grafton & Co.

Chang, Yu-Wei. (2012). Duwei shijin fenlei fa yu zhongwen tushu fenlei fa zhi wenxue lei fenlei bijiao yanjiu. [A comparison study of the literature class in Dewey Decimal Classification and the New Classification Scheme for Chinese Libraries]. *Journal of Library and Information Science Research, 6*(2), 115-37.

Cutter, Charles A. (1891). *Expansive classification*. Boston: Cutter.

Dewey, Melvil& Fellows, Dorcas (1927). *Decimal clasification and relativ index: For libraries and personal use in arranjing for immediate reference, books, pamflets, clippings, pictures, manuscript notes and other material* (Ed. 12., rev. and enl. under direction of Dorcas Fellows, editor. Semi-centennial ed.). Lake Placid Club, N.Y.: Forest Press.

Doo, Dingyou (1925). *Tu shu fen lei fa = Doo's Universal Classification*. Shanghai: Shanghai Library Association.

Fan, Fan (2011). *Minguo shi qi tu shu guan xue zhu zuo chu ban yu xue shu chuan cheng* [Scholarly publication and transfer of library studies in the Republic of China era]. 1st edition. Beijing: Guo jia tu shu guan chu ban she.

Huang, Liyou (1926). *Jin shi shu mu: [10 juan]: Fu mei shu shu lei* [Bibliography of epigraphy: [10 volumes]: including arts category]. Beijing: Huang shi wan bei guan.

Idrees, Haroon (2012). Library classification systems and organization of Islamic knowledge. *Library Resources & Technical Services*, *56*(3), 171–82.

Jiang, Yuanqing (1957). Zhongguo tu shu fen lei zhi yan ge. [The development of Chinese bibliographic classifications] Taipei: Taiwan Zhonghua shu ju.

Lee, Hur-Li (2012). Epistemic foundation of bibliographic classification in early China: A Ru classicist perspective. *Journal of Documentation, 68*(3), 378–401.

Library of Congress (1910). *Library of Congress Classification*. Washington, GovtprintOff.

Lindlof, Thomas R. & Taylor, Bryan C. (2011). Sensemaking: Qualitative data analysis and interpretation. In *Qualitative Communication Research Methods*. (3rd ed.). Thousand Oaks, California: SAGE.

Liu, Kwoh-Chuin (1929). *Zhongguo tu shu fen lei fa = A System of Book Classification for Chinese Libraries.*Nanking: Jinling da xue tu shu guan.

Mcllwaine, Ia C. (1997). The Universal Decimal Classification: Some factors concerning its origins, development, and influence. *Journal of the American Society for Information Science*, *48*(4), 331–39.

Miki, Choji (1928). *Romaji Nihon choshahyo: 2 sujishiki* [Japan author marks of Romanization: 2 number style]. *Toshokan kenkyu*, *1*(4).

Mori, Kiyoshi (1929). *Nihon jisshin bunruiho: Wakan yosho kyoyo bunruihyo oyobi sakuin* [Nippon Decimal Classification scheme: Classification table and index for Japanese, Chinese, and foreign books]. Osaka: Mamiya Shoten.

Wan, Guoding (1929). Han zi pai jian wen ti [The issues of arranging and retrieving Chinese characters].*Tu shu guan xue ji kan*, *3*(1-2), 109-22.

Wang, Yunwu (1928). *Zhong wai tu shu tong yi fen lei fa* [System for the uniform classification of Chinese and foreign books]. (Chu ban.). Shanghai: Shang wu.

Wen Yuan Ge (China). (1987). *Si ku quan shu* [the Complete Library of the Four Treasuries]. (Di 1 ban.). Shanghai: Shanghai gu ji chu ban she.

Zhang, Zhidong. (1936). *Shu mu da wen* [Answers to inquiries into bibliography]. Shanghai: Shang wu yin shu guan.

Carlos H. Marcondes and Maria Luiza de Almeida Campos

Searching for a Methodology to Define Culturally Relevant Relationships between Digital Collections in Archives, Libraries and Museums

Abstract

The emergence of Semantic Web and LOD - Linked Open Data - technologies enable that digital objects representing the holdings of archives, libraries and museums collections be semantically interlinked throughout the Web. What are the different types of cultural relevant relationships that may exists between digital objects of collections in archives, libraries and museums throughout the Web? How discover, organize and formalize these relationships to be used by curators in LOD applications? A methodology to analyze the holdings of archives, libraries and museums is proposed based on onomasiologic perspective. Such methodology is applied to a hypothetical competency question that might be proposed by a digital curator. Results indicated that conceptual models such as FRBR, CIDOC CRM and EDM may provide a rich repertoire of semantic relationships that may be used in LOD applications to interlink collections in heritage institutions.

1. Introduction

For centuries institutions as archives, libraries and museums have the mission of curators of the memory and cultural heritage of societies where they are inserted in. Although their common mission each of these institutions are specialized in different facets of this legacy, thus creating specific methodological procedures and value added criteria concerning the memory and cultural heritage holdings each kind of institution is responsible for. The emergence of Semantic Web and LOD - Linked Open Data - technologies enable that digital objects representing the holdings of archives, libraries and museums collections be semantically interlinked throughout the Web thus creating unexpected meaning and contextual networks, empowering their synergies, their complementarities, their educational and curatorial potentialities.

A digital curator, with the aim of a formalized vocabulary of such relationships, could produce culturally rich virtual expositions, accessible by anybody, which explores the increasing number of memory and cultural heritage collections now available throughout the Web.Such technologies enable that a digital curator discovers and makes sense of,or propose new, unforeseen, cross-domain semantic relationships between digital cultural heritage objects.

> A curator (from Latin *curare* meaning "take care") is a manager or overseer. Traditionally, a curator or keeper of a cultural heritage institution (i.e., gallery, museum or archive) is a content specialist responsible for an institution's collections and involved with the interpretation of heritage material. The object of a traditional curator's concern necessarily involves tangible objects of some sort, whether it is artwork, collectibles, historic items or scientific collections. (CURATOR, Wikipedia).

Cultural heritage objects hold different types of relationships. A film may be inspirited in a literary work, a work of art illustrates an edition of a literary work famous painters who created the scenario and costumes of ballets or assembly plays.

There are different versions of Da Vinci's Mona Lisa made by artists such as Marcel Duchamp, Andy Warhol and Fernando Botero. In KO literature such relationships are the associative relationships.

However, due to a long time tradition of independent, self-contained collections, the adoption of different standards, the possibilities of interoperability between such diverse collections are beyond technological issues. In recent years the Documentation as a domain has used conceptual models to identify, make explicit, standardize, and semantically integrate its objects. This paper aims at investigating methodological approaches that enables the discovery of such culturally relevant semantic relationships. Such semantic relationships are no more within the scope of a specific collection's domain but between the content of the digital representations of archives, libraries and museum objects, certainly with holdings in different subject domains.

LOD technologies enable cross-searching such collections. Thus such technologies enable digital objects of different collections to be mobilized by curators in specific domains as Art, Culture, Literature, History, Journalism, Education, Scientific Divulgation, etc., in order to create a new, unique, curated, digital resource, as a virtual exhibition. Curatorial work is multidisciplinary, hard to delimit, personal, authorial. The exploitation of curatorial potentialities of such LOD resourcescan be considered domain specific or problem oriented. Consider, for example, exhibitions as "Leonardo Da Vinci: The Mechanics of a Genius" [1], or "Human Bodies The Exhibition"[2].

This paper communicates some initial findings and insights of an ongoing research that aims at addressing the following questions. From a curator standpoint what are the different types of cultural relevant relationships that may exists between digital objects of collections in archives, libraries and museums throughout the Web? How to discover, organize and formalize these relationships in a vocabulary to be used by curators in LOD applications in Culture? What should be the role of curatorial activities, what should be the role of knowledge organization activities, related to publishing LOD datasets? What Methodology should lead to such aims?

The specific aim of this paper is to discusses methodological approaches to knowledge discovery, organization and representation in order to identify, define and formalize cultural relevant relationships that may exists between digital objects of collections in archives, libraries and museums aimed to be used in LOD applications.

This paper is organized as follows: introductory section, section 2 revises theoretical approaches in KO to classify relations and make methodological assumptions about how to identify, define and formalize cultural relevant relationships. Section 3 proposes a case in which a digital curator with the task of organize a thematic virtual exhibition needs to find throughout the Web relations between themes relevant to develop such a virtual exhibition. Section 4 discuss how and where - from what sources - one can retrieve and reuse a repertoire of semantic relationships and proposes new curated relationships. Then, the results are evaluated and draws some initial conclusions.

2. Theoretical bases

Information Science, Knowledge organization and Terminology literature exhaustively discusses the nature and meaning of associative relations.Although the vast amount of literature these relations are still controversial.

Within Information Science the so called associative relations rise within the development of thesaurus in 1970 decade. Unlike hierarchical relations, who are logical and abstraction relations, associative relations, including whole-part relations, are *ontic* since they occur between objects. Dalhberg (1978) claims that the establishment of associative relations may depend on specific contexts. Although not adding further details she suggests bases to identify such relations.

Terminological Theory (WÜSTER, 1977) classifies relation in logical and ontic. Within the further there are two subclasses, contact and causality relations. Wüster deeply analyzes these relations. Within the former are whole-part and associative relations. Contac relations are the most important subclass of ontic relations. They are self explanatory from their species, coordination and concatenation relations. The main coordination relation is whole-part. This type of relation may occur between a whole and its parts or between the parts of a whole. It is considered as a relation occurring within a specific spatial location, stressing the fact that a whole and its parts exist simultaneously in a spatial (and time) location. Coordination relations as mentioned by Wüster are the inclusion and integration relations. Concatenation relations are conditioned by time and have as its subclasses precedence and succession relations.

Causality relations are parent relations and have two subclasses: ascending or descending parent relations between two different generations; and phase relations expressing different phases of an individual or substance life. Accordingly phase relations are classified in phylogenic, ontogenic and substance-substance relations (Sales, Campos & Gomes, 2008).

Within Terminology Sager (1990) proposes another classificatory schema to relations, namely: generic, whole-part, polyvalent and complex relations. Generic relations are equivalent to logic relations. Whole-part relations are the same as those proposed by Wüster. Polyvalent relations are equivalent to polyhierarchical relations found in ISO 2788 (1986) and in Aitchison (1987) when discussing hierarchical relations.

Been (2008, p. 156) claims that “Associative relations come into a variety of flavors”. Peters and Weller (2008 , 101) claim that “They are unspecified connections of concepts that can have any kind of relation”, “Thesauri make use of (entirely undifferentiated) associative relations”, and that “In addition, associative relations can be split into a diversity of domain-dependent, specified paradigmatic relations”.

Also associative relations are sometimes defined by exclusion of hierarchical or paradigmatic relationships. According to the EuroVoc Thesaurus: “The associative relationship is a relationship between two concepts which do not belong to the same

hierarchical structure, although they have semantic or contextual similarities". Marcia Lei Zeng (2005) also defines them in a similar way: "This relationship covers associations between terms that are neither equivalent nor hierarchical, yet the terms are semantically or conceptually associated".

Associative relationships in Knowledge Organization literature are thus dubious and semantically inaccurate (imprecise). They are also highly context dependents. In order to be useful and enable computational inferences these relationships must be specialized and have a clear, unequivocal and formal semantics.

3. Methodological assumptions

According to ICOMOS (2002) "Cultural Heritage is an expression of the ways of living developed by a community and passed on from generation to generation, including customs, practices, places, objects, artistic expressions and values. Cultural Heritage is often expressed as either Intangible or Tangible Cultural Heritage".

UNESCO, within the scope of Cultural Heritage sites, defines "World Heritage is the designation for places on Earth that are of outstanding universal value to humanity.

A tentative definition of culturally relevant relationships claims that those are relationships holding between representation of cultural heritage objects - digital objects - in archives, libraries and museum collections that are supported, cited, mentioned, discussed, exposed by a socio-cultural event, that is, cited in a publication, in a conference, in a lecture, in a film, in a law, or in any other socio-cultural event.

Heritage objects from collections in archives, libraries and museums all have intrinsic cultural value recognized and attributed by curatorial activities developed within the scope of these institutions. Thus the cultural value of such objects is not an essential property, instead, is socially attributed by these institutions. Socially attributed properties turn out to be incorporated to what such objects *are*, to their essence. Searle (1995) presents an exhaustive discussion about the process of social attribution of properties to objects, especially to artifacts, that is insightful to the understanding of how heritage institutions as archives, libraries and museums attribute cultural values to objects, to what is the social and cultural value of an archive, library or museum object.

Thus the socially attribution of social, cultural and heritage values to objects is a mission and a mandate of these institutions. These facts have consequences on the way these objects are represented as digital and information objects. This deserves further discussion on the values attributed by the society to archival documents, publications and museum objects; however this is beyond the limits of this paper.

Identify culturally relevant relationships is within the role of the cultural curator. Their job is to interpret, attribute new meanings, re-contextualize, etc., the different kinds of cultural manifestations. If a specific interpretation of a cultural manifestation will become culturally relevant or not, it will depend upon its recognition and

acceptance as one of the accepted interpretation theories for that cultural manifestation. It seems to be a phenomenon similar to citation in scientific communication.

The search for a methodology to identify, define and formalize such relationships is based on the assumptions presented in sequel.

1 Cultural heritage objects belonging to collections in archives, libraries and museums have an "intrinsic" cultural value, attributed by (Searle, 1995) local curators.
2 Due to the subjective character of cultural interpretations, which are individual and authorial we opted not to act within the scope of cultural curatorial activities, which are specific of domains as previously mentioned.
3 Any possible relationship holding between two cultural heritage objects is a consequence of what these object are or were, of its creation, i.e., of its (social) life.

These assumptions suggest the adoption of a methodology which enables the discovery of what a cultural object is or was, of all its possible attributes and relationships, in all the contexts in which such an object has existed during all its life span. Following the previous proposal of Wüster (1978) there may be two types of relationships, logic and ontic. Logic relationship are epistemic, they depend on how someone knows and classifies an object; therefore they are abstract, occurring between concepts. Instead ontic relationships depend on what an object is or was, the occur between objects that are contiguous in space and/or time, ie. have had contact during their life span. These last relationships are fundamental to the methodological approach we are proposing.

A methodology with such a focus is based on the onomasiologic perspective. According to this perspective Language has as one of its functions to build the referent object to a speech community. Language terms reflect and agreed conception of a social object. Accordingly, Language terms reflect the building of a concept, that is, by selecting and highlighting, but also by hiding, different aspects of such a social object. The onomasiologic perspective aims at capturing the "version" of an object that corresponds to how a speech community interacts with such an object.

This proposal aims at enlarging and maximizing the potential cultural value of digital heritage objects. Despite the fact that onomasiologic perspective was developed to other aims it may be useful in describing cultural objects in order to enable possible links between.

Thus, following the methodological perspective outlined, ccultural objects must be exhaustively described, discovering and/or making explicit all its attributes and theontic relationships between them. If, in doing this, we make explicit and standardize the descriptive items of each object, we are thus potentializing the use of these descriptive items as anchors to semantic links using LOD technologies This may

enlarge interoperability between different digital collections in order to answer the competency question posed by curators and have their cultural potential realized.

4. Case analysis

In sequel is presented a hypothetical competency question that serves to guide the inquiry proposed by this paper.

Suppose a curator is in charge of the development of a virtual exhibition dedicate to Brazilian 19th and early 20th century writer, Machado de Assis, and possible relations between his literary style and European literature. He/she needs to be aware of every digital resource about the writer, of his life, of social and literary, historical and social context in which Machado wrote his romances, available throughout the Web and formatted according to LOD technologies. Suppose this curator is using a LOD web browser and virtual exhibition editor that enables the navigation and recovery of records from different LOD information resources and the recording of semantic links established ("curated") by the curator.

What should be the relationships between the digital representations of two cultural heritage objects beyond associative relationships, mere "semantically or conceptually associated" relationships?

As domain specific, the culturally relevant relationships, as defined, are excluded of the analysis, according to the assumption 1, ontic relationships must be examined. In KO literature these relations are also known as syntagmatic relations, as they are occur in space-time realms, independent of how it is organized according to specific paradigms (paradigmatic relations).

Conceptual models such as the FRBR (1998), CRM CIDOC (2013) or Europeana Data Model (2016) exhaustively analyze digital objects from collections in archives, libraries and museums according to onomasiologic perspective and provide a semantically rich and exhaustive inventory of such relationships. The following is a list of some relationships drawn from these conceptual models.

Digital representations of two cultural heritage objects are related if, for ex.:

- both are assigned the same subject, *frbr:has_as_subject*;
- one has as subject the other, *frbr:has_as_subject*;
- both have the same creator, *frbr:is_created_by*;
- both have the same producer, *frbr:is_produced_by*;
- both are or were owned by the same person or corporate body, *frbr:is_owned_by*;
- both share the same context, *frbr:context* for the work;
- one was influenced by the other, *crm:P15* was influenced by;
- one object is illustrated by the other, *crm:P65 shows visual item* (*is shown by*);
- one object is inspired on the other, edm:isDerivativeOf.

If relationships between digital objects in heritage institutions collections and other Web resources are also considered, as for example, between a digital object and an entry in an Authority file or in Wikipedia/DBpedia, additional relationships must be considered as, for example:

- one object *edm: Has Met* an authority, meaning for example that some object was used by a person in some event.

A search for Machado de Assis in LOD datasets with content provided by Brazilian heritage institutions may retrieve records and full-text of different books written by Machado de Assis, or books of different authors analyzing his works, from library collections. Also may retrieve records and digital images of different photos of Machado or his portraits drawn by different artists and published in newspapers, from archives collections. He/she may also retrieve records and digital images from a writing desk with an ink-glass and a pen that belonged to Machado, and with which he wrote some of his works; these objects belong to the collection of Brazilin Literary Academy's museum. Also a video may be retrieved from a digital film library, of an adaptation of Machado masterpiece, Capitu, to Brazilian television.

The authority and bibliographic records retrieved, about Machado de Assis as author or about his works from library catalogs may inform he is a Brazilian 19th and early 20th century author associated with Realism literary movement, or may frbr:has_as_subjectRealism. Also Machado de Assis's entry in the DBpedia informs he is a Brazilian author associated with Realism literary movement; there is an additional entry in DBpedia, "Realism in Brazil" that outlines the role of Machado as the main expression of this literary movement in Brazil. These entries refer to another, "LiteraryRealism", which list the main authors associated with this literary movement in the Americas and Europe.

The relationships founded by the curator are provided by LOD information resources from Brazilian heritage institutions in association with LOD information provided by DBpedia.

Lehmann, Schüppel & Auer (2007) describe a Web application, the RelFinder [3], which enables graphic visualization of the direct and indirect relationships, found in RDF/LOD records, between two concepts typed by a user. This application uses DBpedia RDF/LOD records. Relfinder application enables one to foresee the operation of a Web virtual exhibition editor to be used by digital curators.

As a digital curator use such a virtual exhibition editor, he/she can establish the relationship (and thus records it), creating a semantic link with the semantic crm:P15 was influenced by, between an authority record of Machado de Assis and the DBpedia entry "Literary Realism", thus establishing that Machado de Assis as author was influenced by European Realism literary movement. He/she can also establish a new, *culturally relevant relationship* between Machado and Magic realism [4] or between Machado and the Portuguese authors as Eça de Queiroz and José Saramago.

5. Concluding remarks

Conceptual models such as FRBR, CIDOC CRM and EDM provide a semantically rich repertoire of relationships that can be used in LOD applications to connect digital objects from collections in archives, libraries and museums. These relationships can be the starting point to more specialized, curated, *culturally relevant relationships*. The competency question proposed could be answered by LOD records from collections in archives, libraries and museums, complemented by records form DBpedia.

The curatorial work in Culture is interested in creators and their works, in cultural movements, tendencies and influences. However archives, libraries and museum hold objects that are partial facets of, or relate to,a creator work or an artistic movement, as the different works of a writer or a painter, personal letters of an individual, objects that belong to him/her. Within this context authority files seems to be relevant. But authority control as traditionally exercised by libraries must be complemented and/or integrated with information provided by entries in Wikipedia/DBpedia.

The methodological approach outlined, the onomasiologic perspective - suggests that instead of searching for, discovering, formalizing and organizing culturally relevant relationships to make sense of the amount of cultural data now available throughout the Web of Data, it is more useful to make explicit and available the features and relationships that exist in and between cultural objects from collections in archives, libraries and museums as a starting point and so these features and relationships can be used by digital curators in doing their job.

Notes

[1] http://www.sciencemuseum.org.uk/visitmuseum/plan_your_visit/exhibitions/leonardo
[2] http://www.humanbodies.eu/en/the-exhibition
[3] http://www.visualdataweb.org/relfinder/relfinder.php
[4] https://en.wikipedia.org/wiki/Magic_realism

References

Aitchison, Jean (1987). *Thesaurus construction*: a practical manual. 2. ed. London: ASLIB.

The CIDOC Conceptual Reference Model. (2013).Version 5.1.12, January 2014. [http://www.cidoc-crm.org/docs/cidoc_crm_version_5.1.2.pdf]

Dalhberg, Ingetraut (1978). A referent-oriented analytical concept theory for Interconcept. *International Classification*, 5(3), 142-50.

Definition of the Europeana Data Model v5.2.7. (2016). [http://pro.europeana.eu/documents/900548/0d0f6ec3-1905-4c4f-96c8-1d817c03123c]

ICOMOS, International Cultural Tourism Charter (2002). *Principles And Guidelines For Managing Tourism At Places Of Cultural And Heritage Significance. ICOMOS International Cultural Tourism Committee.*

IFLA (1998).*Study Group on Functional Requirements for Bibliographic Records*: final report. München: K. G. Saur. (UBCIM Publications New Series).

ISO 2788-1986: documentation guidelines for the establishment and development of monolingual thesauri. (1986). 2.ed. [S.l.]: ISO, 32p.

Lehmann, Jens, Schüppel, Jörg & Auer, Sören. (2007). Discovering Unknown Connections-the DBpedia Relationship Finder. The Social Semantic Web: Proceedings of the 1st Conference on Social Semantic Web (CSSW). Leipzig, Germany, 26-28 September, 2007.Bonn: Köllen Druck+Verlag GmbH. Pp. 99-110. [http://ftp.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-301/proceedings.pdf#page=99]

Peters, Isabella & Weller, Katrin (2008). Paradigmatic and syntagmatic relations in knowledge organization systems. *Information Wissenschaft und Praxis* 59(2): 100-7.

Sager, Juan C. (1990). *A practical course in terminology processing*. London: John Benjamins Publishing Company.

Sales, Luana F., Campos, Maria Luisa de A.& Gomes, Hagar E. (2008). Ontologias de domínio: um estudo das relações conceituais. *Perspectivas em Ciência da Informação*, 13: 62-76.

Searle, John R. (1995). The construction of social reality. New York: Free Press.

Wüster, Eugen (1981). L'étude scientifique générale de la terminologie, zone frontalière entre la linguistique, la logique, l'ontologie, L'informatique et les sciences des choses. In *Textes Choisis de Terminologie:* I. Fondéments théoriques de la terminologie, edited by Guy Rondeau and Helmut Felber. Québec: GIRSTERM. Pp. 57-114.

Zeng, Marcia Lei. (2005). *Construction of controlled vocabularies: a primer (based on Z39.19).*Tutorial. [http://marciazeng.slis.kent.edu/Z3919/index.htm]

Aline da Silva Franca and Naira Christofoletti Silveira

The Bibliographic Representation of Authorship of Autochthonous Communities

Abstract

This proposal discusses the bibliographic representation of authorship of Brazilian autochthonous communities because we consider that knowledge organization and representation are at the center of the circle of knowledge. Knowledge organization and representation are responsible for preparing documents for preservation, access and use.To understand the authorship of autochthonous communities is the first step to contribute to the valorization and dissemination of this knowledge, respecting its origin and the cultural identities.

1 Introduction and objectives

The focus on this paper is the Knowledge Organization and Bibliographic Representation of Brazilian autochthonous communities. In Brazil the term "indigenous peoples" is very often used to refer to native groups and, by the way, is the term used in the IFLA Indigenous Matters Section (IFLA, c2016). However, the expression "autochthonous communities" is used in this paper because the term “indigenous people” could generate prejudices in the English language. It is understood that there are many words that show prejudices in different languages and this is a problem for Knowledge Organization (Pinho & Milani, 2013). Following the political correctness suggested, in this paper we use “autochthonous communities” to avoid inconveniences.

The initial point of this paper is the concept of authorship presented by Foucault (2006) and Barthes (1988). One needs to consider authorship as a social construction in order to promote a real representation, to add value to cultural identities, and to understand who the author of a work is.

The wide context of this paper is the social, cultural and ethical matters in knowledge organization. Based on studies by Guimarães and Pinho (2007), Berman (1993), Olson (2002), Capurro (2004), López-Huertas (2013), Miranda and others (2012), we will discuss some questions concerning the social contexts and prejudices about autochthonous knowledge.

The “IFLA Statement on Indigenous Traditional Knowledge” (IFLA, 2014) recommends promoting research and learning about autochthonous communities,publicizing their knowledge values to themselves and to non-autochthonous people, and encouraging the recognition of intellectual property of autochthonous traditional knowledge and products derived from it.

One is considered to commit prejudice or to be naïve and incorrect when they believe that a people belongs to a homogeneous group and have the same culture and principles. The last census published in 2010 by the Brazilian Institute of Geography and Statistics found that Brazil has more than 896,000 autochthonous people living

inside and outside of autochthonous lands, comprising 240 communities who speak about 180 different languages (IBGE, 2010).

Autochthonous groups were kept in a secondary condition in the Brazilian social dynamics during the period that extended from the European colonization until the first half of the twentieth century. Only after 1988 did the Brazilian Constitution recognize formally the autochthonous peoples' rights.

It is observed, therefore, that the advancement of school practice in the autochthonous villages, initialized in the late twentieth century, was the first step towards the emergence of the autochthonous authorship in bibliographic materials. Later on, the autochthonous authorship in bibliographic materials expanded to different regions of the country and the world, being present in several schools and different libraries.

Thus, the focus of this paper is the representation of authorship of materials produced by autochthonous communities. At this point, two types of "autochthonous authorship" are considered: the "authorship of the autochthonous person" as an individual (or more individuals, co-authors, for example) and the "authorship of autochthonous communities" for any creation elaborated by collective people, like institutions, groups or communities. The latter is more relevant for this paper.

Following the increasing number of books and other materials produced by Brazilian autochthonous people, the objective of this research is to discuss some questions about the authorship of autochthonous knowledge. We will study how autochthonous works have been represented, especially in library's catalogs, and how the authorship of autochthonous communities is understood by autochthonous authors.

2 Methodological procedures

This paper concerns an exploratory investigation based on bibliographic, documental and comparative research whose object of study is the concept of autochthonous authorship in bibliographic representation. To elucidate this point, this research interviewed some Brazilian autochthonous authors to discover what they think about authorship and used the documental analysis of autochthonous books to explain and exemplify some matters about autochthonous works. This research was developed at professional master's degree in Library Science at Federal University of the State of Rio de Janeiro and the University's Ethics Committee approved the interviews. The interviews were conducted individually with four Brazilian autochthonous authors, each of them belonging to a particular autochthonous community: Wapichana, Maraguá, Sateré-Mawé and Munduruku.

Two lists of autochthonous bibliographic materials were used for the analysis of the bibliographic representation. One of them, by Freire (2005), compiled 853 items of diverse types, published and unpublished. Another list was compiled by Santos (2014), with 563 items, all of them published. In addition to the two lists aforementioned, some

books were selected to assist the analysis process in order to discuss and exemplify the authorship of autochthonous communities.

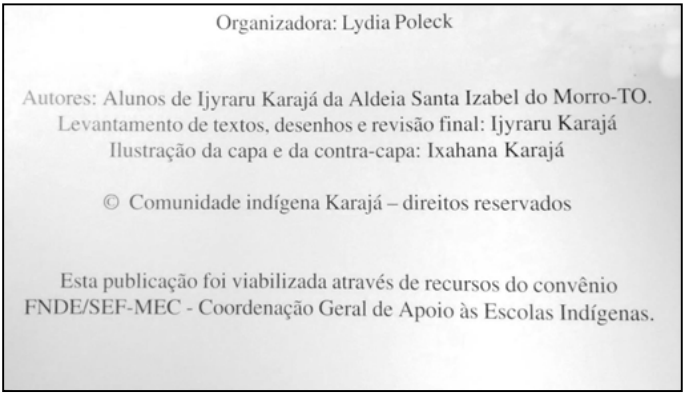
3 Main results

The construction of autochthonous knowledge has particularities, for example, in most of the cases, the origin of the tales lies within autochthonous communities, but their expression in a book is the responsibility of one individual, or more. In this case, the interviews showed that autochthonous authors can create a new version of a tale, at the same time preserving almost 80% of the original theme. In other cases, the same autochthonous author can create a new story which is 100% original.

There are books with a clearly identified authorship, with the name of the autochthonous community on the title page. In other books the cataloguing librarian must read the introduction and verso of the title page to find the copyright. Some books show the copyright of a specific autochthonous community, not necessarily its entire list of authors.

The name of the autochthonous community and the autochthonous people are commonly found on the title page or on the verso of title page. Thus the author of the book acknowledges that the work is not only his or hers as an individual.

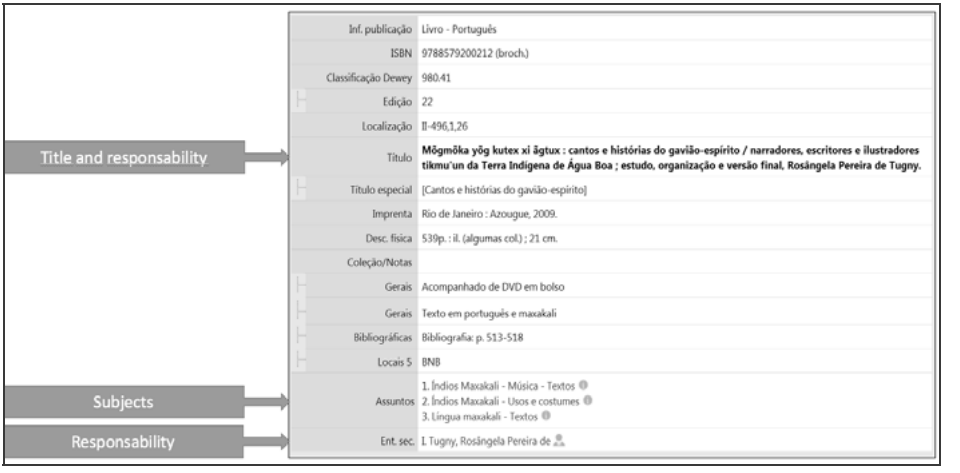
Figure 1. Verso of title page of book *Iny Rybe-my Ijyy* (Comunidade Indígena Karajá, 1998)



This book (Figure 1) shows the name of the organizer, "Lydia Poleck", and attributes, as authors, the "students" from "Santa Izabel do Morro", which is the name of autochthonous village. The autochthonous community is called "Karajá", but only one group from the Karajá autochthonous community is responsible for this book and has the copyright. This case is only one example of many books and different situations that the construction of bibliographic representation comprises.

