

Gil Leiva, Isidoro

La automatización de la indización,
propuesta teórico- metodológica:
aplicación al área de biblioteconomía
y documentación

Universidad de Murcia
Servicio de Publicaciones

A mis padres por su
confianza y apoyo sin peros

AGRADECIMIENTOS

Quiero agradecer a Javi, Gregorio y Pedro el apoyo que me han prestado durante el largo período en el que he llevado a cabo este trabajo. Asimismo, quiero agradecer a Vivina los continuos consejos y ánimos desde que inicié esta labor.

También doy las gracias a Yolanda, Juani, Pepita y Antonio por su ayuda.

Finalmente, agradezco a los doctores Rodríguez Muñoz y Vera Luján la dirección de esta tesis doctoral.

Has de saber que esta vida es
el minúsculo chapoteo
de una gota de agua.
Una bella criatura que desaparece
en el mismo momento
en que empieza a existir.
Por lo tanto, márcate tu meta,
y aprovecha al máximo cada día
y cada noche para alcanzarla.

Tsong-khapa

RESUMEN

Se expone un marco conceptual sobre la automatización de la indización concretado en su delimitación, los posicionamientos de los investigadores en Biblioteconomía y Documentación con respecto a estas indagaciones, el desarrollo diacrónico ocurrido en esta automatización, y en la explicitación de la interdisciplinariedad inherente a este proceso. Se presenta una propuesta teórico-metodológica para diseñar un procedimiento semiautomático para la indización de documentos sobre Biblioteconomía y Documentación constituido por cuatro módulos. En los tres primeros se preparan las fuentes utilizadas, se seleccionan los términos candidatos a descriptores y se valoran y ponderan dichos términos, mientras que en el cuarto módulo el usuario ejecuta una validación y edición interactiva de los resultados propuestos. El sistema se fundamenta en el uso de un vocabulario controlado sobre Biblioteconomía y Documentación construido para tal fin. La consistencia media obtenida entre la indización de cincuenta artículos analizados por indizadores de la Base de datos ISOC y por nuestra propuesta es de 25,93%.

ABSTRACT

A conceptual framework is described for the automatization of indexing involving its delimitation, the positioning of researchers in Library Science and Documentation with respect to these investigations, the diachronous development that has occurred in this automatization, and specifying the inherent interdisciplinary nature of the process. A theoretical-methodological proposal is presented to design a semi-automatic procedure for indexing Library Science and Documentation documents. It consists of four modules. In the first three modules, the sources to be used are prepared, the terms to be candidates for descriptors are first selected, and then evaluated and assigned weights. In the fourth module the user interactively edits and convalidates the proposed results. The system is based on the use of a controlled Library Science and Documentation vocabulary constructed to this end. The mean consistency obtained for the indexing of 50 articles analyzed by ISOC data base indexers and by our proposal was 25.93%.

ÍNDICE

0.- INTRODUCCIÓN	1
1. Aspectos formales	1
2. Motivaciones	2
2.1 Motivaciones científicas.....	2
2.2 Motivaciones personales	3
3. Metodología	4
4. Objetivos	10
5. Estructuración y presentación.....	11
1.- LA INDIZACIÓN	14
1.1. El proceso documental	14
1.2. La indización.....	16
1.2.1. La indización. Definición.....	16
1.2.2. Las etapas de la indización	19
1.2.3. Las zonas de extracción de los conceptos.Tiempo dedicado	20
1.3. Las características de la indización	22
1.3.1. La exhaustividad en la indización.....	22
1.3.2. La especificidad en la indización	25
1.3.3. La corrección de la indización	25
1.3.4. La consistencia de la indización	26
1.4. La recuperación documental: Evaluación del resultado de la respuesta documental.....	29
1.4.1. La exhaustividad y la precisión en la recuperación	31
1.5. Los sistemas de indización	34
1.5.1. La indización por materias.....	34
1.5.2. La indización por unitérminos.....	35
1.5.3. La indización por descriptores.....	35
1.5.3.1. Los enlaces entre los descriptores.....	37
1.5.3.2. Los operadores utilizados en las preguntas docu- mentales.....	40
1.5.3.3. Las características generales de los descriptores	42
1.6. Los lenguajes documentales	42
1.6.1. La tipología de los lenguajes documentales.....	43
1.6.1.1. Las listas de palabras clave	44
1.6.1.2. Las clasificaciones	44
1.6.1.3. Las listas de encabezamientos de materias	45

1.6.1.4. Los tesauros.....	48
1.7. La normalización de la indización	50
2.- LA AUTOMATIZACIÓN DE LA INDIZACIÓN	53
2.1. Introducción	53
2.2. La indización humana versus indización automática	54
2.2.1. Argumentaciones en contra de la automatización de la indización	54
2.2.2. Argumentaciones a favor de la automatización de la indización	57
2.3. El desarrollo diacrónico de la automatización de la indización	60
2.3.1. Los métodos estadísticos	61
2.3.2. Los métodos lingüísticos	65
2.3.3. El uso de tesauros.....	74
2.3.4. El uso de sistemas híbridos.....	76
2.3.5. La comparación de la eficacia de la indización automática versus manual	79
2.4. La interdisciplinariedad en la automatización de la indización.....	81
2.4.1. Lingüística	85
2.4.2. Terminología	87
2.4.3. Informática.....	88
2.4.4. Lingüística computacional	89
2.4.5. Estadística	90
2.4.6. Sistemas expertos.....	91
2.5. La automatización de la indización para información no textual.....	92
2.6. El nivel de implantación de sistemas para la automatización de la indización	93
2.7. Esquema representativo de las herramientas utilizadas en la automatización de la indización	98
3.- PROPUESTA TEÓRICO-METODOLÓGICA PARA LA AUTOMATIZACIÓN DE LA INDIZACIÓN EN EL ÁREA DE BIBLIOTECONOMÍA Y DOCUMENTACIÓN	100
3.1. Introducción	100
3.2. La elección de las fuentes utilizadas en la automatización de la indización	101

3.2.1. Los antecedentes de estos estudios	101
3.2.2. Ensayo en las Bases de datos del CSIC	103
3.2.2.1. Material y métodos	104
3.2.2.2. Resultados	105
3.2.2.3. Conclusiones.....	106
3.3. La elección de un vocabulario controlado.....	107
3.3.1. La justificación de la elección de un vocabulario controlado	108
3.3.2. La elaboración de la lista de términos autorizados	111
3.4. Propuesta para la automatización de la indización.....	115
3.4.1. Los módulos del sistema	117
3.4.1.1. Módulo 1: Preprocesamiento	118
3.4.1.2. Módulo 2: Procesamiento	123
3.4.1.3. Módulo 3: Valoración y ponderación	132
3.4.2. Análisis de un documento en función de la propuesta precedente.....	135
3.4.3. La evaluación de la propuesta	149
3.4.4. Los problemas detectados	156
4.-CONCLUSIONES	157
5.-BIBLIOGRAFÍA	165
6.-ANEXOS	191
Anexo 1: Glosario	191
Anexo 2: Definiciones sobre indización.....	195
Anexo 3: Etapas en la indización	197
Anexo 4: Proceso íntegro de la indización	199
Anexo 5: Palabras vacías.....	200
Anexo 7: Índices de consistencia resultantes.....	202
Anexo 8: Vocabulario controlado.....	220

INDEX

0. INTRODUCTION	1
1. Formal aspects	1
2. Motivations.....	2
2.1 Scientific motivations.....	2
2.2 Personal motivations	3
3. Methodology	4
4. Objectives	10
5. Arrangement and presentation	11
1. INDEXING	14
1.1 The documentary process	14
1.2 Indexing	16
1.2.1 Indexing. Definition.....	16
1.2.2. The stages of indexing	19
1.2.3. The zones of extraction of concepts. Dedicated time.....	20
1.3. The characteristics of indexing	22
1.3.1. Completeness in indexing	22
1.3.2. Specificity in indexing.....	25
1.3.3. Correction of indexing	25
1.3.4. Consistency of indexing	26
1.4 Information retrieval: evaluation of the result of the documentary response	29
1.4.1. Completeness and accuracy in retrieval.....	31
1.5. Indexing systems	34
1.5.1. Indexing by subject.....	34
1.5.2. Indexing by uniterms	35
1.5.3. Indexing by descriptors	35
1.5.3.1. Links between descriptors	37
1.5.3.2. The operators used in documentary questioning...	40
1.5.3.3. General characteristics of the descriptors.....	42
1.6. Documentary languages	42
1.6.1. Typology of documentary languages.....	43
1.6.1.1. Keyword lists.....	44
1.6.1.2. Classifications.....	44
1.6.1.3. Subject-heading lists.....	45
1.6.1.4. Thesauri.....	48
1.7. The normalization of indexing.....	50

2. THE AUTOMATIZATION OF INDEXING	53
2.1. Introduction	53
2.2. Human indexing versus automatic indexing.....	54
2.2.1. Arguments against automatization of indexing.....	54
2.2.2. Arguments in favour of automatization of indexing	57
2.3. The diachronic development of the automatization of indexing	60
2.3.1. Statistical methods	61
2.3.2. Linguistic methods.....	65
2.3.3. The use of thesauri	74
2.3.4. The use of hybrid systems	76
2.3.5. Comparison of the efficacy of manual versus automatic indexing.....	79
2.4. Interdisciplinarity in the automatization of indexing.....	81
2.4.1. Linguistics.....	85
2.4.2. Terminology.....	87
2.4.3. Computer science	88
2.4.5. Computational linguistics.....	89
2.4.5. Statistics	90
2.4.6. Expert systems.....	91
2.5. The automatization of indexing for non-text information	92
2.6. The level of implantation of systems for the automatization of indexing	93
2.7. Representative scheme of the tools used in the automatization of indexing	98
3. THEORETICAL-METHODOLOGICAL PROPOSAL FOR THE AUTOMATIZATION OF INDEXING IN THE AREA OF LIBRARY SCIENCE AND DOCUMENTATION	100
3.1 Introduction	100
3.2. The choice of sources used in the automatization of indexing	101
3.2.1. Antecedents of these studies	101
3.2.2. Trial with CSIC data bases.....	103
3.2.2.1. Material and methods	104
3.2.2.2. Results.....	105
3.2.2.3. Conclusions	106
3.3. The choice of a controlled vocabulary	107
3.3.1. The justification of the choice of a controlled vocabulary ..	108
3.3.2. Construction of the list of authorized terms	111

3.4. Proposal for the automatization of indexing.....	115
3.4.1. The modules of the system	117
3.4.1.1. Module 1: Preprocessing	118
3.4.1.2. Module 2: Processing	123
3.4.1.3. Module 3: Evaluation and weight assignation.....	132
3.4.2. Analysis of a document as a function of the preceding proposal	135
3.4.3. Evaluation of the proposal.....	149
3.4.4. Problems detected	156
4.-CONCLUSIONS	157
5.-REFERENCES	165
APPENDICES.....	191
Appendix 1: Glossary	191
Appendix 2: Definitions concerning indexing	195
Appendix 3: Stages in indexing.....	197
Appendix 4: Integral process of indexing	199
Appendix 5: Stop words	200
Appendix 7: Resultant consistency indices	202
Appendix 8: Controlled vocabulary	220

0.- INTRODUCCIÓN

1 ASPECTOS FORMALES.

El Doctorado, según el Real Decreto 185/1985 del 16 de febrero, constituye la condición esencial para el progreso científico, social y económico de una comunidad. La formación de los investigadores depende de la profundidad de sus contenidos y la seriedad en su planteamiento. Por ello, la Ley de Reforma Universitaria se plantea cuatro grandes objetivos en los estudios de postgrado:

1. Disponer de un marco adecuado para la consecución y transmisión de los avances científicos.
2. Formar a los nuevos investigadores y preparar equipos de investigación que afronten con éxito el reto que suponen las nuevas ciencias, técnicas y metodologías.
3. Impulsar la formación del nuevo profesorado.
4. Perfeccionar el desarrollo profesional, científico, técnico y artístico de los titulados superiores.

La Ley señala como requisitos para la obtención del título de Doctor, la necesidad de estar en posesión del título de Licenciado, Arquitecto o Ingeniero, para: a) realizar y aprobar los cursos y seminarios del programa de Doctorado correspondiente con una duración de, al menos, dos cursos académicos, y b) presentar y aprobar una Tesis Doctoral consistente en un trabajo original de investigación, ambas fases bajo la supervisión y responsabilidad académica de un Departamento¹.

Tras la obtención del título de Licenciado, y con posesión del título de Diplomado en Biblioteconomía y Documentación, comenzamos los Cursos de Doctorado en el

¹ Real Decreto nº 185/1985, publicado en el Boletín Oficial del Estado nº 41 del 16 de febrero de 1985 por el que se regula el tercer ciclo de estudios universitarios, la obtención y expedición del título de Doctor y otros estudios postgraduados.

programa «TÉCNICAS Y MÉTODOS ACTUALES EN INFORMACIÓN Y DOCUMENTACIÓN»², coordinado por los Dres. Vivina Asensi y José V. Rodríguez Muñoz, correspondiente al bienio 94/96. Estos cursos nos sirvieron para perfilar e iniciar la presente Tesis doctoral.

2 MOTIVACIONES.

2.1 Científicas.

Entre las razones que justifican las investigaciones para la automatización de la indización destacan:

La subjetividad está presente en el proceso de la indización. El grado de coincidencia entre los términos de indización asignados por indizadores profesionales diferentes suele oscilar entre el 30% y 60%. Sobre estos y otros aspectos se manifestó Cleverdon [1984] cuando expresó que si dos indizadores expertos analizan separadamente un mismo documento sólo convergen en el 30% de los términos propuestos; si dos personas o grupos construyen un tesoro solamente concuerdan en el 60% de los términos incluidos; si dos profesionales interrogan una base de datos con la misma cuestión sólo el 40% de la información recuperada es común; y por último, si se pregunta a dos científicos sobre la relevancia de un conjunto de documentos, para una determinada cuestión, el acuerdo entre ambos no excede del 60%. (Factor subjetividad).

Las publicaciones periódicas, en la actualidad, son el vehículo de transmisión de ingentes cantidades de información científico-técnica. La comunidad científica necesita mantenerse al corriente de los continuos avances

² Los treinta y dos créditos se completaron con los siguientes cursos: Programación lógica y lenguaje natural; Concepto e historia de la archivística; Evaluación de sistemas de información y documentación. Contraste de los lenguajes de recuperación empleados; Análisis de sistemas de información: propuesta de la metodología métrica; Tendencias actuales en los sistemas de recuperación de información (I); Indicadores de actividad científica y modelos bibliométricos; Tendencias actuales en los sistemas de recuperación de información (II); Líneas de investigación en bibliotecas de instituciones educativas; El desarrollo informativo de la literatura gris en los distintos campos de la ciencia; Proceso y puesta en marcha de un Centro de documentación; Aplicaciones estadísticas en información y documentación; y Bases de datos lingüístico-gramaticales. Cursos impartidos tanto por doctores del departamento de Información y Documentación como por otros como Lengua Española y Lingüística General, Sociosanitarias o Matemáticas.

ocurridos, y para ello, dispone de las bases de datos. Para el almacenamiento de un documento en éstas se ha de indizar previamente. En la Base de datos española ISOC se incorporan unos veintiseis mil artículos al año; en el Centro de Documentación ruso ICSTI hasta 1992 se indizaban anualmente casi cincuenta mil nuevos documentos; en la Biblioteca Nacional de Agricultura de los Estados Unidos entre setenta y ochenta mil; mientras que en la Base de datos alemana PHYS unos ciento veinticinco mil. Si tenemos en cuenta que una indización adecuada de un documento requiere unos diez minutos -si bien es muy difícil precisar este dato-, supone que un profesional en siete horas de trabajo al día "sin descanso" indizará menos de cincuenta documentos.

Sin embargo, si un profesional sólo logra indizar ese número de documentos, algunas unidades de información y/o productores de bases de datos necesitan un gran número de indizadores si pretenden que sus clientes permanezcan al tanto de las últimas novedades científicas. En cambio, con la automatización de esta operación se consigue mayor rapidez. Así por ejemplo, en el Getty Conservation Institute de los Estados Unidos que produce boletines de resúmenes sobre Arte y Arqueología, antes de aplicar la automatización se indizaban 3,3 resúmenes a la hora. Después se pasó a 16,8. (Factor tiempo y económico).

En definitiva, un sistema de indización asistida, semiautomática o automática interesa para alcanzar una mayor consistencia en la indización, para efectuarla siempre bajo los mismos parámetros, y para reducir el tiempo y el coste de ejecución. Obviamente, estos elementos repercuten en la calidad de los resultados y en la productividad de la institución.

2.2 Motivaciones personales.

La razón personal que nos mueve a emprender esta tesis es triple. En primer lugar, para la adquisición de un método de trabajo que nos conduzca al conocimiento de unos saberes que culminen nuestro aprendizaje universitario. En segundo lugar, guiados por el fin mismo de la investigación científica, es decir, la producción de unos frutos que redunden de algún modo en la comunidad en la que vivimos. Y en tercer lugar, el interés por este tema del Análisis del contenido de la Información

arranca desde nuestra etapa como alumno de la Escuela de Biblioteconomía y Documentación de esta Universidad. Estas inquietudes nos llevaron a realizar un trabajo sobre los Orígenes del Análisis, Almacenamiento y Recuperación de la Información, donde se estudiaron aspectos de los incipientes modos de indización sobre los primeros soportes documentales en la Antigüedad.

3 METODOLOGÍA.

El método científico según Sierra Bravo [1994, p. 29] es:

“una forma de realizar una actividad; el camino o proceso que la actividad en cuestión ha de seguir para alcanzar su objetivo [...]. En el método científico se pueden distinguir su contenido o método propiamente dicho, formado fundamentalmente por la serie de etapas sucesivas a seguir para alcanzar el resultado pretendido y su base racional, constituida por el conjunto de ideas que sirven de fundamento y de orientación al método propiamente dicho”.

Por tanto, según estos principios, necesitábamos un método de trabajo con el que guiar la investigación que pretendíamos iniciar. Por consiguiente, la metodología adoptada no fue otra que la típica del método científico dividido en estas etapas:

1.- Descubrimiento de los problemas de la investigación.

La indización y sus aspectos circundantes han sido temas de interés para los investigadores en las últimas décadas debido a que esta operación es la clave para el almacenamiento y la recuperación de la información. La automatización de esta tarea ha centrado numerosos trabajos desde finales de los años cincuenta hasta la actualidad. La mayor parte de ellos han pretendido llegar al mismo fin, pero en ocasiones han seguido diferentes metodologías.

Al acercarnos a la automatización de la indización se echa en falta trabajos donde se ofrezca una visión global del desarrollo ocurrido en este proceso. Por otro lado, se ha reconocido que la Documentación es un área interdisciplinaria, e incluso se ha puesto de manifiesto cuales son las disciplinas y las ciencias que prestan fundamentos tanto teóricos como prácticos para su constitución. Pero en cambio, no se ha tratado la interdisciplinariedad inherente a la automatización de la indización.

Por último, también hay carencias en cuanto a propuestas dirigidas a la automatización de la indización de textos en español.

Estas ideas de partida tratamos de corroborarlas en la siguiente fase del método científico, que no es otra que la documentación de la investigación.

2.- Documentación de la investigación.

La investigación científica debe partir de los descubrimientos científicos antes conseguidos y, por tanto, exige una labor de documentación y de lectura para conocer su existencia y recogerlos si fuera preciso. En virtud de este principio, emprendimos la labor de documentación por medio de una revisión bibliográfica. Esta fase de documentación la dividimos en cuatro momentos bien diferenciados pero complementarios y, en la mayoría de las ocasiones, coincidentes en el espacio y en el tiempo. Consideramos doblemente justificado detallar de forma minuciosa esta fase de búsqueda y de revisión documental, por nuestra condición de doctorando, lo que nos lleva a explicitar la labor documental llevada a cabo, y por nuestra categoría de documentalista.

PRIMERA FASE:

- Análisis y estudio de monografías españolas y extranjeras, que dedicaran parte de su contenido a la automatización de la indización, con el fin de obtener una primera visión del asunto. Esto dio lugar a la confección de un conjunto de fichas temáticas en donde se recogían aspectos que nos parecían interesantes, así como citas textuales.

SEGUNDA FASE:

- Análisis y estudio de publicaciones periódicas por medio del vaciado de revistas:

· Revisamos revistas españolas sobre Biblioteconomía y Documentación en busca de trabajos publicados. Algunas de las revistas fueron:

Revista Española de Documentación Científica
Revista General de Información y Documentación
Ciencias de la Documentación
Boletín de la ANABAD
Boletín de la Asociación Andaluza de Bibliotecarios
Boletín de la Sociedad Española del Procesamiento del Lenguaje Natural
ITEM. Revistade Biblioteconomia i Documentació
Métodos de Información

- Vaciado de las Actas publicadas con motivo de los principales Congresos y Jornadas celebradas en España sobre Biblioteconomía y Documentación:

Jornadas Españolas de Documentación Automatizada
Jornades Catalanes de Documentació

- Vaciado de Actas publicadas en Congresos afines a Biblioteconomía y Documentación:

Sociedad Española para el Procesamiento del Lenguaje Natural
Lenguajes Naturales y Lenguajes Formales

- Vaciado de las bases de datos (en papel) disponibles en la Biblioteca de la Escuela de Biblioteconomía y Documentación de la Universidad de Murcia:

LISA(1991-1992-1993-1995-1996-1997)
PASCAL (1991-1992)

- Vaciado de las publicaciones periódicas extranjeras disponibles en la Biblioteca de la Escuela de Biblioteconomía y Documentación de la Universidad de Murcia, que trataran, directa o indirectamente, nuestro asunto:

Documentaliste-Sciences de l'Information
Documentation et Bibliothèques
Encyclopedia of Library of Information Science
Information Sciences Applications an International Journal
International Forum on Information and Documentation
Journal of Documentation

Journal of Information Science
Knowledge Organization

· Vaciado de otras revistas no ubicadas en Murcia cuya revisión era interesante. De este modo, examinamos publicaciones y números determinados en la Biblioteca de la Escuela de Biblioteconomía y Documentación de Granada y en la Biblioteca del Centro de Información y Documentación Científica (CINDOC) en Madrid. Las fuentes fueron:

Annual Review of Information Science and Technology
Aslib Proceedings
Automatic Documentation and Mathematical Linguistics
Information Processing and Management
Journal of the American Society for Information Science (JASIS)
Library Hi Tech
Library Software review
Pascal Thema
The Indexer

TERCERA FASE:

Consultas a Bases de datos nacionales e internacionales sobre:

Artículos Científicos:

Base de datos ISOCdel CSIC
Base de datos Medline (CD-ROM)
Base de datos Silver Platter (CD-ROM), (Área Biblioteconomía y Doc.)

Tesis Doctorales nacionales y extranjeras:

TESEO: En febrero de 1996 se consultó esta base de datos del Ministerio de Educación y Ciencia, dedicada a Tesis doctorales españolas. Los descriptores utilizados fueron:

INDIZACIÓN
INDIZACIÓN AUTOMÁTICA
LINGÜÍSTICA COMPUTACIONAL
PROCESAMIENTO DEL LENGUAJE NATURAL
LINGÜÍSTICA INFORMÁTICA

Sobre la automatización de la indización no se encontró ninguna Tesis. Esta base de datos se volvió a consultar en febrero de 1997 y el resultado fue el mismo.