Figure 2 shows a bibliographic record and some observations about de autochthonous communities' representations.

Figure 2. Bibliographic record (Fundação Biblioteca Nacional, 2016)



It is possible observe in this bibliographic record that the autochthonous communities isn't represented as author or responsible for this work. The autochthonous appear only as "subjects" in this bibliographic record.This absence of the autochthonous person as the author or principal responsible for the work is very common in the bibliographic record, appearing as subjects in most of case (Franca & Silveira, 2014).

The omission of the actually responsible community in this bibliographic record would be the beginning of the devaluation and non-recognition of the autochthonous knowledge. In this case, the bibliographic record does not respect the IFLA Statement on Indigenous Traditional Knowledge (IFLA, 2014), especially the recommendation #3: "Publicize the value, contribution, and importance of indigenous and local traditional knowledge to both non-indigenous and indigenous peoples"; and the recommendation #6: "Encourage the recognition of principles of intellectual property to ensure the proper protection and use of indigenous traditional knowledge and products derived from it." (IFLA, 2014, 1).

Looking at the figure 2, responsibility is attributed only to the "person", not the "community". About the rules of bibliographic representation, the "community" can be considered a "corporate body" who represents the group or institution and similar situations.

Sometimes, we felt that something was missing when using "corporate body", for example, some communities do not have a formalized complex cultural structure. We understood that "corporate body" can be used by groups, but not by all autochthonous communities. More studies about communities (not only autochthonous) must be conducted in order to create a new concept that aggregates social, ethnical or special groups and communities, such as autochthonous people, Basques, and Kurds.

At this point, the concepts “person”, “corporate body” or “family” and the “work” and “expression” from FRBR (Functional Requirements for Bibliographic Records) and FRAD (Functional Requirements for Authority Data) can be helpful, but we consider that these concepts are rather simple to be applied to complex autochthonous communities.

It is necessary to understand who the responsible person for the “work” is and who has the responsibility for the “expression”. For the Brazilian autochthonous knowledge, it is possible to observe that there are many “works” from the community or corporate body and these "works" have many “expressions” by person (or individual autochtonous author). These perspectives must be considered for a bibliographic representation.

4 Final considerations

The autochthonous works (as abstract creations in the author’s mind) are somehow considered collective authorship, because they are inherited. It is collective, but is not developed by a group of people. The authorship will be updated and transformed through oral transmission. Thus, the cataloguing librarian must comprehend the context of production of that authorship in order to attribute it to autochthonous works. The one who writes the work is just responsible for the expression, making that abstract knowledge into something tangible, in this case readable. The other responsibilities would be other people involved in the work production such as illustrators, translators, etc.

The cataloguing librarian must comprehend the context of works creation and expressions to attribute the authorship and other responsibilities on the bibliographic record before establishing guidelines. The authorship is a complex topic and it depends on the origin of the knowledge. The scientific knowledge, for example, is different from common knowledge.

It is very important that the Knowledge Organization and bibliographic representation acknowledge the cultural identity of the autochthonous communities by themselves, respecting the circle of knowledge and cultural identities. Autochthonous communities have been suffering prejudices in Brazil, whose communities have to fight for their rights. We believe that attributing authorship to autochthonous communities can be a way to non-autochthonous people stop committing prejudice to autochthonous people and recognize the value of their knowledge.

In this perspective, to study about Brazilian autochthonous people is the beginning of more complex studies to promote autochthonous knowledge (including autochthonous botanical knowledge and others) and to acknowledge their value for all societies, especially to construct a Brazilian identity.

References

Barthes, Roland (1988). *O rumor da língua*. São Paulo: Brasiliense.

Berman, Sanford (1993). *Prejudices and antipathies*: a tract on the LC subject heads concerning people. Jefferson: McFarland & Company.

Capurro, Rafael (2004). Intercultural information ethics. In: International Center For Information Ethics Symposium, 2004, Karlsruhe. *Proceedings*... Karlsruhe: Center for Art and Media, 2004. [http://www.capurro.de/iie.html]

Comunidade indígena Karajá (1998). *Iny Rybe-my Ijyy*. Goiania: FNDE/SEF-MEC.

Franca, Aline & Silveira, Naira Christofoletti (2014). A representação descritiva e a produção literária indígena brasileira. *Transinformação*, 26(1): 67-76. [http://www.scielo.br/pdf/tinf/v26n1/a07.pdf]

Foucault, Michel (2006). *O que é um autor?* 6. ed. Lisboa: Vega Passagem.

Freire (2005). *Catálogo de livros indígenas*. Rio de Janeiro: UERJ. Material não publicado.

Fundação Biblioteca Nacional (2016). *Catálogos online*. [http://acervo.bn.br/sophia_web/]. Acessed 17 May 2016.

Guimarães, José Augusto C. & Pinho, Fabio A. (2007). Desafios da representação do conhecimento: abordagem ética. *Informação & Informação*, Londrina, 12(1): 1-21, jan./jun. 2007.

IFLA (c2016). Indigenous Matters Section. *Scope*. [http://www.ifla.org/indigenous-matters]

IFLA (2014). *Statement on Indigenous Traditional Knowledge*. [http://www.ifla.org/publications/ifla-statement-on-indigenous-traditional-knowledge]

López-Huertas, María José (2013). Reflexions on multidimensional knowledge: its influence on the foundation of Knowledge Organization. *Knowledge Organization*, 40(6): 400-7.

IBGE (2010). *Censo Demográfico 2010*: características gerais dos indígenas: resultados do universo. Rio de Janeiro: IBGE.

Miranda, Marcos Luiz C. de et al. (2012). Organização e representação do conhecimento em religiões yorubanas na Library of Congress Subject Headings. In, José Augusto C. Guimarães and Vera Dodebei. *Desafios e perspectivas científicas para a organização e representação do conhecimento na atualidade*. Marília: FUNDEPE. Pp. 153-63.

Olson, Hope A. (2002). *The power to name: locating the limits of subject representation in libraries*. Dordrecht: Kluwer Academic Publishers.

Pinho, Fabio A. & Milani, Suellen O. (2013).Metáfora e ortofemismo na representação de assunto. In*Complexidade e organização do conhecimento: desafios de nosso século,* edited by Vera Dodebei and José Augusto C. Guimarães. Marília: FUNDEPE. Pp. 246-51.

Santos, Waniamara (2014). *Daniel Munduruku: contador de histórias, guardião de memórias, construtor de identidades*. Dissertation. Master's Degree in Estudos da Linguagem, Universidade Federal de Ouro Preto, Minas Gerais, Brazil.

Evelyn Goyannes Dill Orrico and Eliezer Pires da Silva

Knowledge Organization in Archives: The Brazilian Case

Abstract

This research aims to investigate how the thematic content representation practices of archival holdings are accomplished within the Brazilian context. The specific goals involve the characterization of the archival work of documental description and analyze information search instruments made available to users of archival institutions. More specifically, this proposal provides the reflection on the disclosure of archival sources under the perspective of their social uses, besides being a fundamental aspect to the exercise of citizenship. From a methodological point of view, it commences with a literature review about categories as archival information, arrangement and archival holdings description, under the perspective of knowledge organization and, from the findings attained in the first phase, the Memories Disclosure Portal Database managed by the Brazilian National Archives was used. The main outcomes are: firstly, we noticed a poor identification among the public archives and the majority of population in Brazil; secondly, we notice the insufficiency of thematic organization strategies prepared by the custodian entities on their archival holdings in order to reach a broad audience that normally would not visit the archives. The result is that the archives remain away from the population.

1 Introduction

The International Society for Knowledge Organization – ISKO events, either domestic or international, constitute a forum for debating and producing knowledge on information issues in plural contexts of cultural, scientific and technologic connections in societies' spaces. This is an interdisciplinary challenge which has also been explored in the intersections with archive phenomenon, according to Orrico and Silva (2011; 2012; 2013). In the intersection with the Archiving Science, we highlight the demand for thematic representation elements in the construction of archives reference instruments, a difficulty to be overcome from a deepen dialogue with the knowledge organization.

According to Alvarenga (2003), the knowledge records contained in documents start to integrate archives collections, libraries, and documentation centers and are a primary representation of such knowledge in a documental support. Thenceforth, they are going to be represented again in those institutional environments (secondary representation), aiming at their inclusion in reference documental systems to users, constituting our perspective on the knowledge organization in the context of this research.

> In the professional performance of both treatment and information organization, archivists, librarians, museum experts and other information professionals, targeting at the intermediation among documents and users, develop different types of representations in this form, involving a replacement of primary information by specific records related to it, aiming at a subsequent recovery. (Alvarenga, 2003, p. 23)

The difficulty is even greater when the archival holdings disclosure takes place to a broad audience, non-professional of the field, who normally would not visit the archives. The access to those collections is fundamental, in the exercise of its social and cultural function, as it is a resource for the social memory construction. This aspect implies in searching to reach an even larger audience for the archives. The study of

accessing conditions to archive documents by the side of the population contributes to the construction of citizenship and the realization of the right to information. Aiming at meeting the assumption that the archives social role increases the usage of their collections and the number of their users, this objective proposal discusses the obstacles to this access increase to archive documents, going through the procedures used by institutions that state this purpose for the collections they organize and preserve. This investigation not only boosts the knowledge about archives as informational resource to the citizen, but also mainly provides knowledge production about a fundamental subject for the emergency of new discussions in academic space involving several social actors: the archive representation knowledge to the lay population.

This research aims to investigate how the representation practices of the archival holdings contents are accomplished within the Brazilian context. The specific goals include developing indicators that evidence archive work of documental description and analyze informational search instruments made available to users of archive institutions. More specifically, this proposal is going to cause a reflection on the scientific disclosure of the archive sources from the social use perspective, further to be a fundamental aspect for the exercise of citizenship.

2 Archival Description

Archival description is the process by which archivists create representations of a particular archival collection, submitting their context and content. It is an intellectual activity which requires text interpretation comprehension, historical knowledge and ability to draft collections' descriptions. The goal is the archive documents control, having the access promotion in mind.

Surely this is the role of the archivist consistent with knowledge organization in the context of archives. The concept of archive description shows a corresponding isolation in relation to development of representation notion in the knowledge organization ambit provides that, both in this and that field, the technical functions are just like around the goal to mediate the access to information by the users.

The archives description instruments are understood as a form to represent the content of documents and thus mediate the occasional consult to the record itself. The notion of archive information is associated to construction of research instruments of which custodian documents index by means of the representation of contents, targeted to allow and guide a research by users.

To Barros (2014) the notion of Archive Representation would be more proper to understand processes and preparations that involve classification, description, as well as the creation of research instruments (handbooks, inventories, catalogues), searching systems and database in the context of archives. It is the creation of substitutive of documental records by means of the representation of its original content and context

to users of archives.

Two general assumptions guide the execution of this proposal: the power of archive information does not lie in itself, it is enhanced with the circulation and use of such information before instruments that favor its meaning, so that the citizens may build knowledge; the archives need to be more popular institutions than they currently are, so that more people might benefit from knowledge stored there, creating a society gradually more educated, more plural and more democratic.

In this line of research, we have been working about the fact that, after the gathering of records to the custody of an archive institution, the technical treatment of objects classification and description is performed in this environment, with the representation and intellectual control of the documents content in the research instruments, and that will be deployed by users in searching for information. Thus, at first, it is observed that it is not the document itself that is offered, but a representation on the content of documents prepared by the institution and its team. It is this mediation, not fully instrumentalized by practices grounded in knowledge organization, which we are questioning in this research (Pret; Cordeiro, 2015).

By considering the construction of research instruments which index archive custody documents, we understand that such instruments may both place in evidence, and fade out archive documents to users. This double possibility provides the theoretic discussion between remember and forget. The social role of the archival institutions is to assure access to documents under their custody by means of instruments which represent the content of those records, making effective the information transfer in archives as a process which goes from the document storage to the disclosure of their contents, considering the research instruments as mediators for this purpose (Simonato, 2015).

The connection between the preparation of the research instruments present in archives and the knowledge organization is a result of the process of documentary analysis which consists in two basic phases: the analysis and summary of the contents present in documents. To Calderon (2003), this search for problems solving of accessing and information recovery in the archives is translated into the information representation, aiming at recovering and disclosing it.

As part of the theoretical assumptions of this research, we understand the conditions of the exercise of the right to information and how they are related to the access to archives as manifestation of citizenship, to further than freedom of expression. Thus, access to archive documents involves knowing practices that need, further to clearly represent the content, meet a logic of control, government and management.

From the methodological point of view, it commenced from the literature review about categories as archive information, arrangement and archival holdings description, normalization of archival processes and, from the findings attained in the first phase of project, the Disclosed Memories Portal Database managed by the Brazilian National

Archive was used. Later on, the analysis of Portal description fields was performed, in order to identify which fields were filled out, as well as evaluate the consistence of information made available, having the guidelines and guidance of international rules of archives description as criterion.

3 The Archival Information Representation in Brazil

We see the archive information representation, embodied in the broad process of communication, as being a mediation of the language in the context of information treatment in the archives. The assumption here is to consider the weak points of the communicative actions by means of oral language, turning them into a challenge to find forms of interfaces between the documental collections and their users. This is a situation which particularly seems to be problematic in terms of knowledge organization, mostly in light of the diversity; not only of the documental types, but also to who produce them and those wishing to access them, in situations herein submitted as follows.

Since 2005, the Brazilian government has been developing actions with aim at the registration of documents produced by public bodies, as well as by people and entities corporations, throughout all national territory, interested in the subject “Political Struggles in Brazil (1964-1985)”. The official speech is the search for the strengthening of an appreciation public policy of documental historical heritage and enhancement of citizenship and democracy in our country.

The Reference Center of Political Struggles in Brazil, also named "Disclosed Memories", was created in 2009 by the so-called Office´s Chief of the Presidency of Republic, Dilma Vana Rousseff, being implemented in the National Archive with the articulation among the members of federation targeting at the strengthening of a military dictatorship period memory policy.

The Disclosed Memories Database is currently available in internet and gathers information of custodian archival holdings by several entities throughout the country, nourished on-line by the partners in the Reference Center of Political Struggles in Brazil, consolidating an announced National Network of Cooperation and Archive Information.

It was noticed from the surveyed data of Disclosed Memories Portal Database:

- The database descriptions are inspired in NOBRADE (Brazilian Rules of Archive Description);
- There are 7 areas of description, divided into 42 fields;
- 70% of collections are from institutional origin and 30% from natural persons;
- 88% of collections are located in the States of São Paulo, Rio de Janeiro and Federal District;
- There are 5.1 million of items and 32 km of described documents;
- 62% of the fields set forth in archives description rules were not filled out;

- To accomplish the research, the reading of the descriptions of information Disclosed Memories Portal Database was done, checking the descriptions informational quality, being classified according to the table below.
- We noticed that 67% of the descriptions provided in the Portal had not enough information to identify the content and the context of collections.

Table 1: Contextualization Area

Fields	2.1 Producer	2.2 Administrative History/ Biography	2.3 Archive History	2.4 Origin
Not Enough	59%	54%	47%	44%
Not Filled Out	8%	13%	20%	23%
Total	67%	67%	67%	67%

Source: Disclosed Memories Portal Database

From those data, we call the attention to the need of stressing the archivist´s theoretical and methodological commitment with the theoretic field of knowledge organization, considering the archive representation under the perspective of access extension to users. Still, it is important not to miss the point on the discursive limits of archive description, according to the signalization of MacNeil (2005), as it is a representation unable to support the totality of informative potential of the object it intends to describe.

Grounded on the empiric observations performed in Disclosed Memories Portal Database, we propose to set out more mandatory fields in the archival holdings description, considering the pertinent need of offering geneses context and documents custody to users' indicators, once information on the context is elements able to broaden both meaning of documents individually, and evidence its authenticity. That is, it was verified a more continuous approach of archive studies with the theoretical premises of knowledge organization, in order to make possible the accessing enhancement to information by a more diversified and plural audience.

One of the possibilities of approach is pursued in the construction of information recovery systems, according to Kobashi & Francelin (2011), constituting a metalanguage, also called documentary language. The task of the representation resources is limited to the use in communication environments user-information system and, even without the expressive power of Natural Language, it must support more precise meaning strategies with the ultimate aim of recovering information of archive content and context.

4 Final considerations

Concerning the technological resources already used by the archival institutions to enhance the uses and users of their collections, two points still need a deepen investigation. Firstly, we notice that there is a poor identification between most of the

population and public archives in Brazil. Secondly, there are insufficient research instruments provided by the custodian entities of archival holdings to reach a broad audience that normally would not visit archives. As a result, the archives remain away from the population (Jardim, 1996, 1998, 1999).

Thus, it is paramount to think about the instruments for mediating the uses in which archives can be made of, but mainly as important resource for the citizen in the construction of its identity and citizenship. This is to say that we believe in the materialization by the side of the citizen right as a transparency path to the State up to civil society, enabling to make the political participation effective to a basic dimension of citizenship and to minimizing social inequality.

The knowledge organization strategies may subsidize the ordering of information embodied in the custodian collections more effectively and thus turn the archival institutions into intermediate element of citizenship construction.

References

Alvarenga, Lídia (2003). Representação do conhecimento na perspectiva da ciência da informação em tempo e espaço digitais. *Encontros Bibli*: revista eletrônica de Biblioteconomia e Ciência da Informação, 8(15).

Barros, Thiago Henrique Bragato (2014). A representação da informação Arquivística: Uma Análise do discurso teórico e institucional a partir dos contextos Espanhol, Canadense e Brasileiro. Thesis. Doctor's Degree in Ciência da Informação, Universidade Estadual Paulista, Marília, Brazil.

Calderon, Wilmara Rodrigues (2003). Instrumentos de pesquisa nos arquivos públicos permanentes: um estudo sob a ótica da análise documentária. 2003. Dissertation. Master's Degree in Ciência da Informação, Universade de São Paulo, São Paulo, Brazil.

Jardim, José Maria (1996). A invenção da memória nos arquivos públicos. *Ciência da Informação*, 25(2).

Jardim, José Maria (1998). *Os arquivos (in)visíveis: a opacidade informacional do Estado brasileiro*. Thesis. Doctor's Degree in Ciência da Informação, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil.

Jardim, José Maria (1999). Transparência e opacidade do Estado no Brasil: usos e desusos da informação governamental. Niterói: EDUFF.

Kobashi, Nair Yumiko & Francelin, Marivalde Moacir (2011). Conceitos, categorias e organização do conhecimento. *Informação & informação,* 16(3): 1-24.

Leão, Flávia Carneiro (2006). *A representação da informação arquivística permanente*: a normalização descritiva e a ISAD(G). Dissertation. Master's Degree in Ciência da Informação, Universidade de São Paulo, São Paulo, Brazil.

MacNeil, Heather(2005). Picking our text: archival description, authenticity, and the archivist as editor. *The American Archivist*, 68(2).

Orrico, Evelyn Goyannes Dill & Silva, Eliezer Pires da (2011). Network of specialists in the archival field and the impact on knowledge organization: the case of Brazil.In *X Congreso ISKO-España.* Ferrol, Spain, 30 June – 01 July, 2011.

Orrico, Evelyn Goyannes Dill & Silva, Eliezer Pires da (2012). Representação do conhecimento arquivístico e a rede de seus pesquisadores no Brasil. In *Desafios e perspectivas científicas*

para a organização e representação do conhecimento na atualidade, edited by José Augusto Chaves Guimarães and Vera Dobedei. Marília: FUNDEPE.Pp. 49-53.

Orrico, Evelyn Goyannes Dill & Silva, Eliezer Pires da (2013). O trabalho de descrição de acervo arquivístico no Brasil. In *Complexidade e organização do conhecimento: desafios de nosso século*, edited by Vera Dobedei and José Augusto Chaves Guimarães. Marília: FUNDEPE. Pp. 211-6.

Pret, Raquel Luise & Cordeiro, Rosa Inês de Novais (2015). A indexação como ferramenta da gestão de documentos nos arquivos. In *Desafíos y oportunidades de las Ciencias de la Información y la Documentación en la era digital: actas del VII Encuentro Ibérico EDICIC 2015*. Universidad Complutense de Madrid, Madrid, Spain, 16 - 17 November 2015.

Simonato, Ana Carolina (2015). *Modelagem conceitual DILAM: princípios descritivos de arquivos, bibliotecas e museus para o recurso imagético digital*. Thesis.Doctor's Degree in Ciência da Informação, Universidade Estadual Paulista, Marília, Brazil.

Mariângela Spotti Lopes Fujita, Paula Regina Dal'Evedove, Franciele Marques Redigolo and Noemi Oliveira Martinho

The Socio-Cognitive Context of the Subject Cataloger and His Professional Experience

Abstract

Studies of human cognition have provided significant contributions about the mind and its understanding capacities. In the thematic treatment of information, particularly in the process of subject cataloging, the professional's performance may be influenced by various factors. In this paper, an attempt is made to characterize the cognitive context of the subject cataloger focusing on his prior knowledge built through interactions within the context of university libraries. References are made to some studies dealing with Verbal Protocol in order to highlight the importance of using this introspective technique of data collection in the professional subject cataloger's performance.

Introduction

Over the years, Information Science research has focused on the development of theoretical and methodological studies about the thematic treatment of information, thus contributing both to theory and to professional education. One of the major shortcomings in this area is related to the contextual performance of the professional, a social being that influences and is influenced by the information system in which he is inserted.

In the thematic treatment of information, content analysis and representation assignment involve cognitive operations that determine the professional's choices, his view of the world, education and intellectual training. However, in various fields of knowledge the thematic treatment of information is still obscure and lacks some fundamental guidelines and procedures. The need to provide new insights into this area has drawn the attention of researchers to study the information professionals' performance from a cognitive perspective. The ever-growing development of information sources has posed many challenges to information scientists. Such challenges were initially of a practical nature, but they have grown more complex over time, thus requiring the development of innovative theoretical and conceptual perspectives. To understand the professional's socio-cognitive process the moment he has to tackle with the treatment of thematic information is one of the problems currently faced by Information Science. There has been a growing body of research using Verbal Protocol technique designed to observe the cognitive context of the subject cataloger to infer as closely as possible the professional's underlying thought mechanism and choices. This paper presents a brief discussion of subject cataloging from the point of view of the thematic treatment of information aiming to pose some reflections on the subject cataloger's socio-cognitive context and practice.

It is believed that the study of the socio-cognitive processes involved in subject cataloging may provide the professionals working in the context of an academic library with more awareness and consistency in dealing with information. The subject cataloger's knowledge of both theoretical and technical aspects contribute to a responsible and effective performance. Considering that Verbal Protocol is an introspective technique that has introduced some innovations regarding the professional's awareness of the subject cataloging process in a university library, some parameters for analysis and observation of the professional's work are discussed in this study.

Thematic treatment of information: emphasis on subject cataloging

In the field of Information Science, the term "information treatment" may comprise all disciplines, techniques, methods and processes involving physical and thematic descriptions of library documents or information retrieval systems, the development of tools used in these descriptions, as well as the implementation of physical structures for information storage (Dias; Naves, 2007). The thematic treatment of information processes have long been considered an art or a result of a special talent to characterize a particular information theme in a document (Campos, 1987; Guimarães, 2003; Mey, 1987).

Until the mid-twentieth century, the methods used for subject analysis and representation were regarded as empirical operations based on common sense; one or more keywords were assigned to a document in order to facilitate its search and retrieval, without any methodological concern with subject attribution. (Cunha, 1990). This revealed a great variety in the criteria of a random and subjective nature, with no paradigms that would confer a clear and scientific status to the procedures used during subject analysis and representation. (Guimarães, 2003). These findings indicated that information treatment should be dealt with in a more specific way due to the great number of specialties in library collections, thus creating new representation forms to enable a faster and more useful information retrieval (Novellino, 1996).

Documentary treatment represents the intermediate stage in the informational cycle and involves some fundamental operations, such as production, treatment or organization, retrieval, dissemination and information use which, in turn, may generate a new production, thus completing the cycle. Documentary treatment covers a dichotomy between thematic treatment (information content) and descriptive treatment (material support of information). From an evolutionary perspective on information treatment, subject indexing and cataloging are thematic treatment operations involving analysis, synthesis and representation. "In this evolutionary context of information treatment techniques, documentary analysis is an extension of the thematic treatment comprising abstract generation and indexing" (Fujita, 2003, 61).

In a library context, the processes can be done by combining both subject classification and cataloging processes into one, which will result in the subject

analysis of a document. This process of subject cataloging is highly complex, since external and internal factors, such as subjectivity, partiality and bias may not only interfere with understanding and interpreting content but also have a direct impact on its representation. According to Silva and Fujita (2004), subject cataloging is characterized by assigning subject headings to represent the entire contents of documents in library catalogs. Alphabetical indexing is determined by subject headings and, in some cases, it is also called ‘subject cataloging’.

Despite thedivergenceofopinionsover the similarities and differences between the terms ‘alphabetical subject indexing’ and ‘subject cataloging’, they can be considered similar because they result from the same process: subject analysis. Subject indexing and cataloging are conceptually equivalent activities performed in distinct environments (Milstead, 1983). Subject cataloging expresses the informational content of a document. The origin of the term is related to the construction of library catalogs, especially the subject catalog, which is organized by determining subject headings that serve as subject statements formed from an ordered composition of words. Subject headings are the first attempt to organizing the alphabetical subject representation (Silva and Fujita, 2004).

From the point of view of information systems, subject indexing and cataloging are fundamental elements of documentary analysis because they determine the results of a search strategy. Therefore, information retrieval depends on the performance quality of indexing activity. AsChaumier (1980) noted, subject cataloging "is the most important part of documentary analysis, since it attributes value to a documentary system. An unsatisfactory indexing represents 90% of the main causes for the appearance of noise or silence". A poor representation will neither contribute to language development nor enable a staff to use such language.

With reference to the effectiveness of an informational unit, Langridge (1989) believes that no information retrieval system can be better when subject analysis is used. Any concept that may be overlooked in the analysis of a document will not be represented in the system language, since it is the subject that makes information retrieval possible. The quality of the products (catalogs, indexes, etc.) will mostly depend on the cataloger’s competence in subject analysis, for this process may be influenced by various factors, such as the professional’s subjectivity, working conditions, background knowledge, expertise, communicative situation and rules.

Information representation plays an invaluable and indispensable role in the development of society, because it is the means by which individuals interact and build relationships. For this reason, the cataloger must be aware that his task should not be carried out at random but be based on reflections and attitudes thatshow his concern forproviding the products and tools used in organizing and representing knowledge. In this manner, subject cataloging in an academic context must follow the same basic principles of product and service generation in all the spheres of the immensely broad

universe of thematic treatment information. Libraries and informational centers require continuous improvements and professional skill specifications to ensure product quality, search optimization and effective information retrieval. From the evolutionary perspective of data processing, subject indexing and cataloging are operations closely linked to thematic treatment of information.

The subject cataloger'ssocio-cognitive context

Subjectivity has been discussed in the context of cognitive information processing since the end of the twentieth century (Linares, 2004). In the mid-80's, Hjørland criticized the cognitive perspective and proposed a socio-cognitive approach by combining the individual and his knowledge with the external environment in order to study and understand the information user (Martin-Lahera, 2004). The human mind is a processor that collects, stores, retrieves, transforms and also transmits information which, like other correspondent processes, can be studied as patterns and pattern manipulations (Koch, 2002). Cognitivism asserts that all human mental activities involve a sequence of operations that are similar to those performed by a computer. Cognition is an action and learning is a negotiation between the system and the environment; therefore, if this interaction fails, there will be no cognition (Fujita and Cervantes, 2005).

Thinking strategies are a set of processes performed by an individual in order to assimilate new information (Bernard, 1995). Reading strategies can be perceived when some disruption hinders understanding; at this point, the reader slows down and uses meta-cognitive reading strategies (Cavalcanti, 1989). In other words, his actions are directed towards an objective or towards the search for a solution to comprehension problems. Cognition involves unconscious mental processes and meta-cognition refers to the conscious management of a cognitive phenomenon. Besides these processes, another determining factor in subject cataloging issue is the study of the cognitive context. According to Koch (2002), a socio-cognitive context implies an interaction between cognitive contexts. All types of knowledge from various sources that are partially shared and stored in the memory are called "social actants". They are essential to verbal exchange and involve linguistic knowledge, encyclopedic knowledge, knowledge of communicative situation and its rules, super-structural knowledge, stylistic knowledge and inter-textuality.An analysis of the context from a socio-cognitive approach shows that the physical context does not directly interfere with language use. The agent that may cause changes is the reader/subject cataloger, a user of that environment. The context that acts as an influential variable in interfering with the mode of a documentary representation in a unit is the set of assumptions brought to the interpretation of a statement (Koch, 2002).

A context is conceived of as a theoretical reconstruction of a series of features in a communicative situation or traits that are part of the conditions that turn statements into speech acts (Van Dijk, 1997). The cataloger should be able to use and develop more

effective reading strategies than the conventional reader, since understanding depends essentially on textual knowledge, which has a linguistic structure and conveys meaning (reading comprehension). In order to grasp the essence of a document, the cataloger has to rely on both cognitive and descriptive interpretations. Understanding a text depends on both its intrinsic characteristics and the prior knowledge shared between the author and the reader. During the process of documentary reading, the subject cataloger interacts with the document in order to identify its concepts and meaning structure (relationships between words), which is achieved only when reading comprehension occurs. The information treatment policy of the institution, which includes the documentary language used in the informational unit, is also part of the socio-cognitive context of the subject cataloger. A term is meaningful only when the situational context is taken into account; in this sense, the subject cataloger must adjust his decisions to the needs required by the environment (Koch, 2002). The cataloger is considered as a professional reader, a generator of the understanding process. As such, he uses his prior knowledge and other cognitive operations in order to be able to analyze the subjects present in the thematic description.

The socio-cognitive approach renders the cataloger an effective support for the development of subject cataloging, as his task is characterized as a social activity: he is responsible for a clear and precise processing and treatment of information designed to generate new knowledge and fulfill the users'expectations and needs. This perspective differs from the cognitive standpoint, which focus on the individual activity in his own context. Continuing education and training should also be considered in terms of how these factors can affect and contribute to the improvement of the professional's theoretical knowledge and operational skills, as well as to broaden his awareness of the importance of his work in a social context. Studies designed to gain information, to interpretandto assess the subject cataloger's cognitive mechanisms and operations, which cannot be directly observed, are essential to gathering important data that may contribute to improve the professional's own knowledge and performance.

Observation studies of the subject cataloger's professional experience through verbal protocol

Verbal Protocol has been extensively used in studies to investigate the cognitive processes of the human mind. This technique is a rich data source of individuals' spoken thoughts that are associated with the mental processes used when working on a task. Ericsson and Simon (1987) were the precursors of this methodology for monitoring reading activity. Protocols are generally defined as verbal reports of the informer's conscious mental processes ("Think Aloud") while he performs any kind of task (Cavalcanti, 1989). Verbal Protocol has been used extensively for eliciting the sequences of thought and decision-making strategies that occur during the reading task. This technique consists of recording individual responses that, after being transcribed and analyzed, generate protocols that enable researchers to infer how the reader deals

with information. A pioneer study using Verbal Protocol technique was carried out in Brazil by Fujita (1999, 2003) in order to observe reading processes in documentary reading for indexing purposes.

Verbal Protocol includes three procedures: a) *procedures prior to data collection:* definition of the research context, selection ofthe reading text, task definition, selection ofthe subjects, informal conversation with the subjects and familiarization with the "*Think Aloud*" task; b) *procedures during data collection:* recording ofthesubjects' *"Thinking Aloud"*during text reading and retrospective interview (optional); c) *procedures after data collection:* literal transcription of the subjects'recorded speech, detailed reading of the data to collect significant and recurrent phenomena to construct analysis categories, construction of the categories and data analysis to select the discussion excerpts illustrating each phenomenon/category.

Verbal Protocol provides a basis for investigating the subject cataloger's intellectual background, the knowledge acquired throughout his professional experience and applied in an individual and/or collective environment. Considering that subject cataloging process is strongly influenced by the professional, some of the research proposals using Verbal Protocol technique have focused on his socio-cognitive performance in informational contexts. The main purpose of this approach has been to identify and analyze the professional's and the user's views and procedures during documentary representation and the use of search and retrieval of information services, a scenario which is discussed by Fujita (1999, 2003), Ruby (2008) and Fujita, Ruby and Boccato (2009). For example, in Redigolo and Fujita (2009) the results revealed the socio-cognitive aspects involved in this last stage of subject analysis, concept selection, as well as the inferring factors in documentary reading and decision-making in term usage, which inherently concerns indexing policies and users. In another paper, the results indicated that at the moment of exploring textual structure, the subject catalogers relied on prior knowledge as a professional strategy in order to understand and analyze the subject of the books (Dal'Evedove and Fujita, 2008). In Martinho (2010) the results showed that three factors interfere with the documentary reading process: a) the cataloger's prior knowledge, b) the institution indexing policy, and c) the user's focus. It was observed that the strategies used during reading may be changed or adapted according to the context.

Considerations

The subject cataloguer is the connecting link between the user and the information. As such, he must carry out his professional task in a conscientious way, taking into account the context of informational and professional interlocution, as well as all the interfering factors. In this sense, the present study aimed to provide some insights into the socio-cognitive context involving the subject cataloguer's performance in the thematic treatment of information in a university library. The application of Verbal

Protocol in the subject cataloger's actual socio-cognitive environment, an innovative research particularly in this area, provides relevant data that demonstrate his procedures and difficulties, the cognitive and meta-cognitive strategies used in reading comprehension. The assessment of these data is useful for service improvement in academic libraries and for understanding this professional as a social being.

It is believed that Verbal Protocol is an adequate technique for observing the cognitive context of the subject cataloger. Investigations using this approach may elucidate the role of the subject cataloging process and its impact on informational products and services. It should be emphasized that studies in this area have become increasingly necessary, since the quality of the subject cataloger's actions are fundamental to guaranteeing that the information be devoid of personal interference and be fully presented to those who need it, namely, users in an academic context.

References

Bernard, Juan Antonio (1995). Análisis y representación del conocimiento: aportaciones de la psicología cognitiva.*Scire*, 1(1): 57-79.

Campos, Astério T. (1987). A indexação. *Revista de Biblioteconomia de Brasília*, 15(1): 69-72.

Cavalcanti, M. C (1989). *I-n-t-e-r-a-ç-ã-o leitor-texto:* aspectos de interpretação pragmática. Campinas: UNICAMP.

Chaumier, Jacques (1980). *Travail et methodes du/de la documentaliste*: connaissance du problème. Paris: ESF/Libraries Techniques. Exposé 3, Chap.3: L`indexation. Pp. 42-7.

Cunha, Isabel M. R. F (1990). *Do mito à análise documentária.*São Paulo: EDUSP.

Dal´Evedove, Paula R.& Fujita, M. S. L. (2008). A Cognição profissional de catalogadores de assunto em contexto de biblioteca universitária. In15 Seminário Nacional de Bibliotecas Universitárias. São Paulo: CRUESP. Pp. 1-15.

Dias, Eduardo W & Naves, Madalena M. L (2007). *Análise de Assunto:*teoria e prática. Brasília: Thesaurus.

Ericsson, K. A & Simon, H. A (1987). Verbal reports on thinking. In *Introspection in second language research*, edited by Faerch Claus & Kasper Gabriele.Clevedon: Multilingual Matters. Pp. 24-53.

Fujita, Mariângela S. L (1999). A leitura do indexador: estudo de observação. *Perspectivas em Ciência da Informação,* 4(1): 101-16.

Fujita, Mariângela S. L & Cervantes, Brígida M. N (2005). Abordagem cognitiva do protocolo verbal na confirmação de termos para a construção de linguagem documentaria em inteligência competitiva. In*Métodos qualitativos de pesquisa em Ciência da Informação*, edited by Marta L. P. Valentim.São Paulo: Polis.

Fujita, Mariângela S. L (2003). A identificação de conceitos no processo de análise de assunto para indexação. *Revista Digital de Biblioteconomia e Ciência da Informação,* 1(1): 60-90.

Fujita, Mariângela S. L.; Ruby, Milena P & Boccato, Vera R. C (2009). O contexto sociocognitivo do catalogador em bibliotecas universitárias: perspectivas para uma política de tratamento da informação documentária. *DataGramaZero - Revista de Ciência da Informação*, 10(2).

Guimarães, José Augusto C (2003). *Aspectos éticos do tratamento temático da informação (TTI):*elementos para sua caracterização a partir da interface dos ambientes de ensino e pesquisa no Mercorsul. Marília (Projeto integrado de pesquisa - CNPq).

Koch, Ingedore G. V. (2002). *Desvendando os segredos do texto*. São Paulo: Cortez.

Langridge, Derek W. (1989).*Subject analysis: principles and procedures*. London: Bowker-Saur.

Linares, Radamés (2004). La presencia cognitive en ciencia de la información y su entorno. *Ciência da Informação*, Brasília, 33(1): 33-7.

Martin-Lahera, Yohannis (2004). Teoria o metateoría? En el domínio usuário. *Ciência da Informação*, Brasília, 33(3): 50-60.

Martinho, Noemi O (2010). *A dimensão teórica e metodológica da Catalogação de Assunto.* Dissertation. Master's Degree in Ciência da Informação, Universidade Estadual Paulista, Marília, Brazil.

Mey, Eliane S. A (1987). *Catalogação e descrição bibliográfica:* contribuições a uma teoria. Brasília: Associação dos Bibliotecários do Distrito Federal.

Milstead, Jessica L. (1983). Indexing for subject cataloguers*. Cataloging e Classification Quarterly,* 3(4): 37-44.

Novellino, Maria Salet F. (1996). Instrumentos e metodologias de representação da informação. *Informação & Informação*, 1(2): 37-45.

Redigolo, Franciele M. & Fujita, Mariângela S. L. (2009). O Uso de Linguagens Documentárias por Indexadores em Contexto de Bibliotecas Universitárias: uma abordagem sociocognitiva com protocolo verbal. *Ibersid:* revista de sistemas de información y documentación, 3: 119-24.

Ruby, Milena P. (2008). *Política de indexação para construção de catálogos coletivos em bibliotecas universitárias.* Thesis. Doctor Degree in Ciência da Informação, Universidade Estadual Paulista, Marília, Brazil.

Silva, Maria dos Remédios da & Fujita, Mariângela S. L. (2004). A prática de indexação: análise da evolução de tendências teóricas e metodológicas. *Transinformação*, 16(2): 133-61.

Van Dijk, Teun A. (1992). *A la ciencia del texto:* um enfoque interdisciplinario. Barcelona: Paidos.

D. Grant Campbell

Classifying in the Context of Disability: Finding Potential Solutions in Existing Schemes

Abstract

This paper uses the concept of options in DDC as the foundation of a paradigm for representing topics relating to disabilities in information classification.The principle in DDC of using certain numbers as both subdivisions and base numbers enables us to imagine a classification system that enables the user to move between concentration and dispersion, representing conditions like Alzheimer's disease either as unique and separate phenomena with their own places in the schedules, or as aspects of a wide range of topics dispersed throughout the bibliographic universe.

1. Introduction

With the growing global awareness that all human institutions and activities must be adaptable, self-sustaining and capable of respecting diversity, ISKO 2016 offers us a significant opportunity to pull knowledge organization research together into a set of methods and research strategies that rise to meet this challenge.To that end, I would like to suggest that buried in the schedules and tables of DDC lies a valuable key to representing diversity in modern systems: the key lies in the principles of number-building design, particularly as manifested in Table II: the table for both geographic areas and classes of persons.The options available for using biography, particularly, as either a root number or a subdivision provide a paradigm for representing the sharp contrasts inherent in the information needs of communities who are marginalized, particularly because of physical or cognitive disabilities.

2. Disability and Marginalization

The treatment of marginalized communities in classification has attracted considerable attention in knowledge organization scholarship: attention devoted both to articulating the problem in a critically-sophisticated fashion, and to finding practical and pragmatic solutions to the problem within our existing infrastructure of information classification systems.Research into colonialism has helped us to understand the specific classification needs of indigenous peoples, and the need for indigenous ontologies (Duarte & Belarde-Lewis 2015, 677), while queer theory has heightened our awareness of the needs of LGBTQ communities (Drabinski 2013, 94; Campbell 2001).Some scholars recommend user education as a means of training users to treat the catalogue and its classification standards dialogically, rather than as neutral and unbiased representations of objective reality (Drabinski 2013, 94).However, other scholars remain committed to addressing the problem of bias in existing schemes.Many proposed solutions involve amplifications to our systems, in the form of expansions to existing tables and schedules (Green 2015, 211), or external supplements to existing schemes (Sahadath 2013, 15; Olson 1998, 246).This presentation proposes to focus, not

on the explicit room given in classification schedules, but rather on the implicit conceptual power of the options built into classification schedules such as DDC.

3. Options in DDC

Biography in *DDC* is represented by the –092 subdivision in Table I.According to the Manual, biographies are typically classified with the subjects with whom the person was most frequently associated.According to this method, the lives of individuals are dispersed throughout the collection, on the sound assumption that most subjects support a biographical treatment, and that interest in a person's life is generally associated with an interest in the subject in which that person was immersed.However, *DDC* provides three options, including the option of classing individual biography in 92: Biography and Genealogy.This simple option enables a library to treat the human life as a primary genre, and subdivide stories of human lives according to the main Dewey classes, giving us the lives of philosophers in 921, of religious leaders in 922, people in social sciences in 923, and so on.

4. The Concept of the "Gyre"

For many years now, classification research has been moving away from the entrenched, enumerative structures that informed early twentieth-century practices.The work of Ranganathan and of the Classification Research Group spearheaded the ongoing development of faceted classification, in which a defined facet order determines the patterns of concentration and dispersion; and information architecture has adopted that principle of facets into the flexible designs of electronic display, whereby users can select their own facet order for a particular purpose.Equally important, researchers are seeing classification schemes themselves as the products of temporal and rhythmic change. Research into the temporal dimensions of classification has highlighted the importance of tracking and investigating the patterns of evolution in classification schemes over time (Tennis, Thornton & Filer 2012).Research into the disciplinary context of knowledge organization has detected both centripetal and centrifugal forces, simultaneously pulling other disciplines into the work of KO, and sending KO outward to affect other disciplines (Campbell 2001).

The biography option in DDC is neither new, nor is it particularly notable in itself.It does, however, embody a principle that I will call "inversion": a process whereby a knowledge order enables us to concentrate knowledge inward to specific and clearly identified areas, and simultaneously to allow knowledge to spread out from a center to be distributed throughout the classification.This principle resembles the paradigm of the double gyre in the poetry of William Butler Yeats: a vision of human history in which antithetical social, spiritual, historical and moral forces interact in the form of cone-shaped vortices, where the widest part of one vortex is the site of the narrowest point of the opposite vortex (Yeats 1978).For Yeats, human history moves between these two gyres, and one cannot exist wholly within one gyre without an awareness of

the opposite gyre which is forming around it.The narrowest perspective engenders a background awareness of the wider perspective, while the most expansive vision can instantly be turned inward and contracted into a point.

Disability studies have made us aware, in recent years, of a similar dynamic acting in human life.Those who are differently abled frequently find themselves caught in binary oppositions: abled vs. disabled, normal vs. abnormal, or healthy vs. unhealthy.These binary oppositions find their way into information searching, and particularly into the ways in which information relevant to those with disabilities is organized and presented.At times, the condition that creates the disability is of primary importance; at other times, the individual, in seeking information, is also seeking to integrate him- or herself within the broader human community.The line between the "dis-abled" and the "temporarily able-bodied" (Breckenridge & Vogler 2000, 349)is a vexed and unstable one that requires analysis in social, economic and political terms as well as in medical terms (Goodley 2014).We must avoid the pervasive able-bodied/disabled binarism, and recognize that issues relating to disability affect us all.

Critical theory suggests ways in which binary oppositions can be exposed as power relationships capable of deconstruction (Olson & Fox 2010, 297).Facet analysis provides the means of giving the user support and agency in selecting a facet order.But disability studies suggest that individuals living with physical or cognitive disabilities need a device that can instantaneously orient information in opposite directions: to be able to move their bibliographic universe either before or behind the prism of the disability according to the immediate need.The biography option in DDC offers us a model for the construction of just such a system.

5. The Case of Dementia: I

Dementia is a complex phenomenon with a multitude of specific conditions and many consequences for individuals afflicted with it, their caregivers, and the health systems of countries with aging populations.Alzheimer's disease, while by no means the only form of dementia, is the particular example used in this discussion.

DDC offers three primary locations in the schedules where materials on Alzheimer's Disease may be collected:

- 616.831: Diseases of nervous system and mental disorders – Other organic diseases of the central nervous system – Alzheimer disease;
- 618.976831: Gynecology, obstetrics, pediatrics, geriatrics – Geriatrics – Geriatric neurology – Senile dementia – Alzheimer disease
- 362.196831: Social problems and social services – Social problems and services to groups of people – People with illnesses and disabilities – Developmentally disabled people; social services – Dementia – Alzheimer disease

In the library system of the University of Western Ontario, of the 26 resources indexed with the LC subject heading **Alzheimer's Disease** published since 2000, and containing MARC 082 fields, 22 had been assigned 616.831; 2 had been assigned

362.196831, 1 at 306.461, and 1 at 918.61.The London Public Library had 30 print resources in its collection with the subject heading **Alzheimer's Disease**; 7 had been assigned 362.196831 and the rest were assigned 616.831.

If we look at Alzheimer's Disease as a centrifugal phenomenon, we see that resources on that subject collect primarily at 616.831, reflecting the preponderance of resources treating the disease as a medical phenomenon, with a lesser proportion reflecting our growing awareness of Alzheimer's Disease as a social phenomenon.If, therefore, we treat the co-extensive indexing of entire resources as one "gyre," we can see the diverse resources available on Alzheimer's disease narrowing in to two points; anyone looking for information on Alzheimer's Disease would be directed, appropriately, to 362.196831 or to 616.831.

6. The Case of Dementia: II.

Examining the reverse gyre requires recognizing that individuals whose lives have been touched by Alzheimer's Disease find themselves dealing with a wide range of issues.If we examine the content of texts on Alzheimer's disease, we find allusions to subjects and disciplines that spread well beyond the two numbers that form the point of the gyre.Take, for instance, *The Experience of Alzheimer's: Life through a Tangled Veil*, by Steven R. Sabat (2001).Indexed at 362.19683, the book addresses, from the perspective of psychology, the many ways in which individuals with Alzheimer's disease preserve their sense of self and worth.The Index contains 40 subheadings, under the entry "AD Sufferers."If we assign DDC numbers to these subheadings, we find the subject matter in this one resource spreading out beyond its original number in at least three phases.

6.1 Phase One: Alzheimer's as Problem

The first phase gives us topics which are closely connected to our understanding of Alzheimer's disease as a medical condition, with consequences that are perceived and addressed in a medical or health-care context.These topics, while lying beyond the numbers specifically allotted to Alzheimer's disease, nevertheless lie near it.

Table 1: Topics Representing AD as "Problem"

Index Entry	DDC Topic	DDC Number
Abnormal proteins in	Proteins	612.01575
Awareness of reality	Conscious mental processes	612.8233
Brain pathology	Brain diseases—Incidents of	614.598
Decision making	Decision-making and use of information	658.5036
Speech problems	Speech and language disorders	616.855
Variability of symptoms in AD sufferers	Alzheimer's disease	616.831
Word-finding problems	Speech and language disorders	616.855
Writing problems	Writing	616.8553

All of these entries suggest a medical approach to Alzheimer's, treating it as a specific medical disorder.Due to the diffuse nature of Alzheimer's effects over time, and the diverse ways in which it affects different individuals, information about Alzheimer's Disease often leads one into discussions of related medical issues, including brain disease, capacity for decision making, and speech and language disorders, including word-finding problems and problems writing.

6.2 Phase Two: Alzheimer's as Challenge

The second phase moves beyond medicine to address ways of meeting the challenges of Alzheimer's disease on a variety of fronts.The concern remains with helping individuals with Alzheimer's disease and their caregivers; but the focus has shifted to the broader experiences of living.

Table 2: Topics Representing AD as "Challenge"

Index Entry	DDC Topic	DDC Number
Cost of treatment for AD sufferers	Health services and financial management	362.10681
Discourse with AD sufferers, as aide to understanding	Languages for special purposes	401.47
Dysfunctional social interaction with AD sufferers	Social interaction between groups	302.4
Reading problems	Students with reading disorders	371.9144
Support groups	Self-help groups	361.43

These headings implicitly acknowledge that a well-rounded and informed perspective on Alzheimer's disease requires the researcher to draw on other disciplines that impinge on the specific problems identified in the first phase.Thus, we have headings that draw on the theory of language, on the principles of financial management, and on social research to apply broader perspectives to the specific phenomena associated with Alzheimer's disease.The headings also suggest that Alzheimer's disease presents us with phenomena that we've encountered elsewhere.Many groups suffer from reading disorders; many support groups face similar challenges and rewards; managing the health care implications for any widespread health challenge involves financial planning on a grand scale.

6.3 Phase Three: Alzheimer's as Life

The third phase moves us beyond the disease itself, to raise concerns and questions that affect everyone.Alzheimer's disease is no longer the specific focus; rather we are concerned with the questions that experience with the disease trigger within us about the nature of life and the human condition.

Table 3: Topics Representing AD as "Life"

Index Entry	DDC Topic	DDC Number
Anxiety	Anxiety	152.46
Critical personalism	Language—Psychological aspects	401.9
Competency	Judgment	153.46
Coordination	Coordination	152.385
Nonverbal communication	Nonverbal communication	302.222
AD sufferers as semiotic subjects	Meaning, interpretation, hermeneutics	121.68
Sense of humour	Wit and humour	152.43
Sense of isolation	Isolation	302.545

At this broadest end of the gyre, we find that an examination of Alzheimer's disease has led us into questions that confront us all: the nature of anxiety, the pain of isolation, the psychology of language, judgment and coordination, the mysteries of nonverbal communication, the nature of meaning, and the nature of wit and humour.These headings implicitly acknowledge that Alzheimer's disease, and those who live with the disease, carry implications that affect us all, and can shed light on mysteries that touch us all.

7. Conclusion

This preliminary study does not attempt to sell DDC as the ideal means of representing inversion in the treatment of marginalized populations or topics relating to disability.Nor, on the other hand, does it intend to arraign DDC for defects in its treatment of such populations or topics.Rather, I want to suggest that the long-familiar pattern of using numbers as both base numbers and subdivisions could serve as a potential model for an information system that uses a similar design principle to create dynamic information displays that can toggle between a centripetal concentration and a centrifugal diffusion.By creating the means of tagging specific phenomena in a way that permits resources to be simultaneously collected and diffused, we may go some way towards representing, not just "disabilities" as such, but the complex relationship that such disabilities have with human knowledge and human communities.While it is sometimes necessary and desirable to concentrate such treatments into recognizable and visible areas, it is often equally necessary to allow marginalized groups to speak across the entire spectrum of human knowledge.

References

Breckenridge, Carol A. & Vogler, Candance A. (2000). The critical limits of embodiment: disability's criticism. *Public culture*, 13(3): 349-57.

Campbell, D. Grant (2001). Queer theory and the creation of contextualized subject access tools for gay and lesbian communities. *Knowledge organization*, 27, 122-31.

Campbell, D. Grant (2002). Centripetal and centrifugal forces in bibliographic classification research. *13th ASIS SIG/CR Classification Research Workshop,* 8-15.

Drabinski, Emily (2013). Queering the catalog: Queer theory and the politics of correction. *Library Quarterly*, 83(2): 94-111.

Duarte, Marisa H.& Belarde-Lewis, Miranda (2015). Imagining: Creating spaces for indigenous ontologies. *Cataloging & Classification Quarterly*, 53(5/6):677-702.

Goodley, Dan (2014). *Dis/ability studies*. New York: Routledge.

Green, Rebecca (2015). Indigenous peoples in the U.S., sovereign nations, and the DDC. *Knowledge Organization*, 42(4): 211-21.

OCLC. (2011). *Web Dewey*. Dublin: OCLC. Accessed through OCLC Connexion. [http://connexion.oclc.org]

Olson, Hope A. & Fox, Melodie J. (2010). Gayatri Chakravorty Spivak: Deconstructions, Marxisit, feminist, postcolonialist. *Critical theory for library and information science*, edited byG. J. Lecki,L. M. Given &J.E.Buschman. Santa Barbara: Libraries Unlimited. Pp. 295-309.

Olson, Hope A. (1998). Mapping beyond Dewey's boundaries: constructing classificatory space for marginalized knowledge domains. *Library Trends*, 47(2): 233-54.

Sabat, Steven R. (2001). *The experience of Alzheimer's disease: Life through a tangled veil*. Oxford: Blackwell.

Sahadath, Catelynne (2013). Classifying the margins: Using alternative classification schemes to empower diverse and marginalized users. *Feliciter*, 59(3): 15-7.

Tennis, J., Thornton, Katherine & Filer, Andrew (2012). Some temporal aspects of indexing and classification: toward a metrics for measuring scheme change. *Proceedings of the 2012 iConference.* Pp.311-6.

Yeats, WilliamB. (1978). *A vision.* London: Macmillan.

M. Tanti, P. Roux, M. P. Carrieri and B. Spire

Exploiting the Knowledge Organization of Health 2.0 to Create Strategic Value in Public Health - An Example of Application to the Problem of Drug Consumption Rooms in France

Abstract

Our paper presents an original new KO-related method where over 200 000 data of Health 2.0 exchanged on Tweeter, blogs, forums and discussion lists have been identified, represented and exploited to create useful strategic value for health decision maker and collective intelligence. This operation allows us to better understand the construction of a complex and highly publicized public decision in France: the long-lasting decision to experiment Drug Consumption Rooms. This KO-related method is an innovative approach in KO because it combines collection of data from various resources, including data from web 2.0, press and social media. It also combines the simultaneous use of analysis tools and knowledge representation software, in an original and iterative cycle of collection, processing and classification of information. These method from the field of KO, applied to public health, question the conventional methods used in this field, such as observation, questionnaires and interviews. It is in our sense a real methodological paradigm that we can describe as "disruptive changes" in this disciplinary field.

Introduction and state of the art

In recent years, the knowledge organization has been profoundly amended by the revolution of Web 2.0 (Hudon, 2010). Many activities from the public and the private sphere were completely upset by this revolution. This upheaval resulted also in a new way to consider the knowledge organization (Le Deuff, 2006). Especially with Web 2.0, the user has moved from the role of passive consumer to that of full player (Durieux, 2010). These opportunities led to the emergence of the principle of shared knowledge in real time over the network, whose most prominent examples are Wikipedia and Facebook (Durieux, 2010). This new mode of writing had a significant impact on the indexing methods and therefore on patterns of knowledge organization (Salaun, 2009). It's particularly in the health field that Web 2.0 has led to important social and technical paradigms: Electronic Medical Record, telemedicine, health apps on Smartphone (Laubie, 2011). According to a study conducted in France, dating from 2010, Internet is the primary source of health information for 78% of respondents, with the search engine Google in the lead (Doctissimo and Europe 1, 2010). The web has expanded the healthcare debate to new stakeholders: patients. Web 2.0 has now a participatory vocation: The patients with their own health. The patient is not only a reader or a writer. He also produces knowledge (Silber, 2009). Indeed, sharing the medical information on the social web, the patient becomes a vector of new medical practices, bringing forth a new paradigm: the Health 2.0. In this paradigm, patients participate in the organization of new forms of knowledge by providing high-value information on their health status. In particular, they exchange symptoms that give much more detailed descriptions than books, thereby discovering new knowledge. In classifying and organizing symptoms, finding a meaning or cause, they create a new set

of shared knowledge. Health 2.0 corresponds to this new form of medical knowledge organization generated by these communities which aggregate and empower the health care debate (Akrich, 2009). The Health 2.0 leads, within these communities, a break via the direct exchange of experience of disease by patients or discussions on global health concerns. In these exchanges of Health 2.0, it's not only the patients that intervene. Health staffs, physicians, experts, citizens and the media are also involved. All these actors, by these exchanges, create new knowledge and allow the construction of a collective intelligence (Levy, 1994) within these communities. This collective intelligence creates itself a strategic value for the decision-maker, a value that is not today "spotted", exploited and used for orienting public health stakeholders and improve community medicine. This value if it was "detected" and reused will permit to solve and understand some health issues throughout these communities. It would allow the decision maker to anticipate dynamic changes and some health crises. It would allow understanding the emergence of certain health concerns and permit to propose innovations for recurring or emerging health problems.

Objectives and originality of the work

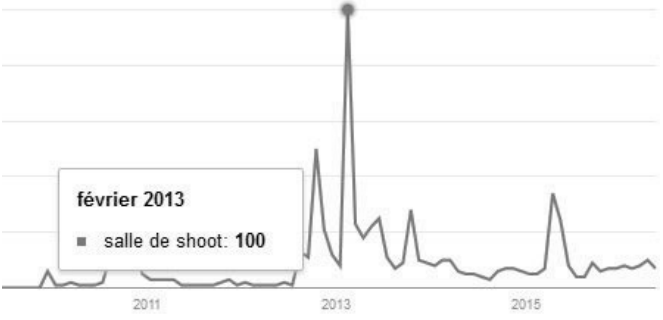
Our paper presents an original new KO-related method where over 200 000 data of Health 2.0 exchanged on Tweeter, blogs, forums and discussion lists have been identified, represented and exploited to create useful strategic value for health decision maker and collective intelligence. This operation allows us to better understand the construction of a complex and highly publicized public decision in France: the long-lasting decision to experiment Drug Consumption Rooms, low-threshold medicalized services which allows people who consume drugs to inject their own products under supervision. Though to manage with such refusal we needed to change the health law, it is crucial to better understand which was the role of people who use drugs PWUD, experts, citizens, media… and communication and how it was possible to change the positions of the French health system concerning this question. The aim is to explore more precisely where, when and how the PWUD- who may be considered as stakeholder of their own health, and not that different than any patient who is involved in decisions about health? What is the position of experts, policy makers, associations… in this debate of Health 2.0? What are the arguments and ideologies that oppose stakeholders hostile to experiment to those who are favorable. The knowledge of such information has a strategic and collective value because it can help French policy makers to better understand who are the actors involved in this construction and what are their games of arguments? It allows helping policy makers to better understand the obstacles of the implementation of innovative health measures. It would allow them to act in particular on the brakes, to shift their positions and imagine new answers to build this experiment and finally to promote debates to find new answers concerning the problematic of drug use in France. We willdescribe this new KO-related method that was used in analyzing the content of the tweets; media… concerning this

particular problematic and we will demonstrate the contribution of this method to the KO field and public health.

Methodology

In this KO-related method, the discourses concerning the problematic ofDrug Consumption Rooms in France are collected on different resources. The extraction took place over two periods: from April 2013 to September 2013 and from March 2016 to April 2016. The evolution of the discourse of the protagonists of this "hypercomplex" construction has been studied during these periods of extraction (Figure 1). For this work, we used the engine Pickanews (www.pickanews.com) that quantitatively tracks the evolution of publication of keywords on media and Tweeter.

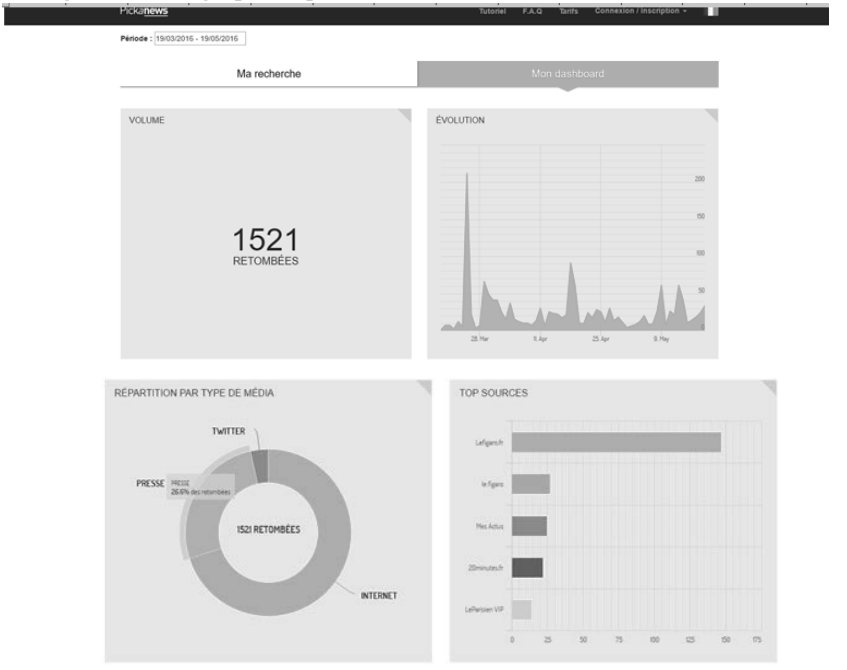
Figure 1: Representation of the evolution of the discourses visualized by pickanews



Keywords “salle de shoot”, “centre de consommation à moindre risque“, “SMCR “,“centre d’injection supervisée“,“salle de drogue“ were chosen for the research because they represent the same conceptThey have been associated with keyword “France”. With these different equations, 80 national and local French media were interviewed, from Europresse.com and Factiva.Com. The main national media interviewed were: LeMonde.fr, Lefigaro.fr, Francesoir.fr, 20minutes.fr and Metronews.fr. The main local media interviewed were: LaProvence.fr, LeParisien.fr, Ledauphine.com, and SudOuest.fr. The weekly press has also been researched with these keywords, including letempsreel.nouvelobs.com, lexpress.fr, lepoint.fr, marianne.net, valeursactuelles.com. All media content and blog posts related to these search equations were collected, cross-checked, analyzed and deduplicated. Exchanges of experts on the issue have been spotted. Knowledge was extracted and indexed. In particular, the analysis of political speeches was made. The arguments were hierarchically, temporally, geographically and thematically classified. For this classification work, we used Pickanews (www.pickanews.com). This engine permits to make quantitative analysis on 22.000 media and Tweeter. The results were collated and analyzed by type of press, media or by date (Figure 2). The engine also permits during

the study to "scan" the defined keywords and access to all identified articles.