Dissertation Abstracts Online: Esta base de datos norteamericana cuenta con más de un millón doscientas mil tesis, principalmente, de Estados Unidos, si bien abarca Canadá y Europa desde 1988. La búsqueda se efectuó en febrero de 1996 y se localizaron 18 Tesis Doctorales, la más reciente de 1995 y la más antigua de 1970.

CUARTA FASE:

En esta última fase el objetivo era buscar información sobre la automatización de la indización en Internet. Para ello, utilizamos los buscadores Lycos, Yahoo, Infoseek y Altavista. Esta opción nos permitió conocer Departamentos universitarios en los que se ha trabajado este asunto y empresas que comercializan productos para el análisis de la información.

En esta fase de documentación solamente se han encontrado dos artículos directamente relacionados con nuestro tema de investigación publicados en fuentes españolas. El primer artículo lo publicó en 1983 Valle Bracero y Fernández García, bajo el título "Automatización de la indización y coordinación de descriptores", en la *Revista Española de Documentación Científica*. El otro artículo corresponde al titulado: "Primeras experiencias sobre el análisis de textos en castellano aplicado a la indexación automática de información", publicado en 1990 por Simón Granda y Lema Garzón, en las *Terceras Jornadas Españolas de Documentación Automatizada*. Estos trabajos presentan varias propuestas para la automatización cuyo fundamento esencial era la extracción de un conjunto de estructuras sintagmáticas preestablecidas o "patrones admitidos" para constituir candidatos a descriptores.

Por otro lado, los manuales sobre Biblioteconomía y Documentación publicados por investigadores o profesionales españoles apenas dedican unas líneas a la automatización de la indización. En cambio, en países como Francia, Brasil pero principalmente, Estados Unidos se ha trabajado en este tema desde los años sesenta.

3.- Definición de los problemas.

La fase anterior de documentación nos valió por un lado, para definir los problemas de partida, y por otro, para concretar aún más la dirección y las líneas de investigación de este trabajo. Por tanto, cabe precisar que:

No se han encontrado trabajos de investigación que presenten de un modo diacrónico el desarrollo ocurrido en la automatización de la indización donde se concreten las metodologías empleadas, los problemas planteados o la misma agrupación de propuestas.

Por otro lado, hemos localizado estudios dedicados a plantear y demostrar la interdisciplinariedad de la Documentación, así como de la relación existente entre la Lingüística y la Documentación, pero no se ha planteado la interdisciplinariedad inherente a la automatización de la indización.

Por último, la automatización de la indización ha sido un asunto poco tratado por los investigadores españoles de Biblioteconomía y Documentación. Esto ha provocado una carencia de propuestas dirigidas en este sentido para textos en español.

4.- Cómo se han abordado los problemas.

Para llevar a cabo los objetivos que se mencionan en el epígrafe siguiente se han seguido diferentes formas de actuar. Por un lado, el análisis crítico de una parte importante de la bibliografía consultada nos ha permitido aclarar y especificar temas tan importantes como las características de la indización, mostrar el desarrollo diacrónico y la problemática de la automatización de la indización, o la interdisciplinariedad inherente a este asunto. Por otro lado, hemos llevado también a cabo un trabajo experimental dividido en dos momentos diferentes de la

investigación. En primer lugar, un ensayo sobre la indización en diferentes Bases de datos donde se comprobó la importancia de los títulos, los resúmenes y los textos en la indización de los documentos. En segundo lugar, se aplicó manualmente el sistema propuesto a un grupo de artículos indizados en una de dichas Bases de datos con el fin de hallar los índices de consistencia entre la indización efectuada por los indizadores profesionales y por el sistema.

4 OBJETIVOS.

Perseguimos tres objetivos principales:

En primer lugar, pretendemos adentrarnos en los aspectos conceptuales relativos a la indización para elaborar un marco teórico del proceso íntegro que contemple desde su delimitación conceptual, fases de actuación, desarrollo diacrónico y el papel fundamental que desempeña la indización como “puente” entre el almacenamiento de los documentos y su recuperación. En definitiva, en esta parte del trabajo buscamos acercarnos al tema de estudio y a ello dedicaremos el capítulo uno.

En segundo lugar, acometeremos la revisión y el análisis del desarrollo diacrónico sucedido en la automatización de la indización desde las primeras propuestas, a finales de los años cincuenta, hasta el presente. Asimismo, comprobaremos cuales son las ciencias y las disciplinas que intervienen, en mayor o menor medida, en automatizar esta operación. El fin de este capítulo segundo es conocer las iniciativas más relevantes en la automatización de la indización, las metodologías empleadas, los resultados obtenidos y el nivel de implantación de estos sistemas en la actualidad.

En tercer lugar, el descubrimiento, la asimilación y la maduración de todos los aspectos estudiados en la consecución de los objetivos anteriores, nos proporcionarán los conocimientos suficientes para diseñar un procedimiento que nos lleve a una propuesta para la automatización de la indización para el español, y más concretamente para el área de Biblioteconomía y Documentación. En consecuencia, el tercer capítulo de la tesis lo dedicaremos al desarrollo teórico del sistema, en el que se explicarán las razones, las herramientas y las metodologías adoptadas.

5 ESTRUCTURACIÓN Y PRESENTACIÓN.

La estructura de este trabajo está compuesta por una introducción, por un cuerpo central integrado por tres capítulos, y por la parte dedicada a las conclusiones. Además, incluye las referencias bibliográficas y el bloque de anexos.

En la introducción, como se ha comprobado, se explica el marco académico en el que se ha desarrollado esta tesis, cuales son las motivaciones que nos han inducido a acometer esta investigación, la metodología seguida y los objetivos marcados.

El contenido de cada uno de los tres capítulos ya se ha comentado en el epígrafe anterior dedicado a los objetivos perseguidos con este trabajo. Y en el cuarto y último capítulo, se exponen las conclusiones alcanzadas en esta investigación.

El anexo está conformado por una serie de gráficos y datos cuya aparición en el cuerpo central de la Tesis no estaba plenamente justificada. Así, contiene los siguientes anexos:

ANEXO 1: Glosario de términos sobre Biblioteconomía y Documentación empleados a lo largo de este trabajo.

ANEXO 2: Tablas con definiciones de la “indización”

ANEXO 3: Etapas de la indización aportadas por diferentes autores.

ANEXO 4: Esquema del proceso íntegro de la indización.

ANEXO 5: Fichero de palabras vacías.

ANEXO 6: Resúmenes de artículos de revista.

ANEXO 7: Resultados de los índices de consistencia obtenidos en la comparación de la indización efectuada por profesionales y la conseguida tras la aplicación manual de la propuesta.

ANEXO 8: Vocabulario controlado sobre Biblioteconomía y Documentación utilizado por la propuesta.

En cuanto a la presentación de la Tesis se ha optado por la colocación de las notas explicativas o aclarativas a pie de página, con la enumeración correlativa a partir de la nota 1. Este mismo proceso se ha repetido en cada uno de los capítulos. El motivo de este planteamiento es evitar el desplazamiento hasta la parte final para consultar alguna nota. Por otro lado, para las citas bibliográficas se ha optado por el método de colocar entre corchetes el nombre o nombres de los autores, el año de publicación de la obra y la página que se cita, recogido en la norma UNE 50-104-94 (equivalente a ISO 690:1987).

En este primer capítulo se revisan los principales aspectos teóricos relativos a la indización. Se comienza con una breve presentación del proceso documental como contexto en el que se efectúa la indización. Posteriormente, se lleva a cabo su delimitación, así como el examen de las etapas sucesivas en la indización de los documentos. Después se analizan sus principales características, los sistemas de indización, los lenguajes documentales, y por último, la normalización de esta tarea.

1.- LA INDIZACIÓN

1.1 EL PROCESO DOCUMENTAL.

El primer paso que juzgamos necesario es situar la indización en el proceso documental. El proceso documental es el conjunto de operaciones dirigidas a la selección, la adquisición, el registro y el tratamiento de los documentos con el fin de posibilitar su almacenamiento y recuperación para su difusión.

La entrada de los documentos en una Biblioteca o un Centro de Documentación se efectúa por medio de dos etapas sucesivas: la selección y la adquisición. La política de adquisiciones debe estar perfectamente diseñada en relación al presupuesto, los recursos y los servicios prestados. La incorporación de fondos al centro se lleva a cabo por compra y de modo gratuito. Los fondos gratuitos se consiguen por el intercambio de documentos entre instituciones o por donaciones privadas.

Por otro lado, como su nombre indica seleccionar es elegir la documentación que debe incorporarse al centro. Por tanto, la selección del material y su posterior adquisición es la primera etapa del proceso que dirige a la constitución de los fondos documentales. Por lo general, en cada centro hay una persona encargada de esta tarea y una cuestión clave en el proceso de la selección es contar con las necesidades, las características y las preferencias de los usuarios. La selección de los documentos se lleva a cabo a través de bibliografías, de catálogos editoriales en papel o en línea, por consultas a bases de datos, a revistas especializadas y de reseñas, entre otras fuentes.

Otra fase en el proceso documental es el tratamiento técnico que reciben los documentos para que sean utilizados por los usuarios. Este tratamiento se divide en: análisis de la forma y análisis del contenido. El análisis de la forma de un documento también se conoce como descripción bibliográfica o catalogación, mientras que el análisis del contenido abarca dos procesos diferentes: el resumen y la indización.

La catalogación se acomete una vez que el documento ha ingresado en el Centro, aunque a veces, la descripción bibliográfica viene en las páginas preliminares del propio documento. El objetivo de la catalogación es:

“suministrar una representación del documento que lo describa de forma única, sin ambigüedades, y que permita luego identificarlo, localizarlo, incorporarlo a los ficheros y catálogos”. [Guinchat y Menou, 1983, p. 101]

Con respecto al análisis del contenido de los documentos (resumen e indización) cabe mencionar lo siguiente. El principal objetivo del resumen es informar a los usuarios del contenido puntual de los documentos a través de los elementos que contribuyan a la toma de decisiones sobre la conveniencia de consultar el texto completo. Moreiro González [1989a, p. 157 y ss.] apunta otros objetivos como reflejar los diferentes bloques de datos con sus características y sus atributos principales, reemplazar la información y expresar el contenido esencial de los documentos no textuales. En cuanto a la tipología, este autor señala que encontramos resúmenes por la mediación informativa que cumplen, la descodificación del texto íntegro, el mediador, el número de fuentes analizadas, el modo de difusión y la calidad de las fuentes.

La indización tiene como principal objetivo el almacenamiento y la recuperación de la información. Al ser la indización un proceso que interviene en dos momentos diferentes del trabajo con los documentos, ofrece dos vertientes enfrentadas. La primera es la indización de los documentos, que tiene por misión el almacenamiento de los mismos una vez que han sido analizados y representados sus conceptos esenciales a través de términos de indización. En segundo lugar, la indización de las preguntas para propiciar la recuperación documental. En consecuencia, el objetivo general de la indización es el almacenamiento y la recuperación de la información.

El proceso de la indización se divide en dos etapas:

1. Análisis de los documentos y las preguntas para la selección de sus conceptos esenciales explícitos o implícitos.
2. Conversión de los conceptos seleccionados en lenguaje natural a un lenguaje documental por medio de herramientas como los vocabularios controlados.

La última fase de la cadena documental es la salida de la información. Todas las operaciones desarrolladas en la fase de entrada y de tratamiento de la información

tienen un objetivo principal: la difusión de la información. La razón de ser de estas operaciones, y de la misma disciplina de la Documentación, es difundir la información que previamente ha sido seleccionada, adquirida y analizada. De este modo, si a la fase de difusión no se da una importancia suprema todo el trabajo desarrollado servirá para poco.

1.2 LA INDIZACIÓN.

En los años setenta se dedicaron numerosos esfuerzos de investigación para la creación de las bases teóricas de la indización y para la búsqueda de una metodología y de unos principios sólidos. A finales de esta década, Cooper [1978] defendió que la indización se había estudiado extensamente pero no se había profundizado en el proceso. Algunos de los estudios hasta esa fecha presentaban cómo eran indizados los documentos por los profesionales, más que cómo se deberían indizar y la posibilidad de automatizar esta operación por medio de normas. En cambio, otros versaban sobre los problemas centrales de la indización: la búsqueda de criterios normativos tanto para la indización humana como para su automatización.

1.2.1 La indización. Definición.

Hay un gran número de definiciones sobre la «indización» y en el anexo 2 mostramos un conjunto de ellas. Aunque no es nuestro objetivo repasar cada una de las definiciones existentes, intentaremos hallar algunas de las principales deficiencias cometidas en su delimitación. En casi todas las definiciones mostradas en el anexo, por lo general, hay coincidencia a la hora de la delimitación del concepto de la indización. No obstante, no estamos de acuerdo con la mayoría de estas definiciones por mostrarse incompletas. Sorprende en primer lugar, la cantidad de verbos empleados para describir la acción que tratamos: retener, asignar, extraer, captar, resumir, describir, caracterizar, escoger, analizar, aislar, identificar, traducir, indicar, interpretar o enumerar. Nosotros recurriremos, generalmente, a los verbos analizar, seleccionar, asignar y convertir para referirnos al proceso de la indización.

Hemos comprobado que para definir la indización solamente se considera el documento como fuente de análisis en la mayoría de las ocasiones. Esto significa

efectuar una delimitación sesgada del proceso, porque se deja de lado la petición del usuario o la pregunta documental. Por estos motivos consideramos necesaria la siguiente distinción para especificar el proceso íntegro de la indización. Por un lado, hablaremos de la indización de los documentos, y por otro, de la indización de las preguntas.

En la indización de los documentos interviene, en un primer momento, el análisis y la selección de los conceptos presentes en el título, el resumen o el texto íntegro (fig. 1).

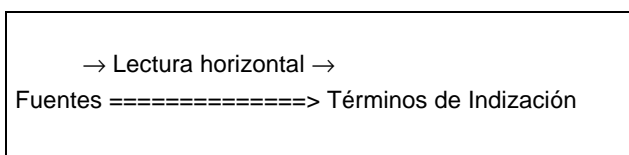


Figura 1. Extracción de términos explícitos

Así como la asignación de los conceptos implícitos en los textos (fig. 2).

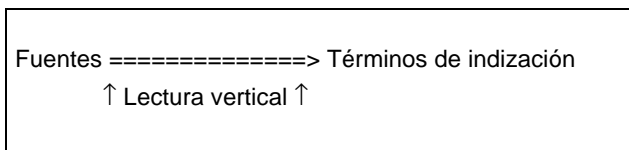


Figura 2. Asignación de términos implícitos

Por tanto, el subproceso de la indización de los documentos queda del siguiente modo:

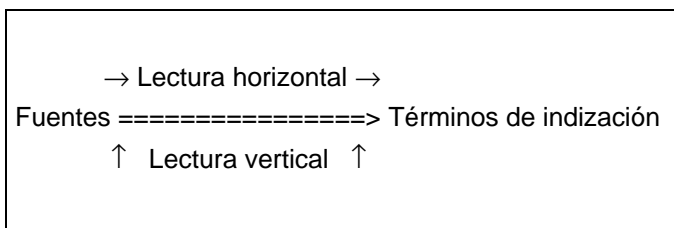


Figura 3. Subproceso de la indización de los documentos

En un segundo momento, los conceptos en lenguaje natural, tanto los extraídos del documento como los asignados, siguen uno de estos caminos: el almacenamiento de esos conceptos a través de términos en lenguaje natural, o la conversión de dichos conceptos en términos normalizados y controlados con la ayuda de un vocabulario controlado. Por último, un aspecto tan importante como los anteriormente descritos, es tener presente, en cada momento del proceso (análisis,

selección, asignación y conversión) cómo solicitaría un usuario ese documento, es decir, realizar una indización de los documentos orientada a las preguntas de los usuarios, o lo que es lo mismo, a la recuperación de los documentos.

En cuanto a la indización de las preguntas cabe mencionar lo siguiente. Cuando llega al documentalista una petición de información documental, tras un diálogo con el usuario, la modela y la materializa en una frase o una pregunta que servirá para interrogar al sistema que tiene almacenada la información. Esta frase o pregunta documental, debe sufrir el mismo proceso expresado para la indización de los documentos. Pero en esta ocasión es preciso estimar qué términos se pudieron manejar en la fase de análisis del documento, es decir, ejecutar una indización de la pregunta orientada al documento.

La observación del esquema del proceso íntegro de la indización (anexo 3) nos induce a definirla como un proceso guiado por el documentalista que permite recorrer tanto a los documentos como a las preguntas un trayecto iniciado desde puntos enfrentados. Este proceso consiste en el análisis y la selección de los conceptos esenciales, así como la asignación de los implícitos -si fuera necesario- y, el almacenamiento de los mismos en lenguaje natural o su conversión en términos normalizados y controlados con la ayuda de una serie de herramientas, que permitan recuperar los documentos en el momento deseado. Es decir, la indización representa a dos objetos en movimiento -documento y pregunta- hacia el mismo punto, cuya unión provoca una respuesta.

Por tanto, de lo mencionado anteriormente se deduce que:

El objetivo de la indización de los documentos es permitir su almacenamiento, mientras que el de la indización de las preguntas encamina la recuperación documental. Por tanto, el objetivo general de la indización es el almacenamiento y la recuperación de la información. Esto lleva a que la indización y la recuperación sean las dos caras de una misma moneda.

Las etapas de la indización las dividimos en dos: el análisis de los documentos y de las preguntas para la selección de los conceptos explícitos o implícitos. Y el almacenamiento de los conceptos por medio de términos en

lenguaje natural, o su conversión a un lenguaje normalizado y controlado (vocabularios controlados).

1.2.2 Las etapas de la indización.

Al igual que ocurre con la definición de la indización, se presentan diferencias en cuanto a la enumeración de los pasos ejecutables. En la literatura revisada, unos autores dividen esta tarea en dos fases mientras que otros establecen hasta ocho. En el anexo 4 han sido recogidas las propuestas de varios autores. La disparidad de criterios a la hora de dividir el proceso que nos ocupa en fases queda patente. No obstante, sin entrar de lleno en este asunto efectuaremos algunos comentarios al respecto.

No participamos de la opinión de considerar una fase independiente la asignación de conceptos que no están explícitos en el documento. Tampoco juzgamos lógico enumerar la etapa de asignación de conceptos explícitos, tras la fase de conversión de los términos en lenguaje natural al de indización. Por último, no coincidimos con la división de Slype en dos de sus etapas propuestas. El primer desacuerdo se produce porque considera como primera etapa la lectura del documento y como segunda la identificación de los conceptos explícitos e implícitos del mismo. No es necesario, a nuestro juicio, la separación de la fase de lectura de la de selección de los conceptos, ya que no son dos etapas sucesivas en el espacio y en el tiempo, sino simultáneas. El segundo desacuerdo llega cuando apunta como cuarta y última etapa, la incorporación de enlaces sintácticos entre los descriptores. A nuestro entender, no hay diferencia si un concepto al convertirlo en término de indización lleva asociado algún enlace, porque el proceso de conversión es el mismo, sólo que más largo.

Según lo alegado en párrafos precedentes existe una estrecha relación entre la indización y la recuperación documental. Por tanto, es conveniente que en los servicios de información sea la misma persona la que lleve a cabo la tarea de indizar y la de atender a los usuarios. Hay casos en los que es imposible que se produzca tal coincidencia como, por ejemplo, la recuperación de la información en las bases de datos en línea. Esta circunstancia es una más de las causas que incitan a conseguir, en la medida de lo posible, la mayor normalización posible en el proceso de la indización y la recuperación documental. Por estas razones, es recomendable

el máximo contacto entre los productores de las bases de datos y aquellos profesionales que, desde cualquier punto distante, las interroga, para asegurar que la fase de recuperación se realice convenientemente.

1.2.3 Las zonas de extracción de los conceptos. Tiempo dedicado.

En cuanto a las zonas de extracción de los conceptos y el tiempo dedicado a ello no existe una coincidencia plena por parte de los profesionales y de los investigadores. Si bien es cierto que se produce una mayor coincidencia en cuanto a las fuentes o zonas más apropiadas para la obtención de los conceptos que finalmente se convertirán en términos de indización.

La norma UNE 50-121-91 señala las siguientes partes de los documentos como importantes a la hora de la indización: título, resumen, sumario o tabla de contenidos, introducción, frases de apertura de capítulos y párrafos, conclusiones, ilustraciones, palabras o grupos de palabras con una tipografía inusual.

En un estudio realizado por Euratom [Dijk y Slype, 1972, p. 105] se llegó a la conclusión de que, los resúmenes de los documentos proporcionaban el triple de los términos de indización que los títulos.

Para comprobar la importancia de los títulos, los resúmenes y los textos para la indización de los documentos, practicamos un ensayo sobre artículos científicos de diferentes áreas del conocimiento en las Bases de datos del CSIC. Este ensayo se detallará en el capítulo tercero, pero podemos adelantar que de cara a la indización de los documentos, en general, los títulos proporcionan un número menor de conceptos que los resúmenes. Y además, en ocasiones, estas dos fuentes se han mostrado insuficientes para la adecuada indización de los artículos, por lo que es necesario recurrir también al texto. En consecuencia, consideramos precisa la utilización de los títulos, los resúmenes y los textos para la indización de los documentos.

Por otro lado, en virtud de las observaciones llevadas a cabo durante este estudio, consideramos que los indizadores (dedicados al análisis del área de Biblioteconomía y Documentación en la Base de datos ISOC) toman como fuentes principales para la extracción de conceptos, en primer lugar, los resúmenes y, en segundo lugar, los

títulos. Y con respecto al texto: la introducción, los epígrafes de los apartados y las conclusiones.

El tiempo medio dedicado, o que debería dedicarse, a la indización es por diversos motivos de difícil delimitación. Señalaremos solamente tres. Primero, depende de las directrices marcadas por el sistema, las que obedecen a su vez, a varios factores como los tipos y las necesidades de los usuarios o simplemente, la cantidad de profesionales dedicados a esta operación. Segundo, en el caso de que la institución no marque las pautas, el tiempo consagrado a cada documento va en función de la profesionalidad de los indizadores. Tercero, está sujeto a las características propias de la indización como la profundidad o la perfección perseguidas, así como a otros factores como la clase de documentos analizados, el tipo de información contenida en los mismos o incluso, el tamaño de los documentos¹.

En un experimento citado en García Gutiérrez [1984, p. 115], y efectuado para comprobar la realidad sobre la indización en Gran Bretaña a principios de los años setenta, se constató que el tiempo para obtener cuatro palabras clave era de tres minutos; cinco minutos para conseguir de cinco a diez; ocho minutos para cosechar de once a veinte, y doce minutos para más de veinte palabras clave. A estos tiempos hay que sumarles el tiempo de convertir las palabras clave en descriptores, si se emplea el término palabra clave para expresar los términos seleccionados y/o asignados en lenguaje natural.

En el estudio ya citado, desarrollado en Euratom, se llegó a la conclusión de que una indización sobre el texto completo de un artículo exige ocho veces más tiempo que la practicada sobre el resumen. Del mismo modo, Farrow [1994, p. 158] toma unos párrafos de Cleverdon [1962] en los que éste último expresaba que, para la indización de informes técnicos, el tiempo óptimo dedicado podía ser de cuatro minutos, más un 60% en función de las condiciones de trabajo. Con respecto a este mismo asunto, Amat [1989, p. 176] mencionó que para una media de unos diez términos se emplea un tiempo medio de veinte minutos.