Figure 2: Cartographic representation and data classification of Pickanews



The exchanges on web 2.0 concerning the problematic were collected particularly on Tweeter (https://twitter.com/?lang=fr) during the period of study. The collected data were analyzed. The keyword used was #salledeshoot (Figure 3). To find influencers and their networks concerning the problematic, we used the software bluenod (http://bluenod.com). As other features, this software allows to manage the twitter lists directly. It permits to export data and get more insight on hashtags. It targets topic-related communities. Finally, the software builds cartographic representation of the information and classifies the data. It helped to determine what were the political protagonists, citizens and family association involved in the problematic. What were their arguments and the evolution their discourses?

Figure 3: Cartographic representation and data classification of BlueNod

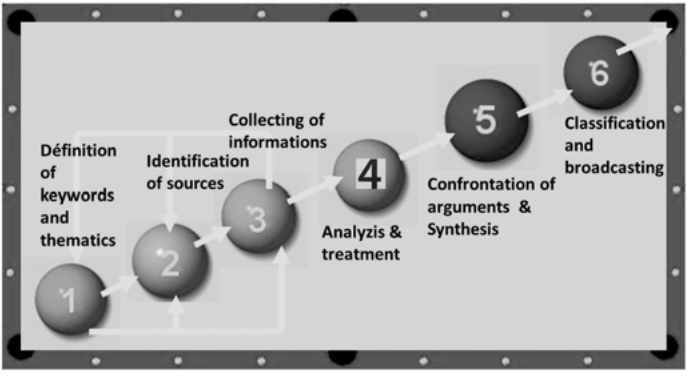
Tweets exchanges were geotagged with the software MapDTweetmap (http://www.mapd.com/demos/tweetmap/) and the top hashtags were determined. The discussions concerning the question were captured on discussion forums Doctissimo (www.doctissimo.fr) and Psychoactif (http://www.psychoactif.org/) (Figure 5). The Keywords: “salle de shoot”, “centre de consommation à moindre risque“, “SMCR “,“centre d’injection supervisée“,“salle de drogue“ were also chosen for research on these forums. The view of PWUD was particularly explored on the forum because the expression of the users was abundant and the messages were multivariate. Data were extracted and evaluated. A sentiment analysis was made and knowledge was represented, including mapping.Finally, all the speeches and arguments of the protagonists collected from these various sources have faced.

Results

The process that we have developed has made contributions in two disciplinary fields: Knowledge Organization and public health.

In KO, we have developed an original cycle of collection, processing and classification of information (Figure 4). This cycle is iterative. It comprises a phase of collection of information with different keywords, on different media resources and social web. It comprises a phase of documentary treatment. It comprises a phase of analysis and confrontations of arguments collected from these resources. It comprises a final step of synthesis for determining the protagonists of this complex construction, the influences of games and information-communication flows. In this cycle, there are feedback loops and adjustments between the different phases. Thus, the choice of keywords and search equations initially defined are refined during the researches. We especially worked by serendipity from the bibliographic references and citations of sections of the media. This has allowed us to navigate hyperlinks to hyperlinks, with navigation deep in the trees. In this process, new needs of documentary information are consistently defined causing iterations and new cycles. The objective of this process is to complete each time the knowledge gained in the previous cycles.

Figure 4: Cycle developed is the KO-related method



The KO-related method is original for another reason. It combines the collection of data from various sources. These data come from national and local media and social web, especially specialized forums and Tweeter. This method also allies the confrontation of data of these various resources. It also allies deduplication and knowledge representation by software. The originality of this method comes also from the fact that a number of tools were used: analysis tools and knowledge mapping software. All this tools were useful to exploit the data collected from Tweeter and media (Bluenod, MapDTweetmap, Pickanews). This operation was made in a concomitant manner. New knowledge from each tool was used for other tools. This use has improved collective intelligence and has created value.

In public health, the exploitation of large volumes of information from Health 2.0 and the knowledge representation are innovative methodological approaches, which are little exploited to create strategic value for making health decision and collective intelligence. For example, in our study, the extraction and analysis showed that in the field of this debate of Health 2.0, PWUD - comparable to patients, experts, policy makers, associations and citizens involved. They revealed, in the sphere of Health 2.0, a general consensus concerning the usefulness of experiment the structures of Drug Consumption Rooms, particularly to reduce the risks and health costs. It has been demonstrated a favorable opinion of the drug users concerning the opening of such structures. The main arguments are the reduction of health risk, principally, risk of overdoses. Examples of successful experiences abroad have largely been taken in testimonies. The expression of citizens and associations was found on blogs associated with daily regional and national press. A discourse analysis was performed. It was highlighted an opinion on the issue unfavorable, particularly linked with the feeling of trivializing risk of drug use and a possible drift towards the legalization of drugs. The analysis of political speeches was revealed a national right / left divergence on the issue. Left Policymakers are rather favorable to the development of such structures. The pitch analysis finds principally the reduction of health risks, including HIV transmission, overdoses and health costs. The analysis and classification of arguments of right policy makers shows a rather negative sentiment. The concept found after reorganization of knowledge, is the potential drift towards the legalization of drugs. The speeches oriented more towards the suppression and prevention, seeming more effective in the eyes of the protagonists. The knowledge representation outcome of the exploration of regional and daily media highlights a general consensus among local political actors concerning the usefulness of these structures to reduce the risks and health costs, whatever the political edge. Finally, the comparison of written media, forums and social media of Health 2.0 clearly shows an incessant legal debate and electioneering issues that push whenever any decisions. Family associations and residents who bring legal proceedings for fear of abuses to legalizing drugs or fear to see the opening of this structure near their residence. Politicians of right who are fiercely and ideologically opposed to any attempts. Politicians of left who want to try an experiment. Local elected officials from all sides that are close of the reality and field problems which have a more pragmatic discourse more focused on risk prevention and in favor of an experiment. There is a judicial evolution. A law on modernization of the health system, pushed by the Minister of Health, has just been promulgated January 26, 2016, and was approved by the State Council. This law provides the experiment of this structure in autumn 2016 in Paris and Strasbourg. In May 2017, France will know new presidential elections. Will the implementation of the experiment again pushed to the approach of this election? And even more in case of change of government? The interest of this new method is principally, for policymakers knowing who are the

stakeholders involved in the construction of national debates, knowing the obstacles to the implementation of innovative health measures. So, they can shift the political decisions to remove these obstacles and devise new health responses including the problem of drug use.

Conclusion

The exploitation of large volumes of information from Health 2.0 and the knowledge representation are innovative methodological approaches in public health, which are little exploited to create strategic value for making health decision and collective intelligence. It's also an innovative approach in KO because it combines collection of data from various resources, including data from web 2.0, press and social media. It also combines the simultaneous use of analysis tools and knowledge representation software, in an original and iterative cycle of collection, processing and classification of information. These method from the field of KO, applied to public health, question the conventional methods used in this field, such as observation, questionnaires, and interviews. It is in our sense a real methodological paradigm that we can describe as "disruptive changes" - in that it leads to question the conventional scientific methods used in this disciplinary field.

References

Akrich, Madeleine & Méadel, Cécile (2009). Exchanges between patients on the Internet. *La Presse Médicale*, 38(10): 1484-90.

Durieux, Valérie (2010). Collaborative tagging et folksonomies, l'Organisation du web par les internautes. In Du web 2.0 au concept 2.0. *Revue Les cahiers du numérique*, 6(1): 69-80.

Europe 1 et Doctissimo (2010). Enquête Les Français, le progrès médical et leur médecin. In *38e Edition du MEDEC*. Paris, 17-18 March 2010.

Hudon, Michèle& Mustafa El Hadi, Widad (2010). Organisation des connaissances et des ressources documentaires- De l'organisation hiérarchique centralisée à l'organisation sociale distribuée. *Revue Les cahiers du Numérique*, 6(3): 9-38.

Laubie,Raphaëlle (2011). Le patient connecté ou les métamorphoses de la santé. *L'Expansion Management Review*, 4(143): 24-31.

Le Deuff, Olivier (2006). Folksonomies: les usagers indexent le web. *Bulletin des bibliothèques de France*, 51(4): 66-70.

Lévy, Pierre (1994). Pour une définition de la notion d'intelligence collective-Intelligence collective: pour une anthropologie du cyberespace. Paris: Editions Découverte.

Salaün, JeanMichel& Arsenault, Clément (2009). *Introduction aux sciences de l'information*. Montréal: Presses de l'Université de Montréal.

Silber, Denise (2009). Médecine 2.0: les enjeux de la médecine participative. *La Presse Médicale*, 38(10): 1456-62.

Rosana Matos da Silva Trivelato and Maria Aparecida Moura

Alterity, Tolerance and Heterotopia: Repercussions on the Religion Science Representation in Bibliographic Classification Systems

Abstract

The orientation religious question behind the classification system has worried the experts for a while. In the specific case of the Religion class, it is observed that the same carries an inheritance Christian bias. The purpose of this article was to analyze the Religion class in the DDC and the UDC from the concepts of alterity and tolerance as constituent elements from heterotopia, as understood by Foucault. Through the religion class updates analysis in both systems, it was sought to understand the alterity recognition religious movement previously included in "Other religions". It was also attempted to identify advances and effective changes in relation to the hospitality of the systems to represent the religious diversity.

1 Introduction

> We do not live in a kind of void that would be filled with diverse shades of light, we live within a set of relationships that delineate irreducible positions to each other and absolutely impossible to be superimposed (Foucault, 2006: 415 p.)

The symbolic and or verbal language construction of the information and knowledge representation, as well as its update, involves complex meaning processes, in which the socio-historical context, the belief systems and culture have an important role. Contemporaneously, the studies that seek to analyze this signic framework repercussion on the information and representation device have been extended.

Michele Drumm's work (2000), in which the author discusses the invisibility and marginalization that the gay and lesbian theme was relegated in the Dewey Decimal Classification to the 2000s, has reflected a lot on the information professional social networks.

The Religion Science representation in the context of bibliographic classification systems is analogous to the theme highlighted by Drumm (2000), where the imbalance and washout of the belief systems of non-hegemonic alterity in knowledge systems is visible. Thus, if the discussion about alterity, multiculturalism and religious tolerance has been accentuated all over the world, it cannot be said the same of its correlate visibility in the bibliographic classification systems.

2 Problem considered

The orientation / religious tradition question behind the classification system has worried the experts for a while. According to Broughton (2000, p. 60), "the major difficulty in constructing a classification for religious literature is that of avoiding bias (real whether or apparent) toward some specific religion or denomination." Both the class ordering heritage DDC and the UDC maintained the structure of Theology/Religion classes by Catholic bias. However, after the Consortium 2000 update, the UDC searched to balance the categorization of the various religions of the world. Broughton (2000, p. 60) points out that the CDU in its new edition: "There is no

concept of value or priority attached to the order of faiths; each is regarded as having equivalent status, even where this is not reflected notationally."

In this context, it is still necessary to highlight the lack of studies on the theological libraries and theological literature area in Brazil. Most of the religious libraries doesn't provide any online catalog, nor belong to the network data unions. Therefore, it is necessary to understand how the information in the Religion Science area is being highlighted in our representation / classification information systems, at the beginning of the 21st century.

3 Theoretical foundation

3.1 Alterity, tolerance and heterotopia

The world has undergone global changes which, among other things reveal the exponential growth of symbolic wars marked by violence, religious fundamentalism and indifference conflicts of all orders.

The globalizing characters of information systems potentiate circulation and access to other cultures and belief systems demand the presence of such knowledge and sociocultural perspectives in information representation devices.

As Olson sticks out (1998, p. 234):

> […] existing literature has critiqued the most widely used classification in the world, the Dewey Decimal Classification (DDC), for its treatment of women, Puerto Ricans, Chinese and Japanese Americans, Mexican Americans, Jews, Native Americans, the developing world (including Africa, the Middle East, and Melanesia), gays, teenagers, senior citizens, people with disabilities, and alternative lifestyles.' To look at these biases with a fresh eye, a theoretical construct capable of revealing the complexities of classification and its social construction was sought.

Heterotopias are concreted or mental spaces which are found within the culture and in which the place of alterity highlights. Foucault (2006) emphasizes that heterotopia has the following characteristics: to fit the culture of any human group, to shelter spaces and positions sometimes conflicting, to adopt different workings of the existing or previous heterotopias, to articulate different times, to have opening and closing systems that isolate and / or make them penetrable and to have functions related to its surroundings.

Libraries and consequently their operating devices, among which the classification systems are understood as heterotopias of time. The author points out that

> […] are heterotopias in which time never ceases to pile up, heaping up on top of its own summit, whereas in the seventeenth century, even until the end of the seventeenth century, museums and libraries were the expression of an individual choice. By contrast, the idea of accumulating everything, the idea of establishing a sort of generalarchive, the will to enclose in one place all times, all epochs, all forms, all tastes, the idea of constituting a place of all times that is itself outside of time, and inaccessible to its ravages, the project of organizing in this way a sort of perpetual and indefinite accumulation of time in a place that will not move – well, all this belongs to our modernity (Foucault, 2006, p. 419).

Alterity is the quality or condition of being another, therefore being indispensable, freedom, social interactions, historical consciousness and cultural diversity. In this

context it is also important to recognize and respect the difference among individuals, groups and society.

Tolerance relates to the peaceful coexistence among alterity, marked by racial, religious or ethnic minorities distinctions and it is guided by the ideal of human civilization.

Intolerance, especially the religious one, is strongly enforced by fundamentalism, considered by Eco (1997) as a hermeneutical principle associated with strict and narrow interpretation of the sacred scriptures. It also has links associated with biological roots, or diffused socio-cultural motivations that frame and repel the different ones and their social representations in relation to the status quo.

Symbolic wars that we are submitted online and we watch on the screens of our smartphones, promote the culture of deterritorialization and the consequent naturalization of indifference. In these terms, knowledge risks to be restricted to the likelihood agreed within the hegemonic groups.

It is noticed then that something similar happens in the technical devices that aim to represent this knowledge. In this case, objectively or subjectively, the heterotopic alterity and their representations tend to get more and more out of these structures and this requires a vertical reflection by KO scholars.

3.2 Bibliographic classification

The bibliographic classification is an object of special interest to the Information Science and can be understood, according to Hudon (2009), as a grouping exercise of similar documents in classes which, by some adopted criteria, will be separated from the documents which they are not related.

Bibliographic classification systems have their origin in the Aristotle's classification process of knowledge and beings, and constitute themselves from a verbal instrument of information representation and how to organize and facilitate free access to the records of knowledge. In this paper it will be addressed the Dewey Decimal Classification (DDC) and the Universal Decimal Classification (UDC).

To answer the organization of information pragmatic issues, Melvil Dewey in 1876 created a practical and easily applied tool, the decimal classification system that emphasized operational means to achieve the functional arrangement of the items in libraries. Over time the system was modified and expanded, nowadays at the 23th edition.

The UDC (1902-1907), was adapted from the Dewey system retaking the DCC main classes and its decimal notation. Designed by Paul Otlet and Henry La Fontaine with the intention to organize a universal bibliographical repertoire of works published anywhere in the world since the invention of printing, it did not claim to order the books on the shelf, but to establish relations between informational data aiming at the information use.

4. Methodology

The organization processes and knowledge representation are born with the prospect of embracing in a balanced way the human knowledge production. Therefore, these processes are impregnated by worldviews as for the subject classificionist as for the librarian and which are condoned on the basis of different values and socio-historical contexts. Besides, these same individuals carry out the representation through the Knowledge Organization Systems (KOS) where concepts and situations arising from other cultures are present and other contexts very different from theirs.

The purpose of this study was to analyze the Religion class in the DDC and the UDC from the concepts of alterity and tolerance as constituent elements from heterotopia, as understood by Foucault. The adopted analytical approach highlights the emergence of religions game in the classification schemes, understanding that the bibliographic classification systems are not immune to the exclusion mode reproduction and religious intolerance of a society. In the specific case of the Religion class, it is observed that the same carries an inheritance Christian bias. In this sense, the popular religiosity theme, for example, is underrepresented in the means that it is categorized in a class entitled "Other religions" in the DDC and "Cults and minor religions" in UDC.

5. Classification and heterotopy, place of another

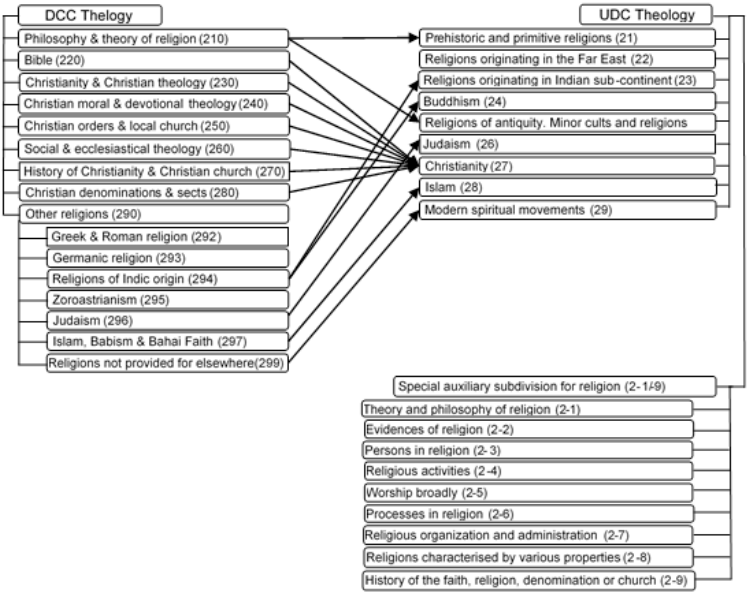
As for the UDC as for the DDC are encyclopedic classifications and, therefore constitute a bibliographic classification scheme which tries, as far as possible, be intended for universal application.

Foskett (1971) suggests that classificationists are the products of their times. In this perspective, the hierarchical structure of a classification can be perceived as a presentation of spatial images and as a mechanism that stores the time heterotopias with the ability to focus on a few discourse and make other rarefied.

Olson (1998) emphasizes that "the construction of classification to the development of spatial imagery as a metaphorical mechanism with the ability to discover the processes by which powerful and privileged discourses shape information and with the potential to inform." In this work we are interested in analyzing heterotopias in the forms of ranking Religion classes on the threshold of the 21st century.

The Religion broader classes were for a long time ranked under the Catholic bias. Seven out of these nine broader classes contemplate exclusively Christianity. A hierarchical structure, normally, enhances inference that a concept can be subordinate or inferior to the other. As in the case of the class "Other Religions", besides typifying a set of religions in a differentiated ordering, where the Christian religions stand out in a hierarchical superior level, the heading "others" show differential treatment of other belief systems.

Table 1: Broader classes



When removing the centrality of Christianity and by incorporating a set of points of views applicable to the several religions through the updates of the UDC Consortium 2000, UDC promotes a break of the DDC in relation. The religions that up to then have been relegated to a secondary plan, started being part of a broader class entitled "Religious systems. Religions and faiths" (21/29), Hinduism, Buddhism, Judaism, and Islam were inserted into the broader class and share from the same highlight as the Christianity.

The notations can be composed with the combination of religion (21/29) and with the use of facets of "Special auxiliary subdivision for religion (2-1/-9)". Broughton (2000) observes that: "When more than one term is compounded with a given religious system, the terms are added in reverse schedule order, (descending numerical order), to maintain the implicit citation order."

Example of combination: 24-788-55Buddhism, monastic orders, funeral ceremonies.

Table 2: UDC Example

Basic class	Facet introducer and principle of subdivision
24 Buddhism	-7 Organization and administration -788 Monastic orders -5 Sacraments -55 Funeral cerimonies

According to Gnoli (2011), the UDC facets appear not to base on the fundamental categories proposed by Ranganathan, expressed by the acronym PMEST (personality, matter, energy, space and time)

> […] facets are attributes that typically occur within a class […]. In more technical terms, they are subdivisions of a class by mutually exclusive criteria, each generating an array (Vickery, 1960; Ranganathan, 1967). Any phenomena can be organized into facets according to certain attributes; however, traditional bibliographic classifications are usually concerned with the facets of a discipline, such as "methods" or "operations" besides those of the objects of study, like "fur colour (Gnoli, 2011,p. 19).

The auxiliary table categories which can be applied to all religions were based on the system ancient structure and, therefore, they still show the Christianity bias. In the case of the "worship" expression, for example, it tends to generate controversy in Protestant and Catholic discourses in the means that, there are those who defend that one should only worship God. In this point of view, it is difficult to find neutral alternatives of different culture concepts, almost impossible.

It is important to point out that primitive religions and the far East religions were also highlighted in this new configuration. Already the former prescription "other religions" gave space to the "Modern Spiritual Movements".

The Modern Spiritual Movements scope note in UDC designate:

> There are a large number of these, predominantly 20th century (most of the preceding array are 19th century phenomena). They often do not have any particular faith associations, although some have grown out of a particular cultural context and may have features of the dominant religion of the culture (e.g. New Age and Paganism). Some have very large numbers of adherents, perhaps because they are 'official' state religions, but even so are not widely recognised outside their own territories (UDC online).

The new category deleted some religions from the "Other Religious" category and ended by promoting failures in recognition of alterity, to the extent that it does not accommodate in the nomination, what results in the erasure of class Religion popular religiosity.

The UDC points out the religion concept as "at the very broadest level, the religion is a system which contains add recognition of the spiritual or supernatural dimension in its view of the world". (UDC online). The principle of the religion class categorization is not clear at all; it is based on the historical time associated to some geographical locations and merged with the indication of religions set, probably the ones with the highest number of followers.

Finally, we have highlighted some elements that show how the two systems deal with the class "Other Religions".

Table 3: DDC X UDC

	Advances	**Setbacks**
DDC	Despite the fading of the non-Christian religions in Other Religion class (290), the instrument accommodates the other religions in the same category. The Religions of African origin which have already been categorized as "Minor no Christian religion" (299) are currently in "Religion not provided for elsewhere".	The Christian bias predominates, the broader classes. The specific elements of Christianity allocates the classes 220 to 280.
UDC	Decentralization of Christianity in the broader classes. The auxiliary table resource can achieve a high degree of specificity with relatively short notations.	The Other Religion class withdrawal lets the class structure without clues to allocate the other religions. The issues related to the Brazilian popular religiosity tend to be categorized as Folklore. Facets still demonstrate the Christianity bias.

Through the religion class updates analysis in both systems, it was sought to understand the alterity recognition religious movement previously included in “Other Religions”. It was also attempted to identify advances and effective changes in relation to the hospitality of the systems to represent the religious diversity in the means that it seems contradictory that religious freedom is one of the main human rights included in international treaties and, at the same time represents the one who is most often threatened by various forms of intolerance and indifference.

6. Conclusions

The carried out study started from the perspective that our task is to critically explore what can be done in the field of knowledge organization systems for the recognition of the differences in a globalised world. The premise that humans are social beings inserted in certain social and historical contexts that give shape to their identity is followed.

The religious experience is present in the whole history of humanity and the religion field is composed of numerous religions and religious practices. The belonging to a community in particular can configure imaginary geographies, as the West in opposition to the East, in the case of the opposition to Muslims, or as the evangelical opposition to the Afro-Brazilian cults.

In the face of the religious plurality differences introduce the possibility of judging and electing some religious referential as valid referential. It is necessary to be alert to the alterities fade, to the place of other which is not implemented. It is highlighted, in this context, the necessary care with the legitimacy of discourses that can lead to categorizations that accommodate erasure, the beliefsystem ranking and the knowledge representation scheme naturalization that, outside the alterities and heterotopias, can be universally applied.

Acknowledgments: Thanks are due to the CNPq and FAPEMIG in the development of this work.

References

Broughton, Vanda (2000). The need for a faceted classification as the basis of all methods of information retrieval.*Journal of Information Management*, 58 (1/2): 49-72.

Hudon, Michèle (2009). Le traitment du document. In *Introduction aux sciences de l'information*, edited by J. M. Salaün and Clément Arsenault. Montréal: Les Presses de la Université de Montréal.

Drumm, Shelly (2000). Naming the love that dare not speak its name: A look at how gays and lesbians are classified in the Dewey Decimal Classification. drum dot info. [http://drumm.info/naming-the-love/]

Gnoli, Claudio(2011). Facets in UDC: a review of current situation. *Extensions & Corrections to the UDC*, 33: 19-36.

Foucault, Michel (2006). Outros espaços. In Ditos e escritos, edited by Michel Foucault. Rio de Janeiro: Forense Universitária.

Olson, Hope A. (1998). Mapping Beyond Dewey's Boundaries: Constructing Classificatory Space for Marginalized Knowledge Domains. *Library Trends*, 47(2): 233-254.

UDC Consortium. UDC online. http://www.udcc.org/

Francisco-Javier García-Marco

Teaching Thesaurus Construction: A Top-Down Approach for LIS Undergraduate Programmes

Abstract

Thesauri constitute a great platform for hands-on learning and teaching of KOS design because they synthetize the alphabetic, terminological approach and the systematic one, flexibly allowing for the different options of hierarchical organization: terminological fields, categories, facets, and disciplines. In this paper, an alternative approach towards thesaurus construction training is discussed, with an emphasis on the difficulties that undergraduate students experience in the programmes of Library and Information Science and Information Studies. In these programmes the teaching of thesauri construction is usually very connected with indexing training and vocabulary control; and their complete design is left for advanced professional or academic courses. In both cases the usual approach is bottom-up, firstly controlling the vocabulary and later organizing the systematic schedules. The proposal in this paper is to offer a top-down approach, focused on the selection and organization of its concepts inside their disciplines, categories or facets, that is, on building its general architecture. Students build a microthesaurus in a field of their choosing, so they can acquire a sense of self-competence, which will allow them to face up confidently their professional challenges in the growing information niches where small- and medium-sized domain oriented KOS are needed.

1. Purpose

Thesauri were incorporated to LIS education programmes very soon, as cutting-edge tools for enhancing information retrieval and knowledge organization (KO). Later, the advances in latent semantic analysis (probabilistic and vector models, relevance ranking…) robbed thesauri of their protagonist role. However, being as they are classic and widely used KO tools, they have continued to be part of main trend LIS teaching programmes as specific subjects, or at least as modules inside them.

Recently, due to the ever-increasing growth of new specialized information and documentation fields and the overload of new systems in the Internet that the semantic web promotes, graduate students are increasingly required to be able to develop new small or medium knowledge organization systems (KOS), suited for emergent or very specific communities of users, and to contribute to the maintenance of existing ones, adding to them or working on their interoperability. So, the need to instruct graduate students in KOS design is becoming again a priority.

In addition to this, thesauri constitute a great platform for teaching KOS design because they synthetize the alphabetic, terminological approach and the systematic one, flexibly allowing for the different options of hierarchical organization: terminological fields, categories, facets, and disciplines…

In this paper, alternative approaches towards thesaurus construction training are discussed, with an emphasis on the difficulties experienced by undergraduate students in the programmes of Library and Information Science and Information Studies.

2. Methodology: two approaches towards thesaurus construction

Knowledge organization teaching in graduate programmes is usually focused on understanding the concepts behind knowledge organization systems (KOS) and on using these tools properly when cataloguing and indexing. This is a very important aim as LIS graduates are supposed to understand catalogues, bibliographies and other reference tools, so that they become qualified to incorporate new items to them and to exploit them for searching and reference on behalf of their users.

As a consequence, the teaching of thesaurus construction is normally very connected with indexing training. Sometimes this occurs because both themes are included in the same curricular course, and thesaurus construction is taught at the same time than indexing with a controlled vocabulary, usually a thesaurus.

In other occasions, thesauri are taught in different, sequential subjects, but the teacher chooses a bottom-up approach towards thesaurus development, more connected with indexing. Thesauri are mainly explained and built as vocabulary control tools, and students learn to interconnect terms and concepts to build up the controlled list into a network of concepts.

But, as presented in the previous section, there is a growing need to develop KOS for specific work teams, small information units and web sites, which, despite their smaller size, need also an ambitious tool capable to organizing their knowledge domain, providing a systematic overview of it. In this sense, there is a need for the students really putting hands on new thesauri from scratch, so that they can become confident to design these tools on which they usually work with as users, not as developers.

However, one of the biggest difficulties that graduate students encounter when building a thesaurus is precisely learning to select and organize its concepts in disciplines, categories or facets, that is, to build its general architecture. Frequently, they become blocked, they need a lot of help from teachers in this part and, after finishing their project, they do not usually find themselves competent enough in organizing the concepts of a specific domain. On the contrary, they come to the conclusion that it is a difficult task that they must leave in the hands of specialists. This can be useful to advocate for the existence of thesaurus specialists, but it can condemn them to be a minority in a context when they could be growing strongly, failing to use it as an opportunity.

In consequence, this paper presents and discusses an alternative approach to thesaurus design that emphasizes a top-down approach, so that students can finish the course with a strong feeling of self-competence, but without disregarding the need to pay attention to detail, carefully referring their work to the current—and excellent—standards.

3. Results

The results of the course development project can be grouped into two big sections: the lessons obtained on the evolution, advantages and disadvantages of the bottom-up and top-down approaches towards thesaurus design, construction and assessment; and the characteristics of the actual course that was designed.

Thesaurus instruction: reasons, pros and cons of the bottom-up and top-down approaches

Excellent courses have been created in thesaurus construction and development. Some are specifically devote to them as Aitchison & Gilchrist (1972) and its many successful editions and follow-up collaboration works—among them Curras (1991) with many editions sequels, translations and a great influence in Latin America—or Lancaster (1985). Others are included inside handbooks of a larger scope, as Lancaster (1972) and its successive editions and translations, or Slype (1987). Many of them were oriented toward professionals, who have a very different profile from graduate students regarding their motivations and their theoretical, practical background, and extension and quality of their encyclopaedic knowledge; but they have been successfully used as text books in undergraduate courses. An excellent revision of the literature on teaching thesaurus is available in Nielsen (2004), and very interesting references and tips are provided in Thomas (2004).