¹ En el estudio que ya se ha mencionado sobre las seis áreas del conocimiento indizadas en Bases de datos del CSIC hemos constatado que no existe relación entre el número de páginas de los documentos y el número de descriptores asignados. Hemos detectado artículos con escasas páginas y descriptores; artículos con bastantes páginas y pocos descriptores; artículos con pocas páginas y

Como se ha observado, resulta difícil la delimitación del tiempo que se debe dedicar a la indización de un documento.

1.3 LAS CARACTERÍSTICAS DE LA INDIZACIÓN.

En este apartado veremos los elementos que confieren o niegan, en gran medida, la calidad al producto resultante de la indización. Nos referimos a la exhaustividad en la indización (los conceptos caracterizadores del contenido de un documento), la especificidad (la precisión para detectar los conceptos más importantes de los documentos), la corrección (la omisión de términos correctos o la inclusión de otros innecesarios), y la consistencia (el grado de coincidencia a la hora de la presentación de los términos de indización por dos sistemas o por dos indizadores diferentes).

1.3.1 La exhaustividad en la indización.

La exhaustividad en la indización está relacionada, como bien se indica en la norma UNE 50-121-91 [p. 156], con la cantidad de conceptos que caracterizan el contenido íntegro del documento. Hemos constatado que la exhaustividad en la indización se suele identificar con el número de términos de indización asignados a un documento. Es decir, cuantos más descriptores se asigna a un documento mayor exhaustividad y viceversa. En esta dirección se han manifestado Sparck Jones [1973], Maron [1979, p. 224], García Gutiérrez [1984, p. 122], Rowley [1988, p. 56], Salton [1989, p. 277], Amat [1989, p. 176] o Pinto Molina [1993, p. 223].

Con respecto a esta concepción de la exhaustividad en la indización, se puede objetar que la exhaustividad tiene que ver con la selección y/o asignación de todos los conceptos esenciales explícitos o implícitos en el documento o en la pregunta documental. Por tanto, si recurrimos al cómputo de los descriptores como medio para medir la exhaustividad, bien en la indización de un documento, o bien entre dos sistemas o dos Bases de datos diferentes, confundimos su significado y utilizamos un solo factor para la medición de la exhaustividad.

una cantidad importante de descriptores y, por último, artículos con un número elevado tanto de páginas como de descriptores.

Para Soergel [1994, p. 591] una entidad² indizada es pertinente para un número de conceptos, por ello, entiende como exhaustividad en la indización que esos conceptos deben estar comprendidos en los descriptores asignados a esa entidad. Y coincidimos plenamente con este autor cuando opina que el número de descriptores empleados por documento no debe ser el único y determinante valor considerado para comprobar el grado de exhaustividad en la indización de los documentos. En este error se ha incurrido más de lo que sería deseable, como hemos comprobado. Este autor proporciona otros factores para la medición de la exhaustividad, tales como el grado de precoordinación, la corrección de la indización, y el criterio de indización.

El grado de precoordinación en los descriptores es un factor para la comparación del grado de exhaustividad en la indización entre diferentes bases de datos o instituciones que analizan los mismos documentos. Imaginemos dos instituciones con diferentes grados de coordinación:

DOCUMENTOS	DESCRIPTORES DE LA UNIDAD DE INFORMACIÓN A	DESCRIPTORES DE LA UNIDAD DE INFORMACIÓN B
Documento 1	FORMACIÓN DE USUARIOS DEMANDA DE INFORMACIÓN SERVICIOS DE INFORMACIÓN	USUARIOS / FORMACIÓN/ DEMANDA/ INFORMACIÓN/ SERVICIOS
Documento 2	PROGRAMAS DE ORDENADOR REDES DE COMUNICACIONES REDES DE INFORMACIÓN	ORDENADOR / PROGRAMAS/ REDES/ COMUNICACIONES/ INFORMACIÓN

² Este autor utiliza "entidad" o "ítem" como términos generales que pueden referirse tanto a documentos como a módulos de programas o a otros elementos.

En la Unidad de información B el número de descriptores asignados a cada documento es de cinco, con un grado de precoordinación nulo, mientras que en la Unidad de información A al emplear un grado mayor de precoordinación solamente se atribuyen tres términos de indización por documento. Por tanto, ¿podemos considerar que la exhaustividad en la Unidad de información A es mayor que en la B?. Evidentemente, no. Esto demuestra que el número de descriptores no es un factor determinante para medir la exhaustividad en la indización.

Otro factor es la corrección de la indización. Es lógico pensar que si a un documento se atribuyen términos de indización incorrectos, por un lado, aumenta el número de términos asignados, pero por otro lado, y éste es el principal problema, no se captan todos los conceptos expresados en el documento.

El tercer factor mencionado por Soergel es el criterio de indización. En algunos sistemas una vez que se han seleccionado los términos de indización se aumenta su número por la asignación de otros relacionados o genéricos. Así por ejemplo, si un documento trata sobre “depresión clínica” podría indizarse por el descriptor «DEPRESIÓN CLÍNICA», que posiblemente fuera uno de los más correctos. Pero a la vez por otros términos más amplios como «PSICOSIS», «PSICOSIS AFECTIVA», o «DESORDEN MENTAL». Sin embargo, estos tres términos no proporcionan un mayor grado de exhaustividad en cuanto a los conceptos presentes en el documento.

Por consiguiente, para un análisis comparativo entre varios sistemas o bases de datos que analicen los mismos documentos para comprobar el grado de exhaustividad, se tendrá que tener en cuenta este factor.

A nuestro juicio existen varias causas por las que algunos sistemas de información o Bases de datos amplían el número de descriptores sin abarcar más conceptos contenidos en el documento:

Por la generalidad o la especialización de la información tratada.

Por las exigencias de los usuarios.

Cuanto más descriptores sean asignados a un documento, mayores serán las posibilidades de recuperarlo, aunque disminuye el grado de precisión en la recuperación.

1.3.2 La especificidad en la indización.

Desde mitad de los años setenta se han publicado trabajos sobre la medición de la especificidad de los términos de indización [Sparck Jones, 1972 ; Robertson, 1972, 1974 ; Yu y Salton, 1976 ; Wu y Salton, 1981 o Wong y Yao, 1992]. Para Van Slype [1991, p. 123], la especificidad estima la calidad en la selección de los términos que equivalen a los conceptos presentes en los documentos, mientras que, para Rowley [1988, p. 56] será la “libertad” que permite el sistema para ser preciso acerca del tema de un documento, por lo que debe fijarse previamente.

El patrón de especificidad deseada lo definen tanto los descriptores recogidos en el lenguaje de indización empleado, como las directrices de indización marcadas por el sistema de información. En consecuencia, la comparación de dos sistemas en función de la especificidad puede resultar difícil y delicado, como hemos visto en párrafos precedentes.

1.3.3 La corrección de la indización.

La corrección de la indización o la ausencia de errores es un factor de suma importancia para el resultado de la recuperación, porque, como señalara Soergel [1994, p. 593], en el proceso de la indización pueden darse dos tipos de errores: por omisión (cuando un término es omitido) y por inclusión (al contrario, se incluye un término sin ser necesario). En cambio, la ausencia de un término correcto y la asignación de uno más genérico, específico o relacionado, es una clase especial de error de omisión y de inclusión a la vez.

Para determinar este tipo de errores se debe conocer qué descriptores son los más apropiados para un documento, para lo que hay que recurrir a un consenso entre varios indizadores y usuarios. Por tanto, la perfección de la indización establece una relación entre la presencia de descriptores correctamente asignados y la ausencia de omisiones.

1.3.4 La consistencia de la indización.

Para designar al concepto que nos ocupará en este apartado hemos descubierto diferentes términos. Nos referimos a “uniformidad”, “coherencia” y “consistencia”. El Diccionario de la Lengua Española los define de la siguiente manera:

«Uniformidad»: “calidad de uniforme”; y «uniforme» como “dos o más cosas que tienen la misma forma. Igual, conforme, semejante”.

«Coherencia»: “conexión, relación o unión de unas cosas con otras”, y

«Consistencia»: “duración, estabilidad, solidez. Trabazón, coherencia entre las partículas de una masa o los elementos de un conjunto”.

Los términos más utilizados en la literatura española, posiblemente en parecidas proporciones, son coherencia y consistencia³. Este último como reflejo del término inglés «consistency», empleado unánimemente en la bibliografía anglosajona. En adelante emplearemos el término “consistencia” para expresar el concepto que ahora referimos.

La consistencia en la indización se puede estudiar como referencia a un único indizador o a varios. En el primer caso, cuando un profesional indiza un mismo documento en diferentes momentos temporales (consistencia intraindizador). En el segundo caso, cuando varios profesionales indizan un mismo documento de manera diferente (consistencia interindizadores).

La consistencia de la indización la definieron Zunde y Dexter [1969, p. 259] como:

“el grado de concordancia en la representación de la información esencial de un documento, por medio de un conjunto de términos de indización seleccionados por cada uno de los indizadores de un grupo”.

Desde principios de los años sesenta hasta el presente se han presentado bastantes estudios tanto teóricos como prácticos sobre la consistencia de la

³ Precisamente en la norma UNE 50-121-91 se utiliza indistintamente el término “coherencia” y “consistencia” referido al mismo concepto.

indización. En el trabajo de Zunde y Dexter se muestra una revisión de los trabajos publicados hasta ese momento. Veamos algunos de ellos:

Rodgers [1961] en un ensayo sobre combinaciones de dos indizadores la consistencia alcanzada fue del 24%.

Jacoby [1962] en la indización de patentes de Química obtuvo una consistencia media del 10%.

Slamecka y Jacoby [1962] obtuvieron unos valores de consistencia del 16,3% para indizadores experimentados y del 12,6% para indizadores sin experiencia.

En otro trabajo Slamecka y Jacoby [1963] presentaron un ensayo con indizadores experimentados que se sirvieron de un vocabulario controlado para indizar un grupo de documentos. La consistencia conseguida osciló entre el 35% y el 45%.

Painter [1963] alcanzó valores de consistencia entre el 40% y el 70% según el sistema de indización y los tipos de documentos.

Korotkin y Oliver [1964] en otra prueba alcanzaron valores que oscilaron entre el 36% y el 59%.

Más recientemente, Sievert y Andrews [1991] han efectuado un análisis en la Base de datos ISA⁴. En este estudio se concluye que la consistencia en los descriptores es del 47,2%, mientras que en los identificadores es del 32,8%.

Otros trabajos* publicados sobre este asunto son los de Leonard [1975 ; 1977] , Rolling [1981], Funk, Reid y Mcgoogar [1983], Markey [1984], Sievert [1987], o más recientemente el de Livonen [1990], Reich y Biever [1991], Tonta [1991] o Ellis et al. [1994].

⁴ Information Science Abstracts recoge casi quinientas revistas, libros, informes y actas de Congresos sobre Biblioteconomía y Documentación.

* Estas referencias no se han consultado directamente, sino que se localizaron durante el proceso de revisión bibliográfica y de consulta de diferentes Bases de datos.

Como se ha podido observar cada investigador llega a porcentajes diferentes. No obstante, la tónica general es que la consistencia no se sitúe por debajo del 25% ni por encima del 60%. Por tanto, es necesario contemplar la inconsistencia en la indización más como un elemento inherente a esta tarea que como una anomalía.

Para averiguar la consistencia entre la indización manual y la automática Salton y McGill [1983, p. 100], propusieron una fórmula que tiene validez también para emplearse entre dos indizadores profesionales. La fórmula, con una leve modificación de los símbolos, es la siguiente:

$$C_i = \frac{T_{co}}{(A + B) T_{co}}$$

donde,

C_i = la consistencia entre dos sistemas o dos indizadores.

T_{co} = el número de términos comunes asignados por los dos sistemas o dos indizadores.

A = el número de términos asignados por el sistema 1 o indizador 1.

B = el número de términos propuestos por el sistema 2 o indizador 2.

T_{co} = el número de términos comunes asignados por ambos sistemas o indizadores.

La explotación de los índices de consistencia como indicador de una indización correcta es problemática porque podemos encontrar una indización consistentemente incorrecta. Esto se produce cuando dos profesionales indizan un documento y cometen el mismo error, o por el contrario, dos indizadores pueden lograr una perfección y pureza en la indización de uno y estar ante una indización consistente.

En definitiva, las características de la indización se pueden esquematizar del siguiente modo:

1. APLICACIÓN DE CRITERIOS DURANTE LA INDIZACIÓN

- Exhaustividad
- Especificidad
- Grado de coordinación
- Ponderación de los descriptores
- Generación de enlaces

2. EVALUACIÓN DE LA CALIDAD DE LA INDIZACIÓN: REINDIZACIÓN

- Exhaustividad de la indización
- Especificidad
- Corrección: Perfección y Pureza
- Consistencia

3. EVALUACIÓN DEL RESULTADO DE LA RESPUESTA DOCUMENTAL

- Exhaustividad en la recuperación
- Precisión en la recuperación

1.4 LA RECUPERACIÓN DOCUMENTAL: EVALUACIÓN DEL RESULTADO DE LA RESPUESTA DOCUMENTAL.

La recuperación documental comprende una serie de etapas divididas en: pregunta, búsqueda y respuesta documentales. Examinemos cada una de ellas.

La pregunta documental se formaliza por medio del lenguaje natural o controlado. En este segundo caso, la tarea inicial es la conversión de la pregunta o de la frase documental en lenguaje natural a un lenguaje controlado a través de vocabularios controlados. El fin es adquirir unos descriptores simples o compuestos para

interrogar una base de datos. Estos descriptores se emplean solos o combinados con operadores booleanos. Veamos dos ejemplos:

Pregunta 1: ELECTROMIOGRAFÍA

Pregunta 2: ELECTROMIOGRAFÍA Y ADOLESCENTES

Por otro lado, las preguntas en lenguaje natural también son posibles. Algunos sistemas de almacenamiento y recuperación de la información permiten interrogar las bases de datos en lenguaje natural.

Ejemplo:

Pregunta: La relación entre los índices pluviométricos y la desertificación.

La búsqueda documental comienza una vez introducida la pregunta en el lugar destinado para ese fin en el programa informático, y proporcionada la orden de inicio. La duración de la búsqueda dependerá de las características del sistema y de la complejidad de la pregunta. Si el sistema es manual, el lapso de tiempo estará sujeto al tipo y al tamaño del mismo, y a la habilidad del profesional.

La respuesta del sistema, o lo que es lo mismo, el resultado de la búsqueda puede ser positiva y se materializa en un número, y negativa cuando no se ha encontrado nada que responda a la pregunta planteada. Veamos la respuesta en dos programas diferentes una vez preguntado por «Fotografía submarina» y «Bases de datos»:

: FOTOGRAFIA SUBMARINA **0**

Buscar: BASES DE DATOS

[**No hay documentos asociados**]

[**a esa petición de búsqueda**]

Por el contrario denominamos respuesta positiva cuando el resultado es ≥ 1 . Del mismo modo, el programa puede responder de maneras distintas:

: TESIS DOCTORALES**17**

Buscar: LINGÜÍSTICA COMPUTACIONAL

Lingüística: **51**Computacional: **22**Lingüística computacional: **6**

Sin embargo, se encuentra muy extendida la utilización de la expresión “búsqueda documental” para aglutinar todo el proceso de recuperación documental, esto es, la pregunta, la búsqueda automática o manual y la respuesta.

1.4.1 La exhaustividad y la precisión en la recuperación documental.

Tradicionalmente, los resultados de las respuestas documentales se han evaluado por la exhaustividad («recall» en inglés) y la precisión. El término «recall» se ha traducido de diferentes modos en español. Así por ejemplo, Coll-Vinent [1990, p. 139] emplea “recogida”, mientras que Slype [1991, p. 271] utiliza “llamada”. Para expresar este concepto Pérez Álvarez-Ossorio [1988, p. 64], García Gutiérrez [1992, p. 11], Gil Urdiciain [1994b, p. 80] o la norma UNE 50-121-91 [p. 158] recurren a “exhaustividad”. En cambio, Gómez Guinovart [1996, p. 7] lo traduce por “cobertura”. Nosotros adoptaremos la expresión «exhaustividad en la recuperación» para no confundirlo con «exhaustividad en la indización».

Cuando se realiza una búsqueda en una base de datos, la colección de documentos almacenada se divide en dos: los documentos recuperados y los no recuperados. Los documentos recuperados se subdividen a su vez, según el criterio del usuario, en documentos pertinentes y no pertinentes. Igual ocurre con aquellos no recuperados. Estos aspectos los esquematizó de este modo Lancaster [1978, p. 171]:

Documento	Pertinente	No pertinente	Total
Recuperado	a	b	a+b
No recuperado	c	d	c+d
Total	a+c	b+d	a+b+c+d

Una búsqueda perfecta es cuando se recuperan todos los documentos contenidos en la base de datos juzgados por el usuario como pertinentes (a+c), y se rechazan los que considera no pertinentes (b+d). En este caso, estamos ante una respuesta con el mayor grado tanto de exhaustividad como de precisión en la recuperación. Para averiguar los grados de exhaustividad y precisión en la recuperación se emplean las siguientes fórmulas:

$$\text{Exhaustividad} = \frac{\text{N}^{\circ} \text{ de documentos pertinentes recuperados}}{\text{N}^{\circ} \text{ total de documentos pertinentes en la colección}}$$

es decir,

$$\text{Exhaustividad} = \frac{a}{a+c}$$

$$\text{Precisión} = \frac{\text{N}^{\circ} \text{ de documentos pertinentes recuperados}}{\text{N}^{\circ} \text{ total de documentos recuperados}}$$

esto es,

$$\text{Precisión} = \frac{a}{a+b}$$

Los cocientes de exhaustividad y precisión de una respuesta documental son variados. En unos casos hay:

Respuestas con una exhaustividad y una precisión muy bajas.

Una exhaustividad muy baja y una precisión muy alta.

Una exhaustividad y precisión muy altas.

Sin embargo, estos dos parámetros aparecen inversamente relacionados en la práctica, porque si se produce una mejora en alguna de las dos, se asocia con un peor resultado de la otra. Además, la precisión se puede considerar como un tipo de factor coste en el tiempo que el usuario necesita para separar las referencias pertinentes de las que no lo son.

Por último, el ruido y el silencio documentales se hallan relacionados con la exhaustividad en la recuperación. Por ruido documental entendemos la información resultante de una búsqueda que no interesa al usuario:

$$\text{Ruido} = \frac{\text{N}^{\circ} \text{ de documentos no pertinentes extraídos}}{\text{N}^{\circ} \text{ de documentos extraídos}}$$

En cambio, el silencio documental es la información que sí interesa al usuarios pero no ha sido recuperada del sistema:

$$\text{Silencio} = \frac{\text{N}^{\circ} \text{ de documentos no pertinentes extraídos}}{\text{N}^{\circ} \text{ de documentos no pertinentes existentes}}$$

1.5 LOS SISTEMAS DE INDIZACIÓN.

En [GIL LEIVA, I., RODRÍGUEZ MUÑOZ, 1996c, p. 9] señalábamos que los orígenes de la indización se encuentran en las tareas desempeñadas por los antiguos escribas de la baja Mesopotamia. En aquella época se comenzó a dedicar salas para la copia de tablillas de barro, la confección de colofones y de etiquetas y la conservación de las planchas. Los textos se almacenaban en estos lugares de varias formas: sobre estanterías de madera o en nichos ubicados en los muros. Además, existía un tercer método de almacenamiento que consistía en introducir las tablillas en cestas de mimbre o cajas de madera. Para saber qué contenían los recipientes, se les adosaba una pequeña etiqueta de arcilla a un lateral, donde escribían el contenido de los documentos depositados. En estas tareas rudimentarias vemos los primeros pasos de lo que conocemos por indización.

1.5.1 La indización por materias.

Los antecedentes más remotos de la indización por materias hay que buscarlos en la baja Mesopotamia, como ya se ha mencionado. Otro paso más importante fueron las clasificaciones elaboradas a lo largo de la historia, principalmente, a finales del siglo XIX. Pero el momento decisivo se produjo en 1876 con la publicación de las «*Rules for dictionary catalog*» de A. Cutter.

La indización por materias se caracteriza por:

Describir los temas principales de los documentos en detrimento de los asuntos específicos.

Dominar la precoordinación entre sus términos.

Proporcionar un almacenamiento y una recuperación de la información secuencial, lo que ofrece poca flexibilidad en el momento de la recuperación y a la hora de combinar los términos.

Por tanto, si reflexionamos sobre estas características comprendemos que con la aparición de lenguajes más especializados y el aumento de las demandas de información, se precisaron sistemas más ágiles, operativos y flexibles para la

indización y para la recuperación. De ahí el desarrollo de la indización por unitérminos.

1.5.2 La indización por unitérminos.

La introducción de la indización coordinada se ha asociado, tradicionalmente, a Mortimer Taube que manejó un pequeño vocabulario de términos simples o unitérminos para organizar una colección de documentos de la Agencia de Información Técnica de las Fuerzas Armadas de los Estados Unidos, a principios de los años cincuenta. [Artandi, 1971, p. 679 ; Moreiro González, 1990, p. 172].

Este método de indización consistía en la extracción de las palabras simples del texto analizado, generalmente formas sustantivas, y su almacenamiento se llevaba a cabo en fichas cuadrículadas sin ningún tipo de normalización o control. Estas fichas se confeccionaban con un apartado horizontal en el que se insertaba el unitérmino y con diez columnas numeradas del 0 al 9. En estas columnas se colocaban los números de registro de los documentos que trataban sobre este unitérmino.

Cuando se deseaba satisfacer una demanda documental que versara, por ejemplo, sobre la “automatización de la indización”, se seleccionaban las fichas encabezadas por los unitérminos «INDIZACIÓN» y «AUTOMATIZACIÓN» y se comprobaba qué números de registro coincidían en las dos fichas. Este modo de indización acarrea una serie de inconvenientes, ya que propiciaba una completa descontextualización de los términos y provocaba falsas combinaciones en el momento de la recuperación⁵.

1.5.3 La indización por descriptores.

La indización por descriptores es una técnica intelectual introducida por Calvin Mooers para la especificación temática de la información y su recuperación, a finales de los años cuarenta. La palabra “descriptor” la ideó Mooers para proporcionar una nueva y distintiva terminología que denotara la incipiente metodología en la recuperación de la información.

⁵ Efectivamente, Dijk y Slype [1972, p. 51] manifestaron que la palabra «servicio» tiene casi una veintena de sentidos diferentes, especialmente servicio militar, servicio religioso, servicio público, servicio doméstico, vajilla o ayuda.

Una característica de los descriptores simples señalada por Mooers [1972], era su capacidad para comunicar ideas elementales, alejándose de usos terminológicos particulares utilizados en los documentos. Este objetivo planteaba dos problemas: cómo representar estas ideas elementales y cómo definir los descriptores.

El primer problema aparecía con la necesidad de consagrar un símbolo para representar cada descriptor. Tal símbolo podía consistir en un número carente de significado como 3A5 ó 5040, que representaba a un término de indización como se hacía en diversos sistemas creados a partir de las fichas de Taube. Pero en este caso, Mooers creyó más conveniente el empleo de una palabra elegida estratégicamente (o a veces, una abreviatura o acrónimo) para representar a cada descriptor. De este modo, el mismo símbolo era capaz de estimular la memoria del usuario acerca de la idea representada por el descriptor, o por términos altamente expresivos.

Ejemplo:

<u>Unitérminos</u>	<u>Descriptores</u>
SERVICIO	ASTROLOGÍA
411	HEMEROTECA
141	NACIONALISMO

El segundo problema planteado era cómo definirlos para delimitar el ámbito de aplicación de cada uno de ellos. En contraste con lo que ocurría a la asignación de unitérminos, la idea asociada a un descriptor era cuidadosamente seleccionada como el modo más perfecto de la recuperación de los documentos. En este sentido, un descriptor no es una palabra del lenguaje ordinario, sino que es un término particular empleado en la recuperación de la información.

A cada descriptor se le podía atribuir una nota de alcance (scope note -SN-) donde se especifica su utilización en el sistema. Asimismo, otra característica era la posibilidad de establecer relaciones entre los descriptores. Relaciones del tipo: término genérico (broader term -BT-), término específico (narrower term -NT-) o término relacionado (related term -RT-).

Ejemplos:

DOCUMENTACIÓN*

SN: Utilícese para designar el material relativo a las técnicas documentales propiamente dichas y no para designar colecciones de documentos.

LINGÜÍSTICA de COMPUTADORA**

BT Aplicaciones de las computadoras
Lingüística

NT Indización automática
Lingüística estadística
Lingüística estadística (indización)
Metateoría informática
Traducción automática

1.5.3.1 Los enlaces entre los descriptores.

Los enlaces entre los descriptores sirven para que no se produzcan falsas combinaciones en el momento de la recuperación de la información y para ayudar al usuario a decidir sobre la pertinencia de un documento en función de sus descriptores. Hay distintos tipos de enlaces entre los descriptores y su utilidad depende del tipo de almacenamiento y recuperación documental perseguido.

· La *yuxtaposición* es el enlace más común, y consiste en colocar los descriptores unos detrás de otros separados bien por una barra (/) o por punto y coma (;).

INDUSTRIAS DE LA LENGUA / LINGÜÍSTICA / DOCUMENTACIÓN /
LENGUA / TRATAMIENTO AUTOMÁTICO / CENTROS DE
DOCUMENTACIÓN / BIBLIOTECAS

* Entrada del descriptor «Documentación» en el MACROTESAURO para el procesamiento de la información relativa al desarrollo económico y social. París: OCE, 1991, p. 74.

** Entrada del descriptor «Lingüística de computadora» en el TESAURO de la Unesco. París: Unesco, 1984, p. 363

· La *ponderación* es otro tipo de enlace para clasificar los términos en principales y en secundarios. Los primeros acogen los contenidos que se consideran fundamentales de los documentos, mientras que los secundarios incluyen aquellos conceptos tratados más superficialmente. El ejemplo presentado anteriormente quedaría del siguiente modo:

Descriptores principales: INDUSTRIAS DE LA LENGUA / LINGÜÍSTICA / DOCUMENTACIÓN / LENGUA / TRATAMIENTO AUTOMÁTICO

Descriptores secundarios: BIBLIOTECAS / CENTROS DE DOCUMENTACIÓN

· Los enlaces de especificación de *punto de vista* se emplean para que los descriptores aporten, además de un concepto, el punto de vista de cómo están tratados en los documentos. El descriptor de punto de vista o auxiliar se enlaza a través de signos de puntuación empleados para estos casos.

Ejemplo:

LINGÜÍSTICA: morfología

BIBLIOTECAS PÚBLICAS: automatización

Cuando reciben los usuarios estos descriptores como respuesta saben inmediatamente de lo que tratan los documentos: uno, de la parte de la lingüística que se ocupa de las formas de las palabras, y otro, de los aspectos relativos a la automatización de las Bibliotecas Públicas.

· Los enlaces de *indicadores de rol* sirven para lograr una mayor especificidad en la indización de los documentos. En la Base de datos PASCAL⁶, automatizada desde principios de los años setenta y que comprende más de once millones de referencias se vale de enlaces del tipo:

!ENT: material sometido a un proceso químico o nuclear y material oxidante

!FIN: material resultante de un proceso químico o nuclear

!ACT: material catalizador o inhibidor o aditivo

⁶ «<http://www.inist.fr/pascal/roles.htm>» [Consulta: 3 diciembre 1996].

!SUB: medio o solvente reaccional
!ANA: material analizado o dosificado
!SEC: material secundario o impuro
!SOL: solvente utilizado

Ejemplo:

HIDRÓGENO!ENT =====> Elemento sometido a un proceso químico o nuclear

AGUA!FIN =====> Elemento resultante de un proceso químico o nuclear

Por último, Mahapatra y Chandra [1986, p. 21] utilizan en el sistema PRECIS indicadores de rol del tipo:

(0) indica localizaciones.

(2) expresa acción/efecto.

(g) indica un concepto coordinado.

(v) conecta la fila de términos donde aparece con la siguiente en sentido descendente.

La aplicación de los indicadores de rol para la siguiente frase de indización queda del siguiente modo según estos autores:

“El impacto de las tasas indirectas y directas sobre la inflación del precio en las industrias manufactureras del Reino Unido”:

(0) Reino Unido
(1) industrias
(q) industrias manufactureras
(2) inflación del precio
(s) impacto \$v of \$w on
(3) tasas directas \$v
(g) tasas indirectas

· El enlace de *especificación de vínculo* entre los descriptores se emplea para evitar falsas combinaciones en el momento de la búsqueda documental, según Slype [1991, p. 121]. Un documento que trate diferentes temas se representa por conjuntos de descriptores, con la posibilidad de que no tengan nada que ver entre ellos. Por este motivo, para evitar en la fase de la recuperación, problemas de coordinación entre los dos grupos de descriptores se le asignan vínculos. El autor propone los siguientes descriptores como ejemplificación de este tipo de enlace:

EPIDEMIOLOGÍA(1); VARICELA(1); ADOLESCENTE(1); DEPORTE(1);
OCUPACIÓN PROFESIONAL(2); PROGENITOR(2)

Un documento que ha sido indizado con estos términos y con enlaces de especificación de vínculo no aparecerá como resultado de una búsqueda con vínculo sobre la epidemiología y las ocupaciones profesionales, mientras que sí se recuperará en una consulta sobre la epidemiología en los deportistas adolescentes.

1.5.3.2 Los operadores utilizados en las preguntas documentales.

Cuando planteamos una pregunta al sistema que almacena la información se desencadena un proceso que concluye con una respuesta documental. Ya se ha señalado que un modo de almacenar el producto de la indización es a través de palabras en lenguaje natural o de descriptores. Por tanto, en el proceso inverso de recuperar la información contenida en la base de datos, tendremos que emplear el lenguaje natural o bien los términos controlados.

El sistema planteado por el matemático Georges Boole permite la combinación de descriptores con operadores lógicos para sacar mayor rendimiento a los términos de indización almacenados en las bases de datos, y por tanto, plantear preguntas más complejas. Esta lógica presenta las siguientes operaciones: unión, intersección y negación.

OPERADORES	ESPAÑOL	INGLÉS
Intersección	.Y	.AND
Unión	.O	.OR
Complementación	.NO	.NOT

El operador de intersección selecciona los documentos que contienen los dos descriptores (BIBLIOTECA Y USUARIOS); el de unión incluye a los documentos que encierra uno de los dos operandos o bien ambos (MORFOLOGÍA O SINTAXIS), y la diferencia selecciona los documentos que incluyen el primer operando pero no contienen el segundo (INDIZACIÓN NO AUTOMÁTICA).

Los sintácticos son otro grupo de operadores que comprenden el operador de adyacencia cuando dos términos se escriben de forma consecutiva o hay palabras vacías por medio (ENERGÍA EÓLICA / RECUPERACIÓN DE LA INFORMACIÓN); y el operador de distancia que recupera los documentos que contienen los operandos conectados por él separados como máximo por un número n de palabras (SERVICIOS 1 BIBLIOGRÁFIC*). Por tanto, recuperará además de «SERVICIOS BIBLIOGRÁFICOS», documentos donde aparezca «SERVICIOS DE ALERTA BIBLIOGRÁFICA».

Otros tipos de operadores son los de párrafo. Localizan los documentos que contienen los términos que le siguen y le preceden en el mismo párrafo (ARCHIVO P EXPURGO). Los operadores de truncamiento que permiten recuperar términos con cadenas de caracteres comunes. Esto simplifica la búsqueda de los plurales, los sufijos y los prefijos, y elimina la necesidad de escribir todas las posibilidades de las palabras similares. Los operadores de truncamiento son: «*» (asterisco) que representa cualquier número de ocurrencias, y «?» (interrogación) que representa una sola ocurrencia.

Ejemplo: AUTOMÁTIC* recupera los documentos con las palabras *automática*, *automático* o *automáticamente*; y AUTOMÁTIC? recupera los documentos con las palabras *automático* o *automática*.

1.5.3.3 Las características generales de los descriptores.

Tras el análisis de algunos aspectos relativos a los descriptores, se pueden señalar como sus características más importantes las anotadas a continuación:

Son expresiones menos complejas que las producidas en la indización por materias, pero más que en la indización con unitérminos.

No condicionan el modo de almacenar y recuperar la información como ocurre en las Listas de encabezamientos de materias o en las Clasificaciones, por lo que ofrecen más agilidad y flexibilidad a la hora de trabajar.

Proporcionan bastantes puntos de acceso a la información.

No existe un número predeterminado para la representación del contenido de un documento, al contrario de lo que se recomienda en las Listas de encabezamientos de materias.

Se puede realizar una clasificación básica de los descriptores por su estructura: simples y compuestos; por su temática: onomásticos, geográficos, de materias o cronológicos; y según su ponderación: primarios y secundarios.

Su presentación en los tesauros es alfabética, sistemática o gráfica.

1.6 LOS LENGUAJES DOCUMENTALES.

Un lenguaje documental es un conjunto de términos normalizados y controlados vinculados entre sí a través de una sintaxis específica, para expresar cómo deben utilizarse los términos en la fase de análisis o para explicitar sus relaciones de sinonimia, de jerarquía, de proximidad temática o de antonimia. Los lenguajes documentales son herramientas empleadas en la indización de los documentos (almacenamiento) y para la indización de las preguntas (recuperación).

A continuación, reproducimos un pasaje de principios de los sesenta, de Calvin Mooers [1963]⁷ donde se señala que el lenguaje de indización es:

“el mecanismo intelectual que permite al sistema [documental] operar. Tras el almacenamiento de la información, es probablemente la parte más importante del mismo. El lenguaje de indización es la manera de mediar entre el pensamiento de los clientes y la información almacenada. Es el puente, la conexión entre los clientes y la información”.

1.6.1 La tipología de los lenguajes documentales.

La tipología de los lenguajes documentales, según Gil Urdiciain [1996a, p. 22], se concreta desde tres puntos de vista: control del vocabulario, coordinación de los términos y estructura.

1. Por el *control* efectuado sobre el vocabulario se dividen en:

Libres: listas de descriptores libres.

Controlados: clasificaciones y tesauros.

2. Por el criterio de *coordinación* en:

Precoordinados: clasificaciones, listas de encabezamientos de materia.

Postcoordinados: listas de descriptores libres, listas de palabras clave y tesauros

3. Por su *estructura*, principalmente, en:

Jerárquica: clasificaciones jerárquicas.

Combinatoria: léxicos documentales y tesauros.

⁷ Information Retrieval Today: Papers Presented at an Institute Conducted by the Library School and the Center for Continuation Study, University of Minnesota, Sept. 19-22, 1962, Wesley Simonton (ed.), Minneapolis: The Center, 1963, p. 21-36. El trabajo que acoge esta definición se reimprimió en Chan [1985, p. 249 y ss.].

1.6.1.1 Las listas de palabras clave.

Estas listas corresponden con el modo de indización ya presentado de los unitérminos de Taube. Es decir, las palabras se extraen literalmente de los textos y sin ningún tipo de normalización y control se destinan como términos de indización.

1.6.1.2 Las clasificaciones.

La clasificación documental según Rolland-Thomas [1995, p. 17], es una manifestación de las actividades clasificatorias de la humanidad. Se distingue de otras prácticas clasificatorias por su doble vocación: ilustrar una sistematización de los conocimientos y proporcionar una ordenación material de los libros y de otros documentos según el tema tratado en los mismos. Si tomamos como base estos principios, las clasificaciones documentales nacieron de la necesidad de organizar, según un orden preestablecido, desde las tablillas de barro de la antigua Mesopotamia hasta las grandes bases de datos actuales.

Las clasificaciones se dividen en enciclopédicas y en especializadas. Las primeras surgieron para la sistematización de todos los conocimientos manejados por el hombre.

Las principales clasificaciones enciclopédicas documentales son la Clasificación Decimal de Dewey (CD), la Clasificación de la Biblioteca del Congreso de Washington (LC), la Clasificación Decimal Universal (CDU) y la Clasificación Colonada de Ranganathan.

En cambio, las especializadas o sectoriales abarcan un área concreta del conocimiento, y su elaboración se lleva a cabo en Instituciones o en Centros de Documentación privados o públicos para la clasificación y la organización de sus fondos. Según Maniez [1992, p. 28] existen varios centenares, de las que destacan la Clasificación del Instituto Internacional de Soldadura, la Clasificación Decimal Astronáutica, las clasificaciones del Instituto Francés del Petróleo, la Clasificación Internacional sobre Higiene y Seguridad en el Trabajo, el Sistema Internacional de Clasificación SfB (sobre productos de la construcción) o la CANDO (Clasificación Alfanumérica de la Documentación médico-farmacéutica).

El principal inconveniente de las clasificaciones enciclopédicas es su compleja actualización, mientras que el de las especializadas es la carencia de uniformidad como producto del origen y del uso particular que las caracteriza. En general, las clasificaciones proporcionan poca flexibilidad y agilidad tanto en los procesos de almacenamiento como en los de recuperación. Además, suministran pocos puntos de acceso a la información y no son instrumentos con los que se alcancen altos índices de consistencia entre indizadores.

1.6.1.3 Las listas de encabezamientos de materias.

En la publicación de las «*Rules for a dictionary catalog*» de Cutter en 1876 se encuentran los antecedentes de las listas de encabezamientos de materia. Las listas de encabezamientos son para Gil Urdiciain [1996a, p. 31]:

“un lenguaje precoordinado, de estructura asociativa o combinatoria que consiste en listas alfabéticas de palabras o expresiones del lenguaje natural capaces de representar los temas de los que trata un documento”.

Las relaciones entre los encabezamientos de materia que conforman las listas, se establecen para fijar un control terminológico entre los mismos. Se trata de referencias del tipo:

Véase (V.) : remite de un término no aceptado al preferente.

Use por (U.p.) : precede a los términos no aceptados y avisa de los encabezamientos que no deben buscarse en la lista.

Véase además (V.a.) : envía a otra información complementaria a la que se busca.

R.e. : remite a información más específica.

En el siguiente ejemplo intervienen varias de las relaciones mencionadas:

Intervención estatal

Úsese también como subencabezamiento

U.p. Control burocrático

Control estatal

V.a. Economía planificada

R.e. Nacionalizaciones

Los principios básicos sobre los que se asientan las Listas de encabezamientos de materia son:

Especificidad : la elección del término más específico en detrimento de los más generales.

Síntesis : la asignación de un encabezamiento de materia es un proceso mental de condensación para la reducción del contenido de un documento al menor número posible de asuntos o materias, y la búsqueda de la expresión con la mayor simplicidad.

Lingüístico : se emplea el lenguaje usual y en el orden natural del idioma, y prevalece el sustantivo frente al adjetivo, como elemento inicial⁸.

Economía : la asignación del mínimo número de encabezamientos de materia, por lo general, de uno a tres para una Biblioteca pública.

Las formas de los encabezamientos de materia pueden ser diversas. Encabezamiento con una sola palabra (Disolventes), encabezamiento de nombre + adjetivo (Histología animal), encabezamiento de nombre + complemento (Transportes por carretera) y encabezamiento de nombre + nombre (Parques y jardines). En cuanto a los subencabezamientos, se unen al encabezamiento por medio de un guión (-). Existen cuatro clases: de materia (Información-Tratamiento), topográficos (Trenes-Francia), cronológicos (Portugal-Edad Media) y formales (Documentación-Bibliografías). En el caso de que concurren varios prevalece el orden en el que se han apuntado.

Algunos autores [Izquierdo Arroyo y Fernández Moreno, 1994, p. 308 ; Gil Urdiciain, 1996b] han manifestado la progresiva “tesaurificación” de las Listas de encabezamientos de materia. Este cambio es fácilmente observable si contemplamos, por ejemplo, la lista de encabezamientos de materia de la red de Bibliotecas del CSIC⁹:

Plantas ornamentales

Plants, Ornamental

Pueden usarse además los nombres de flores, arbustos y árboles. Ej.:

Rosas, Rosales, Pinos, etc.

U.p. Plantas de jardín

T.g. Plantas cultivadas

T.r. Plantas ornamentales-Industria

T.e. Crocus

Petunias

Plantas de invernadero

Saxifraga

Como veremos a continuación, las referencias (T.g.), (T.r.) y (T.e.) se emplean comúnmente en los tesauros para remitir al usuario de un término a otro más general, relacionado o específico.

En definitiva, las características e inconvenientes de las Listas de encabezamientos de materia son:

Cobertura generalmente sectorial.

Uso del mínimo número de encabezamientos de materia por documento, por lo que a veces, se escapa parte del contenido del documento analizado.

⁸ Solamente se procede a invertir los términos con la ayuda de la coma (,) cuando dicho elemento inicial es muy general y es posible que en la fase de búsqueda de ese documento se escoja un término más específico.

⁹ LISTA de encabezamientos de materia de la red de Bibliotecas del CSIC. Madrid: Consejo Superior de Investigaciones Científicas, 1995.

Predominio de la precoordinación, lo que conduce a un trabajo controlado y guiado en la fase de almacenamiento y recuperación de la información.

Dificultad para lograr altos índices de consistencia entre indizadores.

Proporcionan un almacenamiento y una recuperación de la información de un modo secuencial.

1.6.1.4 Los tesauros.

Un tesoro está compuesto por terminología normalizada y controlada con el fin de convertir el lenguaje natural empleado en los documentos y en las preguntas a dicha terminología. Los términos incluidos en los tesauros se relacionan entre sí por medio de una serie de símbolos (sintaxis) que establecen relaciones de equivalencia, jerárquicas y asociativas (semántica). La cobertura de los tesauros es de carácter sectorial, es decir, cubren un campo o disciplina del conocimiento como por ejemplo la Economía, el Arte o el Urbanismo.

Los descriptores que aparecen en los tesauros establecen entre ellos una serie de relaciones de naturaleza semántica que facilita el desarrollo de un esquema lógico-conceptual [Rodríguez Muñoz et al. 1992 ; Rodríguez Muñoz, 1994].

Las relaciones terminológicas semánticas entre los términos de un tesoro se explicitan del siguiente modo:

Relación de equivalencia: este tipo de relación se manifiesta con los símbolos UP y USE¹⁰. UP Significa “usado por”. El término que sigue a este símbolo es un sinónimo¹¹ o cuasisinónimo¹² no preferente. Un término no

¹⁰ Estos dos símbolos de equivalencia abarcan varios tipos de relaciones entre las que destacan: la sinonimia total (Organización Internacional de Normalización USE ISO), la cuasi-sinonimia (Estimación de costes USE cálculo de costes) y la antonimia (lengua muerta USE lengua viva) [Eurovoc, 1995].

¹¹ Un término es sinónimo cuando tiene el mismo sentido o casi que otro, y es, prácticamente, intercambiable (Bibliobús y Biblioteca itinerante ; Evaluación y Estimación ; Congreso y Symposium) [NF Z 47-100, 1981, p.8].

¹² Términos cuasisinónimos son aquellos en los que la significación es, generalmente, diferente en el uso normal de dicho término, pero que se puede considerar como sinónimos para las necesidades particulares de la indización (Secretaría de Estado = Ministerio ; Legislación = Ley) [NF Z 47-100, 1981, p.8].

preferente, también denominado “no-descriptor” no se asigna a los documentos, pero proporciona un punto de acceso a partir del cual el usuario es dirigido mediante una instrucción (por ejemplo USE) al término preferente (descriptor). En cambio, el término que sigue al símbolo USE (utilícese) es el preferente entre varios términos sinónimos o cuasisinónimos.

Relación de jerarquía: esta relación se indica con los símbolos TE (Término Específico) o NT (en inglés Narrower Term), y TG (Término Genérico) o BT (Broader Term). Según la norma UNE 50-106-90 estas relaciones entre los términos marcan la diferencia entre un tesoro sistemático y una lista no estructurada de términos como un glosario o un diccionario. Un término superordenado representa un todo o una clase y los términos subordinados corresponden a sus miembros o sus partes. El símbolo TG precede al término superordenado mientras que el símbolo TE precede al término subordinado.

Relación asociativa: Con el símbolo TR (Término asociado) o RT (Related term) se asocian términos que no son equivalentes y no manifiestan ninguna relación jerárquica, pero sí “mentalmente hasta el punto de que la conexión entre ellos debe hacerse explícita en un tesoro” [PNE - 50 106, p. 608].

Los tesauros encierran una estructura jerárquica¹³ y combinatoria. El proceso de búsqueda en los tesauros, al contrario que en las clasificaciones y en las listas de encabezamientos de materia, se realiza de modo combinatorio y no secuencial. Los costes de elaboración, en conjunto, son mayores en los tesauros que en las clasificaciones y en las listas de encabezamientos. No obstante, en estas dos últimas herramientas el mantenimiento y la puesta al día son más complejos.

Los conceptos representados por términos de indización [AENOR UNE 50-106-90] pertenecen a las categorías de entes concretos, entes abstractos y entes individuales. Las formas más comunes de los términos son los sustantivos y las frases sustantivadas bajo dos formas: frases adjetivadas y frases preposicionales. A la hora de elegir entre el singular y plural se recomienda, si bien no siempre se

¹³ Tradicionalmente, para la división de los lenguajes documentales se ha utilizado, entre otros, los términos «jerárquicos» y «combinatorios» para incluir en los primeros las clasificaciones y en el segundo, los tesauros, principalmente. A este respecto, en un trabajo acerca de la terminología metalingüística en los lenguajes documentales, Izquierdo Arroyo y Fernández Moreno [1995, p. 56] han expuesto que el carácter jerárquico también se manifiesta en los tesauros, por lo que identificar “jerarquía como elemento diferenciador de los lenguajes documentales conduce a situaciones equívocas, por su falta de rigor”.

cumple, que los sustantivos cuantificables se expresen en plural, mientras que los sustantivos no cuantificables y los nombres de conceptos abstractos sea en singular. Por último, hay diferentes formas para la presentación de los descriptores y sus relaciones: a) alfabética, b) sistemática, y c) gráfica con una sección alfabética. La mayoría de los tesauros tiene alguna de estas tres presentaciones aunque los hay que incluyen varias.

1.7 LA NORMALIZACIÓN DE LA INDIZACIÓN.

García Gutiérrez [1989, p. 227] reprodujo unas frases de Sanders para definir el término normalización como:

“la forma de aplicar y establecer reglas con el fin de poner en orden un campo de actividad determinado, con el interés y concurso de todos los sectores afectados”.

Efectivamente, según esta definición, en la normalización de la indización deben intervenir investigadores, profesionales y usuarios para alcanzar de esta manera, el entendimiento perseguido.

La normalización en las tareas de análisis del contenido documental es tremendamente complejo debido a los elementos que intervienen en esta operación: aspectos cognitivos, formativos-intelectuales y subjetivos. No en vano, desde la mitad de los años sesenta hay trabajos que tratan de explicar cómo indizar los documentos. Algunos de estos trabajos citados por Dijk y Slype [1972, p. 105], proceden de iniciativas particulares [Speight, 1967] o de entidades gubernamentales, como el caso del Centro de Documentación de la NASA [Branhorst y Eckert, 1966] o la Biblioteca Nacional de Medicina de los Estados Unidos [MEDLARS indexing manual, 1967] y [Marcetich y Schuyler, 1981]. Las directrices pretendían, entre otros aspectos, la aplicación de criterios homogéneos en las reglas de indización, la indización de todos aquellos conceptos presentes en los documentos o, la indización de todos los conceptos en su correcto nivel de análisis.