In general, these courses follow an inductive, bottom-up approach. For example, the handbook created by F. W. Lancaster for the instructors of the General Information Programme of the UNESCO in 1985—based on a course in Buenos Aires in 1978—is organized into 14 chapters:(1) purpose of vocabulary control; (2) major components of a controlled vocabulary; (3) gathering terms; (4) organizing terms; (5) the hierarchical relationship; (6) the associative relationship; (7) characteristics of descriptors; (8) the entry vocabulary; (9) scope notes and identifiers; (10) thesaurus format and display; (11) growth and updating; (12) computer use; (13) vocabulary factors affecting the performance of information systems; and (14) natural language systems. So, after an overall presentation of thesauri in the frame of vocabulary control, organizing the terms comes in fourth place after gathering terms (3rd chapter).

So, this course follows an inductive, bottom-up approach. And this is also the case with the main standards on thesaurus construction and some more recent publications on teaching thesaurus construction, as the very practical and useful by Shearer (2004) or the previous by Cabero & Castro (1997).

This approach is coherent with the origin and evolution of the use of thesauri for information retrieval. Thesauri for information retrieval evolved from post-coordinated indexing, as a way of controlling concepts and terms, and grouping them to allow for easier term selection by the indexers and for enhanced search expansion and refinement. Only after some years the systematic presentation was introduced, first hierarchical (mainly in disciplines or more specific categories, depending on the scope

of the thesaurus), and thereafter faceted (Aitchison & Clarke, 2004; Garcia-Marco, 2016).

This structure has also the advantage of going from the more elemental units − concepts, terms and relations − towards the more complex subjects, like organizing the thesaurus, editing it for publication and maintaining it.

But it has also some serious disadvantages when we consider the pedagogical context of undergraduate programmes. The topic does not have usually many hours assigned, and, as the inductive procedure takes a lot of time, undergraduate students finish with a feeling of incompetence in dealing with the more abstract, difficult task of organizing a knowledge domain. As a consequence, one of the course's main objectives − promoting self-competence − is missed.

Therefore, an innovative education project was set on to research if a top-down approach to thesaurus learning and teaching could serve to promote student self-competency in a more efficient way, inspired in the concept of Gestalt or cognitive closing. As a result, a complete new schedule and group of activities was programmed for the subject "Thesaurus construction and assessment" that is actually being taught in the Information and Documentation Graduate Programme of the University of Zaragoza (Spain), which has a student load of 6 ECTS [1]: It is an optional subject that students can choose in their third or fourth courses [2], so they are advanced graduate students, with previous knowledge of indexing, cataloguing and classification.

Course design

The course has been designed according to the philosophy and requisites of the Higher European Education Space: accounting the total student workload in 10-hour credits, establishing a set of generic and specific competences as the final educational goals, determining the corresponding learning results as operational and measurable variables, and setting a series of activities as the tool to achieve the learning results.

According to the configuration of the subject in the graduate study curriculum, students must devote a total of 150 hours, though different recounts have showed that the medium student does not reach this level of commitment, only the best ones.

The competences proposed for the course are a subset of those defined for the whole graduate programme [3]. Four competences are generic: developing autonomous learning, an orientation towards continuous improvement and innovation, better personal organization and planning skills, and promoting an ethical engagement with users and their work environment. And two competences are specific: analysis and representation of information, and organizing and storage of information. As it can be seen, thesauri construction was selected when the programme was designed as one of the subjects where creativity, innovation and ethical engagement can be better taught. This specific challenge was strongly assumed when designing the learning-teaching process.

The competences are unfolded into twelve learning results. According to them the students should be able to:

1) identify, analyse and describe the objectives of a thesaurus, its components, structure and procedures for its creation, maintenance, dissemination and use;
2) organize a knowledge domain to facilitate the retrieval of documents pertaining to it;
3) detect and argue about the implications of the selection of indexing and retrieval terms;
4) taking into account the information dissemination needs when designing a thesaurus;
5) plan and manage the construction of a thesaurus as a project;
6) assess thesauri;
7) understand and use the ISO 25964 standard;
8) build specialized thesauri using appropriate software;
9) organize their work schedule;
10) analyse the ethical implications of their decisions;
11) plan and execute their work autonomously; and
12) develop and improve their thesaurus by taking innovative decisions.

Infrastructure

ISO 25964-1 is used as the focus context and reference.

Diego Ferreyra's TemaTres (TemaTres, 2006-; Ferreyra, 2016) was selected as the helping application: a PHP open source software that can be used in networked environments and allows for many export formats: Skos-Core, Zthes, TopicMap, Dublin Core, MADS, BS8723-5, RSS, SiteMap, txt, SQL.

The software was implanted in an OS X Server 10.11, with MySQL 5.6.21 and PHP 7.0.6. Each student has a complete TemaTres installation and can invite other students to cooperate with him. The students' projects are available at ibersid.unizar.es.

Learning activities

Students must work around with selected references to build a categorized chronology of the evolution of thesauri. In this way, they gain a general overview of their place among the ecology of KOS. The suggested references are partially changed from year to year to avoid copy and paste from previous students' assignments. After completing this assignment, they are provided with additional references in case they would want to widen their state of the question (García-Marco, 2016).

Later, they are given a presentation of the previous decisions that must be taken before beginning a thesaurus project, and the main available alternatives. Whilst, they must find several potential topics that can become the subject of their project, and choose one among them. For this, they use the common vocational perspective of finding intersections among their personal interests and capabilities, and the needs of

people to whom they are related (“clients”). Thereafter, they must do an assessment of potential problems and difficulties, so they can choose a project that fits the duration of the course and their background knowledge. As thesaurus construction can be a very technical and time-consuming activity, we try at least that the subject they choose supports their motivation instead of becoming an obstacle, a common problem when there are provided with a list of compulsory subjects.

Thirdly, the students are offered a detailed exposition on concepts and terms, while they continue their work on their thesaurus preliminary decisions, which will eventually become the introduction of their thesaurus. They also have a couple of sessions to present the projects to their peers and discuss them. To increase motivation and get some approximation to real environments, students are asked to consider themselves a KOS firm, give a brand name to it, and prepare a business meeting where they will present their projects as professionals. They usually enjoy this part very much.

While they are being explained the conceptual relations, they begin their work with the sources, which must be chosen not only because they offer potential concepts, but mainly because they also provide alternative organization perspectives for the domain. While this is a departure from which is advised in ISO 25964-1, it has been found to be highly pedagogical, because students can work with the more general layers of the domain they are working with, putting them in relation with the abstracter knowledge organization tools, e.g., facets, categories and disciplinary trees. Generally, a simple Excel sheet is enough to swiftly sketch the thesaurus general structure, and enter descriptors in other languages, non-preferred terms, related terms and notes.

Thereafter, the presentation and disposition of thesauri is introduced, with some relevant, selected examples. In between, they finish the sketch of their hierarchy, and begin their work with the thesaurus application, TemaTres, entering concepts, terms, relations and notes, and working with the different presentations that are provided. Batch import using an amended Excel file is encouraged and assisted.

Finally, the students are offered with an introduction to interoperability regarding its importance, context with an emphasis on the semantic web, tools and problems, so they learn which is the next step they are expected to take to improve their thesaurus construction and maintenance capabilities. They practice exporting their thesaurus in at least one of the semantic web formats, and the student group analyses the files.

The course finishes with four deliverables: an oral presentation; a traditional thesaurus report with its introduction and the two basic presentations, alphabetical and systematic; its online version supported by TemaTres; and a detailed task report with the timespans devoted to the project, the problems encountered and the solutions provided.

4. Future developments

The current results are very satisfactory, and both the student satisfaction reports and the student projects show that the key objective of promoting student self-competence in thesaurus design is achieved. However, there is much room for improvement.

At the present moment, the practice of interoperability—which would be relatively easy using TemaTres—has not been addressed in depth, mainly because of a lack of time. At least two weeks should be devoted to explain the basics of KOS interoperability according to ISO 25964-2, and do some mapping between the thesauri that have been designed by the students. At this stage, introducing Protégée seems a very good option that should be taken into account, as Zeng (2005) has effectively shown for postgraduate courses. Protégée is powerful, well proven, interoperable, professional, and provides the greater context of ontology development. Introducing it in less advanced courses would be very formative. Thesaurus teachers must evaluate their students and context to see it this software can be adopted in undergraduate courses without sacrificing more basic aims.

Also, to allow for better teamwork and integration with the rest of the learning activities of the LIS programme, it would be very relevant to provide for import and export outputs in the most common formats for bibliographic authorities. In this way, students would be able to better connect their cataloguing practices with their thesaurus design classes.

Besides these specific improvements, the integration of alternative teaching strategies should be explored. More directional programmes as the one proposed by Irving (1995) are complementary to a project-based educational approach. They can be very useful to ensure that students do not miss any important point, and that they have completed their conceptual learning before addressing the next step.

Finally, it is well known that learners unfold different styles of learning and thinking when dealing with the subjects they must master (Sternberg, 1997). These learning styles seem to be very connected with the personality traits of the students. In particular, some persons prefer an analytical, step-by-step approach, while others need to obtain a gestalt of the field (pregnancy) to deal later with the details. This could be connected with a preference for a bottom-up or top-down approach towards thesaurus development learning. So, further research must be overtaken in this respect to inquiry if such an important personality trait has a real impact in thesaurus construction teaching and learning. Of course, such studies should be done in a controlled manner, by obtaining objective and subjective measures about the learning styles of the students.

Acknowledgements: This paper has been developed in the frame of the project "Implantación de un servidor de tesauros para el apoyo al desarrollo de metodologías activas y colaborativas en el Grado de Información y Documentación", supported by a grant of the University of Zaragoza (PIIDUZ_15_031).

Notes

[1] ECTS stands for European Credit Transfer and Accumulation System. In the Spanish case, each credit typically results into ten hours of class attendance, both practical and theoretical, and a total of 25 hours of student load per credit, including the ten an-hour face-to-face classes.

[2] Currently, general graduate programmes in Spain are four-year ones, with the exception of Medicine and Surgery. Their total learning load is 240 ECTS. Each yearly course is typically 60 ECTS.

[3] Most library and information graduate programmes in Spain follow a white book built by consensus among the existing ones in 2003-4 under the guidance of the Agencia Nacional de Evaluación de la Calidad y Acreditación (2005). Its professional competence analysis follows the results of the DECIDoc project, developed under the Leonardo da Vinci programme of the European Union (Euroguide…, 2000).

References

Agência Nacional de Evaluación de la Calidad y Acreditación (2004). *Título de Grado en Información y Documentación: Libro Blanco*. Madrid: Agencia Nacional de Evaluación de la Calidad y Acreditación. [http://www.aneca.es/media/150424/libroblanco_jun05_documentacion.pdf].

Aitchison, Jean & Clarke, Stella Dextre (2004). The thesaurus: a historical viewpoint, with a look to the future. In*The Thesaurus: Review, Renaissance, and Revision*, edited by S K Roe and A. R. Thomas. New York: Haworth Press. Pp. 5-21.

Aitchison, Jean& Gilchrist, Alan (1972). *Thesaurus construction: a practical manual*. London: Aslib.

Cabero, Manuela Moro& Castro, Carmen Caro (1997). Propuesta metodológica para la enseñanza de la utilización y elaboración de tesauros [Methodological proposal for teaching thesaurus use and construction] In*Organización del Conocimiento en Sistemas de Información y Documentación*. Pp.159-67.

Currás, Emilia, Aitchison, Jean& Gilchrist, Alan (1991). *Thesaurus: lenguajes terminológicos*. Madrid: Paraninfo.

Euroguide LIS: the guide to competencies for European professionals in library and information services. Association for Information Management. [http://www.aslib.co.uk/pubs/2001/18/01/foreword.htm]

Ferreyra, Diego (2016). *TemaTres*. [http://www.vocabularyserver.com/blog/contact]

García-Marco, Francisco Javier (2015). *25731 — Construcción y evaluación de tesauros: Guía docente para el curso 2015–2016*. Zaragoza: Universidad. [http://titulaciones.unizar.es/asignaturas/25731

García-Marco, Francisco Javier (2016). The evolution of thesauri and the history of knowledge organization: between the sword of mapping knowledge and the wall of keeping it simple.

Brazilian Journal of Information Studies: Research Trends, 10(1). [http://www.bjis.unesp.br/revistas/index.php/bjis/article/view/5786]

Irving, H. (1995). CAIT: Computer-assisted indexing tutor, implemented for training at NAL.*Agric.Libr.& Inform.Notes*, 21(4-6): 1-5.

Lancaster, F. W.(1985). *Thesaurus Construction and Use: A Condensed Course*. Paris: United Nations Educational, Scientific and Cultural Organization, General Information Programme. [http://unesdoc.unesco.org/images/0007/000703/070359EB.pdf]

Nielsen, Marianne Lykke (2004). Thesaurus Construction: Key Issues and Selected Readings. *Cataloging & Classification Quarterly*, 37(3-4): 57-74.

Shearer, James R. (2004). A Practical Exercise in Building a Thesaurus. *Cataloging & Classification Quarterly*, 37(3-4): 35-56.

Slype, George van (1987). *Les langages d'indexation: conception, construction et utilisation dans les systèmes documentaires*. París: les éditions d'organisation.

Sternberg, Robert J. (1997). *Thinking styles*. New York: Cambridge University Press.

TemaTres (2006). TemaTres: controlled vocabulary server. *Sourceforge*. [https://sourceforge.net/projects/tematres]

Thomas, Alan R. (2004). Teach Yourself Thesaurus: Exercises, Readings, Resources. *Cataloging & Classification Quarterly*, 37(3-4): 23-34.

Zeng, Marcia Lei (2005). Using software to teach thesaurus development and indexing in graduate programs of LIS and IAKM. *Bulletin of the ASIS&T*, 31:6: 11-3.

Katarzyna Materska

Knowledge Organization in University Repositories in Poland

Abstract

The overall purpose of this study is to depict the current state of Polish university repositories in the context of terminological tools used for subject access, from the knowledge organization point of view. This research adopts an inductive method. For the purpose of the research the group of 12 academic (institutional) and 3 departmental repositories were selected as the initial objects of study. The author attempts to considerthe quality of keywords representing the repository content and its implications. Most of this terminology is proposed by the researchers who are the authors of the deposited documents. On one hand, keywords are accurate and precise terms created by the professionals in their domain. On the other hand, one can easily identify a lot of inconsistencies and information noise. This paper highlights the necessity of improving the content descriptions and search functions of repositories as a next step in the development of Polish university repositories.

Introduction

The computer networks have created an excellent environment for the transfer of knowledge. Nowadays, institutional repositories are indicated as key tools of the scientific and academic policy of the university. They have become an essential infrastructure for exposing the intellectual inventory of research institutions (Lynch, 2003).

The development of institutional repositoriesbegan with researchers within the field of computer science, making theoretical and technical fundaments for that type of knowledge platforms. Next, practitioners and researchers within the information science and library science became more involved with the development and use of institutional repositories (Stevenson, Zhang, 2015). Current literature in the information and library science identifies institutional repositories as a growing phenomenon, with an increasing number of libraries planning to implement this kind of specialized digital collections - environments for the transfer and exchange of knowledge.

"The Institutional Repository (IR) is understood as an information system that collects, preserves, disseminates and provides access to the intellectual and academic output of the university community" (Guidelines, 2007, 61). Access to full texts of digital objects (documents) makes the repository a fundamental support tool for teaching and research, and at the same time multipliesthe visibility of a given institution in the international community (global arena), enhances the educational competitiveness of the university and offers preservation features for future generations.

The repository supports faculty efforts to discover and communicate new knowledge as well as helps students, PhD students and faculty to build impressive, lasting digital portfolios. Sapa (2010) writes that the institutional repositories play similar role as any KO system in academic organizations, that is, the preservation and dissemination of the entire knowledge produced by an organization. Scientific data repositories offer a unique environment where knowledge organization and personal information management converge.

Most universities have created their own customized repositorieswith information generated by their academic teachers and researchers (Almuzara, 2012). Usually there is a

variety of methods used in repositories to represent the content of the documents. Both main (conflicting?) tendencies in KO: standarization and semantization (Sosinska, 2015) are present in academic repositories.

Information professionals create repositories that are based on traditional knowledge organization theories and schemas, which rely on the unity and order. However, they do not reflect true life and reality, and there has been an increasing support for the idea that disorder, as opposed to order, is a more realistic representation of the way the real world is organized (Weinberger, 2007; White, 2012).

"Because the repository's primary mission is to disseminate the university's or research institution's primary output, some repositories have seemed to forget that researchers - rather than institutions - are the most important users of repositories" (Aschenbrenner et al., 2008).

Aims

The overall purpose of this study is to depict the current state ofPolish university repositories in the context of terminological tools used for subject access, from the knowledge organization point of view. The success of the repositories usually depends on well-organized content collections for effective dissemination, maintenance and preservation of resources.

Repositories have a common set of metadata defined with Dublin Core - the schema for metadata most often employed for the description of digital objects held by Polish institutional repositories. The creation of metadata should facilitate retrieval from the collection, which is more than only browsing through hundreds or thousands of digital resources. In the following discussion the author attempts to considerthe quality of keywords representing the repository content and its implications. Most of this terminology is proposed by the researchers who are the authors of the deposited documents. On one hand, keywords are accurate and precise terms created by the professionals in their domain. On the other hand, one can easily identify a lot of inconsistencies and information noise. It is worth commenting that a growing understanding of knowledge dependence on social, cultural and pragmatic contexts of its production and application has its adherents (Hjørland, 2013; Kwasnik, 2010). The systems discussed in the paper are thus always more or less based on a certain view of the domain being organized.

The final goal of the paper is to propose some recommendations for Polish university repositories that would enable overcoming the afore-mentioned issues.

Methods

This research adopts an inductive method. For the purpose of the research the group of 12 academic (institutional) and 3 departmental repositories were selected as the initial objects of studyfrom CeON Aggregator national directory (a single access point to Polish open access repositories). All those repositories are registered in OpenDOAR international directory as well.

There was an invitation to participate in a call survey sent to the repository managers (repository editors, responsible for the approval of document descriptions). In order to accomplish the research taskthe keyword-based activities of two groups of repository users were analyzed: those of the experts, i.e. researchers (self-archiving and describing their own publications) and those of the end-users (searching, browsing and retrieving the repository content). In many cases academics were both the content creators and theend-users. The qualitative and comparative analysis of the collected material is the basis for the recommendations and conclusions.

Scholarly communication users and content creators

"Digital repositories have been with us for more than a decade, and despite the considerable media and conference attention they engender, we know very little about their use by academics." (Nicholas at al, 2012, p. 195).If one browses Polish professional literature on this subject, the observations are very similar. There is little detailed information available on the use and usefulness of repositories in general and resource discovery in particular.This paper is a step towards filling this gap in some sense.

Repositories have numbers of various potential users (both human and machine ones)who navigate via the appropriate institutional websites and/or rely on global search engines such as Google Scholar to locate relevant objects. Academics are both authors and content creators of repositories as well as the largest group of users of their content (via searching and resource discovery).

The practice of auto-archiving, where an author uploads a copy of his/her paper at an open accesswebsite,has significant implications for the way of describing and searching objects in the repository. In the majority of cases, keywordsare the option selected by the authors who deposit their publications.

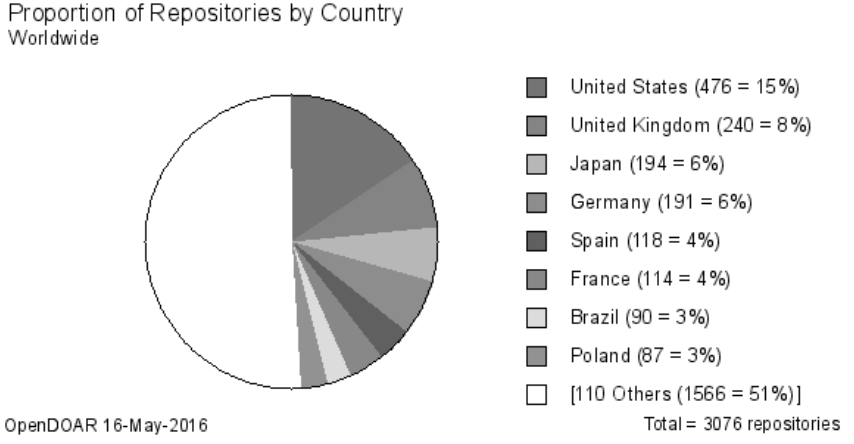
In the traditional search (keywords-based search) users search for the occurrences of particular text strings within available metadata. "Keywords-based queries can often be ambiguous, due to the polysemy of terms used in the query, thus leading to the low precision; they can also be affected by synonymity between query terms and the contents of the search corpus, thus leading to low recall." (Solomou & Koutsomitropoulos, 2015). These problems are clearly visible in the case of Polish university repositories.

University repositories in Poland – research context

More and more Polish universities are involved in the process of creating repositories for e-prints authored by the academic staff of their parent academic institutions. The traditional paradigm of scholarly communication (with scholars creating new knowledge and other professional parties responsible for its editing, distribution, collecting, retrieval, preservation and availability) gradually fades. There is a growing interest and pressure from scholars to have the possibility for scholarly publishing under the *greenroad OA* policy (self-archiving papers published in non-OA journals).

OpenDOAR, which is an authoritative directory of academic open access repositories existing all over the world, lists 87 Polish open access repositories (Fig.1), which, in fact, in the majority of cases, are digital libraries focused on preserving national and cultural heritage as well as collecting regional publications and objects used for teaching and learning processes. For that reason they mostly contain historical materials, digitized valuable items from the library special collections. They are not focused on supporting the scholarly communication or providing access to the intellectual output of institutional (university) communities.

Fig 1. Repositories by country



Afore-mentioned 87 open access repositories/digital libraries include 12 university (multidisciplinary) digital repositories and 3 departmental repositories. All of them are registered in Polish national directory called CeON Aggregator (http://agregator.ceon.pl/). All repositories selected for the research purposes are presented in tables 1 and 2.

Table 1: Polish university repositories (as of May 20, 2016)

	Repository	Number of documents
1	Repozytorium Uniwersytetu Jagiellońskiego (Jagiellonian University Repository)	25245
2	Repozytorium Uniwersytetu Łódzkiego (RUŁ) University of Lodz Institutional Repository	16206
3	Repozytorium Uniwersytetu im. Adama Mickiewicza w Poznaniu AMUR (AMUR - Adam Mickiewicz University Repository)	13103
4	Repozytorium Politechniki Krakowskiej (Tadeusz Kościuszko Cracow University of Technology Repository)	8075
5	Repozytorium Uniwersytetu w Białymstoku (University of Bialystok Repository)	611
6	Repozytorium Uniwersytetu Kazimierza Wielkiego w Bydgoszczy	3242

	(Repository of Kazimierz Wielki University in Bydgoszcz)	
7	Repozytorium UMK Nicolaus Copernicus University Repository	3083
8	Repozytorium Instytucjonalne Krakowskiej Akademii imienia Andrzeja Frycza Modrzewskiego (Andrzej Frycz Modrzewski Krakow University Repository)	866
9	Repozytorium instytucjonalne WSB NLU (WSB NLU Institutional Repository)	785
11	Repozytorium Uniwersytetu Przyrodniczo Humanistycznego w Siedlcach - RepoS (Repository of Siedlce University of Natural Sciences and Humanities – RepoS)	356
12	Repozytorium Publikacji Naukowych Politechniki Śląskiej (Silesian University of Technology Digital Repository)	298
13	Repozytorium Dolnośląskiej Szkoły Wyższej oPUB (Repository of University of Lower Silesia oPUB)	254

Table 2: Polish departmental repositories (as of May 20, 2016)

	Repository	Number of documents
1	Repozytorium Eny Politechnika Wrocławska (Electrical Engineering Repository, Wroclaw University of Technology)	31
2	Repozytorium IBB PAN IBB PAS Repository (Institute of Biochemistry and Biophysics – Polish Academy of Sciences)	636
3	ECNIS Repository (Environmental Cancer Risk, Nutrition and Individual Susceptibility Repository)	604

All university repositories in Poland use international standards for describing digital materials (Dublin Core), and provide them in globally accepted formats (e.g. DjVu, PDF). All of them were developed and are run by the academic libraries of their parent universities.

Searching and browsing – organizing access to repository items

One of the fundamental questions of this research concerns the organization of access to repository resources.The ease of finding relevant information depends on the effectiveness of knowledge organization. All Polish repositories offer the possibility of browsing. The most common options are: titles, authors, publication date, item type, affiliation, keywords. Cracow University of Technology does not provide keyword browsing (options available: collection type, faculty, recently published, with search options: author, title, anywhere). The main focus of this paper is not on “descriptive metadata” (such as title, author, and date), but on keywords, that is, subject terms selected to describe the aboutness feature of the scientific objects in the repository.

One should note here that there are two terms used in Polish repositories – keywords and subjects – although both refer to the same Dublin Core metadata field, i.e. dc.subject.

Keyword/subject indexes usually include mixed Polish and English terms. Only the Jagiellonian University Repository distinguishes three types of indexes: keywords in Polish, keywords in English, keywords in other languages; the latter are provided in the language of the document (respectively in Dublin Core: dc.subject.pl, dc.subject.en, dc.subject.other).

In Polish university repositories there are two methods of adding keywords to the document descriptions: 1) keywords are added by the scholars while depositing their papers; 2) keywords are added by the repository managers (editors).

The consequences of those methods are visible at a first glance. The lack of any authority control of the keywords added by the scholars results in:

- the presence of numerous forms of the same term, sometimes even misspelt, e.g.:
- biznesplan [10] biznes plan [45] biznes-plan [4] – WSB NLU Institutional Repository;
- współpraca przedsiębiorstw [3], wspólpraca przedsiębiorstw [1] (the latter expression misspelt) – WSB NLU Institutional Repository;
- the presence of the same term in singular and plural form:
- academic libraries [2] academic library [2] - AMUR - Adam Mickiewicz University Repository;
- the extensive scattering of the vocabulary: most often each term is used to index 1-2 titles.

If some control is provided by the repository manager, the consistency of the vocabulary is higher and obvious mistakes are corrected immediately.

So far only one university repository (Andrzej Frycz Modrzewski Krakow University Repository) has introduced an additional (next to the keyword one) index, that is, tematy (topics), reflecting dc.subject.other Dublin Core field. The topics index is a quasi-controlled vocabulary covering the names of study fields offered at the university. It contains general terms such as architecture, education, ecology, economics, philosophy, etc. Topic terms are created and added by the librarian managing the repository to the descriptions of documents based on the analysis of keywords suggested by the academics depositing their papers and the document summary. The vocabulary used in the topic field is expanded by the librarians when there is a growing number of publications on a given subject.

The lack of a controlled vocabulary does not appear problematic to all librarians working for the repository. Their observations and computer statistics show that the most important access point is title (the suitable and informative title plays extremely significant role).

All survey participants agree that the majority of non-university users find and access the repositories from external sources and tools such as Google or Google Scholar search engines. This observation is further confirmed by the repository statistics. This means that in their opinion the development of controlled vocabularies (thesauri) would bring more losses than benefits, in particular when considering the institutional repositories with universal vocabularies covering a variety of science fields.

The need for controlled or semi-controlled vocabularies is more visible in the case of narrow-scope repositories. IBB PAS Repository (Institute of Biochemistry and Biophysics – Polish Academy of Sciences) can be searched both with uncontrolled keywords and Library of Congress subjects: G Geography. Anthropology. Recreation; L Education; Q Science; R Medicine; S Agriculture; T Technology; Z Bibliography. Library Science. Information Resources.

Search options in all discussed repositories include full-text searching. Very broad results of that type of search rarely are ordered by relevance, which may be problematic for end-users. Moreover, most repositories offer advanced search with the possibility of selecting indexes. Yet if the user does not know the name of the author or the title of the publication, he/she may have troubles with keyword search, in particular due to the lack of the synonymy relationships and mistakes/errors discussed earlier.

The drawbacks of the searching apparatus result in non-university users accessing the repositories occasionally, usually driven by the scientific or general search engines. The university repositories are far more efficiently explored by the researchers who know the names of authors (their colleagues) and the titles of their papers.

Discussion

The university repositories have been developing recently in Poland as new platforms for knowledge dissemination. Numerous factors confirm that the librarians' observations and experience as well as the postulates formulated by the researchers would tend towards the change of knowledge organization in the repositories. For many years some researchers have been arguing that the way knowledge is currently organized by the information science is not useful to the expert researchers and that new approaches should be taken (Szostak, 2004, Weinberger, 2007).

Who should create subject terms? This question seems to be an important one and should be addressed during the repository design phase – should they be created by the information professionals or the authors of the papers or should they rather be automatically generated. Information professionals have more experience using metadata, while scientists are better at applying detailed scientific terms.

On one hand, there is a lot of evidence that the creation of subject terms based on personal preference or professional authors' long-term habits is a good direction to follow. Subjects move from very broad and unprofessional to more specific ones (sometimes too specific, if one considers a non-university recipient). On the other hand, it is clear that there is a need for some systematic control run by the information professionals. Thus the conciliation and cooperation appear to be the best solution, irrespective of whether one considers general or domain-specific repositories.

Conclusion

Though the university repositories growth rate has increased, the research on repositories through the terminological perspective is very modest. This state of affairs is rather not

surprising, as the introductory stages of the repository development were focused on the organizational, legal, promotional and other practical dimensions, not the terminological aspects and possibilities for content browsing, searching and discoverability.

The empirical analysis of subject description of the university digital collectionsindicates that Polish repositories are not excellent in the sense of knowledge organization, navigation and effective retrieval. Simplification tendencies in contemporary organization of digital knowledge resources are evident. One of common features of nearly every repository system is the use of very simple structure of the university hierarchy (faculties and institutes)and the navigation system through its content (with titles, authors, publication date, item type, affiliation, keywords). As the corpus of deposited documents increases in size, it gets progressively harder for a user to find the needed information (and documents) using uncontrolled vocabulary (keywords) as it do not ensure a precise representation of indexed and classified content.

This paper highlights the necessity of improving the content descriptions and search functions of repositories as a next step in the development of Polish university repositories. Repository services should focus on the stewardship of content that has long-term value to the university and facilitate current access for all users from the global network, irrespective of their expertise in the domain and searching methods. Yet one "should forget about universalistic systems for everybody, and try to build consistent KOSs instead, which would be a picture of communication in various domain and discourse communities" (Wozniak-Kasperek, 2014, 310).

Is there any good method of implementing such a task? The search for more effective methods and tools for organizing access to the digitally recorded knowledge of the university becomes imperative for knowledge organization.

References

Almuzara, Leticia, Díez M. Luisa & Bravo Blanca (2012). A Study Of Authority Control in Spanish University Repositories. *Knowledge Organization,* 39 (2): 95-103.

Aschenbrenner, Andreas,Blanke, Tobias, Flanders, David, Hedges, Mark &O'Steen, Ben (2008). The Future of Repositories? Patterns for (Cross-)Repository Architectures. *D-Lib Magazine,* 14(11/12). [http://www.dlib.org/dlib/november08/aschenbrenner/11aschenbrenner.html]

CeON Aggregator [http://agregator.ceon.pl/]

Guidelines for the creation of institutional repositories at universities and higher education organisations (2007). Alfa Network Babel Library. Valparaiso: Columbus: Europe Aid Co-Operation Office: Babel Library. [http://eprints.rclis.org/13512/2/Guidelines_IR_english.pdf]

Hjørland, Birger (2013). Theories of Knowledge Organization – Theories of Knowledge. *Knowledge Organization,* 40 (3): 161-81.

Kwasnik, Barbara H. (2010). Semantic Warrant: A Pivotal Concept for Our Field. *Knowledge Organization,* 37(2): 106-10.

Lynch, Clifford A.(2003). Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age. *ARL: A Bimonthly Report*, 226: 1-7.

Nicholas, David, Rowlands, Ian, Watkinson, Anthony, Brown, David, Jamali Hamid R. (2012). Digital repositories ten-years on: what do scientific researchers think of them and how do they use them? *Learned Publishing,* 25 (3): 195-206.

Open DOAR - *The Directory of Open Access Repositories.* [http://www.opendoar.org/]

Sapa, Remigiusz (2010). Access to Scholarly Output of Academic Staff: Bibliographic Databases and Institutional Repositories in Polish Academic Libraries. *Libri*, 60(1): 78-91.

Solomou, Georgia &Koutsomitropoulos, Dimitrios (2015). Towards an evaluation of semantic searching in digital repositories: a DSpace case-study. *Program: Electronic Library & Information Systems*,49(1): 63-90.

Sosinska, Barbara (2014). Semanization and standardization – cooperative or conflicting trends in knowledge organization? In*Knowledge Organization in the 21st Century: Between Historical Patterns and Future Prospects: Proceedings of the Thirteenth International ISKO Conference,* edited by WiesławBabik. Krakow, Poland, 12-22 May 2014.Würzburg: Ergon Verlag. Pp. 580-7.

Stevenson, Jennifer Ann & Zhang, Jin (2015). A temporal analysis of institutional repository research. *Scientometrics*, 105: 1491–525.

Szostak, Rick (2004). *Classifying science: phenomena, data, theory, method, practice*. Dordrecht, Netherlands: Springer.

White, Hollie C. (2012). *Organizing scientific data sets: Studying similarities and differences in metadata and subject term creation*. Chapel Hill. [https://cdr.lib.unc.edu/indexablecontent/uuid:f00840f9-ace4-4920-b3cf-233ac149fbad]

Weinberger, David (2007). *Everything Is Miscellaneous: The Power of the New Digital Disorder*. New York: Henry Holt and Company.

Wozniak-Kasperek, Jadwiga (2014). Terminology as a picture of knowledge organization in a scientific discipline. In*Knowledge Organization in the 21st Century: Between Historical Patterns and Future Prospects: Proceedings of the Thirteenth International ISKO Conference,* edited by WiesławBabik. Krakow, Poland, 12-22 May 2014.Würzburg: Ergon Verlag. Pp. 305-11.

Camelia Romero Millán and Catalina Naumis Peña

Representation of Contents on Female Participation in Salaried Work

Abstract

The research presented here was carried out to learn about the behavior and response of controlled languages within the field of Gender Studies (GS). It focuses particularly on the subject of female participation in salaried work in Mexico by means of analysis of a sample of 110 documents on the topic that were written in Spanish and published between 2000 and 2010. The documents were culled from library catalogs in Mexican institutions of higher education that undertake GS research, thus ensuring results. In an initial phase, to identify the terminology most frequently used by authors, the automatized indexing method and Zipf mathematical model were applied and the Goffman transition point was calculated. The second phase involved using the comparative method to identify terms not found in the controlled languages. The results are meant to contribute to the enrichment of controlled languages, closing terminological voids that, in this case, exist in Mexico with regard to what is known as emerging topics, within indexing tools.

1. Introduction

This research is focused on finding out how far the organization of knowledge through construction of controlled languages (subject and thesauri headings) has responded to the progress and expansion of Gender Studies (GS) as a field of study. The purpose of language in representing contents is to respond to the phenomena studied in science within an environment where catalogs tend to become generalized and incorporated through interconnected data and interoperability of languages.

In general, terminological exploration is carried out through recovery of documents within a subject area. In this case, the terminology used by authors is contrasted with the controlled languages and thesauri used in library catalogs in order to define the existence or lack thereof of local or specialized terms that are or are not reflected in the catalogs of Mexican libraries specialized in the topic.

In Mexico, the libraries of the Instituciones de Educación Superior (IES, Institutions of Higher Education) have developed GS collections in response to interest that the subject has stimulated within the academic community.

Given the new terminology constantly being coined by authors to express themselves and that turns into specialized language, it is essential to consider incorporating it into controlled languages, which are tools for indexing information sources and making them retrievable.

2. Characteristics and Theoretical Foundation of the Research

The study concentrates on observation, analysis and evaluation of the language used to classify documents in Spanish dealing with the participation of women in salaried work in Mexico. Considered one of the most recurrent topics in GS, the subject was further chosen given the impossibility of covering all aspects of the discipline.

The research is of a theoretical-qualitative and experimental nature, its scope defined as exploratory, descriptive, comparative and explanatory. Descriptive studies are useful "…to precisely show the angles or dimensions of a phenomenon, event, community, context or situation" and describe tendencies of a group or population (Hernández-Sampieri, Fernández Collado and Baptista-Lucio, 2010, p. 84-85). In the case of this research, the qualitative properties of the terminology used in the discourse of the authors are described. Such terminology was compiled in previously selected documents to compare them with indexing terms used in library catalogs.

From the perspective of the examination of a little studied or novel topic, the exploratory element enables gradually appropriating and interpreting the scanty specialized literature on it, unlike previous studies dealing with terms that already exist in the controlled languages. The latter studies show that the problem lies in the lack of adequate terminology within the linguistic indexing tools used to describe document content, in line with what the authors who developed it expressed and conclude on the existence of a "terminological void" (López-Huertas and Torres, 2005).

The theoretical basis behind this research lies with the principle of literary guarantee upon which the work is justified, in addition to the importance of recovering local terminology to describe, compare and explain studies done on the topic (Maniez, 1987, Lancaster 1986, Svenonius, 2003, Naumis, 2000). What follows is a brief discussion of the main texts used as support: In his *Principles of Book Classification,* published in 1950, Wyndham Hulme proposed bases for literary guarantee and explained that validation of headings should be the result of verifying their existence in books within the specialty and not of philosophical or theoretical models (Lancaster, 1986).

In 1951, Mortimer Taube and Albert Thompson introduced the selection of individual terms from an original document to represent different facets of its content. Meanwhile, Hans Peter Luhn applied statistics to identify recurring words in texts, and from there to create Key Words in Context (KWIC). One of the pioneers in the field of information representation and retrieval, Calvin Mooers coined the term "information retrieval" in 1950 (as cited in Chu, H. (2007).

Gerard Salton, creator of the System for the Manipulation and Retrieval of Text (SMART), separated relevant words from the title and abstract of a sample of documents –words from free language– and determined the value of the words as a function of the number of appearances in the document, which was called automatic selective indexing. In short, Salton proved that the frequency of free-language words determines their value as a function of the number of appearances in a document and facilitates identification of semantic nuclei around which thematically similar words relate (as cited in Chu, H., 2007 and Maniez, 1987).

Another principle that supports this paper is what terminology has called the theory of doors, in other words that for the sake of uniformity and in order to ensure univocity,"…The denominative and conceptual diversity of reality is silenced." (Cabré,

T. 2002) Cabré mentions that formerly, general terminology theory was limited to ensuring univocity in professional communication as of normalization, which allowed for having a common language to overcome the restrictions of internationally extended language.

Within the culture of each country, language is determined by such factors as age and the speaker's gender, education and daily activity (Lara and Ham, 1979). Hence, the terminology used in Mexico on the topic being studied reflects precisely the way in which women are positioned with respect to work, education and legislation, among other factors. As a result of this, the new theory of doors that conceives the term as "…a unit formed by three different aspects: a semiotic and linguistic aspect, a cognitive aspect and a communicative aspect" was taken into account. Such units convey specialized knowledge but occur through the practice of speakers; it is natural language (Cabré, T., 2002). This principle is complemented by the approach of Marshall; Olson and Schlegl (1999), and López-Huertas and Torres (2005), in relation to the sacrificing of identity for univocity.

López-Huertas and Torres (2005) researched terminological dispersion in women thesauri, dictionaries and glossaries and identified the difficulty of translating terms into Spanish, first due to cultural differences and second because of the focus on the use of masculine and feminine in both languages, similar to the situation in Mexico.

Common to the studies done by Olson and Schlegl in the 1990s regarding representation analysis of documents with controlled language is examination of the Dewey Decimal Classification System and the Library of Congress subject headings. They showed there was open marginalization on topics of race, ethnicity and gender. In 1999, Olson co-authored an article titled "Bias in Subject Access Standards: A Content Analysis of the Critical Literature" with Schlegl, and it was published in *Information Science, Where Is It Going?* For the study, 93 documents were selected for analysis as to degree of adequate representation in terms of headings and assigned classification. The results show a slant toward GS, as well as marginalization on topics of race and religion.

3. Objetive

The overall objective is to come up with a methodological proposal for analyzing documents in the specialty and local terms coined by authors in their works, contrasting the extracted terms with headings or descriptors from controlled languages, in order to show those that have not been incorporated into the indexing tools and may be considered candidates for headings or descriptors.

4. Methodology

The first element in the method that was developed is selection of GS-specialized documents: on compensated work done by women as a contribution to the economy,

where the importance of their participation has been shown. Documents with the following characteristics were selected:

- Works that address the subject of the participation of women in salaried work in Mexico
- Works published in Mexico between 2000 and 2010 (books, articles from peer-reviewed journals and masters and doctoral theses)
- Works that are in IES libraries in Mexico and develop research lines associated with the topic or offer academic courses and/or masters or doctoral programs in GS

Searches in IES library catalogs were carried out using the words *woman* and *work* to produce greater results, given that when the terms *genderstudies* and *work* were used exclusively, the results obtained did not cover the subject. This is due to the interpretation of *gender* as a synonym of *woman* (Scott, 1999). Since phrases such as *sexual discrimination at work* or *violence in the workplace* are more precise expressions, they yield few results but exclude documents dealing with the topic because the assigned headings are more general.

Once the sample was defined –a total of 110 recovered works–, the automated indexing methodology was applied, using WordSmith software to get the most frequently used words and eliminate the "empty words", in other words prepositions, articles and some nouns (Gil Leiva, 2008, pp. 329-331). For document analysis, the Zipf mathematical model and Goffman's mathematical model were applied to the 110 documents selected (Gorbea, 2005; Urbizagastegui and Restrepo 2011).

Following the Zipf principle of word economy, those used most frequently were identified. To delimit the range of relevant words, as well as their frequency, the Goffman mathematical model was applied, which enables identifying the cut off in each case. Since the software extracts words only, Notepad (open access) software was used to obtain representative terminological phrases within their context.

Applying the comparative method enabled identifying, describing and explaining the language used to classify the subject in the catalogs of the libraries that included the works selected and analyzed in the previous phase. A comparison was made between the results of the terms used in the literature on the subject and those used in the controlled languages. This made it possible to verify the presence or absence of the terms in the controlled languages and identified the candidates for headings or descriptors that might contribute to representing the content of documents more precisely.

The controlled languages that were taken as reference in the comparison are both pre coordinated and post coordinated languages. Of the pre coordinated ones, the *Lista de Encabezamiento de Materia para Bibliotecas* (Subject Heading List for Libraries), assembled by the Luis Ángel Arango Library, revised and published in 1998, was used, and the equivalent headings in Spanish are from the Library of Congress Subject

Headings, an open access database in which Mike Kreyche of the University of Kent participates. Regarding the post coordinated languages, two GS-specialized thesauri were used: the*Tesauro de Género: lenguaje con equidad* (Gender Thesaurus: Language with Equality), by the Instituto Nacional de las Mujeres de México (2006) and the *Tesauro de Género de la Red de Centros de Documentación y Bibliotecas de Mujeres del Estado Español* (2014, Gender Thesaurus of the Network of Spanish State Women's Documentation Centers and Libraries). The results of the comparison made it possible to verify the presence or absence of the terms in the controlled languages and to identify candidates for headings or descriptors that contribute to describing documents on the subject.

5. Results

The results of the research show that when seeking universalization of specialized language terms, a tendency of the controlled languages used in the comparison with expressions of the authors, in literature on the subject generated in Mexico, the conditions of discrimination experienced by women in the realm of salaried work become invisible, and while some social and cultural contexts share common linguistic units, as in the case of the term "glass ceiling" (Galeana,2013, p. 11), others are situations that take place exclusively in Mexico, such as the term "*mujeres tortilleras"* referring to female tortilla makers.

As for the results produced so far, it can be seen that when making a "calque" –as it is termed– from encyclopedic controlled languages, the idiosyncrasy reflected through language is lost in different cultures. Similarly, this research found that certain terms concern women in general, while others respond solely to conditions in Mexico, and insofar as that terminology obtained directly from author-generated contents is not used, the situation in which gender relations are reproduced is not known.

Although women contribute to the economy and GS have shown the importance of their participation through decades of research, terms that designate such progress have still not been inserted into controlled languages, to describe and make them visible through indexing.

When terms taken from the literature were compared with headings and descriptors used in catalogs, the former did not appear in the catalogs because they were not in the controlled languages.

1. In the workplace: *hostigamiento sexual en el lugar de trabajo (sexual harassment in the workplace), hostigamiento sexual laboral* **(***job-related sexualharassment)*or*violencia de género en el trabajo (gender violence at work)*
2. In agriculture: *asalariadas en la agroindustria (salaried agro-industrial workers)*, *empleo rural femenino (rural female employment)*, *feminización de la producción agrícola (feminization of agricultural production)*, *jornalera agrícola (female farmworker)*, among others

3. In the sex trade, initially called *prostitución (prostitution),*which is not an activity exclusive to women,other terms are detected in GS literature: *comercio sexual*, (*sex trade), compra de mujeres (purchase of women), sexual exploitation (explotación sexual), prostitución femenina (female prostitution)*
4. In factories where females participated prior to the 19th century: *mujeres en las fábricas*(*women in factories), obreras fabriles (female workers in factories)*, *presencia de mujeres en la fábrica (presence of women in the factory)*,*trabajadoras de las fábricas (female factory workers)*,*mujeres en la fuerza de trabajo industrial (women in the industrial work force)*, *feminización del trabajo en la maquiladora* (*feminization of maquiladora work),* among others
5. In the textile sector, which also had early involvement of women: *integración de la mujer al sector* (*incorporation of women into the textile sector textil*,*mujeres en la industria textil (women in the textile industry)*,*mujeres indígenas en la producción textil (indigenous women in textile production)*,*participación de mujeres en microempresas textiles* (*participation of women in micro textile companies)*
6. On women workers' rights: *derechos laborales de las mujeres (women workers' rights)*, *protección de las trabajadoras (protection of female workers)*
7. On women in management positions: *jefas de empresa (female company bosses)*, *mujeres en posiciones de dirección (women in management positions)*, *mujeres en puestos de jerarquía (women in hierarchy positions)*, *mujeres en puestos directivos (women in directorial positions)*

Review and comparison of the controlled languages studied confirms the focus of the theory of doors, where the terminological value in a varied field opens the doors to countless applications (Cabré, 2002). It confirms what authors Olson and Schlegl (1999) maintain regarding the universal controlled language being inattentive to minorities. Eating corn tortillas is part of Mexican culture, yet attention is not given to the working conditions of women whose jobs involve their production. While research on this population group of female workers does not abound, three of the 110 papers reviewed deal with *condiciones de vida y salud de las tortilleras* (*living and health conditions of tortilla makers),*and by not having the adequate headings or descriptors, they get lost in the overall collection.

6. Conclusions

One feature of specialized language is that it constantly evolves. The former search for term universalization was meant to facilitate communication among professionals and prevent natural language polysemy. However, given social changes, the opening up of new lines of research and the contributions of scientific production, controlled languages must be updated.

Contributions of terminological theory that has also supported the construction of controlled languages, such as the theory of doors, should also be considered to propose more flexible solutions, such as, for example, inclusion of abundant synonymy corresponding to the preferred descriptor.

Once again the approaches set out by Olson and Schlegl, on the one hand, and López-Huertas and Torres, on the other, have been confirmed, as have several research studies carried out internationally that are very accurate in terms of the loss of visibility of marginalized groups, as is the case presented here.

This research led to locating terms that describe living conditions of female workers in general, others that deal exclusively with conditions experienced by Latin American women and finally those that solely concern Mexican women. The terms detected in this paper are just as they appear in the literature and comprise candidates for analysis as possible documentary language descriptors.

Though progress in incorporating GS-responsive terminology is slow, this paper adds to the efforts made to date, and it is an invitation to replicate the model in other emerging subjects that seek resonance. Controlled language is an auxiliary indexing tool, and it is necessary to contribute to providing legitimacy and visibility to a field of studies that has proven its validity and contributions, which is why it has earned a space in the libraries of the Instituciones de Educación Superior.

References

Biblioteca Luis Ángel Arango. (1998). *Lista de encabezamientos de materia para bibliotecas*. Santafé de Bogotá: R. Eberhard.

Cabré, M. Teresa (2002). Terminología y lingüística: la teoría de las puertas. *Estudios de lingüística del español*, v. 16. [http://elies.rediris.es/elies16/Cabre.html]

Chu, Heting (2007). *Information Representation and Retrieval in the Digital Age*. Medford, New Jersey: Information Today, Inc.

Galeana, Patricia (2013). *Rompiendo el techo de cristal: las mujeres en la ciencia, en la educación y en la independencia financiera*. Mexico City: Federación Mexicana de Universitarias AC.

Gil Leiva, Isidoro (2008). *Manual de indización: Teoría y práctica*. Gijón: Trea.

Golubov Figueroa, Nattie L. (2006). *Tesauro de género: lenguaje con equidad.* Mexico City: Instituto Nacional de las Mujeres.

Gorbea, Salvador (2005) *Modelo teórico para el estudio métrico de la información documental*. Gijón: Trea.

Hernández Sampieri, Roberto, Fernández Collado, Carlos& Baptista Lucio, Pilar (2010). *Metodología de la investigación*. Mexico City: McGraw-Hill.

Lancaster, F. W. (1986). *Vocabulary Control for Information Retrieval*. Arligton: Information Resources Press.

Lara, LuisFernando& Ham Chande, Roberto (1979). Base estadística del diccionario del español de México. In *Investigaciones lingüísticas en lexicografía*, edited byLuis FernandoLara, RobertoHam Chande and Ma. Isabel García Hidalgo. Mexico City: El Colegio de México. Pp. 5-39.

LCSH-ES.ORG (2014). LC *Subject Headings in Spanish* [http://lcsh-es.org/about.html?l=es]

López-Huertas Pérez, Maria José and Torres Ramírez, Isabel de (2005). Terminología de género. Sesgos, interrogantes, posibles respuestas. *DataGramaZero* 6(5).[http://www.dgz.org.br/out05/Art_03.htm]

Maniez, Jacques (1987). *Les langages documentaires et classificatoires: Conception, construction et utilisation dans les systèmes documentaires*. Paris: Éditions d'Organisation.

Marshall, Joan K. (1977). *On Equal Terms: A Thesaurus for Nonsexist Indexing and Cataloging.*New York: Neal-Schuman.

Naumis Peña, Catalina (2000). Análisis de la confluencia entre término y descriptor en la elaboración de tesauros. *Investigación bibliotecológica*, 14(29): 95-113.

Olson, Hope A. and Schlegl, Rose (1999). Bias in Subject Access Standards: A Content Analysis of the Critical Literature. Congress of the Canadian Association for Information Science. In *Information Science: Where Has It Been, Where Is It Going?*CAIS. Pp. 236-47.

Olson, Hope A. and Schlegl, Rose (2001). Standardization, Objectivity and User Focus: A Meta-Analysis of Subject Access Critiques. *Cataloguing and Classification Quarterly*, 32(2): 81-90.

Red de Centros de Documentación y Bibliotecas de Mujeres (2014). *Tesauro de Género: Tesauro de la Red y Centros de Documentación y Bibliotecas de Mujeres del Estado Español*. Sevilla, Spain. [http://www.juntadeandalucia.es/iam/catalogo/doc/web/tesauro_genero.pdf]

Scott, Joan W. (1999). *Gender and the Politics of History.* New York: Columbia University Press.

Scott, Mike (2012). *WordSmith Tools 6.0*: United Kingdom: Oxford University Press. [http://www.lexically.net/wordsmith/version6/]

Svenonious, Elaine (2003). Design of controlled vocabularies. In *Encyclopedia of Library and Information Science*. New York: Marcel Dekker. Pp. 822-38 [http://polaris.gseis.ucla.edu/gleazer/260_readings/Svenonius.pdf]

Urbizagastegui, Rubén& Restrepo, Cristina (2011). La ley de Zipf y el punto de transición de Goffman en la indización automática. *Investigación bibliotecológica*, 25(54): 71-82.

Rita Costa Veiga Zamboni and Marivalde Moacir Francelin

The Location of Classification: Between the Local and the Global

Abstract

This article discusses, through literature review, cultural bias in classification from the standpoint of knowledge organization in an interdisciplinary dialogue with social studies, feminist positions and post-colonial theory. Information is a key aspect to understand the social, cultural, political and economic relations intertwined in the map of the globalized world. The role of classification as an epistemological tool that promotes a culturally biased use of knowledge. Classification systems can be ameliorated to encompass knowledges in a context of cultural diversity.

Introduction

This article discusses, through literature review, a view of classification from the standpoint of social studies, feminist positions, and post-colonial theory. Information is a key aspect to understand the social, cultural, political and economic relations intertwined in the map of the globalized world.

Information can be seen as a key element to understand the socio-cultural, political and economic relations which design the map of our globalized world. The claim to universality and objective representations of reality, and the search for universal laws and truths stripped from context seem to have been replaced, in the 19th and 20th centuries, by a perspective in which context was again taken into account in several disciplines (Olson 2011, p. 114). The ideal of universalism imposes one single viewpoint to all social groups around the world by obliterating difference and making it hard for groups to maintain their social (political and cultural) traits in a globalized world. Postmodern, postcolonial and feminist theories have addressed the challenge of rethinking the world as a multitude of social groups that should coexist on the same level. Santos identifies the need to counteract negative universalism via a consensus around the fact that "no struggle, objective or agent has the overall recipe for the social emancipation of humanity" (Santos 2010 p. 237). Social emancipation is "an ethical and political exigency, perhaps more pressing than ever in the contemporary world" (Santos 2010, p. 237).

Such standpoint makes it relevant to discuss the role of classification as a core tool in knowledge organization. In his article "Declassification in knowledge organization: a post-epistemological essay", García Gutiérrez (2011) sets out to analyze the prevailing epistemological position in knowledge organization in a complex, culturally diverse world. García Gutiérrez argues that it is necessary to form communication networks with other areas of study, as a way to overcome positivist barriers that may have been imposed by a strict view of areas of knowledge. He mentions postcolonial theories and feminist positions as possible areas of overlap which could "promote the in-depth revision of the conceptions, procedures, relationships and actions revolving around KO" (García Gutiérrez 2011, p. 6).

This revision of conceptions might be particularly relevant at a time when information seems to be a key element to understand what is at stake when the contact among cultures is elevated to an unprecedented level in a globalized world. Santos (2000, p. 38) argues that the authoritarian use of information is a revealing characteristic of the present time. In his view, the technical conditions which could allow for the collective enhancement of knowledge in the planet in an egalitarian and democratic way were appropriated by a restricted group of actors.

In this sense, the perspective of interculturality might offer insights as to how to understand these questions within the framework of knowledge organization. From this perspective it is possible to perceive the classification of knowledge as a construct that is not without cultural bias. The ordering of knowledge is a political and ideological act which has a profound influence on the ordering of the world. It can determine which groups produce knowledge and which groups are subjected to knowledge produced by others; which group gets to write and preserve its history, memory and cultural traits and which groups gets it done for them (or not done at all).

A commonly heard criticism to an intercultural approach is that it does not appear to have a clearly defined epistemological or methodological stance, bordering on relativism. Social emancipation demands the construction of an ethical and political position which should not be grounded on an absolute principle or embrace relativism. Thus it poses the challenge of "knowing how to maximize interculturality without subscribing to cultural and epistemological relativism" (Santos 2010, p. 238).

Similarities and differences: classification thought

Classification is present in most areas of our lives, be it in the form of everyday classification or the most elaborate knowledge organization system. The omnipresence of classification may be one of the reasons why classification (and its implications) may seem invisible to us. Classification can be regarded as "the quintessential core of knowledge organization" (Smiraglia 2014, p. 57). Organizing knowledge is a human activity and as such it carries the assumptions, interests and motives of the society where it takes place. Classification is then a powerful means to enforce a determined economic and political view of the world. When the views of one (or few) social groups are privileged, cultural diversity is not being taken into account.

In a globalized world, cultural diversity seems to be taken for granted. The ubiquity of the concept might be disguising the fact that the concept has far-reaching implications which deserve to be examined. As regards the field of knowledge organization, cultural diversity can be seen as the underlying concern behind discussions about classification bias.

In a culturally diverse society, the notion of the universality of knowledge must be reviewed. Mai (2013, p. 242) describes the relationship between classification and bias as a reflection of reality. Reality is biased, therefore classifications must also be biased.

Mai points to the fact that the modern hope for universality in classification has been reconsidered in more recent conceptualizations.

In "Classification and universality: Application and construction" (2006), Olson argues that classification is seen as an essential (and natural) aspect in the process of creating knowledge. This view of classification as an innate process has contributed to its acceptance as a "natural" and universally applicable process. The terms "similarity"/ "sameness" and "difference" are the guiding principles of classification in Western culture, and such conceptions carry with them cultural bias. The duality sameness/ difference disregards the fact that what is defined as similar is culturally determined by creating a hierarchy that presents knowledge in categories that are supposedly essential, natural and universal.

According to Olson, "discipline – as the primary facet in our classifications – is the fundamental sameness" (Olson 2001a, p. 117). As such, classifications depend on specialists to decide what content is or is not included in its classes. In such a rigid framework, incorporating new knowledge is not an easy task. Knowledge that does not fall into a pre-determined category might be incorporated into the main structure or not, depending on a number of social, cultural, political and economic interests. In the context of a culturally diverse society, the viewpoint of the "majority" may be imposed on the "minorities" by not acknowledging the fact that knowledge organization systems are cultural and political constructs themselves. As Olson explains it, "effective searching for marginalized topics will require greater ingenuity and serendipity that searching for mainstream topics" (Olson 2001b, p. 639)

A classification has a representational function: it is a set of categories *and* a system (Olson 2007, p. 380). A classification based on hierarchy and the principle of mutual exclusivity leaves out anything that cannot be ascribed to a specific place in the scheme. If different knowledge organization systems can be built from the perspective of different cultures, it follows that these systems, as cultural constructs, are biased towards the culture from which they stem. In other words, knowledge organization systems are constructs that tend to present themselves as invisible (or apparently neutral) to their users (Bowker; Star 2000).

In this sense, to research knowledge produced outside of the legitimized circuits in mainstream knowledge organization systems may prove difficult. Institutions which organize information "reflect the marginalizations and exclusions of the society they serve." (Olson 2001b, p. 639).

García Gutiérrez (2011) notices the classification process is often viewed as a neutral and non-ideological element, even if it does produce ideology and culture. Classification is based on metonymic reduction as a tool to create bias. García Gutiérrez (2011 p. 6) describes classification as a "first-order gnoseological and epistemological operation that impregnates totality, and totally our relationship with the world". Whereas classification occurs in all cultures, epistemology is a product of

Western culture. Classifications are a compelling strategy impose an ordering of the world "by means of essentialist demarcations and ontological purifications in an illusion of universalism and consistency" (García Gutiérrez 2014 p. 393).

These processes of classification, re-classification and re-signification are intensified by the digital technology which amplifies their reach and impact. At the same time, information and communications technology offers unprecedented capabilities for the development of new arrangements which could allow for more social groups to negotiate knowledge production, distribution and access on equal grounds.

The confluence of ICT and cultural diversity present ethical challenges. Capurro (2010) proposes an "intercultural comparative critical reflection" as a means to problematize the bias behind the use of technology in informational processes. Knowledge organization studies on cultural diversity show a distinct connection ethics as a way to create pathways to reduce cultural bias.

The voices in classification

Studies in the field of knowledge organization in an interdisciplinary dialogue with social studies, feminist positions and postcolonial theories suggest that it is necessary to find ways to let the voices that have been obliterated by the presumption of universality be heard.

To Mignolo (2000), the subalternization of knowledge (the colonial epistemic difference) is the origin of the dichotomy between the societies that produce knowledge and the societies that merely absorb that knowledge, or the ones about which knowledge is produced (objects of study). In this manner, the location from where knowledge is produced will determine how (or if) it will be incorporated by institutions that organize knowledge such as libraries. Geographical space and knowledge are both human constructs which are inextricably linked.

Grosfoguel (2008) emphasizes the fact that Western philosophy and sciences have been influenced by Eurocentric paradigms that have not catered culturally diverse perspectives. According to Grosfoguel (2008, p. 119), the locus of enunciation is hidden in Western philosophy and sciences creating the illusion of universal knowledge.

García Gutiérrez proposes the introduction of pluralism in classification through the process of declassification, a process that requires the awareness of incompleteness, of bias and of subjectivity. It does not reject classification but introduces the principle of contradiction to it, acknowledging that "a thing is also another thing" / "a thing could always be another thing" (García Gutiérrez 2011, p. 11). Consequently, declassification operates within open categories.

The notion of pluralism is further developed by Mai (2011 p. 723) as a dynamic concept which "[…] is not something that can be set aside as simply something that has to do with culture, society and language, but it is *also* something that has to do with the individual." In this view, a classification system should veer away from the idea of

consensus and instead embrace the idea that "[…] any document and any domain could be classified from multiple equal correct perspectives" (Mai 2011 p. 723).

The idea of multiple perspectives may be daunting to groups accustomed to the certainty of "one truth", which would tend to view such an approach as lacking in epistemological and methodological foundations. According to Mai (2009 p. 639), accepting the idea of plurality in classification systems does not mean that "everything goes". Rather, it poses the challenge or dealing with bias in a transparent and critical way.

Bowker and Star (2001 p. 324-25) propose three features that could facilitate dealing with the ethical dimensions of classification in a culturally diverse society. They highlight the importance of: 1. recognizing the balancing act of classifying, namely being aware of cultural diversity by incorporation of ambiguity; 2. making voice retrievable by making the system politically flexible; 3. being sensitive to exclusions.

The positions taken by the authors mentioned appear to have in common the notion that acknowledging cultural diversity is essential to practices which involve knowledge in all its facets. As far as knowledge organization is concerned, the importance of recognizing classification as a culturally-sensitive activity is clear. It is less clear, however, in which fashion cultural diversity is supposed to be integrated into knowledge organization practices.