Instituciones como la FID, la IFLA y la UNESCO han colaborado activamente con la ISO en asuntos de normalización documental. Una de estas normas ha sido la ISO

5963: 1985, relativa al análisis del contenido de los documentos para seleccionar los términos de indización. Esta norma es técnicamente equivalente a la española UNE 50-121-91 recogida en la Recopilación de Normas UNE publicadas por AENOR en 1994. Esta norma española recomienda, acerca de las fases pertinentes en la indización, las zonas más apropiadas para la selección de los conceptos y las características de la indización como la exhaustividad, la especificidad y la consistencia.

En este segundo capítulo se presentan las posiciones y razonamientos de los investigadores y profesionales ante la automatización de la indización; las principales metodologías adoptadas para acometerla, así como sus mayores inconvenientes. Asimismo, se repasan ensayos llevados a cabo para verificar la calidad de los sistemas automáticos con respecto a los indizadores profesionales. También se testimonian las principales aportaciones que ciencias o disciplinas ajenas a la Documentación incorporan en dicha automatización.

2.- LA AUTOMATIZACIÓN DE LA INDIZACIÓN

2.1 INTRODUCCIÓN.

Stevens [1965] definió la indización automática como la utilización de ordenadores para extraer o asignar términos de indización sin intervención humana, una vez establecidos programas o normas relativas al procedimiento. Bastos Vieira [1988, p. 48] entiende que es la operación que identifica palabras o expresiones significativas de los documentos para describir su contenido de forma condensada por medio de programas de ordenador.

La revisión de la literatura sobre la automatización de la indización nos ha permitido descubrir una amplia variedad terminológica para la designación de conceptos semejantes. Hemos detectado una veintena de expresiones como: «automated assisted indexing», «automated indexing», «automated support to indexing», «automatic support to indexing», «computer aided indexing», «computer assistance in indexing», «computer assisted indexing», «computer indexing», «computerized indexing», «indexing by computer», «indexing program», «indexing software», «machine aided indexing», «machine indexing», «machine-assisted indexing», «mechanical indexing», «mechanized indexing», «microcomputer-based indexing», «semi-automatic indexing», y «automatic indexing». Esta última expresión es, con diferencia, la más utilizada.

Esta importante variedad de expresiones solamente hace referencia a tres conceptos diferentes:

1. Programas que asisten en el proceso de almacenamiento de los términos de indización, una vez obtenidos de modo intelectual. Estos sistemas proporcionan, por medio de pantallas de ayuda, notas explicativas sobre la utilización de un término, sus términos relacionados, permiten la asignación de términos sin necesidad de teclearlos, o incluso, la consulta en línea de documentos anteriormente indizados para comprobar cualquier aspecto. (Indización asistida por ordenador durante el almacenamiento).
2. Sistemas que analizan los documentos de modo automático, pero, los términos de indización propuestos los valida y edita (si es necesario) un profesional. (Indización semiautomática).

3. Programas que no necesitan ningún tipo de validación, es decir, los términos propuestos se almacenan directamente como descriptores de dicho documento. (Indización automática).

2.2 LA INDIZACIÓN HUMANA VERSUS INDIZACIÓN AUTOMÁTICA.

Con los primeros intentos para automatizar la indización surgieron posturas en contra y a favor. Estas discrepancias se iniciaron en los años sesenta. Los primeros defensores fundamentaron sus ideas en los avances informáticos producidos en los años sesenta y setenta y en el desarrollo de herramientas para el procesamiento del lenguaje natural (PLN), y los sistemas expertos dentro de la Inteligencia Artificial. Asimismo, la automatización de esta operación siempre ha estado sujeta a los avances y la disponibilidad tanto de hardware como de software.

En cambio, los detractores de la automatización de la indización se han apoyado en el pensamiento de que un programa de ordenador no es capaz de indizar los documentos con una mínima calidad. Su principal argumento ha sido la impotencia de un programa para simular el conocimiento de un indizador profesional. Es decir, que un programa informático no tiene la capacidad para analizar, detectar, relacionar y seleccionar los conceptos explícitos e implícitos de los documentos que mejor representen su contenido.

Con la finalidad de profundizar en este debate mostramos a continuación un conjunto de afirmaciones de investigadores que se han pronunciado al respecto.

2.2.1 Argumentaciones en contra de la automatización de la indización.

La indización automática se puede considerar como un caso especial de degeneración de la indización humana, en la que el indizador se convierte en un mero observador de los indicadores estadísticos del programa informático. Cooper [1978, p. 108].

Los sistemas automáticas muestran su incapacidad para el reconocimiento de diferentes significados de una palabra o una frase aparecida en diferentes contextos, y por tanto, con significados, en ocasiones, totalmente diferentes. Fugmann [1990, p. 65], citado en Quinn [1994, p. 142]

Mientras la indización automática reconoce, ante todo, cadenas de caracteres que constituyen palabras no vacías, la indización intelectual distingue conceptos, es decir, representaciones mentales de objetos del conocimiento. Slype [1991, p. 116].

Una indización perfecta no puede ser completamente automatizada. Por tanto, es defendible una indización asistida en el proceso de análisis. Karetnyk, Karlsson y Smart [1991, p. 25].

En el proceso de indización, ya sea manual o automático, se mantienen una serie de problemas como la lectura lineal, palabra a palabra o enunciado a enunciado, que impide la comprensión del texto y, por tanto, su representación. García Gutiérrez [1992, p. 34].

La búsqueda de una indización totalmente automática es un esfuerzo inútil porque los términos de indización propuestos necesitan una edición humana, si se busca una indización útil y aceptable. Wellisch [1992].

La indización manual habilita una mayor profundidad de análisis y evita problemas de polisemia. Además, proporciona una mayor racionalización del proceso y posibilita la extracción de conceptos no expresados de forma explícita por el autor del texto. Gil Urdiciain [1992, p. 199].

Los indizadores humanos tienen capacidad para percibir los conceptos implícitos del texto. Albrechtsen [1993], citado en Palma Villalón [1995, p. 224].

En la actualidad, los sistemas basados en normas de decisión así como los basados en métodos estadísticos, sintácticos y semánticos intentan imitar los resultados producidos por los indizadores humanos, pero su éxito está limitado por las lagunas existentes en la comprensión del proceso de la indización. Milstead [1994, p. 578].

La indización automática o asistida, en la mayoría de las ocasiones, está restringida a áreas específicas del conocimiento. Por otro lado, la indización automática no podrá sustituir a la intelectual para el análisis de los documentos importantes. Bonura [1994].

Otros autores expresaban sus dudas pero dejaban una puerta abierta a los probables avances en este sentido. De este modo, Coll-Vinent [1982, p. 178] fue uno de los primeros investigadores españoles en expresar sus ideas con respecto a este asunto.

Por último, recogemos las recapitaciones practicadas por García Gutiérrez [1992, p. 37], al hilo de la función que juegan las normas o recomendaciones en el ejercicio del análisis del contenido documental en el trabajo de los documentalistas: “¿es posible regular un procedimiento general o particular de análisis de contenido en documentación?”. A lo que responde afirmativamente. Pero también se preguntaba: “¿son favorables los gestores de la información a que tal procedimiento exista y podamos, en consecuencia, programar máquinas para que lo lleven a cabo con la mayor objetividad posible?”. A lo que responde negativamente.

Una vez mostrados los juicios en contra de la automatización de la indización nos vemos en la obligación de efectuar algunas aclaraciones para rebatir parte de estas ideas:

1. Se ha mencionado que el indizador se convierte en un mero observador de los indicadores estadísticos. Sin lugar a dudas, ésta sería una de sus misiones pero no la única. El indizador se constituirá en un elemento activo en la detección de errores del sistema. De este modo, su participación en la mejora del programa o en la construcción de otros futuros queda garantizada. Por otro lado, una vez que este tipo de técnicas y otras similares (elaboración automática de resúmenes, por ejemplo) lleguen a aplicarse en los Centros de Documentación, el tiempo y el esfuerzo dedicado al análisis del contenido de la información pasará a consagrarse a los usuarios-clientes, y en definitiva a la difusión de la información. Tarea que constituye la razón de ser de la Documentación.
2. También se ha afirmado que en la automatización de la indización se reconocen palabras y en la intelectual conceptos. En realidad esta afirmación es correcta pero caben algunas precisiones. La indización semiautomática o automática debe perseguir la captación no de las palabras sino de la terminología de los textos. La terminología, como tendremos oportunidad de comprobar en el tercer capítulo, cumple una triple función: representativa, cognitiva y comunicativa, porque tras la terminología se esconden los conceptos y por tanto, el conocimiento.

3. Otra de las alegaciones presentadas argumenta que la indización automática precisa una fase de postedición para la validación de los conceptos propuestos. La indización de profesionales, como mostraremos más adelante, se mueve en unos niveles de inconsistencia altos. Y estos valores, no son fruto de casos aislados, sino más bien se trata de una característica permanente en la indización. Por tanto, si partimos de esta premisa, y se comparan los resultados de una indización intelectual con los de un sistema semiautomático o automático y los índices de consistencia resultan similares no hay necesidad de tal validación, sino que los términos propuestos pasarán, directamente, a convertirse en los términos de indización asignados al documento.

Por último, sólo resta mencionar varios problemas no menos importantes de la automatización de la indización que no han sido referidos:

La indización de la información en formato no electrónico requiere su escaneo con la consiguiente pérdida de tiempo.

La corrección ortográfica de los textos a indizar es decisiva para el correcto análisis y la selección de los términos, independientemente, de la metodología practicada.

2.2.2 Argumentaciones a favor de la automatización de la indización.

La indización humana es subjetiva, lenta y costosa. García Gutiérrez [1984, p. 114], Bertrand-Gastaldy y Pagola [1992, p. 85].

El desarrollo continuo de la producción y la publicación de información en formato electrónico favorecerá el desarrollo de investigaciones y propuestas para la obtención de la indización automática. Sparck Jones [1986, p. 12], citada por Pinto Molina [1989, p. 365], Ward [1996, p. 225].

El ordenador constituye una herramienta eficaz que no sustituye el esfuerzo inteligente del hombre en su trabajo, sino por el contrario, lo aumenta y mejora. La indización automatizada representa un ahorro de tiempo y evita el trabajo de la lectura de los textos. Coll-Vinent [1990, p. 142].

La indización automática basada en el PLN ofrece alternativas atractivas para la indización de los documentos. Además, si a estas técnicas se le une el uso de un tesoro para refinar las estructuras lingüísticamente válidas, se puede desafiar a la indización humana por su perfección, consistencia y precisión. Evans et al. [1991, p. 108].

La indización de la base de datos MEDLINE de la Biblioteca Nacional de Medicina de los Estados Unidos, acarrea dos problemas principales: la inconsistencia de la indización humana y el coste. La consistencia en la asignación de los conceptos principales no supera el 61%, mientras que la combinación de encabezamientos de materia y subencabezamientos era del 38%. Por otro lado, la indización de esta base de datos por profesionales supone un gasto de más de 2 millones de dólares por año. [Hersh et al., 1992, p. 292].

Las bases de datos documentales se alimentan de la indización realizada por diferentes indizadores con distintos y variables criterios en determinados momentos. De ahí, la necesidad de unificar procedimientos con la automatización de estas operaciones. García Gutiérrez [1992, p. 128].

En el estado actual de las investigaciones, cabe proponer una indización asistida por ordenador en detrimento de una indización totalmente automática. Karetnyk, Karlsson y Smart [1991, p. 25], Bertrand-Gastaldy y Pagola [1992, p. 85] y Humphrey [1994, p. 161].

La indización automática es más rápida, económica, consistente y efectiva que la manual. Anderson [1994, p. 632].

A estas argumentaciones a favor podemos añadir las siguientes:

Mayor objetividad puesto que se aplican siempre los mismos parámetros. Además, se evita la inconsistencia producida por un mismo indizador o la provocada por diferentes profesionales en el análisis del mismo documento.

La disminución de los errores repercute positivamente en las bases de datos en el momento de la recuperación de la información.

Permite una recuperación de los documentos más rica, si bien es cierto que la indización intelectual parece ser más precisa.

Para comprobar las ideas de los profesionales y los investigadores españoles y latinoamericanos de Biblioteconomía y Documentación con respecto a este asunto, lanzamos en abril de 1996 un mensaje a la lista de correo electrónico IWETEL, foro de discusión compuesto por, aproximadamente, un millar de abonados. Con el siguiente mensaje se pretendía iniciar un debate sobre la idoneidad de la automatización de la indización:

"Quisiera plantear un debate alrededor de este tema: ¿debemos mantener, como hasta el momento, la indización humana de documentos, o por el contrario, investigar, y por tanto, perseguir una indización automática?".

Este interrogante dio lugar a un total de treinta y dos mensajes para ofrecer opiniones al respecto. En las intervenciones se ofrecieron razonamientos sobre la polémica planteada, se manifestaba el desacuerdo con los criterios expuestos con anterioridad o se patentizaba que estábamos ante un debate muy interesante. Esta treintena de mensajes fueron enviados por once abonados.

Los tertulios se posicionaron a favor o en contra de la automatización de la indización. La razón principal ofrecida por los partícipes en contra de este proceso, fue la imposibilidad, en el estado actual de la investigación, para conseguir una indización totalmente automática. Los argumentos a favor fueron el mayor coste, la subjetividad y la lentitud de ejecución de la indización humana.

A lo largo del debate surgieron aspectos de interés, alrededor del tema de discusión, de los que mencionamos los siguientes:

Una indización automática o humana serán igual de buenas si se consideran elementos como el fondo con el que se trabaja, el contexto de dicho fondo y las necesidades de los usuarios reales y potenciales.

La necesidad de emplear un vocabulario controlado en los sistemas de indización automática.

En el proceso de creación, análisis y difusión de la información intervienen autores, indizadores, profesionales dedicados a la recuperación de dicha información (ya que no siempre coinciden éstos últimos con los indizadores) y los usuarios. Este grupo de actores no siempre tiene estructuras mentales, concepciones, conocimientos e intereses semejantes.

En definitiva, la subjetividad, la lentitud y el coste son importantes argumentaciones en contra de la indización intelectual, mientras que los defensores de su automatización alegan la consistencia, la rapidez y la exhaustividad, que originan una mayor productividad y calidad en la indización.

2.3 EL DESARROLLO DIACRÓNICO DE LA AUTOMATIZACIÓN DE LA INDIZACIÓN.

Durante los años cincuenta y sesenta se produjo un incremento exponencial de la información científica. Esto causó un cambio en la manera de concebir y mantener los sistemas de información dedicados a abastecer a los científicos de información. Como expresamos en Gil Leiva y Rodríguez Muñoz [1996a] el ordenador comenzó a contemplarse como una herramienta útil para el manejo y el tratamiento de la información, y en especial para la indización de los documentos. Con la incorporación del ordenador a la tarea de indizar se buscaba la agilización de los análisis de la información, la consecución de mejores índices de consistencia y la reducción de los costes, y en definitiva, una mayor calidad en los sistemas de información.

El acercamiento a la automatización de la indización se hizo en los primeros momentos, desde el cálculo de la frecuencia de las palabras en los textos, y más tarde, se incorporaron técnicas del procesamiento del lenguaje natural. Esto provocó que algunos de las propuestas para la automatización de esta tarea mezclaran las dos metodologías.

A continuación, veremos las bases teóricas sobre las que se asentaron las propuestas estadísticas para la automatización de la indización. De la misma forma, se comentarán brevemente algunos ensayos al respecto.

2.3.1 Los métodos estadísticos.

G. K. Zipf [1949] llegó a la conclusión de que en la comunicación hablada o escrita se produce el llamado «principio del mínimo esfuerzo». Este principio está relacionado con el recurso de los hablantes y los escritores en una lengua a la repetición de ciertas palabras en lugar de utilizar otras diferentes. Este investigador estableció la siguiente fórmula tras el estudio estadístico de Ulysses de Joyce:

Frecuencia x clasificación \approx constante

que representa el valor constante que tiene la relación entre la frecuencia de aparición de las palabras y el rango o puesto que éstas ocupan en el orden frecuencial.

A partir de estas ideas, Hans Peter Luhn [1957] fue el primero en sugerir que la frecuencia de aparición de las palabras en un documento o en una colección tenía que ver con la utilidad de éstos para la indización. Las palabras de frecuencia muy alta (aquellas que se manifestaban en casi todos los documentos) no aportaban carga informativa debido a su carácter general. Por tanto, si se empleaban en la recuperación de la información provocaban una escasa precisión. Por el contrario, los vocablos de frecuencia muy baja eran muy específicos y causaban una baja exhaustividad en la recuperación. Para Luhn los mejores términos eran los que detentaban una frecuencia media, es decir, los que no se presentaban ni en pocos ni en un gran número de documentos.

En virtud de estos fundamentos, Luhn expresó en 1958 las siguientes consideraciones con respecto a su aplicación en la automatización de la indización [Salton y McGill, 1983, p. 60]:

1. Dada una colección de n documentos se calcula la frecuencia de aparición de las palabras para cada documento.
2. Se determina para cada palabra su frecuencia en toda la colección $TOFREQ_k$ por la suma de las frecuencias en cada documento.

3. Una vez ordenadas las palabras en orden decreciente según su frecuencia en la colección, se eliminan todas las palabras que sobrepasan un umbral de frecuencia determinado. En esta fase se eliminan palabras como: «el», «de», «y», «para», «a», «en», entre otras. También se prescinde de aquellas poco frecuentes en la colección, por medio de un umbral previamente establecido.
4. Las palabras restantes, con una frecuencia media, se asignan como términos de indización para los documentos analizados .

A este método se le plantearon críticas [Salton y McGill, 1983, p. 62], puesto que si se eliminaban todas las palabras con una frecuencia alta y baja se producían pérdidas en la recuperación, tanto de exhaustividad como de precisión. Además, era necesario no errar en la elección de los umbrales adecuados para obtener las palabras con la frecuencia media. Y por último, la utilización obligada de la frecuencia relativa para la identificación de palabras con carga informativa en un documento concreto de la colección.

A partir de estas consideraciones efectuadas por Luhn se concretaron otros métodos de ponderación de los términos como: la función de frecuencia inversa en un documento propuesta por Sparck Jones [1972] o el valor de discriminación de los términos de Salton y Yang [1973].

La ponderación de la frecuencia inversa del documento (inverse document frequency weight) propuesta por Sparck Jones [1972], mide la escasez de aparición de un término en una colección. En cierto modo, se toma la idea de que la frecuencia de aparición de una palabra está en relación inversa a su capacidad informativa. Estas ideas están recogidas en la siguiente fórmula:

$$\text{IDF}_i = \log_2 \frac{N}{n_i} + 1 = \log_2 (N) - \log_2 (n_i) + 1$$

donde N es el número de documentos de la colección y n_i es el número de documentos que contienen el término i en la colección.

Esta forma de ponderación también se ha utilizado conjuntamente con la frecuencia de aparición de un término en un documento, tanto en experiencias de automatización de la indización como de recuperación de la información.

Posteriormente, en Salton, Yang, Yu [1975] se presentó un nuevo método para conferir el peso o el valor más alto a aquellos términos que causaban la máxima separación posible entre los documentos de una colección. Este método lo denominaron el valor de discriminación de los términos. El valor de discriminación de un término lo definieron como la medida de los cambios manifestados en la separación espacial cuando un término cualquiera es atribuido a una colección como término de indización. Estos investigadores consideraron que si atribuían a cada término sus densidades espaciales se podrían clasificar de acuerdo a sus valores de discriminación.

Por tanto, en razón a estos principios los términos de indización participan de unas características aproximadas:

Los términos con un valor positivo de discriminación, si se consideran como términos de indización propicia que decrezca la densidad espacial de los documentos.

Los términos con valores de discriminación indiferentes, si se suprimen o suman los términos con un valor de discriminación cercano a cero no cambia la similitud entre los documentos.

Los términos con valor de discriminación pobre, si se utilizan proporcionan mayores semejanzas entre los documentos, lo que produce un aumento de la densidad espacial de los documentos.

A continuación veremos varios ensayos que utilizaron algunos de estos métodos:

Damerau [1965] defendió acumular la frecuencia de aparición de las palabras de una colección específica de documentos. A la hora de indizar un documento calculaba la frecuencia de aparición de cada palabra en dicho documento y las comparaba con las frecuencias obtenidas previamente para toda la colección. Las

palabras con los datos estadísticos más semejantes a los almacenados se convertían en términos de indización.

Paralelamente se introdujo el uso de la probabilidad para la búsqueda de mejores resultados, tanto en la indización intelectual como en los procesos para su automatización. Ejemplos de estas propuestas son las de Carrol y Roeloffs [1969], Rosenberg [1971], Bookstein y Swanson [1975], Harter [1975] o Robertson y Harding [1984].

Los problemas de los métodos estadísticos en la automatización de la indización son los siguientes:

En primer lugar, se muestran incapaces del reconocimiento de relaciones semánticas como:

La sinonimia: Los distintos significantes con el mismo significado. Ejemplo: un concepto representado a través de una sigla puede representarse de modos distintos: de forma desarrollada (procesamiento del lenguaje natural), como sigla (PLN) o como sigla pero con cada letra separada por puntos (P.L.N). En este caso, estos dos términos (PLN y P.L.N) se contabilizarán de forma independiente.

La anáfora: La función de ciertas palabras -ésta, lo, allí, entre otras- para asumir el significado de una parte del discurso ya emitido. Ejemplo: "Hay diferencias entre la indización manual y la indización automática. Ésta la ejecuta un programa y la primera la acomete una persona". En este caso, «ésta» toma el significado de "indización automática" mientras que «primera», el de "indización manual".

La elipsis. La omisión de una o más palabras en una oración sin distorsionar el sentido de la misma. Ejemplo: "El mismo documentalista realiza (i) y (ii) corrige la indización de los artículos". En esta oración se ha omitido "la indización de los documentos" (i) y "el mismo documentalista" (ii).

En segundo lugar, por lo general, los términos compuestos se consideran como si fueran simples, porque los algoritmos examinan los textos como una secuencia de

caracteres separados por espacios. Por tanto, conceptos representados por términos compuestos como «Crecimiento celular», «Combustibles líquidos» o «Patrimonio bibliográfico» son considerados como términos simples para la aplicación de cálculos estadísticos. Esto significa que se contabilizará «Crecimiento», «Combustibles» y «Patrimonio» por un lado, y «celular», «líquidos» y «bibliográfico» por otro lado. De este modo, se destruyen los conceptos simbolizados por los términos compuestos. El uso de los resultados de estos sistemas para la indización de los documentos conlleva un alto grado de ambigüedad en la indización, que posteriormente, se concreta en la recuperación en una alta exhaustividad y una baja precisión.

En tercer lugar, los métodos estadísticos carecen de la posibilidad de normalizar los términos. Esto propicia que «documental» o «documentario» se computen por separado en términos compuestos sinónimos como «análisis documental» o «análisis documentario».

En definitiva, si los aspectos mencionados en los párrafos precedentes no se consideran en los cálculos de frecuencia de las palabras, los resultados no serán los correctos, y en cierto modo, emergerán “falseados”.

2.3.2 Los métodos lingüísticos.

A principios de los años sesenta se inició la asociación entre las técnicas del PLN y la automatización de la indización. El procesamiento del lenguaje natural consiste en el estudio y el análisis de los aspectos lingüísticos de un texto mediante la utilización de programas informáticos. Un sencillo ejemplo de PLN es el corrector ortográfico de un procesador de textos, si bien hay otras herramientas más complicadas como veremos más adelante. El lenguaje natural se distingue de los lenguajes artificiales por su riqueza (en vocabulario y construcciones), flexibilidad (reglas con múltiples excepciones), ambigüedad (diversos significados de una palabra o una frase según el contexto), indeterminación (permite referencias y elipsis) y posibles interpretaciones del sentido literal según la situación en que se produce [Verdejo Maillo, 1994, p. 5]. Por tanto, las ventajas para la comunicación humana se convierten en problemas a la hora de su tratamiento informático.

Esta autora ha expresado en el mismo contexto que el PLN surgido en la década de los cincuenta entrelaza su historia con las investigaciones que sobre el lenguaje se llevaban a cabo en otras disciplinas. Estos dominios eran Lingüística Formal, Psicología Cognitiva, Lógica, Informática e Inteligencia Artificial, lo que dio lugar a una disciplina denominada Lingüística Computacional. La Lingüística Computacional es la intersección de la Lingüística y la Informática con el fin de procesar o generar las lenguas. Veamos esquemáticamente las distintas etapas en el despliegue teórico y práctico del PLN:

De los años cuarenta a mitad de los sesenta. La aparición de los ordenadores extendió la idea de que el procesamiento del lenguaje se lograría en muy poco tiempo, pero paulatinamente surgieron las incógnitas que conllevaban los intentos en este sentido. Por estos motivos se abandonó la financiación de proyectos encaminados a la traducción automática ruso-inglés, entre otros, a la vez que se iniciaron experimentos para comprender el lenguaje en ámbitos muy específicos.

Desde principios de los setenta hasta comienzos de los ochenta. El tratamiento de la sintaxis se consolidó en un tema importante de investigación por la dedicación de esfuerzos en cuanto a su formalización y al desarrollo de algoritmos para su tratamiento automático.

De la década de los ochenta hasta la actualidad. Se llevó a cabo la unión entre las teorías lingüísticas y los mecanismos de parsing, a la vez que se han ampliado los estudios del PLN a nivel de la semántica y del discurso. Por otro lado, se comenzó a hablar de las llamadas "Industrias de la lengua", que propiciaron la aparición de productos en el mercado, en donde se integra la informática y la lingüística. Se trata de correctores ortográficos incorporados a los procesadores de textos, de programas de traducción automática, de reconocimiento del habla y de análisis y recuperación de la información. En definitiva, productos en los cuales se escudriña aún hoy para perfeccionarlos.

Las técnicas del procesamiento del lenguaje se organizan a través de diferentes análisis. Cada uno de los análisis ocupa distintos niveles relacionados directamente con la complejidad de éstos.

Un análisis morfológico de modo automático persigue la segmentación de la palabra ortográfica con el fin de obtener la gramatical y determinar su estructura y propiedades. Este tipo de herramientas, según Verdejo Maillo [1994, p.7], se enfrenta con problemas como la separación de los enclíticos (dámelo), las contracciones (del) o la unión de las palabras separadas (sin embargo). Y a nivel de la palabra habrá que considerar posibles composiciones (prefijos, infijos o sufijos) o derivaciones (normalizaciones a partir de verbos). Los programas que practican estas operaciones de modo automático reciben el nombre de analizadores o etiquetadores morfológicos.

En cambio, la sintaxis es la “encargada del estudio de las relaciones sintagmáticas contraídas por las distintas unidades lingüísticas”, según Vera Luján [1994, p. 56]. Un analizador sintáctico detecta, por tanto, las relaciones sintácticas entre las palabras de una frase y resuelve los problemas no determinados en el analizador morfológico con respecto a la ambigüedad gramatical de las palabras.

Por último, la semántica estudia la significación de los signos lingüísticos y sus combinaciones. Un análisis semántico trata de averiguar el significado de las oraciones de un texto, y por extensión la comprensión del mismo. Uno de los obstáculos en el procesamiento del nivel semántico es la gran cantidad de conocimiento necesario acerca de las palabras y su significado en el universo del discurso con el fin de formalizar tales interpretaciones. De ahí que la Lingüística computacional no haya ofrecido aún mecanismos definitivos.

A continuación, comentamos brevemente algunas propuestas de sistemas para la automatización de la indización que emplean diferentes aspectos del PLN.

Sistema SMART

El sistema SMART fue desarrollado por Gerard Salton a principios de los años setenta. Este programa analizaba documentos de forma automática con el propósito de atender peticiones de información. En aquellos momentos supuso una importante aportación para el proceso iniciado años antes de automatización de la indización. El sistema SMART utilizaba como principal herramienta los cálculos estadísticos, si bien incorporó procesamientos lingüísticos básicos en relación a la forma de las palabras, la estructura de las oraciones y su significado.

Como apuntamos en Gil Leiva y Rodríguez Muñoz [1996a, p. 281] las raíces y los sufijos de las palabras se lograban a través de un diccionario compuesto por dos listados: uno con las raíces de las palabras ordenadas alfabéticamente (ej.: ecom-) y otro con sufijos (-ist, -ists, -ical) aplicado a la descomposición de las palabras como «economist», «economists», o «economical». Se introdujo también la posibilidad de reconocer como equivalentes una palabra en singular o en plural («location» y «locations»), las cuales poseían un único código de identificación. Por tanto, los vocablos con la misma raíz se trataban como semejantes («automaton», «automation» o «automatic»). Por otro lado, se incorporaron herramientas para la localización de sinónimos por medio de un diccionario, así como la comparación de palabras y frases presentes en los documentos ya analizados con peticiones de búsqueda. Para ello, se contaba con estructuras semánticamente equivalentes, pero construidas de modo diferente desde el punto de vista sintáctico.

El sistema de Faraj et al.

Faraj et al. [1996, p. 2] han manifestado que las investigaciones para obtener términos compuestos de los textos con el fin de aplicarlos en la automatización de la indización no han producido resultados estimulantes. Esta afirmación la sustentan en virtud de los trabajos realizados por Fagan [1987], Croft, Turtle y Lewis [1991], Lewis [1992] y Blosseville, et al. [1992]. Por estas razones Faraj y sus colegas han optado por considerar los términos compuestos como si fueran simples a la hora de la ponderación.

El sistema de indización automática propuesto por Faraj et al. [1996] para textos en francés, se apoya en métodos lingüísticos y estadísticos. El proceso lingüístico de los textos es el primer paso acometido en el programa denominado Termino¹. Este programa practica un análisis sintáctico que permite la resolución de las ambigüedades lexicales y la generación de 'synapsies'². Está constituido por varios módulos, cada uno de ellos con funciones diferentes. Veamos los principales:

¹ Termino es un programa de análisis terminológico asistido por ordenador desarrollado por el grupo Recherche et Développement en Linguistique Computationnelle (RDLC) del Centro de análisis de textos por ordenador (ATO) de la Universidad de Québec de Montréal.

² Los autores denominan 'synapsie' a una unidad nominal polilexical formada por varios términos construidos sintácticamente. Se trata de grupos preposicionales, nominales o adjetivales. El programa

EDITO (Tratamiento de las marcas de edición). Lleva a cabo la descomposición del texto en palabras y en frases. Además, reconoce los nombres propios.

LCMF (Lematización y Caracterización Morfológica). Proporciona la categoría gramatical para cada palabra del texto y su caracterización morfológica (género, número, persona, tiempo y modo).

ALSF (Analizador Léxico-Sintáctico). Resuelve las posibles ambigüedades halladas en el módulo anterior con respecto a las posibles categorías gramaticales de una palabra.

MRSF (Módulo de Reconocimiento de las 'synapsies'). La determinación de las 'synapsies' consiste en una exploración de la representación sintáctica proporcionada por ALSF. Cuando se ha analizado una frase, primero se buscan todos los grupos nominales que la constituyen, y después, se aíslan los construidos con la ayuda de complementos adjetivos o preposicionales. Y cada 'synapsies' queda asociada con una representación de su estructuración y marcación sintácticas en la frase. El programa Termino produce cuatro categorías diferentes contenidas en distintos ficheros: 'synapsies', adjetivos, verbos y sustantivos. Para la indización se eligen todas estas clases o una combinación de ellas.

Al resultado del módulo de PLN efectuado por el programa Termino se aplican métodos de ponderación estadística. El peso de los términos simples lo han calculado a partir de las propuestas que razonaron Salton y Buckley [1988], en base a tres elementos: su frecuencia en el documento, su frecuencia en la colección y un factor de normalización.

Una vez obtenidos los distintos cálculos, utilizaron para la evaluación del sistema el modelo vectorial³. Los autores han llegado a la conclusión de que, al menos para la

Termino localiza estructuras sintácticas del tipo «software integrado de gestión», cuya forma es T GA GP, donde T es el núcleo de la 'synapsie', GA es el grupo adjetival, y GP es el grupo preposicional.

³ En el modelo vectorial cada documento está representado por un vector de la forma $D_i = (p_{i1}, p_{i2}, p_{i3}, \dots, p_{it})$ donde t_k representa el peso del término k en el documento i. Para efectuar una búsqueda en el sistema, el usuario proporciona una pregunta en lenguaje natural. La demanda se analiza y se calcula el peso para cada una de los términos según los criterios descritos, y se representa por el

colección de documentos y preguntas evaluadas, la utilización de términos compuestos cosechados por el análisis sintáctico producen mejoras sistemáticas con respecto al empleo de términos simples.

A continuación, revisamos varias propuestas para la automatización de la indización presentadas por investigadores españoles. Las iniciativas de Valle Bracero y Fernández García [1983], Simón Granda y Lema Garzón [1990] se fundamentan en la localización de estructuras sintagmáticas en los textos, previamente aceptadas como idóneas para seleccionar los conceptos representados con dichas estructuras. Esta misma línea de actuación se sigue también en Seo [1993].

A grandes rasgos la propuesta de Simón Granda y Lema Garzón queda establecida de la siguiente manera:

1. Segmentación del texto en unidades inferiores de la oración (frases comprendidas entre los signos de puntuación).
2. Verticalización de las frases comprendidas entre los signos de puntuación.
3. Eliminación de todas las palabras vacías tras su comparación con un fichero.
3. Análisis morfológico del resto de palabras para la adjudicación de las posibles categorías gramaticales.
4. Selección de las estructuras localizadas en los textos previamente establecidas.
5. Eliminación de los términos repetidos para establecer con los restantes una jerarquía de más general a más específica.
6. Presentación y Validación de los términos.

vector $R_j = (t_{j1}, t_{j2}, t_{j3}, \dots, t_{ji})$. Posteriormente se calcula el coeficiente de similitud entre los términos asignados al documento con los de la pregunta.

La propuesta de Valle Bracero y Fernández García [1983] se asienta en un método mixto entre la indización semántica (selección de términos tras su comparación con vocabulario controlado) y la sintáctica (reconocimiento de estructuras sintagmáticas preestablecidas en los textos). El área elegida para el análisis es la Metalurgia. El proceso se concreta a grandes rasgos en estas etapas:

1. Búsqueda de estructuras sintagmáticas tras el análisis morfológico y sintáctico del tipo: sustantivo + adjetivo y sustantivo + participio.
2. Singularización y eliminación de preposiciones de los descriptores autorizados.
3. Comparación de los términos candidatos formados automáticamente con el vocabulario controlado. Si coincide plenamente se toma como descriptor y de lo contrario sólo se selecciona el unitérmino coincidente.

El fundamento esencial de estas propuestas era la extracción de un conjunto de estructuras sintagmáticas preestablecidas o “patrones admitidos” para constituir candidatos a descriptores. Los autores utilizaron diferentes estructuras sintagmáticas⁴. Si tras la fase de etiquetado de las palabras, algunas de las estructuras sintagmáticas no se encuentran entre los patrones admitidos el programa las rechaza. Y se inicia la búsqueda del próximo sustantivo que pueda encabezar un posible descriptor.

Hemos tenido la oportunidad de comprobar y presentar [Gil Leiva y Rodríguez Muñoz, 1997b] la diversidad de estructuras sintagmáticas adoptadas por los descriptores. Para ello analizamos la estructura interna de 1284 descriptores de las Bases de datos del CSIC (ISOC, IME e ICYT) para seis áreas del conocimiento:

⁴ En el caso de Simón Granda y Lema Garzón son las siguientes: 1. Sust ; 2. Sust + Adj ; 3. Sust + (Comentario) ; 4. Sust + Sust_Adj ; 5. Sust + Frase Prep ; 6. Sust + Frase Prep + Frase Prep³ ; 7. Sust + Part ; 8. Sust + Adj + Frase Prep ; 9. Sust + [Adj + Sust + Prep] ; 10. Sust + Cardinal ; 11. Sust + Ordinal ; 12. Frase Prep⁴ ; 13. Siglas.

Biblioteconomía y Documentación, Medicina, Química, Biología, Psicología y Física. Algunas estructuras que dan muestra de esta multiplicidad son las siguientes:

1. Sust + Conjunción + Sust
2. Sust + de + Sigla
3. Voces inglesas
4. Cardinal + Sust
5. Sust + de + Sust + no + Sust + Adj
6. Sust + Adj + de + Sust + Adj
7. Sigla + Cardinal
8. Sust + Voz latina
9. Sigla + Cardinal
10. Voz latina + Adj + Adj
11. Sust + de + Nombre propio
12. Sust + Sigla
13. Sust + Nombre propio + Adj
14. Sust + Conj + Sust + de + Sust + de + Sust
15. Sust + de + Sust + Nombre propio
16. Sust + de + si + Adj
17. Sust + de + Sust + Adj + de + Nomb propio
18. Voz latina
19. Sust + no + Adj

Asimismo, se revisaron los descriptores de diferentes áreas temáticas incluidos en seis tesauros. Se obtuvieron al menos una treintena de estructuras sintagmáticas diferentes en cada tesoro examinado. Ejemplos:

Tesoro Spines [1988]

CRÉDITO PARA LA I+D
INSTITUCIONES QUE OTORGAN SUBVENCIONES

Tesoro Eurovoc [1995]

FINANCIACIÓN A MUY CORTO PLAZO
COMITÉ PARLAMENTARIO MIXTO EEE

Tesoro de la Unesco [1982]

INDUSTRIAS CON FUERTE DENSIDAD DE MANO DE OBRA
INCITACIÓN AL ODIOS Y A LA VIOLENCIA

En definitiva, el estudio de la posible variedad de estructuras sintagmáticas en los descriptores es de suma importancia para un sistema de indización semiautomático o automático basado en estas premisas. Por tanto, si se dejan estructuras sin incluir en los patrones admitidos, el programa no localiza dichas estructuras como candidatos a descriptores, lo que provoca la exclusión de posibles términos presentes en el texto.

Los inconvenientes achacables a los sistemas para la automatización de la indización fundamentados en el procesamiento del lenguaje natural son los siguientes:

Los etiquetadores morfo-sintácticos consumen gran cantidad de recursos teóricos (bases de datos léxicas, bases de datos con reglas gramaticales) y de proceso (tiempo de ejecución).

Cuando una propuesta de indización semiautomática o automática utiliza técnicas del PLN, al resultado obtenido se aplican bien cálculos estadísticos o bien se compara con un vocabulario controlado. Es decir, en última instancia, el PLN acometido se ve supeditado a los mandatos estadísticos o a la presencia de estructuras sintagmáticas en dichas herramientas terminológicas.

La inexistencia de una biblioteca básica de programas informáticos de libre distribución para el PLN. Esta biblioteca básica de programas debía contener desde corpus de diferentes dominios y niveles y lexicones hasta analizadores morfológicos y sintácticos. Estas herramientas son necesarias para el desarrollo de investigaciones que requieren la utilización del PLN.

Los problemas para el desarrollo de herramientas dedicadas al procesamiento del lenguaje natural, según Verdejo Maillo [1994, p. 19], son de carácter:

Científico (la diversidad de teorías y metodologías para el abordaje de una misma tarea).

Técnico (la complejidad de los algoritmos).

Comerciales (el elevado coste de recursos técnicos y humanos encarece los productos, unido a un mercado escaso).

En definitiva, estos y otros problemas han propiciado que esta situación precaria se mantenga hasta la actualidad, porque Carretero y Rodríguez [1997, p. 39] se siguen lamentando de la “carencia” de estos instrumentos para el español, con respecto a otras lenguas menos extendidas que la nuestra, pero con mayor influencia tecnológica.

2.3.3 El uso de Tesoros.

La mayor parte de los sistemas operativos encaminados a la automatización de la indización utilizan un tesoro, ya sea en instituciones privadas (productores de bases de datos) o en públicas (Centros de Documentación o Bibliotecas especializadas).

En los trabajos o lugares referenciados a continuación, han utilizado un tesoro en los sistemas desarrollados para automatizar la indización: Martínez, Lucey y Linder [1987]^{*} ; Biebricher, et al. [1988]^{*} ; Lovtsov [1990]^{*} ; Coret, Ducloy y Menillet [1991]^{*} ; Silvester, Genuardi y Klingbiel [1994]. Y además, en las Bases de datos TULSA⁵ ^{*} e INIS⁶ ^{*}, en el Centro de Información Técnica de la Defensa de Alexandria (en Vancouver)^{*}, y en las Bibliotecas Nacionales de Agricultura y de Medicina de los Estados Unidos^{*}.

Otras propuestas fundamentadas en los tesoros son las siguientes: Strode [1977], Dillon [1982], Scheele [1983], Valle Bracero y Fernández García [1983], Ciganik [1990], los sistemas PASSAT, SINTEX y ALEXDOC [Slype, 1991], Evans [1991], Hersh et al. [1992] y Wan [1995].

Como tendremos la oportunidad de explicar en el siguiente capítulo la terminología constituye un requisito primordial para la existencia y el funcionamiento de un

* Referencias obtenidas de Hodge [1992].

⁵ Base de datos perteneciente a la Universidad de Tulsa que contiene literatura técnica y patentes en los campos de la exploración y producción de derivados del petróleo y gas natural.

⁶ Base de datos internacional con sede en Viena preparada por la Agencia Internacional de la Energía Nuclear. Esta base de datos, consultable desde diferentes países, contiene más de un millón y medio de registros bibliográficos sobre aplicaciones pacíficas de la Tecnología y la Energía nuclear.

lenguaje especializado utilizado por profesionales de un determinado campo. La terminología cumple diversas funciones y su captación de los textos supone la aprehensión de conceptos y por tanto, del conocimiento. En base a estos fundamentos, los términos que representan conceptos no sólo nos permiten conocer el tema y el contenido de los documentos sino también transmitirlos.

Para verificar como actúan los sistemas que persiguen la automatización de la indización con la ayuda de un tesoro vamos a comentar brevemente varias experiencias.

Con la idea de reducir los costes y agilizar los procesos de la indización se inició en el Centro de Información Aeroespacial de la NASA un proyecto para diseñar un sistema de indización semiautomático. El sistema está compuesto por tres módulos [Silvester, Genuardi y Klingbiel, 1994]. En el primero se realizan diversas funciones entre las que destacan: la identificación de las fuentes que van a ser procesadas, la limitación de las series de palabras del texto (título y resumen) y las llamadas al segundo módulo para ejecutar ciertas operaciones. Además, a este primer módulo llegan finalmente los términos de indización propuestos para su validación. En el segundo módulo se efectúa la búsqueda de frases significativas del texto con un máximo de cinco palabras, para lo cual, recurren a la base del conocimiento o red conceptual (tercer módulo) constituida por un vocabulario controlado con más de 115.000 candidatos a términos de indización. Desde esta base del conocimiento se establecen las posibles relaciones existentes entre los términos desde el punto de vista jerárquico o incluso, la desambiguación de los mismos. Según los autores, cuando los resúmenes son de calidad, en la fase de validación se acepta el 60 % de los términos propuestos por el sistema.

El sistema SAPHIRE [Hersh et al., 1992] es un sistema de análisis y recuperación de documentos de Biomedicina, cuyo núcleo es un algoritmo de comparación de términos entre un documento o una pregunta de un usuario y un tesoro.

El programa consagra un tesoro de la Biblioteca Nacional de Medicina de Estados Unidos llamado Meta-1. Este metatesauro⁷ contiene dos tipos de entradas: los

⁷ Meta-1 es un metatesauro producto de un proyecto emprendido en la Biblioteca Nacional de Medicina de Estados Unidos a partir de 1986. La finalidad era obtener una herramienta que enlazara un gran número de vocabulario médico, para lo cual unificaron vocabularios como el MeSH (manejado para indizar MEDLINE), DSM-III (American Psychiatry Association), SNOMED (American College of Pathologists), ICD-9 (World Health Organization), y LCSH (Library of Congress).

términos aceptados y sus sinónimos (los sinónimos son tanto a nivel de término («alto» y «elevado») como de conceptos «hipertensión» y «tensión alta»). Meta-1 utiliza 28.423 términos, 78.244 sinónimos y 28.603 raíces de palabras. Una vez comparado el texto con el tesoro, a cada término retenido se le aplican cálculos estadísticos. En concreto, la frecuencia inversa del documento multiplicada por la frecuencia de cada término en dicho documento.

La principal objeción atribuible a los sistemas de indización semiautomática o automática asentados en este método es la siguiente. La incapacidad de seleccionar de un texto un candidato a término de indización que no se encuentre en el vocabulario controlado. Este mismo inconveniente aparece cuando se indiza un documento de modo manual, ya que generalmente, no se suelen asignar términos que no aparecen en el vocabulario controlado utilizado.

2.3.4 El uso de sistemas híbridos.

En realidad, a excepción de las primeras propuestas de los años sesenta fundamentadas íntegramente en métodos estadísticos, la mayor parte de los prototipos utilizan al menos dos de las metodologías presentadas en los párrafos precedentes. Veamos unos ejemplos:

Sistemas fundamentados en métodos estadísticos y PLN

- SMART -aunque fundamentalmente la estadística- [Salton, 1980]
- Andreewsky y Ruas [1982]*
- SPIRIT [Xivry, 1993]
- Faraj (et al.) [1996]

Sistemas fundamentados en métodos estadísticos y tesoro

- SAPHIRE

* Esta referencia no se ha consultado directamente, sino que se localizó durante el proceso de revisión bibliográfica y de consulta de diferentes bases de datos.

Sistemas fundamentados en PLN y tesoro

- Valle Bracero y Fernández García [1983]

Sistemas fundamentados en métodos estadísticos, PLN y tesoro:

- PASSAT
- CLARIT

En SPIRIT⁸ la indización automática, como ya hemos expresado, reposa sobre fundamentos lingüísticos y estadísticos. El procesamiento lingüístico [Gil Leiva y Rodríguez Muñoz, 1996a, p. 284] se acomete en las siguientes fases: división de las cadenas de caracteres en términos simples; análisis morfológico con un diccionario formado por 500.000 palabras; identificación de términos compuestos que representan un concepto (ej.: «caballos de vapor»); análisis sintáctico por el que se detectan las ambigüedades gramaticales y las resuelve; reconocimiento de palabras compuestas en base a criterios sintácticos con la identificación de estructuras nominales y verbales en las frases; eliminación de las palabras vacías a partir de criterios gramaticales (pronombres) y morfológicos (lista de palabras vacías). Y por último, la normalización de las palabras restantes. Esta normalización convierte una palabra flexionada en su canónica (livre / sustantivo = sustantivo ; livre / verbo = livrer).