Szostak (2014) poses a crucial question in this regard: "what exactly should we want a classification to do in order to respect and support diversity?" (Szostak 2014, p. 160) Szostak points out that the efforts which have been made to make classification more open to social diversity have only had a limited effect. Szostak suggests that the reason for such limited progress is related to the fact that classifications are hierarchical structures. Szostak refers to Olson's "How we construct subjects: a feminist analysis" (2007) to introduce the idea that hierarchies are more "reflective of a masculine perspective, and that a classification that blended hierarchy with a web-of-relations approach would be more gender-neutral" (Szostak 2014, p. 164).

Beghtol (2002) proposes cultural hospitality as an ethical warrant for knowledge organization systems. The concept of cultural hospitality "refers to the ability of a classification notation to incorporate new concepts and to establish appropriate semantic and syntactic relationships among the old and the new concepts." (Beghtol 2002, p. 518) According to Beghtol (2002), adding different cultural warrants could contribute to making the concept of cultural hospitality more efficient in catering to the information needs of individuals within their cultures in an ethical way.

Ethical concerns about knowledge organization systems stem from the concept of cultural warrant, as they provide "the rationale and authority for decisions about concepts and what relations among them are appropriate for a particular system" (Beghtol 2005, p. 904). A cultural warrant is related to the notion that the culture in which a knowledge organization system is based may facilitate or hinder the access to

information: the users belonging to the culture in which the system has been built would be at an advantage in relation to the users not pertaining to that culture. In this way, a more "visible" and "permeable" system has better chances of catering to a socially-diverse society.

Smit (2012) suggests that disclosing information about the processes involved in knowledge organization and institutional policies to users is a necessary step to instill information ethics.

Developing strategies (techniques, policies, approaches) to make knowledge organization systems and institutions more amenable to cultural diversity is a great challenge. As Olson (2001b, p. 659) points out, techniques to make knowledge organization systems more permeable ("redemptive technologies") involve *relinquishingpower* to the other and might prove difficult to develop.

Conclusion

The ethical dimension of knowledge organization in relation to cultural diversity is brought to the fore by a number of authors. An ethical relationship with the "other" should allow for the inclusion of marginalized knowledges and cultures.

Classification systems are an intrinsic part of information and communications technology. An intercultural information ethics, such as proposed by Capurro (2010), may allow for the establishment of a cultural ethos through local and global intercultural networks. The notions of cultural warrant and cultural hospitality also seem to have a potential to make classification systems more amenable to deal with the challenges brought about by culturally-diverse world.

The new information and communication technologies have a potential to become the cornerstone to foster changes in knowledge organization systems that would be more open to cultural diversity. Identifying bias and subjectivity in classification systems and dealing with such bias in a critical and ethical way is the first step.

References

Beghtol, Clare (2002). A proposed ethical warrant for global knowledge representation and organization systems.*Journal of Documentation*, 58(5): 507-32.

Beghtol, Clare (2005). Ethical decision-making for knowledge representation and organization systems for global use.*Journal of the American Society for Information Science and Technology*, 56(9): 903-12.

Bowker, Geoffrey C. & Star, Susan L. 2000. *Sorting things out: classification and its consequence*s. Cambridge: The MIT Press.

Capurro, Rafael (2010). Desafios teóricos y prácticos de la ética intercultural de la información.*Conferencia inaugural en el I Simpósio Brasileiro de Ética da Informação*, João Pessoa, March 2010. [http://www.capurro.de/paraiba.html]

Grosfoguel, Ramón (2008). Para descolonizar os estudos de economia política e os estudos pós-coloniais: transmodernidade, pensamento de fronteira e colonialidade global.*Revista Crítica de Ciências Sociais*, 80: 115-47.

García Gutiérrez, Antonio (2011). Declassification in Knowledge Organization: a post-epistemological essay.*TransInformação*, 23(1): 5-14.

García Gutiérrez, Antonio (2014). Declassifying Knowledge Organization. *Knowledge Organization*, 41(5): 393-409.

Mai, Jens-Erik. 2010. Classification in a social world: bias and trust.*Journal of Documentation*, 66(5): 627-42.

Mai, Jens-Erik (2011). The modernity of classification.*Journal of Documentation*, 67(4): 710-730.

Mai, Jens-Erik(2013). Ethics, Values and Morality in Contemporary Library Classifications.*Knowledge Organization*, 40(4): 242-53.

Mignolo, Walter D. (2000). *Local Histories/ Global Designs. Coloniality, Subartern Knowledges, and Border Thinking*. Princeton: Princeton University Press.

Olson, Hope A. (2001a). Sameness and difference: a cultural foundation of classification. *Library Resources & Technical Services*, 45(3): 115-22.

Olson, Hope A. (2001b). The Power to Name: Representation in Library Catalogs. *Signs*, 26(3): 639-68.

Olson, Hope A. (2006). Classification and universality: application and construction.*Semiótica*, (139): 377–91.

Olson, Hope A. (2007). How we construct subjects: a feminist analysis.*Library trends*, 56: 509-41.

Santos, Boaventura de S. (2010) From the Postmodern to the Poscolonial – and Beyond Both. In *Decolonizing European Sociology: Transdisciplinary Approaches*, edited byEncarnación Gutiérrez Rodrigues and Manuela Boatcã and Sérgio Costa. Farnham: Ashgate. Pp. 225-42.

Santos, Milton (2000). *Por uma outra globalização*. Rio de Janeiro: Record.

Smiraglia, Richard P. (2014). *The Elements of Knowledge Organization*. New York: Springer.

Szostak, Rick (2014). Classifying for Social Diversity.*Knowledge Organization*, 41(2): 160-70.

Patrick Keilty and Richard P. Smiraglia

Gay Male Nomenclature

Abstract

In the case of the colloquial language of regular populations, very little research in the KO domain exists. This paper presents data derived from the sexual posts of men seeking sexual contact with other men on Craig's List on one day in five different North American cities. This report is a first analysis of that domain for the purpose of extracting the ontology of online sexual attraction. Results show not only a high degree of unreality among expectations, but also a fair amount of correspondence with the "closeted" community. This brief analysis shows that in the vocabulary of male homosexual sex posts on Craig's List there is a core set of categories topped by sex roles and activities, but clearly critical are specific fetishized body parts and fantasy roles and ethnicities.

1.0 Ontological concepts of sexualities

A primary goal of the science of knowledge organization (KO) is the extraction, or capture, of ontological concepts for every domain. Although much of the effort of knowledge organization in the past has been expended on creating systems for organizing documents, it still is incumbent on the science of KO to provide the methodological tools and epistemological points of view for extracting domain-centric ontologies. In the case of the colloquial language of regular populations, very little research in the KO domain exists. This paper presents data derived from the sexual posts of men seeking sexual contact with other men on Craig's List on one day in five different North American cities.

Probably the most vocal author on the language of male homosexuality in the KO domain is Keilty, whose two papers cited here both make the call for analysis of "sexual nomenclatures within specific information institutions at particular points in history" (2012a, 428; 2012b, 323). Most likely Keilty had major bibliographic classifications in mind when he made reference to information institutions. Elsewhere Smiraglia (2014) has posited the role of large social networks as information institutions, and Craig's List is one such very influential institution. In the male homosexual subculture, Craig's List is one of the major venues for sexual communication.

All sexualities have a place on Craig's List. We have isolated our study to male homosexual posts simply as a matter of limiting the scope of our study. But a case could be made that for male homosexuals, Craig's List is a powerful information institution providing potential meetings as well as a constant scan of the sexual atmosphere in specific locales. Craig's List differs from the mobile app "hook-up" sites Grindr, Scruff and Growlr, mainly because it is utterly anonymous. It is possible that this perceived anonymity invites posts from sexually-experimental, sexually-oppressed and sexually-closeted individuals to a far greater extent than the mobile social circles, which depend on the open "outness" of members.

Keilty (2012a, 422) refers to a sort of "disciplinary power" that is maintained in subcultures by which the members regulate behavior that eventually becomes perceived as ideological. Christensen (2008) refers to the representation of homosexuality in major bibliographic knowledge organization systems, and warns that it might be difficult for a domain to cohere among a group of people who have nothing in common but sexual preference. On the other hand, Keilty's analysis of tags on Xtube, an amateur pornographic social network, demonstrates both a distinct vocabulary and a form of "socio-citational relations" (427). In one of the earliest papers in the KO domain on homosexual ontology, Huber and Gillaspy (2000) demonstrate the oppressive effect of dysfunctional indexing systems on the homosexual subculture. They include a litany of social problems consistent with membership in the homosexual subculture including fear of health-care and concomitant declines in health due to the perceived necessity to hide their sexuality. Conversely, Huber and Gillaspy's results show explicit references in nursing literature to care approaches that could resolve these social difficulties. Pinho and Guimarães analyze male homosexuality in Brazilian indexing languages and reach much the same set of conclusions—that there is not sufficient domain-centric vocabulary available to inform KOSs, but also that the subcultural aspects of homosexuality tend to suppress the evolution of workable ontology.

All of these are the hallmarks of a domain (Smiraglia 2012). As perplexing as it might seem, the teleology of achieving sexual fantasy combined with a common vocabulary (ontology), and social semantics demonstrated both by substantial reference in nursing literature and within community hyperlinking shows the coherence of the male homosexual poster on Craig's List as a domain. This report is a first analysis of that domain for the purpose of extracting the ontology of online sexual attraction. Results show not only a high degree of unreality among expectations, but also a fair amount of correspondence with the "closeted" community.

2.0 Methodology

On Saturday, February 1, 2014 Craig's List was searched for Los Angeles, Milwaukee, New York City, Toronto, and Vancouver and all titles of men-for-men personal posts for January 31, 2014 were downloaded (in Craig's List this includes "men seeking men" and "casual encounters: m4m"). The posting titles were downloaded as hyperlinks into spreadsheets. The analysis reported here is limited to the posting titles. It could be considered a limitation of this study that we have analyzed only the posting titles. Obviously rich nomenclature populates the post texts as well. However, this is a preliminary survey analysis to discover initial indicators about the breadth of the nomenclature. Furthermore, the posting titles constitute a form of signaling in the domain and are worthy of analysis on their own. Regardless, the data gathered are rich in detail.

It is worth noting that we have the population of one day's posts rather than a sample of any sort. On the other hand, we cannot know how the events of the day might have shaped these posts or how posts on any other day might have been different. Thus our results are not generalizable in any meaningful way, but they are indicative. To provide context we can say that the weather on January 31, 2014 in each location was:

New York City, New York 6 - 4 no precipitation
Los Angeles, California 17 -13 trace precipitation
Milwaukee, Wisconsin -6 to -8 no precipitation
Toronto, Ontario 1 - -2C 10cm snow expected overnight
Vancouver, British Columbia .2 - -4 with a trace of precipitation anticipated

Some major headlines in the local newspapers included a headline in the *Milwaukee Journal-Sentinel* about "Counting the homeless provides a glimpse of their stories." A side-bar noted a car-school bus crash, a TV news reporter departing for Detroit, and an Appleton Wisconsin church cutting ties with the boy scouts following its decision to accept openly gay adult leaders. The *New York Times* ran a story on "January ends with another decline on Wall Street." In the *Toronto Star* an editorial was "A claim that can't be ignored," concerning notorious former mayor Rob Ford's alleged involvement in a violent act.The *Los Angeles Times*' fulsome archive of the day's news includes "Former Chris Christie aide says governor knew of bridge lane closures," "Suicide rate among active-duty soldiers falls sharply," "California drought could force key water system to cut deliveries," "California flu deaths for those under 65 running far ahead of last year," and on and off rain was expected for the weekend. A letters column in the *Vancouver Sun* carried reader reactions to a story about a Mexican woman who hanged herself in jail. In sum, it was winter in North America and no particular news story predominated.

In Craig's List the form of post requires certain particulars to be included in the post (as for example, date of the post, place of the poster, or age of the poster, and the ubiquitous "m4m" which appears in front of the poster's parenthetical place); these were stripped out during data-cleaning. The remaining term clusters were entered into Provalis Suite's Simstat software, and once the data were properly identified and normalized, they were analyzed using the WordStat software module.

3.0 Results

Results of co-word analyses reveal a very narrow spectrum of homosexual desire presented in these personal post titles. There were 34,440 terms in the dataset, of which 2679 were unique, and 310 had a frequency greater than 10. Our analysis is focused on this cluster of terms that occur 10 or more times, which constitute 16.5% of the total number of keywords. Among those terms there were 21,006 phrases of which 645 two-word terms, 195 three-word terms, 94 four-word terms and 54 five-word terms occurred more than 3 times. Specific keywords in the top frequency tier can be analyzed as part of two-word terms. "Top" for example, occurred 573 times, "guy"

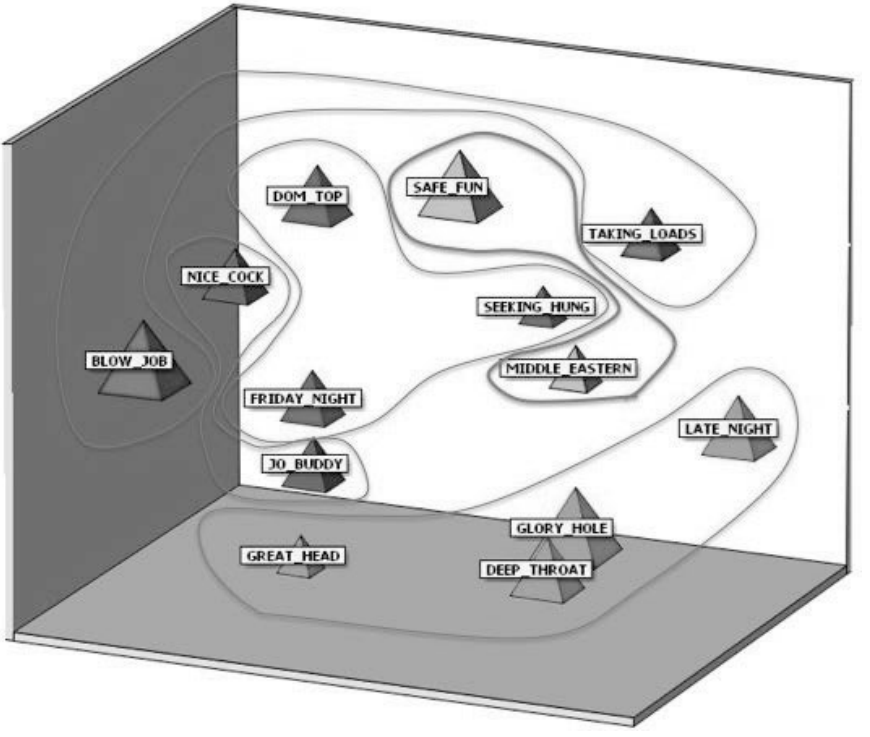
529, “bottom” 487, and so on. By combining the most frequently-occurring keywords with the most frequently-occurring phrases, a small taxonomy including 161 terms in 61 clusters was built. This is shown in Figure 1.

Figure 1. Most frequently occurring terms.

- ASIAN
  - ASIAN BOTTOM
  - ASIAN BOY
  - ASIAN GUY
- BI
  - BI BOTTOM
  - BI CURIOUS
  - BI DISCREET
  - BI DUDE
  - BI GUY
  - BI GUYS
  - BI MASC
  - BI TOP
  - BI WHITE
- BIG
  - BIG BLACK
  - BIG BLACK COCK
  - BIG COCK
  - BIG DICK
  - BIG LOAD
  - BIG THICK
- BLACK
  - BLACK COCK
  - BLACK DICK
  - BLACK GUY
  - BLACK OR LATINO
  - BLACK TOP
- BLOW JOB
- BODY MASSAGE
- BOTTOM
  - BOTTOM BOY
  - BOTTOM FOR TOP
  - BOTTOM GUY
  - BOTTOM HOSTING
  - BOTTOM LOOKING FOR A TOP
  - BOTTOM SEEKING TOP
  - SMOOTH BOTTOM
- BUBBLE
  - BUBBLE ASS
  - BUBBLE BUTT
- BUST A NUT
- COCK
  - COCK SUCKED
  - COCK SUCKER
  - COCK TO SUCK
  - COCKSUCKER HOSTING
  - NICE COCK
  - THICK COCK
  - UNCUT COCK
- COLLEGE
  - COLLEGE BOY
  - COLLEGE GUY
- DAD SEEKS
- DAY LOAD
- DEEP THROAT
- DICK
  - DICK SUCKED
  - DICK TO SUCK
- DISCREET
  - DISCREET BOTTOM
  - DISCREET FUN
  - DISCREET GUYS
  - DISCREET MASC
  - DISCREET MASCULINE
  - DISCREET TOP
- DL
  - DL BI
  - DL GUY
- DOM TOP
- EAST INDIAN
- FAT GUY
- FREE MASSAGE
- FRIDAY NIGHT
- FUCK BUDDY
- FUCK MY TIGHT HOLE
- FUN TONIGHT
- GL MASC GUY
- GLORY HOLE
- GREAT HEAD
- GUY LOOKING TO SUCK COCK
- GUYS TO HANG
- HORNY
  - HORNY BOTTOM
  - HORNY GUY
  - HORNY TOP
- HOT
  - HOT ASS
  - HOT BOTTOM
  - HOT COCK
  - HOT DAD
  - HOT FUN
  - HOT GUY
  - HOT GUYS
  - HOT JOCK
  - HOT MOUTH
  - HOT TOP
  - HOT WHITE
- HUGE COCK
- HUNG
  - HUNG BI
  - HUNG BLACK
  - HUNG COCK
  - HUNG DOM
  - HUNG TOP
- ISO BLACK MEN
- JO BUDDY
- JOCK
  - JOCK ASS
  - JOCK BUDDY
- LATE NIGHT
- LATINO
  - LATINO BOTTOM
  - LATINO TOP
- MARRIED
  - MARRIED DUDE
  - MARRIED GUY
  - MARRIED MAN
  - MARRIED TOP
- MASCULINE
  - MASC BI
  - MASC BOTTOM
  - MASC DUDE
  - MASC FIT
  - MASC GUY
  - MASC GUY WANTS SOME DICK
  - MASC TOP
  - MASCULINE BOTTOM
  - MASCULINE DISCREET
  - MASCULINE GUY
  - MASCULINE TOP
- MIDDLE EASTERN
- NORMAL GUY
- NSA
  - NSA BJ
  - NSA FUN
- NUDE MASSAGE
- ORAL
  - ORAL BOTTOM
  - ORAL FUN
  - ORAL SERVICE
- PHONE SEX
- PUMP AND DUMP
- REG JOE U B TOP
- SAFE FUN
- SEEKING HUNG
- SEX PARTY
- STRAIGHT
  - STRAIGHT GUYS
  - STRAIGHT JOCK
- SUCK
  - SUCK AND FUCK
  - SUCK AND SWALLOW
  - SUCK COCK
  - SUCK DICK
  - SUCK MY COCK
  - SUCK MY DICK
  - SUCK YOUR COCK
  - SUCK YOUR DICK
  - SUCKING COCK
  - SUCKING DICK
- TAKING LOADS
- TIGHT
  - TIGHT ASS
  - TIGHT BOTTOM
  - TIGHT HOLE
- TOP
  - TOP DAD
  - TOP FOR BOTTOM
  - TOP GUY
  - TOP HOSTING
  - TOP HOSTING DL LATIN
  - BOTTOMS
  - TOP LOOKING FOR BOTTOM
  - TOP SEEKING
  - TOTAL TOP
- UR
  - UR COCK
  - UR LOAD
- WANNA
  - WANNA FUCK
  - WANNA SUCK
- WHITE
  - WHITE BOTTOM
  - WHITE GUY
  - WHITE JOCK
  - WHITE TOP

It is clear that the core words are role designators “top” and “bottom,” and also that the stem “guy” occurs in combination with many other keywords. These terms are consistent with generally perceived male gender roles. There are many fetishized body parts and particular sex acts, which frequently are combined with modifiers such as “hot” and “fun.” “Discreet,” “married,” and “straight” are adjectival descriptors that occur in conjunction with the role, part, act and modifying terms. All of these suggest the men posting (in this case, signaling) are not part of the main-stream “out” gay community. We also see racial or desired type modifiers, and “free massage,” which likely is a coded signal for paid sex (Craig’s List ceased support of erotic services posts in 2009). The taxonomy was used as a filter to analyze term co-occurrence throughout the dataset, resulting in a multi-dimensionally-scaled (MDS) plot shown in Figure 2.

Figure 2. MDS plot of term co-occurrence (stress=0.25216 R^2=0.9210).



To improve goodness of fit of the plot only the largest clusters are pictured. We see five regions, which seem to be clustered according to sexual tastes variously described (or perhaps deconstructed is another way to explain the clusters). Most of the granular terms, which are not shown, fall into the region represented in red by “blow job” and “taking loads” The five regions are not connected to each other by term co-occurrence.

We wanted to know whether this language was ubiquitous or whether regional variations might be involved. The next group of preliminary results (Tables 1 and 2) show the keywords and phrases most often occurring in each locale.

Table 1. Most frequently occurring keywords by locale.

Milwaukee		New York		Toronto		Vancouver		Los Angeles	
SUCK	1.80	*TOP*	1.70	*GUY*	2.10	CUTE	0.30	*TOP*	1.70
BOTTOM	1.70	*GUY*	1.50	*TOP*	1.50	DADDY	0.30	*BOTTOM*	1.50
GUY	1.40	*BOTTOM*	1.40	*BOTTOM*	1.30	DUDE	0.30	*GUY*	1.30
TOP	1.30	*COCK*	1.10	*COCK*	1.20	MAN	0.30	*SUCK*	1.20
HEAD	1.20	*HOT*	1.00	*SUCK*	1.10	SUCKED	0.30	*COCK*	1.10
COCK	1.00	*BI*	0.90	*FUN*	1.10	BUDDY	0.20	*HUNG*	0.80
DICK	1.00	*SUCK*	0.90	*BI*	0.80	HAIRY	0.20	*HOT*	0.80
BLOW	0.80	*HUNG*	0.80	*HORNY*	0.70	MEN	0.20	*DICK*	0.70
FUN	0.80	*FUN*	0.80	*ASIAN*	0.70	SERVICE	0.20	*FUCK*	0.70
HOT	0.80	*MASC*	0.70	*GUYS*	0.70	BLOW	0.20	*MASC*	0.70
NSA	0.80	*ASS*	0.70	*STRAIGHT*	0.60	CAR	0.20	*BI*	0.60
SUCKING	0.70	*MASSAGE*	0.60	*DISCREET*	0.60	CUM	0.20	*WHITE*	0.60
BOY	0.60	*HOSTING*	0.60	*HUNG*	0.60	DICK	0.20	*FUN*	0.60
BUDDY	0.60	*DISCREET*	0.60	*FIT*	0.60	FREE	0.20	*MASSAGE*	0.60
CUM	0.60	*MARRIED*	0.60	*HOT*	0.60	SEEKS	0.20	*BIG*	0.60
HORNY	0.60	*JOCK*	0.60	*MASC*	0.50	TIGHT	0.20	*DISCREET*	0.60
SEEKING	0.60	*DICK*	0.50	*TONIGHT*	0.50	UNCUT	0.20	*DL*	0.60
STRAIGHT	0.60	*BIG*	0.50	*DICK*	0.50	CD	0.20	*GUYS*	0.50

Although the order of specific terms varies, there is remarkable consistency across the locales with the exception of Vancouver, where we see a completely different priority sequence. In the Vancouver keywords list, the focus at the top of the distribution is less on generic sex role and more on specific desired descriptors. Table 2 shows the comparative distributions of phrases across the locales.

Table 2. Most frequently occurring phrases by locale.

Vancouver		Toronto		Milwaukee		New York		Los Angeles	
BI GUY	0.20	BI GUY	.40	SUCKING YOUR LOADS	0.10	BI GUY	0.80	WHITE GUY	0.90
TOP GUY	0.20	HUNG TOP	.30	BLACK DICK	0.10	JOCK BOTTOM	0.60	MASC GUY	0.80
HUNG TOP	0.20	STRAIGHT GUY	.30	HOT TOP	0.10	COCK SUCKER	0.60	SUCK COCK	0.80
HORNY GUY	0.20	STRAIGHT GUYS	.30	LATIN DICK	0.10	BOTTOM HOSTING	0.60	BIG DICK	0.60
FREE MASSAGE	0.20	COCK SUCKER	.20	ATHLETIC GUY	0.10	BIG DICK	0.50	GLORY HOLE	0.70
EAST INDIAN	0.20	FRIDAY NIGHT	.20			WHITE GUY	0.50	ISO BLK MEN	0.50

There is some distortion from an attempt to restrict the size of this table; notice how different the percentages are from place to place. Milwaukee's list is blank at the

bottom because there were only five phrases in the entire distribution. This time the Vancouver list is comparable to the others.

4.0 Discussion and conclusions

Beaudoin and Ménard generated a study in 2015 of vocabulary used in pornographic video websites. They used a categorization developed by Ogas and Gaddam (2011) to compare vocabulary used by site designers with search vocabulary and tags contributed by users. Their conclusion was that the two vocabularies do not match (98). User-generated search terms conformed to the formal vocabulary of porn websites only with regard to physical characteristics and particular body parts. Activities, settings, and genres were particularly not present among search terms.

What is perhaps relevant for this research is the schema adapted by Beaudoin and Ménard (94), which included: activities, physical characteristics, age, sexuality, roles, ethnicities, various fetishized body parts and fetishized settings, as well as terms related to video production. Our data are dominated by sex roles and physical characteristics, but there also is a hefty emphasis on fantasy roles, ethnicities and fetishized body parts. An obvious conclusion is that within the nomenclature of sex there is a difference between any sort of formal, descriptive language and the language of desire, as expressed by Beaudoin and Ménard's user search terms and as seen in the taxonomy derived from data in this study. This is echoed by the disconnection among the clusters in the MDS plot in Figure 2, where we see desire expressed by clustered categories of desirable roles, desirable potential partners and fantasized circumstances. It is not apparent, however, what the slight difference among results in different locales suggests. Two options among many are that either there could be regional cultural differences, or it might be that greater divergence in terminology emerges in larger, more diverse populations. Further study incorporating data from more locales is clearly called for.

This brief analysis shows that in the vocabulary of male homosexual sex posts on Craig's List there is a core set of categories topped by sex roles and activities, but clearly critical are granular terms that specify fetishized body parts, fantasy roles and ethnicities. The signaling language of posting titles is used to attract attention with a few powerful keywords; only when a title leads to follow-through by clicking on the link to the full post does an actual match become possible. It is worth considering whether this function restricts the vocabulary of the posting titles either to core terminology or to terms that might be thought to be more exciting. Perhaps we are viewing a sort of two-tier differentiation system consisting of fact (top seeks bottom, college seeks guy seeks hot bottom, etc.) and fantasy (safe fun, late night, pump and dump, etc.).

Both Beudoin and Ménard and Ogas and Gaddam cited age as a critical category, but this research so far contains no analysis of age-based content, although the ages of persons posting and of their desired contacts are among our data for further study.

Acknowledgment: The authors would like to thank Mr. Peter Turner of Kennett Square, Pennsylvania, for assistance in data-gathering.

References

Beaudoin, Joan & Elaine Ménard (2015). Objects of Human Desire: The Organization of Pornographic Videos on Free Websites.*Knowledge Organization*, 42: 90-101.

Christensen, Ben (2008). Minoritization vs. Universalization: Lesbianism and Male Homosexuality in *LCSH* and LC*C.Knowledge Organization*, 35: 229-38.

Halperin, David(2012). *How to be Gay*. Cambridge, Mass.: Harvard University Press.

Huber, Jeffrey T. & Mary L. Gillaspy (2000). An Examination of the Discourse of Homosexuality as Reflected in Medical Vocabularies, Classificatory Structures, and Information Resources. In *Dynamisn and Stability in Knowledge Organization: Proceedings of the Sixth International ISKO Conference, 10-13 July 2000, Toronto,Canada*, ed. Clare Beghtol, Lynn C. Howarth, Nancy J. Williamson. Advances in Knowledge Organization v. 7. Würzburg: Ergon-Verlag. Pp. 219-23.

Keilty, Patrick (2012a). Sexual Boundaries and Subcultural Discipline.*Knowledge Organization*, 39: 417-31.

Keilty, Patrick (2012b). Tagging and Sexual Boundaries.*Knowledge Organization*, 39: 320-4.

Ogas, Ogi & Sai Gaddam (2011). *A Billion Wicked Thoughts: What the Internet Tells Us About Sexual Relationships*. New York: Dutton.

Pinho, Fabio Assis & José Augusto Chaves Guimarães (2012). Male Homosexuality in Brazilian Indexing Languages: Some Ethical Questions.*Knowledge Organization*, 39: 363-9.

Smiraglia, Richard P. (2012). Epistemology of Domain Analysis. In *Cultural Frames of Knowledge*, ed. Richard P. Smiraglia and Hur-Li Lee. Würzburg: Ergon-Verlag. Pp. 111-24.

Smiraglia, Richard P. (2014). *Cultural Synergy in Information Institutions*. Cham: Springer.

Francisco Arrais Nascimento, Francisco F. Leite Junior and Fabio Assis Pinho

What Gender Is This? Challenges to the Subject of Representation about the Gender Boundaries

Abstract

The domain of human sexualities is configured as a complex task under the aspect of classification studies and indexation, given the multifaceted performance of them. It is understandable that sexualities in modern times have earned space in the discussions of the fields of the Information Organization and Knowledge, where the notion of identity is evoked in the context of discussions in an attempt to embrace such domain. We sought to identify and organize the nomenclatures of non-binary gender variants that emerge from the borders of genders and has earned social space through discursive group of Lesbian, Gay, Bisexual, Transgender, Intersex (LGBTTI) and other sexual identity variations that deviate from the heteronormativity and consequently the gender binarism guided by current social proselytizing. The research of nature both qualitative and exploratory leaned on the theme, through documentary research, where the genre boundaries, from which emerge identity sexual categories that are configured as hybrids and/or previously unreleased productions, approaching and distancing of the binarism depending on the subject that builds, blurring the boundaries and creating something new in the context of gender identities already documented. Given the marginal nature of some sexual practices and gender performativities allocated to the bordering space, where no gender binarity emerges in order to "blur" the boundaries established by current social standards and certified by the social proselytizing, which demands an ethical concern in the indexing systems and classifications with the goal of represent the full breadth of this theme coming from the area of human sexualities in a way the best represent it, earning the space required by the idiosyncratic universe of sexualities. With this information about the bordering identities should obey a deepening in deviating universe to better understanding and representation to configure plausibly, ethical and impartial.

Introduction

The domain of human sexualities is configured as a complex task under the aspect of classification studies and indexation, given the multifaceted performance of them. It is understandable that sexualities in modern times have earned space in the discussions of the fields of the Information Organization and Knowledge, where the notion of identity is evoked in the context of discussions in an attempt to embrace such domain, despite still focus on a binary perspective and where arise eventual prejudices and antipathies (Berman, 1993).