En el módulo estadístico⁹ se atribuye a cada término normalizado (simple o compuesto) una ponderación según su valor de discriminación. Este peso servirá tanto para la indización como para la recuperación del documento. El algoritmo asigna el valor informacional de cada término en relación al conjunto de los términos contenidos en la base de datos. De este modo, el sistema concede mayor relevancia

⁸ Système Probabiliste d'indexation et de Recherche d'Informations Textuelles. Se trata de un programa de Gestión Documental comercializado en Francia que permite la adquisición, indización, almacenamiento, búsqueda y difusión de la información. Es el resultado de más de 20 años de investigaciones teóricas y aplicadas realizadas en aquél país en el seno de la Comisaría de la Energía Atómica (C.E.A) y la Facultad de Lingüística d'Orsay. Analiza diversos tipos de textos (jurídicos, científicos, técnicos, comerciales, informáticos) y en diferentes idiomas (francés, inglés y alemán). Además, la interrogación en la base de datos, que se formaliza en lenguaje natural, se puede realizar en estos mismos idiomas.

⁹ En ninguna de las fuentes consultadas [Slype, 1991; Xivry, 1993; documento informativo obtenido de la marca comercial T.GID] para conocer este sistema se concreta la base y el funcionamiento del módulo estadístico.

cuanto menos común sea un término. Por tanto, en una base de datos sobre Odontología, se atribuirá más peso al concepto «caries» que a «odontología» puesto que éste último estará presente en casi todos los documentos.

El sistema CLARIT es una propuesta para la automatización de la indización completamente interdisciplinar. Este prototipo puede tomarse como 'modelo' de los sistemas híbridos ya que utiliza un tesoro, herramientas para el PLN y métodos estadísticos. Está compuesto [Evans et al., 1991] por tres módulos donde desarrolla diferentes tareas de formateado del texto, procesamiento del lenguaje natural y filtrado de los términos.

En el primer módulo efectúa el formateado del texto con la colocación de las marcas de comienzo y final de oraciones y párrafos. En el segundo módulo se lleva a cabo el PLN en dos etapas: el análisis morfológico y el sintáctico. El léxico está compuesto por más de cien mil raíces de palabras inglesas. Y en el tercer módulo se formalizan varios procesos:

i) los cálculos estadísticos sobre los términos y las frases nominales ofrecidos por el segundo módulo en cuanto al número de veces que aparece una palabra en el documento. Este dato estadístico se considera junto a otros del tipo: número de documentos en los que una palabra ha aparecido, frecuencia de aparición de una palabra en el corpus específico en cuestión y frecuencia de aparición de una palabra en un corpus general.

ii) los términos que alcanzan una determinada ponderación se comparan con los términos del tesoro y se clasifican en tres categorías: los términos exactos (coinciden con los del tesoro), los términos nuevos (sobrepasan un determinado umbral de aparición) y los términos generales (no están entre los exactos ni entre los nuevos).

Según la exposición hilada a lo largo de este capítulo, la selección de términos en la indización semiautomática o automática se ha realizado mediante la utilización de diferentes metodologías y principios:

Estadísticos (el cálculo de frecuencias de aparición de las palabras tanto en los documentos como en las colecciones).

Sintácticos (la búsqueda de las estructuras sintagmáticas adoptadas por los términos de indización).

Semánticos (la comparación de textos con un vocabulario controlado).

2.3.5 La comparación de la eficacia de la indización automática versus manual.

En todas las propuestas para la automatización implantadas, aunque haya sido a nivel de laboratorio, se ha intentado comprobar la calidad del producto en relación a la indización intelectual. Tanto es así, que la revisión bibliográfica sobre indización automática de Bastos Vieira [1988] recoge brevemente varios ensayos encaminados a la comparación de estas diferentes formas de indizar. Este autor no proporciona información sobre las características y los principios de los sistemas automáticos en la mayoría de los casos, por lo tanto, mostramos solamente las conclusiones alcanzadas en dichos ensayos:

Carroll y Roeloffs [1969] concluyeron que la indización automática era más viable porque los resultados obtenidos fueron semejantes a los conseguidos por profesionales, a lo que habría que sumar los costes de contratación de personal y la inconsistencia de los indizadores humanos.

Boyce y Lockard [1975] practicaron sus ensayos con documentos de Medicina. Según los resultados alcanzados, con la indización automática la exhaustividad en la recuperación era superior tanto con el empleo de preguntas generales como específicas. En cambio, la precisión fue mayor para la indización manual cuando las preguntas incluían términos específicos, pero cuando comprendían términos generales la automática conquistaba mejores índices de precisión. Finalmente juzgaron que la indización automática era tan eficiente como la manual.

Klingbiel y Rinker [1976] comprobaron la eficacia de los dos métodos de indización con títulos y resúmenes del programa Machine-Aided Indexing. En este ensayo la indización manual y la automática obtuvieron índices de exhaustividad y precisión similares en la recuperación, pero la primera se mostró inferior en la precisión.

Meulen y Janssen [1977] utilizaron los títulos y los resúmenes de documentos para llevar a cabo su experimentación. Crearon dos bases de datos con los mismos documentos pero indizados de forma manual y automática respectivamente. Después, interrogaron cada base de datos con dos preguntas. La indización manual presentó mejores coeficientes de exhaustividad y precisión, cercanos al 20 %, en relación a la automática. Según los autores se obtuvieron estos resultados por el escaso número de preguntas utilizadas.

Barnes, Costantini y Perschke [1978] tomaron también títulos y resúmenes. El ensayo se realizó sobre 5000 documentos. En la recuperación, los términos propuestos por el sistema automático se mostraron más eficientes que los asignados por profesionales. El coeficiente de exhaustividad en la recuperación con la indización automática fue del 90%.

Otros ensayos para comparar la validez de los métodos automáticos frente a los intelectuales han sido los siguientes:

Gerard Salton analizó los resultados de la indización con el sistema SMART con respecto a los reportados por profesionales. Para ello, tomó los datos de un estudio de Lancaster [1969] y los comparó con los resultados proporcionados por SMART. En la tabla facilitada [Salton y McGill, 1983, p. 104] se comprueba que cuando el sistema automático utiliza técnicas de selección de términos simples desde los resúmenes, la exhaustividad y la precisión en la recuperación es inferior en un 16% y 19% respectivamente en relación a MEDLARS. En cambio, cuando SMART emplea la técnica del valor de discriminación de los términos, esta diferencia queda reducida al 8% en la exhaustividad y al 4% en la precisión. Aún más, cuando en el proceso de indización automática se utiliza un tesauro para reconocer términos sinónimos y relacionados desaparecen estas diferencias. La exhaustividad con SMART supera a la de MEDLARS en un 4% mientras que la precisión es la misma en los dos modos de indización.

Hersh et al. [1992] para comprobar la efectividad de SHAPIRE (sistema automático de indización y de recuperación) efectuaron el siguiente ensayo: indizaron 200 resúmenes tanto por expertos como por el sistema SAPHIRE (sólo títulos y resúmenes). El sistema automático asignó un total de 4552 términos, con una media de 22,8 por resumen, por el contrario, los indizadores

humanos asignaron un total de 1966 términos con la MeSH con una media de 9,8. Posteriormente, tras seleccionar una tanda de preguntas se acometieron búsquedas con términos de MeSH¹⁰ en los títulos y los resúmenes de los documentos. La conclusión a la que se llegó es que con SAPHIRE se produjo una menor exhaustividad y precisión en la indización con respecto al modo tradicional. Según estos autores, las causas de estas diferencias se debieron a las lagunas de sinonimia del tesoro Meta-1, por lo que con una mejora substancial en el mismo, repercutirá fructuosamente en la indización.

Por último, en una búsqueda en la base de datos Dissertation Abstracts Online localizamos varias tesis doctorales presentadas en el Instituto de Tecnología de Illinois, donde se ha tratado de comprobar la efectividad de la indización automática frente a la intelectual. En uno de los trabajos [Hmeidi, 1995] se emplearon resúmenes de documentos en árabe. Estos documentos se indizaron de forma automática (por palabras, temas y raíces de palabras), y de modo manual. Tras los análisis sobre los coeficientes de exhaustividad y precisión en la recuperación, se concluye que la indización automática es tan efectiva como la manual. A las mismas conclusiones se llegó en [Wan, 1995] y [Seo, 1993]¹¹. El primero trabajó sobre resúmenes de documentos en chino, mientras que Seo destinó su investigación a resúmenes en coreano.

En definitiva, según los resultados alcanzados en estos experimentos, los sistemas automáticos y la indización intelectual logran índices semejantes en cuanto a la exhaustividad y la precisión en la recuperación de la información.

2.4 LA INTERDISCIPLINARIEDAD EN LA AUTOMATIZACIÓN DE LA INDIZACIÓN.

La interdisciplinariedad es un aspecto inherente a la automatización de la indización. Es más, consideramos que no se podría hablar de indización semiautomática o automática sin el aporte de una serie de disciplinas que ahora mostraremos. Una

¹⁰ Lista de encabezamientos de materias de Medicina utilizada en la Biblioteca Nacional de Medicina de los Estados Unidos.

¹¹ En los resúmenes de los trabajos de Hmeide y Wan, recogidos en la Base de datos Dissertation Abstracts no quedan explicitados los métodos y los principios de indización automática utilizados.

revisión superficial de las propuestas en esta dirección basta para comprobar la intervención de diferentes disciplinas en la resolución de los problemas planteados.

Ribeiro Pinheiro y Matheus Loureiro [1995] y Saracevic [1995] consideran que los problemas complejos necesitan abordajes interdisciplinares y soluciones multidisciplinares. Idea que compartimos. Por otro lado, no aportamos nada nuevo si expresamos que la indización es una operación complicada. También hemos aceptado [Gil Leiva, 1997a ; 1997b] que esta complejidad se torna doble cuando se persigue la automatización de esta tarea. De ahí que se busquen sólidos argumentos, tanto teóricos como prácticos, en otras disciplinas, así como el trabajo conjunto con profesionales de dichas áreas, para la subsanación de las dudas suscitadas. Las principales ciencias o disciplinas a las que se recurre son la Lingüística, la Informática, la Estadística, la Lingüística computacional y la Terminología.

El desarrollo científico y técnico ha provocado el nacimiento y la consolidación de un grupo de disciplinas o áreas de trabajo con un tremendo carácter interdisciplinar como es el caso de la Terminología, la Inteligencia Artificial, la Lingüística Computacional o la misma Documentación.

La Terminología es impensable sin los fundamentos lingüísticos, normalizadores o informáticos.

La Inteligencia Artificial no tiene razón de ser sin las aportaciones de la Informática, la Ciencia Cognitiva o la Psicología.

La Lingüística computacional no se entiende sin sus apoyos en la Lingüística, la Lógica, la Matemática, la Programación o la Psicolingüística.

La Documentación está impregnada de ciencias y disciplinas que le otorgan principios para la constitución de sus fundamentos y herramientas de trabajo. Algunas de las ayudas recibidas proceden de la Lingüística, la Informática, la Terminología, la Semiótica, la Lingüística Computacional, la Ciencia cognitiva, la Inteligencia Artificial, o la Lógica.

Las aportaciones más importantes que recibe la Documentación, y en particular la automatización de la indización, son de la Lingüística y la Informática. Esta idea está completa y magníficamente recogida en unas frases escritas a mitad de los años setenta por Susan Artandi [1976, p. 235]. Aquí juzgaba que el uso de los ordenadores permite a la Documentación la manipulación de textos en lenguaje natural para su utilización en la recuperación, puesto que la recuperación de información es de vital interés para las, cada vez más, sofisticadas necesidades de información por parte de la sociedad. Efectivamente, la Documentación se dedica a la recopilación, análisis, almacenamiento, y difusión de la información para lo que demanda, conocimientos e instrumentos de otras disciplinas como medio de alcanzar su misión.

Por otro lado, la asistencia de la Lingüística y la Informática en la automatización de la indización no son menos destacadas. Ésta última, persigue la selección de una serie de conceptos de los textos escritos en lenguaje natural. Por tanto, conlleva que el ordenador detecte signos que representan conceptos o comprenda los diferentes niveles implicados en el lenguaje natural. A partir de aquí, las propuestas para la automatización de la indización, si descansan sobre métodos lingüísticos, se asientan en uno o varios niveles y componentes del lenguaje.

Diversos autores han reconocido no solo la interdisciplinariedad de la Documentación en general y de la automatización de la indización en particular, sino que consideran un paso importante la creación de grupos interdisciplinares para avanzar en el área. Estos autores son [Baranow, 1983 ; García Gutiérrez 1989b; 1992; 1995 ; Coll-Vinent, 1990 ; Moreiro González, 1993b,c ; Gil Leiva y Rodríguez Muñoz, 1996a; 1996b ; Gil Leiva, 1997c].

Esta interdisciplinariedad, tanto de especialidades o técnicas como de investigadores de diferentes áreas que intervienen en la automatización de la indización, se puede comprobar en la práctica en diversos trabajos:

En Jones et al. [1990] los autores pertenecen unos al Departamento de Ciencias de la computación, y otros, a la School of Library Science, ambos de la Universidad de Louisiana.

En la propuesta de Foltz y Dumais [1992] el primero es del Departamento de Psicología de la Universidad de Colorado y la segunda, es una investigadora en Biblioteconomía y Documentación.

También se detecta esta interdisciplinariedad en [Karetnyk, Karlsson y Smart, 1991] cuyos integrantes del proyecto SIMPR son lingüistas computacionales, documentalistas e informáticos.

O en los planteamientos de Schuegraf y Bommel [1993] que incluye a miembros del Departamento de Matemáticas y Ciencias de la Computación de la Universidad de St. Francis Xavier de Antigonish (Nova Scotia).

Otras iniciativas interdisciplinares las encontramos en [Fagan, 1989], adscrito a la vez al Departamento de Lengua Moderna y Lingüística y al de Ciencias de la computación de la Universidad de Cornell (Estados Unidos).

En Woodruff y Plaunt [1994], el primer autor pertenece al Departamento de Ciencias de la computación y el segundo, a la Library and Information Studies, ambos de la Universidad de California en Berkeley.

Más reciente aún es el trabajo de [Faraj et al., 1996], donde sus participantes son del Departamento de informática y del Centro de análisis de textos por ordenador de la Universidad de Québec de Montréal (Canadá).

La interdisciplinariedad presente en la automatización de la indización propicia que en la formación de los grupos interdisciplinares cada investigador o grupos de investigación aporten sus bagajes, traducido no sólo en conocimientos sino, a veces, también en herramientas construidas por ellos mismos. Esta circunstancia hemos tenido la oportunidad de observarla en Karetnyk, Karlsson y Smart [1991] y en Faraj, et al. [1996].

A continuación, veremos las aportaciones efectuadas a la automatización de la indización por las ciencias y las disciplinas ya mencionadas. En algunas ocasiones, las contribuciones de disciplinas al asunto que nos ocupa resultan decisivas, y en otras ocasiones, los sistemas toman principios conceptuales o prácticos. Por tanto, mostraremos algunos de los acervos teóricos o procedimentales adquiridos.

Para concluir con esta introducción a la interdisciplinariedad de la Documentación y en particular de la automatización de la indización reproducimos unas reflexiones de García Gutiérrez [1995, p. 14] acerca de estos asuntos:

“En los últimos años, ora la Documentación se ha abierto a otras disciplinas, ora otras disciplinas han hecho escala en la nuestra. Pues bien, aun siendo defensor de la investigación interdisciplinar como única vía para otorgar estatuto científico a la Documentación, he de expresar una gran preocupación, desprendida de la lectura de algunos libros o de la asistencia a congresos, por el extrañamiento del objeto documentológico, aún embrionario, en beneficio de otras disciplinas tradicionales y consolidadas que, por motivo de saturación en muchos casos, optan por aplicaciones en Documentación. Así, estudios inicialmente documentológicos se transforman y consideran estudios lingüísticos, lógicos, tecnológicos u otros”.

2.4.1 Lingüística.

Las reflexiones sobre la relación entre la Lingüística y la Documentación y de los aportes de la primera comenzaron a principios de los setenta. La automatización de la indización, en particular, se ha asentado principalmente sobre el cálculo de la frecuencia de las palabras, en el análisis automático del lenguaje de los textos o en ambos a la vez. Por tanto, la relación entre la automatización de la indización y Lingüística queda justificada.

Un lenguaje es un sistema [Marques Cintra, 1983 p.7], una organización relacional en donde cada elemento existe en la medida en la que se relaciona con otro u otros del mismo conjunto. Además, un lenguaje documental es un lenguaje porque concreta la capacidad simbólica del hombre a través de la organización de sus términos y reglas. Por tanto, un lenguaje documental tendrá que resolver complejidades presentes en el lenguaje natural como la polisemia, sinonimia, homonimia o antonimia.

Los niveles y los componentes de la lingüística que intervienen en la automatización de la indización son la morfología, la sintaxis, y en menor medida, la semántica:

La morfología se ha empleado para la descomposición de las palabras en sus raíces para agrupar aquellas con una raíz común. El fin es calcular la aparición de las palabras en base a sus raíces. Este nivel proporciona también las categorías gramaticales de las palabras.

La sintaxis, se ha encargado, principalmente, de los problemas de desambiguación gramatical no resueltos en el nivel morfológico. Algunas propuestas de indización semiautomática o automática seleccionan estructuras sintagmáticas de los textos como candidatas a términos de indización.

La semántica se utiliza, en algunos sistemas, para distinguir si un mismo término que se puede emplear en diferentes áreas del conocimiento pertenece a una u otra, según en el contexto en el que aparece. Ejemplo: banco = entidad financiera ; banco = mobiliario ; banco = pesca.

La idea de que la Lingüística es esencial para conceder sentido a la Documentación tanto teórica como aplicada está presente en [Coyaud, 1972* ; Montgomery, 1972* ; Sparck Jones y Kay, 1973* ; Gardin, 1973* ; Natali, 1978* ; Basilio, 1979* ; Baranow, 1983 ; García Gutiérrez, 1990 ; Moreiro, 1993b,c ; Díez Carrera, 1994].

Por otro lado, las relaciones entre la Lingüística y la automatización de la indización ha quedado reflejada en los trabajos de Smit [1974*]; Artandi [1976] ; Noceti y Figueiredo [1978*]; Michell [1979*]; Long [1980] ; Walker [1981] ; Marques Cintra [1983] ; Crystal [1984] ; Korycinski y Newell [1990] ; Lopes Gínez de Lara [1993]. Incluso se ha estudiado el impacto de la Lingüística en la Documentación en la tesis doctoral de Warner [1991]¹².

* Estas referencias no se han consultado directamente, sino que se localizaron durante el proceso de revisión bibliográfica y de consulta de diferentes bases de datos.

¹² El material utilizado para esta tesis fue el análisis de referencias bibliográficas citadas en artículos de un conjunto de fuentes con gran impacto internacional sobre Biblioteconomía y Documentación. Los resultados obtenidos refieren a que pocos investigadores en Documentación han utilizado el trabajo de un número reducido de lingüistas de la sintaxis y semántica de los años sesenta y setenta (Chomsky, Fillmore o Lakoff). Y unas de sus conclusiones es que hasta el momento la Lingüística no ha tenido un gran impacto en la Documentación.

2.4.2 Terminología.

La terminología es el “conjunto de conceptos de un campo del saber sistematizados, explicados por medio de definiciones, explicaciones e ilustraciones y hechos comunicables mediante términos” [Picht, 1996, p. 9]. La terminología está concebida de modo transdisciplinario puesto que mantiene vínculos con la Lingüística, la Lingüística computacional, la Ciencia Cognitiva y la Documentación, entre otras [Cabré, 1993, p. 71 ; Arntz y Picht, 1995, p. 22]. Las relaciones fundamentales entre la Terminología y la Documentación derivan porque la Documentación es un campo especializado con terminología propia, utiliza la Terminología y además, la Documentación hace accesibles los conocimientos terminológicos.

Coincidimos plenamente con estos autores [Baranow, 1983, p. 32 ; Espelt, 1995, p. 126 ; Esteban Navarro, 1995, p. 206] que han defendido la inclusión de la enseñanza de la Terminología en las Escuelas y Facultades de Biblioteconomía y Documentación. La difusión del conocimiento, ha sostenido Esteban Navarro, son dos objetivos tanto de la Terminología como de la Documentación. La Documentación controla y difunde los documentos que contienen conocimientos, mientras que la Terminología persigue la normalización y difusión de lenguajes especializados para permitir una buena comunicación científico-técnica.

La relación entre la Terminología y la Documentación aflora con más fuerza cuando se aplican principios de la Terminología en la construcción, el mantenimiento y el uso de los lenguajes documentales. La terminología está presente en los tesauros porque, como ya hemos visto, la indización es la búsqueda de los conceptos más representativos tanto de los documentos como de las preguntas para almacenar y recuperar la información. Por tanto, esos conceptos representados por términos se pueden convertir en descriptores. Y para que estos conceptos se conviertan en descriptores sufren un proceso de normalización y control con la ayuda de los tesauros dedicados a un área especializada.

El fin último de la utilización de los tesauros, es decir, de los lenguajes especializados, es la representación y la recuperación de los conceptos de un modo unívoco. Las relaciones principales entre la terminología de un tesoro son de sinonimia (ej.: BACTERIAS U.p Bacilos // Bacilos Use BACTERIAS), de jerarquía que comprende relaciones de términos genéricos o específicos (ej.: PLANTAS

ORNAMENTALES T.g. Plantas cultivadas // INTERVENCIÓN ESTATAL R.e. Nacionalismo), y de relación (ej.: NACIONALISMO T.r. minorías).

La relación existente entre la Terminología y la Documentación ha quedado patente en los párrafos precedentes. Pero si cabe alguna duda, basta mencionar que especialistas en Terminología como R. Arntz y H. Picht [1995] y M.T. Cabré [1993], han dedicado en sus obras epígrafes para establecer la conexión entre estas dos disciplinas denominados, en ambos casos «Terminología y Documentación».

2.4.3 Informática.

La informática se ocupa del tratamiento automático de la información y ha permitido desde los años cincuenta su representación, procesamiento, así como su almacenamiento y su recuperación de forma automatizada. En la actualidad, se emplea, directa o indirectamente, en todas las ciencias y técnicas. No obstante, la Documentación sin la Informática, pierde parte de su razón de ser, por lo que el desarrollo de la primera, durante las últimas décadas, ha sido paralelo al de la Informática.

La mayor parte de las tareas ejecutadas por un documentalista y/o bibliotecario se efectúa por medio de los ordenadores. Estas labores comprenden desde la entrada de la información al Centro, hasta su difusión a los usuarios. La disponibilidad tanto de hardware como de software específico ha provocado que cometidos como la selección o la adquisición de documentos y el análisis, pero, principalmente, el almacenamiento y la recuperación de información se desempeñen de modo automático.

La informática no solamente se utiliza en la indización para lo que podríamos denominar análisis de la información sino también para el proceso de almacenamiento de los términos de indización seleccionados de modo manual. En este último caso el indizador puede seleccionar términos o códigos de ficheros de autoridades, validar los términos introducidos con el fichero de autoridades, averiguar el desarrollo completo de una sigla o viceversa, y acceder a otros documentos indizados previamente. En definitiva, aunque la indización se realice de modo intelectual los ordenadores constituyen herramientas importantes para facilitar

las labores de los indizadores y agilizar los procesos de control de calidad en la indización.