In a simple way, it conceives that all of construction, an indiovidual or an identity, involves a certain degree of normalization, whose effect, produces exclusion. Such individuals occupying the exclusion zone are defined by (Butler, 2002) as abject bodies. (Prins; Meijer, 2002) recognize the contradiction between the "not to be" in this definition of "being contemptible," and his own existence as a "materializable" by an exclusion discourse. For (Butler, 2002) the abject relates to all kinds of bodies whose lives are not considered lives and whose materiality is understood as not important. Urging the reader to think that the configuration of the normative model required by the society leaves the margin particularities that escape the normal classification and

excludes sexuality as a multiplicity of combinations. Making realize the existing limits, both material and discursive.

Nowadays multiple experiences are lived that break away with the binary pattern of sex/gender not restricting in any way the sex and heteronormativity, but relates instead to all kinds of bodies whose lives are not considered lives and whose materiality is understood as not important, beyond the central bodies. The treatment aimed at certain forms of life within a disparity relation in the social sphere is described by (Butler, 2010, p. 13), below:

[...] Certain lives do not qualify as lives, or from the beginning are not designed as living within certain epistemological references, then, such lives will never consider lived or lost in the fullest sense of both words.

So if the materiality of the subject, if his own humanity and his rights are impractical, deconstructed and curtailed respectively, the silences are aimed as a way to control such subjects, which are not worthy of recognition as human under the social perspective, that takes root his precepts in a sanitized, elitist and heteronormative model. Thus was elected the border identities as a focus for this study. These perceived as thing, behavior, phenomenon and above all a complex and plural area to be studied because of its "no" importance, can be glimpse that what was built on and around the indentity/sexual behaviors, They guided up since the nineteenth century by the scientific medical discourse that prevailed until the last decades of the twentieth century and in recent years has been facing crumbling the deconstruction of the stereotypes that figured marginalization spaces, pathologization and exclusion beyond the issue of consolidation within the research social nature.

In other words, the space earned by the field of human sexuality in a broader spectrum of social practices, brings out a universe behind the imagetic and textual surface of the speech, the unsaid, which was built about this domain.

Segundo (Guattari and Rolnik, 1996, p. 12):

> [...] we find ourselves ordered all the time and from all sides to invest the powerful manufactures serialized, producer these men that we are, reduced the value support condition - and this even (and especially) when we occupy the most places privileged in the hierarchy of values. Everything leads to this kind of economy. Often there is no other way out. Is that when the disassembly, perplexed and devoid we weaken ourselves, the tendency is to adopt purely defensive positions. For fear of marginalization in which we run the risk of being confined when we dare create any natural territory, that is, independent of subjective serialisations; for fear of this marginalization reach to compromise even the very possibility of survival (which is entirely possible), we end up claiming a territory in the building of recognized identities. We become so - often at odds with our consonance - producers of some sequences of desire assembly line.

Thus approaching the identity notion thought by Stuart Hall and discussed by (Silva, 2007), where the idea of identity as temporary attachment points highlights the importance of context situated between elements that antecede and elements proceed. Noticing a wide range of influences that come from epistemologies that claim as truth their views of human and world.

The idea of gender has taken on positions and marking ways of living the sexuality that break away from the patterns, emphasizing the importance of performativity, where (Butler, 2010) wonders if "the sex" would have a story or if it is a given structure, free of questions in view of his unarguable materiality. This new positioning becomes a complex element within a social context from which emerge specific demands within the representations, of the classification studies and indexing itself.

In a similar move to (Scott, 1995), Butler plans to historicize the body and sex, dissolving the dichotomy sex x gender, that it provides to feminists limited possibilities for problematisation of the "biological nature" of men and women. For Butler, in our society we are facing a "compulsory order" that requires full coherence between a sex, a gender and a desire/practice that are obligatory heterosexual. That way the role of gender would be produce a false sense of stability in the heterosexual matrix would be provided by two fixed and consistent sexes, which are opposed to all the binary oppositions of occidental thought: male x female, man x woman, masculine and feminine, penis and vagina and etc. Besides a whole complex mechanism where the speech that leads to the maintenance of such compulsory order, it is about a social control device that transcends the said and the unsaid, from a Foucault perspective.

We sought to identify and organize the nomenclatures of non-binary gender variants that emerge from the borders of genders and has earned social space through discursive group of Lesbian, Gay, Bisexual, Transgender, Intersex (LGBTTI) and other sexual identity variations that deviate from the Heteronormativity and consequently the gender binarism guided by current social proselytizing.

Methodological Aspects

Aligned with the objective of the research and guided by qualitative exploratory nature, it was investigated by means of documentary research, the genre's borders, where emerging identity categories that constitute hybrids and/or unreleased productions, approaching and distancing of the binarism depending on the subject that builds, blurring the boundaries and creating something new in the context of gender identities already documented.

Is understood that the actions of organizing information that are originated and are intended for public LGBTTI and other sexual identity variations that deviate from the heteronormativity and gender binarism should consider the terminology used by them so that the works are properly classified and indexed, thereby generating reflections about their universe idiosyncratic as producer community and consumer information. These terminology studies for LGBTTI domain consider the literary, cultural and use guarantees (Beghtol, 1986; 2002, 2005; barite; Fernández-Molina, José Guimarães & Moraes, 2010), respecting a transcultural ethics of mediation (García Gutiérrez, 2002) and imposing limits to a "power to nominate" from the indexer (Olson, 2002) in their possible prejudices and antipathies (Berman, 1993).

For the study in question was identified thirty-five terms that designate identity sexual categories of non-binary genders, namely: Agender, Aliagender, Ambigender, Androgine, Bigender (female-male), Butch Non-binary, Cristaline, Demigender, Denboy, Demigirl, Efêmere, Femme Non-binary, Genderfluid (female-male), Genderflux, Genderfuck, Genderpivot, Genderqueer Non-binary, Graygender, Male Non-binary, Intergender or Intersex, Female Non-binary, Nan0gender, Nan0boy, Nan0girl, Nan0-menine, Negative, Neutrois, Pangender, Poligender, Positive, Third Gender, Transfemale or Male to Female - MTF, Transmale or Female to Male - FTM , Transvestite Non-binary e Trigender.

Results and Discussions

The results clarify that gender identities are constituted while a complex and subjective universe, which reveals how multiple are gender performances nowadays, guiding by the questions of (Butler, 2006), "Is there any way to link the issue of the materiality of the body with the performativity of gender? And how the category "sex" figure within such a relationship?". It's possible to realize clearly that the author indicates the coherence inability of gender identity, which, if considered in a binary and linear structure presupposes a need for adjustment to the norm by those who do not fall into such structures. (Butler, 2006) convokes thinking the configuration of the behavioral model required by society leaving aside anatomical or psychological particularities that escape the normal classification, and excludes sexuality as a multiplicity of combinations that do not arise from the psychosocial imposition. Quickly the idea of performativity in (Butler, 2006) brings and presents individuals excluded by the standard at the same level as the dominant genders, ie the normative ideal is illusory character and can not be decisive in the classification of sexual identities when normal or pathological. The body does not completely accepts the rules imposed its materialization. In that sense, the body resists both the intentions of the subject and to social norms. According to (Lauretis, 1994) the idea that the speech in its entirety, help to perpetuate the stereotypical differences imposed to distinguish male and female. Under that north, some points emerge: First, the "Gender is a representation" and is concretized in people's behavior. Second: "The representation of the genre is its your construction" and evolves as society evolves too. Third: the construction of gender is uninterrupted. Finally, claims that the construction of gender also makes through its deconstruction.

Closing Remarks

Given the marginal nature of some sexual practices and gender performativities allocated to the bordering space, where no gender binarity emerges in order to "blur" the boundaries established by current social standards and certified by the social proselytizing, consolidating what (Rich, 2010), defines as being compulsory

heterosexuality, allocating the same as "normal" and "natural", not being perceived the construction from the normative crossings pronounced by the institutions.

Since all the subjects are normatized in order to be adequate to the society, accepting while natural obligatoriness, which demands an ethical concern in the indexing systems and classifications with the goal of represent the full breadth of this theme coming from the area of human sexualities in a way the best represent it, earning the space required by the idiosyncratic universe of sexualities maintaining the necessary impartiality. With this information about the bordering identities should obey a deepening in deviating universe to better understanding and representation to configure plausibly, ethical and impartial.

References

Albuquerque JR, Dúrval Muniz de (2009).*A invenção do Brasil e outras artes.*4. ed. São Paulo: Cortez.

Barité, Mario, Fernández Molina, Juas Carlos, Guimarães, José Augusto C. & Moraes, JoãoBatista E. (2010). Garantia literária: elementos para uma revisão crítica após um século. *Transinformação,* 22: 123-38.

Beghtol, Clare (1986) Bibliographic classification theory and text linguistics: aboutness analysis, intertextuality and the cognitive act of classifying documents. *Journal of Documentation,* 42(2): 84-113.

Beghtol, Clare (2005). Ethical decision-making for knowledge representation and organization systems for global use. *Journal of the American Society for Information Science and Technology,* 56(9): 903-12.

Beghtol, Clare (2002) A proposed ethical warrant for global knowledge representation and organization systems. *Journal of Documentation,* London, 58(5): 507-32.

Beghtol, Clare (2002). Universal concepts, cultural warrant, and cultural hospitality. In *Challenges in knowledge representation and organization for the 21st century: integration of knowledge across boundaries*, edited by Maria José López-Huertas.Würzburg: Ergon Verlag. Pp. 45-9.

Berman, Sanford (1993). *Prejudice and antipathies:a tract on the LC subject heads concerning people.*Jefferson: McFarland & Company Inc. Publishers.

Butler, Judith (2002). *Cuerpos que importan: sobre los límites materiales y discursivos del "sexo".* Buenos Aires: Paidós.

Butler, Judith (2006). Corpos que pesam. Sobre os limites discursivos do sexo. In*O corpo educado: Pedagogias da sexualidade*, edited by Guacira Lopes Louro. Belo Horizonte: Autêntica.

Butler, Judith (2010). *Problemas de gênero: feminismo e subversão da identidade.* Rio de Janeiro: Civilização Brasileira.

García Gutiérrez, Antonio (2002). Knowledge organization from a “culture of the border”: towards a transcultural ethics of mediation. In *Challenges in knowledge representation and organization for the 21st century: integration of knowledge across boundaries*, edited by Maria José López-Huertas.Würzburg: ErgonVerlag. Pp. 516-22.

Guattari, Felix& Rolnik, Suely (1996). *Micropolítica: cartografias do desejo.* Petrópolis: Vozes.

Lauretis, Teresa de (1994). A tecnologia do gênero. In*O feminismo como crítica cultural,* edited by Heloisa Buarque Holanda.Rio de Janeiro: Rocco. Pp. 206-42.

Olson, Hope A. (2002*). The power to name: locating the limits of subject representation in libraries.* Dordrecht: Kluwer Academic Publishers.

Petry, Analídia R. & Meyer, Dagmar E. E. (2011). Transexualidade e heteronormatividade: algumas questões para a pesquisa.*Textos & Contextos,* 10(1): 193-8.

Prins, Baukje & Meijer, Irene Costera (2002). Como os corpos se tornam matéria: entrevista com Judith Butler. *Revista Estudos Feministas,* 10(1): 155-67.

Rich, Adrienne (2010). *Heretossexualidade compulsória e existência lésbica [1981]*. Bagoas.Pp. 17-44.

Rolnik, Suely (1989). *Cartografia sentimental: Transformações contemporâneas do desejo.* São Paulo: Estação Liberdade.

Scott, Joan W. (1995). Gênero: uma categoria útil de análise histórica. *Educação e Realidade,* Porto Alegre, 20(2): 71-99.

Silva, Tomaz Tadeu, Hall, Stuart & Woodward, K. (2007). *Identidade e diferença: a perspectiva dos estudos culturais.* 7 ed. Petrópolis: Vozes.

List of Contributors and Author's Index

Adagunodo, Emmanuel Rotimi ***88, 420***
Obafemi Awolowo University
Ile-Ife − Nigeria
eadagun@yahoo.com

Aderibigbe, Stephen Ojo ***88, 420***
Lagos State Polytechnic
Lagos − Nigeria
aderibigbe2000@gmail.com

Aderonke, Kayode Anthonia ***88***
Osun State College of Technology
Esa-Oke – Nigeria
kayodeaa2@gmail.com

Afolabi, Babajide Samuel ***88, 420***
Obafemi Awolowo University
Ile-Ife − Nigeria
afolabib@gmail.com

Akhigbe, Bernard Ijesunor ***88, 420***
Obafemi Awolowo University
Ile-Ife − Nigeria
benplus1@gmail.com

Almeida, Tatiana ***350***
Rio de Janeiro State University
Rio de Janeiro − Brazil
tatiana.almeida@unirio.br

Alves, Bruno Henrique ***227***
São Paulo State University
Marília SP − Brazil
brhenriquealves@gmail.com

Andrade, Tesla Coutinho ***171***
Federal University of Rio de Janeiro
Rio de Janeiro − Brazil
teslacoutinho@uol.com.br

Apoorva, K. H. ***342***
PES Institute of Technology
Bangalore − India
apoorva.khp@gmail.com

Araujo, Andre Vieira de Freitas ***59***
Federal University of Rio de Janeiro
Rio de Janeiro − Brazil
armarius.araujo@gmail.com

Araújo, Paula Carina ***36***
São Paulo State University
Marília − Brazil
paula.carina.a@gmail.com

Araújo, Webert Júnio ***326***
Federal University of Minas Gerais
Belo Horizonte − Brazil
webertaraujo@ufmg.br

Arave, G. ***308***
Indiana University Bloomington
Bloomington - United States
garave@indiana.edu

Arboit, Aline Elis ***193***
Court of Auditors of the State of Paraná
Curitiba − Brazil
aarboit@yahoo.com.br

Babik, Wiesław ***451***
Jagiellonian University
Krakow − Poland
w.babik@uj.edu.pl

Barbosa, Nilson Theobald ***350***
Federal University of Rio de Janeiro
Rio de Janeiro − Brazil
nilson@tbarbosa.org

Barité, Mario ***146***
Universidad de la República
Montevideo – Uruguay
mabarite@gmail.com

Barros, Camila Monteiro ***164***
Federal University of Santa Catarina
Florianópolis − Brazil
camila.c.m.b@ufsc.br

Bernstein, Jay H. ***51***
Kingsborough Community College
Brooklyn −United States
jay.bernstein@kbcc.cuny.edu

Bezerra, Lilian Miranda ***403***
University of São Paulo – Brazil
São Paulo − Brazil
lilianmiranda@usp.br

Bhargav, K. N. ***411***
PES Institute of Technology
Bangalore − India
bhargavkn.1996@gmail.com

Bräscher, Marisa ***211***
Federal University of Santa Catarina
Florianópolis − Brazil
marisa.brascher@gmail.com

Brito, Marcilio ***265***
University of Brasília
Brasília − Brazil
mdbrito@unb.br

Budd, John M. ***142***
University of Missouri
Columbia − United States
buddj@missouri.edu

Bufrem, Leilah Santiago ***227***
Federal University of Pernambuco
Recife – Brazil
santiagobufrem@gmail.com

Café, Lígia Maria Arruda ***164***
Federal University of Santa Catarina
Florianópolis – Brazil
ligia.cafe@ufsc.br

Campbell, D. Grant ***523***
University of Western Ontario
London – Canada
gcampbel@uwo.ca

Campos, Maria Luiza de Almeida ***493***
Fluminense Federal University
Niteroi – Brazil
marialuizalmeida@gmail.com

Carrasco, Laís Barbudo ***317***
São Paulo State University
Marília – Brazil
laiscarrasco@hotmail.com

Carrieri, Patrizia ***530***
University of Marseille
Marseille – France
pmcarrieri@aol.com

Carvalho, Lidiane S. ***376***
Oswaldo Cruz Foundation
Rio de Janeiro – Brazil
Carvalho.ldn@gmail.com

Casarin, Helen de Castro Silva ***469***
São Paulo State University
Marília – Brazil
helenc@.marilia.unesp.br

Castanha, Renata Cristina Gutierres ***219***
São Paulo State University
Marília – Brazil
regutierres@gmail.com

Choi, Inkyung ***116***
University of Wisconsin-Milwaukee
Milwaukee – United States
ichoi@uwm.edu

Crippa, Giulia ***59, 179***
University of São Paulo
Ribeirão Preto – Brazil
giuliac@ffclrp.usp.br

Cunha, Ana Maia ***350***
Fluminense Federal University
Niterói – Brazil
aninhamaia@vm.uff.br

Dal'Evedove, Paula Regina ***515***
Federal University of São Carlos
São Carlos – Brazil
dalevedove@ufscar.br

Deng, Jun ***275***
Jilin University
Jilin – China
dengjun@jlu.edu.cn

Dodebei, Vera ***171***
Federal University of Rio de Janeiro
Rio de Janeiro – Brazil
dodebei@gmail.com

Esteves, Monnique S. P. A. ***350***
Fluminense Federal University
Niterói RJ – Brazil
monniquespa@gmail.com

Franca, Aline da Silva ***502***
Federal University of Rio de Janeiro
Rio de Janeiro RJ – Brazil
francaaline@yahoo.com.br

Francelin, Marivalde Moacir ***572***
University of São Paulo
São Paulo – Brazil
marivalde@usp.br

Fujita, Mariângela Spotti Lopes ***29, 234, 257, 515***
São Paulo State University
Marília – Brazil
fujita@marilia.unesp.br

García-Marco, Francisco-Javier ***105, 546***
University of Zaragoza
Zaragoza – Spain
jgarcia@unizar.es

Gil Leiva, Isidoro ***243***
University of Murcia
Mucia – Spain
isgil@um.es

Gnoli, Claudio ***368***
University of Pavia
Pavia - Italy
claudio.gnoli@unipv.it

Grácio, Maria Cláudia Cabrini ***219***
São Paulo State University
Marília – Brazil
cabrini@marilia.unesp.br

Graf, Ann M. ***125***
University of Wisconsin-Milwaukee
Milwaukee – United States
anngraf@uwm.edu

Guimarães, José Augusto Chaves ***36, 193***
São Paulo State University
Marília – Brazil
guima@marilia.unesp.br

Helfer, Olivia ***333***
University at Buffalo
Cheektowaga – United States
oliviahe @buffalo.edu

Henry, Joshua A. ***156***
University of Wisconsin-Milwaukee
Milwaukee – United States
henryja@uwm.edu

Iyer, Hemalata ***384***
State University of New York
Albany – United States
hiyer@albany.edu

Jacob, Elin K. ***308***
Indiana University Bloomington
Bloomington – United States
ejacob@indiana.edu

Jivrajani, Aarti ***342***
PES Institute of Technology
Bangalore – India
aartij17@gmail.com

Karanth, Pallavi ***437***
PES University
Bangalore – India
karanthpallavi@gmail.com

Keilty, Patrick ***579***
University of Toronto
Toronto – Canada
p.keilty@utoronto.ca

Kleineberg, Michael ***133***
Humboldt-Universität zu Berlin
Berlin – Germany
michael.kleineberg@hu-berlin.de

Laplante, Audrey ***164***
University of Montréal
Montréal – Canada
audrey.laplante@umontreal.ca

Lee, Hur-Li ***116***
University of Wisconsin-Milwaukee
Milwaukee – United States
hurli@uwm.ed

Lee, Wan-Chen ***485***
University of Washington
Seattle – United States
leew23@uw.edu

Leite Junior, Francisco Francinete ***587***
University of Fortaleza
Fortaleza – Brazil
freud.g@bol.com.br

Lima, Gercina Ângela Borém de Oliveira ***300, 326***
Federal University of Minas Gerais
Belo Horizonte – Brazil
glima@eci.ufmg.br

Lima, Vânia Mara Alves ***283***
University of São Paulo
São Paulo SP – Brazil
vamal@usp.br

López-Huertas, Maria José ***13***
University of Granada
Granada – Spain
milopez@ugr.es

Maculan, Benildes Coura Moreira dos Santos ***300***
Federal University of Minas Gerais
Belo Horizonte – Brazil
benildes@gmail.com

Mahesh, Kavi ***437***
PES University
Bangalore – Índia
drkavimahesh@gmail.com

Marcondes, Carlos H. ***350, 493***
Fluminense Federal University
Niterói – Brazil
marcon@vm.uff.br

Materska, Katarzyna ***555***
Cardinal Stefan Wyszynski University
Warszawa – Poland
katarzyna.materska@gmail.com

Martínez-Ávila, Daniel ***142***
São Paulo State University
Marília – Brazil
dmartinezavila@marilia.unesp.br

Martinho, Noemi Oliveira ***515***
São Paulo State University
Marília – Brazil
gleanom@yahoo.com

Mattos, Nayara Bernardo ***469***
São Paulo State University
Marília – Brazil
naybmattos@gmail.com

Messa, Joyce ***350***
Brazilian Public Work Ministry
Brazil
joyce.messa@mpt.mp.br

Moraes, João Batista Ernesto ***201***
Sao Paulo State University
Marília – Brazil
jota@marilia.unesp.br

Moura, Maria Aparecida ***251, 538***
Federal University of Minas Gerais
Belo Horizonte – Brazil
mamoura@ufmg.br

Mustafa El Hadi, Widad ***265, 392***
University of Lille 3
Villeneuve d'Ascq – France
widad.mustafa@univ-lille3.fr

Nascimento, Francisco Arrais ***587***
University of Pernambuco
Recife – Brazil
francisco.arrais.nascimento@gmail.com

Naumis-Peña, Catalina ***564***
Universidad Nacional Autónoma de México - México City – México
naumis@unam.mx or cnaumis@gmail.com

Neves, Dulce Amélia de Brito ***234***
Federal University of Pernambuco
Recife – Brazil
damelia1@gmail.com

Ohly, H. Peter ***460***
Leibniz Institute for the Social Sciences
GESIS
Bonn – Germany
peter.ohly@gmx.de

Oliveira, Elaine Diamantino ***300***
Federal University of Minas Gerais
Belo Horizonte – Brazil
nanadiamantino@yahoo.com.br

Oliveira, Ely Francina Tannuri ***227***
São Paulo State University
Marília – Brazil
etannuri@gmail.com

Orrico, Evelyn Goyannes Dill ***508***
Federal University of Rio de Janeiro
Rio de Janeiro – Brazil
orrico.evelyn@gmail.com

Paula, Lorena Tavares de ***251***
Federal University of Minas Gerais
Belo Horizonte – Brazil
lorena.ltp@gmail.com

Pereira, Diogo Alves Cândido ***429***
Federal Center for Technological Education - Belo Horizonte – Brazil
diogoac.pereira@gmail.com

Pierozzi Júnior, Ivo ***326***
Embrapa Agricultural Informatics
Campinas – Brazil
ivo.pierozzi@embrapa.br

Pika, Jiri ***75***
UDC Editorial Team
Zürich – Switzerland
pika@gmx.ch or jpika@ethz.ch

Pinheiro, Lena Vania Ribeiro ***29***
Brazilian Institute of Information in Science and Technology
Rio de Janeiro – Brazil
lenavania@ibict.br

Pinho, Fabio Assis ***587***
Federal University of Pernambuco
Recife – Brazil
fabiopinho@ufpe.br

Pires, Ana Silvia ***403***
University of São Paulo
São Paulo – Brazil
anapires18@usp.br

Portugal, Rosana ***350***
Fluminense Federal University
Niterói – Brazil
rosanabiblio@gmail.com

Raghavan, K. S. ***342, 411***
PES Institute of Technology
Bangalore – India
ksragav@hotmail.com

Rao, I. K. Ravichandra ***411***
PES Institute of Technology
Bangalore – India
ikrrao@hotmail.com

Redigolo, Franciele Marques ***515***
Federal University of Pará
Belém – Brazil
franbiblio@gmail.com

Ridenour, Laura ***43, 477***
University of Wisconsin-Milwaukee
Milwaukee – United States
ridenour@uwm.edu

Rodriguez, Sonia Troitiño ***234***
São Paulo State University
Marília – Brazil
stroitino@globo.com

Romero-Millán, Camelia ***564***
El Colegio de México
Mexico City – Mexico
cromero@colmex.mx

Rosas, Fábio Sampaio ***219***
São Paulo State University
Marília – Brazil
fabio@dracena.unesp.br

Roszkowski, Marcin ***392***
University of Warsaw
Warsaw – Poland
m.roszkowski@uw.edu.pl

Roux, Perrine ***530***
University of Marseille – France
Marseille – France
perrine.roux@inserm.fr

Sabba, Fiammetta ***59***
University of Bolonha
Ravenna – Italy
fiammetta.sabba@unibo.it

Sabbag, Deise ***179***
University of São Paulo
Ribeirão Preto – Brazil
deisesabbag@usp.br

Saldanha, Gustavo ***186***
Federal University of Rio de Janeiro
Rio de Janeiro – Brazil
gustavosaldanha@ibict.br

Sales, Rodrigo de ***67***
Fluminense Federal University
Niterói – Brazil
rodrigosales@id.uff.br

Santis, Rodrigo de ***368***
Federal Institute of Parana
Irati – Brazil
rodrigodesantis@gmail.com

Schmidt, Clarissa Moreira dos Santos ***403***
Fluminense Federal University
Niterói RJ – Brazil
clarissaschmidt@id.uff.br

Silva, Carlos Guardado da ***290***
University of Lisboa
Torres Vedras – Portugal
carlosguardado@campus.ul.pt

Silva, Edson Marchetti da ***429***
Federal Center for Technological
Education - Belo Horizonte – Brazil
emarchettisilva@decom.cefetmg.br

Silva, Eliezer Pires ***508***
Federal University of Rio de Janeiro
Rio de Janeiro – Brazil
eliezerpires@gmail.com

Silva, Márcia Regina ***179***
University of São Paulo
Ribeirão Preto – Brazil
marciaregina@usp.br

Silveira, Naira Christofoletti ***186, 502***
Federal University of Rio de Janeiro
Rio de Janeiro – Brazil
naira.silveira@unirio.br

Smiraglia, Richard P. ***43, 97, 156, 579***
University of Wisconsin-Milwaukee
Milwaukee – United States
smiragli@uwm.edu

Smit, Johanna Wilhelmina ***403***
University of São Paulo
São Paulo – Brazil
johannawsmit@gmail.com

Soergel, Dagobert ***275, 333***
University at Buffalo– United States
dsoergel@buffalo.edu

Solis Y., Marisol ***283***
University of São Paulo
São Paulo – Brazil
msolis@ime.usp.br

Souza, Renato Rocha ***243, 429***
Getúlio Vargas Foundation
Rio de Janeiro – Brazil
renato.souza@fgv.br

Spire, Bruno ***530***
University of Marseille – France
Marseille – France
bruno.spire@inserm.fr

Suenaga, Cynthia Maria Kiyonaga ***201***
Sao Paulo State University
Marília – Brazil
cmksuenaga@gmail.com

Szostak, Rick ***359***
University of Alberta
Edmonton – Canada
rszostak@ualberta.ca

Tanti, Marc ***530***
University of Marseille – France
Marseille - France
mtanti@gmx.fr

Tartarotti, Roberta C. Dal'Evedove ***257***
São Paulo State University
Marília – Brazil
roberta_tartarotti@yahoo.com.br

Tennis, Joseph T. ***84***
University of Washington
Seattle – Unites States
jtennis@uw.edu

Tognoli, Natália Bolfarini ***201***
Sao Paulo State University
Marília – Brazil
nataliatognoli@marilia.unesp.br

Trivelato, Rosana Matos da Silva ***538***
Federal University of Minas Gerais
Belo Horizonte – Brazil
rosanatrivelato@ufmg.br

Vargas, Marli Marques de Souza ***403***
University of São Paulo
São Paulo – Brazil
marlimsv@usp.br

Vianna, William Barbosa ***211***
Federal University of Santa Catarina
Florianópolis – Brazil
william.vianna@ufsc.br

Vidotti, Silvana Borsetti Gregório ***317***
São Paulo State University
Marília – Brazil
vidotti@marilia.unesp.br

Vieira, Simone Bastos ***265***
University of Brasília - Brazil
sbastosvieira@gmail.com

Wassermann, Renata ***283***
University of São Paulo
São Paulo – Brazil
renata@ime.usp.br

Weiss, Leila Cristina ***211***
Federal University of Santa Catarina
Florianópolis – Brazil
eilacw@gmail.com

Zamboni, Rita Costa Veiga ***572***
University of São Paulo
São Paulo – Brazil
ritazamboni@gmail.com

Zeng, Marcia Lei ***444***
Kent State University
Kent – United States
mzeng@kent.edu

Žumer, Maja ***265, 444***
University of Ljubljana - Slovenia
maja.zumer@ff.uni-lj.si

Scientific Committee

Abdelkrim Meziane, Research Centre on Scientific and Technical Information, Algeria
Alan Gilchrist, The Cura Consortium, United Kingdom
Amos David, Université de Lorraine, France
Antonio Luis García Gutiérrez, Universidad de Sevilla, Spain
Barbara H. Kwasnik, Syracuse University, United States
Blanca Rodríguez Bravo, Universidad de León, Spain
Carlos Cândido de Almeida, Universidade Estadual Paulista, Brazil
Christian Wartena, Hochschule Hannover, Germany
Claudio Gnoli, Universitá di Pavia, Italy
Clément Arsenault, Université de Montréal, Canada
Dagobert Soergel, University of Buffalo, United States
Daniel Martínez-Ávila, Universidade Estadual Paulista, Brazil
Elias Sanz Casado, Universidad Carlos III de Madrid, Spain
Emmanuelle Chevry-Pebayle, Université de Strasbourg, France
Fabio Assis Pinho, Universidade Federal de Pernambuco, Brazil
Fernanda Ribeiro, Universidade do Porto, Portugal
Francisco Javier García Marco, Universidad de Zaragoza, Spain
H. Peter Ohly, German Social Science Infrastructure Services, Germany
Hur-li Lee, University of Wisconsin-Milwaukee, United States
Isidoro Gil-Leiva, Universidad de Murcia, Spain
Jens-Erik Mai, Royal School of LIS, Denmark
Johanna W. Smit, Universidade de São Paulo, Brazil
Jonathan Furner, University of California, United States
José Antonio Moreiro-González, University Carlos III of Madrid, Spain
José Augusto Chaves Guimarães, São Paulo State University, Brazil
Joseph T. Tennis, University of Washington, United States
Juan Carlos Fernández-Molina, University of Granada, Spain
Judi Vernau, Metataxis Ltd, United Kingdom
K. S. Raghavan, PES Institute of Technology, India
Kathryn La Barre, University of Illinois at Urbana-Champaign, United States
Lynne Howarth, University of Toronto, Canada
M. P. Satija, Guru Nanak Dev University, India
Maja Žumer, University of Ljubljana, Eslovania
Malek Ghenima, Université de la Manouba, Tunisia
Marcia Zeng Lei, Kent State University, United States
Maria Manuel Borges, University of Coimbra, Portugal
Marilda Lopez Ginez de Lara, University of São Paulo, Brazil
Mario Barité, Universidad de la República, Uruguay

ISBN 978-3-95650-221-7
ISSN 0938-5495