2.4.4 Lingüística computacional.

En la introducción del libro de Grishman [1991] se expresa, claramente, la definición y las pretensiones de la Lingüística computacional. La Lingüística computacional es una disciplina entre la Lingüística y la Informática que trata de los aspectos computacionales de las lenguas naturales. Su principal objetivo es la comprensión de los procesos lingüísticos desde el punto de vista de su procesamiento para que los sistemas informáticos generen e interpreten las lenguas. Por tanto, los productos de la Lingüística computacional ofrecen para nuestros intereses una gran aplicación: el procesamiento automático de los textos, de donde tomamos los fundamentos y los procedimientos para el análisis (indización) y la recuperación de la información.

Uno de los objetivos de la Lingüística computacional es el análisis de los distintos niveles de las lenguas (morfológico, sintáctico, semántico y pragmático) a través de programas informáticos. Este proceso implica un análisis a nivel de la palabra, de la oración y del texto, respectivamente. Las herramientas de la lingüística computacional más utilizadas en la automatización de la indización son los etiquetadores morfológicos y los sintácticos.

Los etiquetadores morfológicos llevan a cabo la descomposición de las palabras para descubrir su categoría gramatical (sustantivo, adjetivo, verbo, u otras) y su caracterización morfológica (género, número, persona, tiempo y modo). En la indización semiautomática o automática el uso más sencillo ha sido para la descomposición de las palabras en su forma raíz, y así contabilizar todas las que posean la misma. Los etiquetadores morfológicos proporcionan a cada palabra su categoría gramatical. Los analizadores sintácticos toman como materia de análisis no la palabra sino la oración. Son capaces de identificar los diferentes componentes de la oración (sujeto, verbos o complementos), de averiguar si la oración es correcta gramaticalmente (concordancias entre sujeto y verbo o entre sujeto y complementos). Y asimismo, proceden a la desambiguación gramatical de la palabra en el caso de que fuera necesario. Los analizadores sintácticos se han utilizado en algunos sistemas para la automatización de la indización para la

desambiguación lexical o para la búsqueda de estructuras sintagmáticas establecidas de antemano.

En la actualidad, se sigue trabajando en este tipo de herramientas para mejorar y ampliar los logros alcanzados, pero aun así, como señalara Moreiro González [1993, p. 44] a este respecto:

“es necesario conjuntar la contribución y avanzar en la aplicación de los aspectos morfológicos, lexicográficos, sintácticos, de representación de los conocimientos en redes semánticas, y de investigación en los procesos inferenciales desde: el contexto, lo no dicho y las anáforas”.

2.4.5 Estadística.

La estadística lingüística aplica estas técnicas a los fenómenos de la lengua, según se ofrece en el diccionario de Lingüística coordinado por Cerdá Massó [1986, p. 106]. Esto ha permitido la determinación de la riqueza léxica de un autor, el enclave de un texto anónimo a una época o a un autor y el cálculo de la frecuencia relativa de uso de las palabras para comprobar su capacidad informativa. Este último uso es el que interesa en el contexto de este trabajo.

Ya hemos señalado en el apartado dedicado a los métodos estadísticos que G.K. Zipf propuso la fórmula:

$$\text{Frecuencia} \times \text{clasificación} \approx \text{constante}$$

para el valor constante que tiene la relación entre la frecuencia de aparición de las palabras y el rango o puesto que éstas ocupan en el orden frecuencial.

H.P. Luhn a finales de los cincuenta fue el primero en tomar la ley de Zipf para aplicarla en la automatización de la indización. Su idea consistía en que si una palabra aparecía en un texto con un determinado coeficiente de frecuencia (ni alto ni bajo) podía ser tomada como término de indización porque transmitía el contenido del texto. Posteriormente, estas ideas fueron retomadas por otros investigadores para ponderar los términos de los documentos con la función de frecuencia inversa

en un documento [Sparck Jones, 1972] o el valor de discriminación de los términos [Salton, Yang y Yu, 1975].

El uso de la frecuencia de las palabras en los textos ha sido la metodología más utilizada en las propuestas para la automatización de la indización. En general, podemos afirmar que hasta principios de los ochenta fue empleada casi de forma exclusiva en las diferentes aproximaciones efectuadas [Luhn, 1958 ; Damerau, 1965 Carrol y Roeloff, 1969 ; Rosenberg, 1971 ; Sparck Jones, 1972 ; Salton y Yang, 1973 ; Bookstein y Swanson, 1975 ; Salton, 1980].

Posteriormente, con los avances producidos en el procesamiento del lenguaje natural se incorporaron, paulatinamente, herramientas lingüísticas para el tratamiento automático de los textos, como instrumentos para la descomposición de las palabras, analizadores morfológicos o incluso sintácticos. Esta incorporación del PLN ha hecho que algunas de las propuestas se puedan denominar "híbridas", es decir, la utilización de métodos estadísticos y lingüísticos [Karetnyk, Karlsson y Smart, 1991 ; Evans (et al.), 1991 ; Hersh (et al.), 1992 ; Schuegraf y Bommel, 1993 Xivry, 1993 ; Faraj (et al.), 1996].

2.4.6 Sistemas expertos.

Los programas informáticos basados en los sistemas expertos incluyen el conocimiento y la experiencia de uno o varios especialistas para ofrecer soluciones a problemas específicos tal como lo haría un entendido. Los sistemas expertos son herramientas potentes que permiten trabajar de forma más rápida, fácil y completa, lo que puede suponer mejoras en la productividad, en el aprendizaje y la comprensión y en la conservación de conocimientos importantes [Frenzel, 1989].

La clave de los sistemas expertos es el conocimiento, de ahí, que a veces, se denominen (knowledge-based systems) sistemas basados en el conocimiento. Por tanto, la diferencia entre una base de datos y una base de conocimientos es importante. Mientras que un registro [Frenzel, 1989, p. 45] de una base de datos es una unidad de información que contiene hechos y figuras, más que conocimientos, una base de conocimientos contiene fragmentos individuales del conocimiento. Una forma habitual de expresar el conocimiento es a través de «reglas». Las reglas tienen la forma (IF-THEN)SI-ENTONCES. La primera parte de la regla va precedida

por la palabra SI seguida de la condición. La segunda parte viene guiada de ENTONCES y establece la acción o respuesta. Algunas de las reglas que se forman son las siguientes:

Si aplicamos los sistemas expertos a la automatización de la indización para la solución del problema semántico de la sinonimia, una regla simple sería:

1. SI préstamo entre bibliotecas ENTONCES PRÉSTAMO
INTERBIBLIOTECARIO

Otras reglas son:

2. SI en un párrafo 'universidad'
Y en el mismo párrafo 'enseñanza'
ENTONCES ENSEÑANZA UNIVERSITARIA

3. SI programas de ordenador
O programas informáticos
ENTONCES SOFTWARE

Los sistemas expertos se han utilizado en Documentación de modo experimental en aplicaciones bibliotecarias como la catalogación y los servicios de referencia [Gibb, 1986* ; Anderson, 1988 ; Cavanagh, 1989*]; en la recuperación de información [Pollitt, 1981* ; Shoval, 1985* ; Belkin, 1988* ; Alberico y Micco, 1990]; y en la automatización de la indización [Martínez, Lucey y Linder, 1987* ; Humphrey, 1989 y 1994 ; Kuntz, 1991 ; Driscoll, et al. 1991 ; Schuegraf y Bommel, 1993].

2.5 LA AUTOMATIZACIÓN DE LA INDIZACIÓN PARA INFORMACIÓN NO TEXTUAL.

A lo largo del capítulo hemos estudiado la automatización de la indización dirigida a la información textual, pero en este epígrafe mencionaremos diversas líneas de investigación abiertas durante la década de los ochenta encaminadas a la indización de imágenes y de sonido de forma automática. Tradicionalmente, se ha utilizado la

* Estas referencias no se han consultado directamente, sino que se localizaron durante el proceso de revisión bibliográfica y de consulta de diferentes bases de datos.

información textual para recuperar la información de las bases de datos, por lo que con este tipo de iniciativas se pretende que otra clase de información como son las imágenes y los sonidos se convierta en clave tanto de almacenamiento como de recuperación.

Las iniciativas dirigidas a interpretar el sonido de un modo automático tienen utilidad no sólo para recuperar esta clase de información de una base de datos sino también se ha empleado en las prospecciones a gran profundidad para el descubrimiento de yacimientos petrolíferos. Un trabajo para facilitar el acceso, el tiempo y el esfuerzo para seleccionar un sonido en una base de datos lo han llevado a cabo Feiten y Gunzel [1994].

Por otro lado, en la indización semiautomática o automática de imágenes hay que tener presente aspectos como la composición, el nivel de reconocimiento y las diferentes interpretaciones que surgen tras la contemplación de los objetos. A estos problemas hay que sumar la complejidad para determinar y representar convenientemente las relaciones semejantes entre los objetos porque varían las estructuras de las imágenes. Propuestas en este sentido han sido las de Bordogna et al. [1990] y Rabitti y Savino [1992].

2.6 EL NIVEL DE IMPLANTACIÓN DE SISTEMAS PARA LA AUTOMATIZACIÓN DE LA INDIZACIÓN.

Hodge publicó un informe en 1992 sobre el estado real de la aplicación de sistemas para la automatización de la indización de documentos. Posteriormente, este informe lo extractó para difundir el estado de la cuestión en un Congreso [Hodge, 1993] y en un artículo [Hodge, 1994]. Para la elaboración de este estudio entró en contacto con los responsables de veintitrés destacadas entidades públicas y privadas como institutos especializados, Centros de Documentación, Bibliotecas especializadas, empresas de servicios de información (indización y resumen), y con productores de bases de datos.

Los distintos sistemas para la automatización de la indización se dividen según lo expuesto en el informe en:

a) Asistencia en la indización y en el control de la calidad.

Estas posibilidades incluyen la localización de un candidato a término de indización pero, principalmente, ayudas en la introducción de los términos en las bases de datos.

Un indizador puede recurrir a diferentes ayudas durante la fase de análisis para localizar un término de indización:

Consulta en línea de un vocabulario controlado.

Acceso a notas explicativas donde se especifica la utilización de un descriptor determinado.

Selección de un término en lenguaje natural del título o el resumen de un documento para que el sistema ofrezca su término controlado del tesoro.

Acceso en línea a documentos previamente indizados para la aclaración de algún aspecto.

Asignación automática de códigos a los nombres de organismos que aparecen en los títulos de los documentos.

Introducción de códigos mnemotécnicos que permiten visualizar el desarrollo completo de nombres de empresas o instituciones.

Asimismo, se emplean diferentes métodos para facilitar y reducir el trabajo en la introducción de los términos de indización de un documento en las Bases de datos:

Selección de los descriptores desde pantallas de ordenador con teclas de función o con el ratón.

Cuando a un grupo de documentos se le asignan un mismo conjunto de términos de indización, cabe la posibilidad de asignarlos automáticamente a todos los documentos, sin teclear cada uno de ellos para conseguir una mayor rapidez y consistencia.

Por último, para garantizar la consistencia de los términos de indización introducidos en las bases de datos, se emplean los ficheros de autoridades. Estos ficheros contienen los términos de indización validados. Los controles se llevan a cabo bien por lotes o bien en línea. En el primer caso, cada cierto tiempo o según un número determinado de documentos analizados, se comprueba, de forma automática, que los términos de indización utilizados están en el fichero de autoridades. El control en línea, supone la validación de los términos en el momento de su introducción en la base de datos por la consulta al fichero de autoridades. En algunos sistemas, cuando se localiza un término invalidado se propone automáticamente un término autorizado.

b) Sistemas automáticos para el proceso de análisis.

Los indizadores disponen de sistemas expertos basados en reglas para analizar los títulos y los resúmenes de los documentos para asignarles términos desde un tesoro. Otros utilizan sistemas expertos basados en correlaciones estadísticas entre los términos en lenguaje natural de los documentos y los términos de indización de los documentos ya indizados. En otros sistemas, los indizadores pueden revisar los documentos almacenados en la base de datos para aclarar dudas durante la indización, cuando les surgen problemas de especificidad, o por el desconocimiento de siglas o de abreviaturas utilizadas por un autor. De este modo, se consiguen mayores índices de consistencia, se mantiene la política de indización en la institución y se logra mayor calidad.

c) Sistemas de indización automática

Varias instituciones utilizan la indización automática para el análisis de los documentos, según los datos ofrecidos en el informe de Hodge. En este caso, la expresión «indización automática» se refiere a indización sin ningún tipo de validación ni de validación por parte de los profesionales. El mismo autor parece sorprendido del número de instituciones que utilizan estos sistemas debido al esfuerzo requerido para su construcción. No obstante, en todos los casos, la indización automática se aplica a áreas o temas concretos, y a determinados tipos de documentos. Veamos brevemente qué instituciones emplean sistemas completamente automáticos:

En el Centro de Información técnica de Defensa de los Estados Unidos utilizan un programa para la seleccionar series de palabras (de una a cuatro) de los títulos y de los resúmenes con un diccionario de reconocimiento. Este diccionario está compuesto por más de ciento treinta mil términos simples y frases. Una vez seleccionados los términos y frases de los textos se comparan con un tesoro para transformar el lenguaje natural en descriptores.

El Centro de Documentación Karlsruhe alemán proporciona la indización a diferentes Bases de datos. En este Centro se utilizan dos sistemas diferentes para la automatización de la indización. Un sistema automático obtiene nombres de elementos químicos de fórmulas localizadas en el texto. En cambio, en el sistema AIR/PHYS se indizan documentos sobre Física a partir de los títulos y los resúmenes.

El sistema AIR/PHYS se basa en un diccionario que contiene relaciones entre los términos en lenguaje natural de las fuentes analizadas y los términos de indización asignados. La relación se establece por medio de un factor de asociación de acuerdo a estimaciones estadísticas sobre la probabilidad que tiene un descriptor de ser asignado a un documento si en su título o resumen contiene el término en lenguaje natural asociado a dicho descriptor.

Estas relaciones se consiguieron tras la indización de casi cuatrocientos mil documentos que dieron lugar a un número elevado de relaciones entre términos en lenguaje natural y descriptor, de las que en el sistema solamente se utilizaron trescientas cincuenta mil. El sistema actúa del siguiente modo: el texto se descompone en palabras, frases y oraciones, y un algoritmo reduce cada palabra a la forma estándar (un verbo a su infinitivo y un plural al singular). Posteriormente, el texto resultante se compara con el diccionario para detectar las relaciones entre los términos del texto y los descriptores. A continuación, se ponderan las relaciones según las estimaciones de probabilidad para asignar los términos de indización al documento¹³.

¹³ Para comprobar el grado de calidad del sistema se estudiaron los primeros veinte mil documentos indizados. Se obtuvo que el 19% de los documentos presentaba una buena indización, la indización para el 64% fue de gran utilidad a excepción de pequeñas correcciones, mientras que para el 13% de los documentos la indización contenía numerosos errores.

En el Centro Internacional para la Información Científica y Técnica rusa se han indizado automáticamente informes técnicos y científicos desde 1978 a 1991. El sistema utiliza un analizador morfológico y otro sintáctico para el procesamiento de títulos y resúmenes. Para este procesamiento se utilizan varios ficheros de términos controlados, de desinencias utilizadas en el análisis morfológico, de palabras vacías, y de abreviaturas como «Fig.» o «Bibliogr.». De acuerdo a estos ficheros los términos significativos se extraen de los textos y se comparan con el tesoro.

Por último, en la Base de datos INSPEC que trabaja con documentos de Física, Electrónica y Computación, se han reindizado automáticamente más de tres millones de registros contenidos en la base de datos desde 1969 a 1989.

En definitiva, como se ha comprobado la automatización de la indización comienza a estar presente en diversas instituciones, pero, en grado y en forma diferente. En todos los casos, son sistemas construidos por los mismos profesionales de las entidades, de modo muy arbitrario, y según las necesidades, los recursos, el entorno de la indización y el tipo de documentos manejados. Por tanto, la revisión de cada uno de los casos presentados en el informe de Hodge nos ha permitido constatar que las entidades:

Analizan principalmente informes técnicos o artículos científicos.

Tratan áreas concretas del conocimiento como Física, Biología, Medicina, Arte o Química.

Toman como fuentes de análisis los títulos y los resúmenes de los documentos.

Todas las instituciones utilizan un vocabulario controlado para la automatización de la indización.

2.7 ESQUEMA REPRESENTATIVO DE LAS HERRAMIENTAS UTILIZADAS EN LA AUTOMATIZACIÓN DE LA INDIZACIÓN.

HERRAMIENTAS	SISTEMAS
Métodos Lingüísticos	SMART ; INDEXD ; CLARIT ; SIMPR ; SPIRIT ; PASSAT ; ALETH ; DARWIN ; ALEXDOC ; INDEXICON ; Valle Bracero y Fernández García [1983] ; Simón Granda y Lema Garzón [1990] .
Métodos Estadísticos	Luhn [1957] ; Damerou [1965] ; SMART ; SHAPIRE ; CLARIT ; SPIRIT ; PASSAT ; SINTEX.
Análisis Probabilísticos	Rosenberg [1971] ; Robertson y Harding [1984] ; AIR/PHYS.
Vocabulario Controlado	Strode [1977] ; Dillon [1982] ; Scheele [1983] ; Valle Bracero y Fernández García [1983] ; Martínez, Lucey y Linder [1987] ; Lovtsov [1990] ; Ciganik [1990] ; Coret, Ducloy, Menillet [1991] ; Wan [1995] ; Sistema NASA ; AIR/PHYS ; CLARIT ; SHAPIRE ; PASSAT ; ALETH ; SINTEX ; ALEXDOC.
Fichero Palabras Vacías	SMART ; SIMPR ; Sistema NASA ; Lovtsov [1990] ; Robredo [1991].
Fichero expresiones idiomáticas	SPIRIT ; ALEXDOC.
Fichero de siglas	DARWIN
Normalización de Términos	SIMPR ; SPIRIT ; Valle Bracero y Fernández García [1983].
Autoreenvío de Conceptos	SMART ; SHAPIRE ; SINTEX ; ALEXDOC.
Validación de Términos	Rosenberg [1971] ; Martínez, Lucey y Linder [1987] ; Simón Granda y Lema Garzón [1990] ; Sistema NASA ; SIMPR ; ALETH ; INDEXICON ; SINTEX ; ALEXDOC.

Este tercer capítulo se dedica a especificar la propuesta para la automatización de la indización en el área de Biblioteconomía y Documentación. También se explica la metodología adoptada y las fuentes utilizadas. Por último, se lleva a cabo un ensayo de modo manual donde se aplica el algoritmo diseñado para conseguir los índices de consistencia con respecto a indizadores profesionales.

3.- PROPUESTA TEÓRICO-METODOLÓGICA PARA LA AUTOMATIZACIÓN DE LA INDIZACIÓN

3.1 INTRODUCCIÓN.

A lo largo del primer capítulo de este trabajo se llevó a cabo una exposición conceptual para enmarcar la indización en el proceso documental, se expuso qué es, cuáles son las etapas sucesivas acometidas durante esta tarea y para qué sirve la indización. También se mostró cuáles son sus principales características, los diferentes sistemas de indización utilizados, así como las distintas herramientas empleadas en la indización de los documentos. Todos estos aspectos se explicaron porque se consideraban básicos para la contextualización del marco teórico de esta tesis doctoral.

En el segundo capítulo, dedicado a los aspectos de la automatización de la indización, se trató de delimitar aún más el campo de actuación. Para ello, comenzamos con la problemática alrededor de la automatización de esta operación por parte de profesionales e investigadores del área. Posteriormente, en otro paso, se mostraron las principales metodologías adoptadas para la automatización de la indización, desde finales de los cincuenta hasta la actualidad. Con la presentación de los diferentes métodos empleados se expusieron los principales inconvenientes y problemas que circundan a cada uno de los métodos de indización semiautomática o automática. Por otro lado, se analizaron diferentes ensayos dirigidos a la evaluación de los sistemas automáticos con respecto a la indización propuesta por indizadores profesionales. Y por último, se patentizó la interdisciplinariedad presente en la automatización de la indización.

Llegados a este punto conviene señalar otro de los objetivos marcados al inicio de este trabajo, esto es, la presentación de una propuesta para la automatización de la indización en el área de Biblioteconomía y Documentación.

A continuación, a la vez que formulamos nuestra propuesta, justificamos cada una de las elecciones. De este modo, razonamos las fuentes utilizadas para extraer los términos de indización aportados por el sistema, y el por qué basamos nuestra propuesta en un vocabulario controlado. También, una vez que hayan sido

presentados los diferentes mecanismos para llevar a cabo esta tarea realizaremos una evaluación del sistema. Para ello, lo aplicaremos, manualmente, a cincuenta artículos de revista, y comprobaremos los índices de consistencia resultantes con respecto a los documentos analizados por indizadores profesionales del CSIC. Y al mismo tiempo, detectaremos posibles problemas y errores con la finalidad de mejorarlo.

3.2 LA ELECCIÓN DE LAS FUENTES UTILIZADAS EN LA AUTOMATIZACIÓN DE LA INDIZACIÓN.

Cabe mencionar que la indización semiautomática o automática toma varias fuentes sobre las que actuar: títulos, resúmenes y textos. Desde finales de los años cincuenta, cuando se concibieron los primeros sistemas automáticos para la creación de índices a partir de los títulos, el empeño de algunos investigadores consistió en demostrar el valor de los títulos de los artículos científicos como único recurso para la indización. Asimismo, el afán de otros investigadores residió en evidenciar que los títulos de los artículos de las ciencias sociales aportaban menos carga informativa, con propósitos de indización y recuperación, que los de las ciencias experimentales o viceversa.

3.2.1 Los antecedentes de estos estudios.

Seguidamente, mostramos una serie de investigaciones dirigidas a constatar la validez de los títulos y los resúmenes de los artículos científicos para la indización de los documentos. Bastos Vieira [1988], en su trabajo ya citado de revisión bibliográfica sobre la indización automática, presentaba brevemente algunos ensayos encaminados a contrastar la validez de estas fuentes con el objetivo marcado:

Maizell [1960] acreditó uno de los primeros estudios sobre la importancia de los títulos de los artículos, y más concretamente sobre artículos de *Physics Abstracts*. De dichos análisis dedujo que el 63% de los títulos contenía información suficiente para la indización.

Montgomery y Swanson [1962] obtuvieron resultados más elevados sobre los títulos del *Index Medicus*. Llegaron a la conclusión de que el 86% de los

títulos poseía un contenido suficientemente significativo para que fuera utilizado en la indización automática.

Ruhl [1964] en publicaciones del área de *Química*, encontró que el 57% de los títulos contenía todos los descriptores asignados en el Chemical Abstracts, y solamente el 12% de los títulos no incluía tres o más palabras significativas.

Kraft [1964], respecto a títulos de *Derecho*, evidenció que el 64% de los mismos englobaba palabras significativas, mientras que tan sólo el 10% no entrañaba ninguna palabra existente en el índice *Index to Legal Periodicals*.

Por último, Svenonius [1972] y Feinberg [1973] afirmaron, apoyados en sus respectivos estudios, que la indización obtenida de los títulos ostentaba una precisión mayor que con la explotación del resumen o del texto.

Por otro lado, hemos identificado otros investigadores [Salton, 1972 ; Dijk y Slype, 1972 ; Kwok, 1974 ; Braga, 1982 ; Salton y McGill 1983] que consideraron los títulos como fuentes insuficientes para la transmisión completa de la carga informativa de los documentos. Por estas razones, juzgaban necesario apoyarse también en los resúmenes. Veamos algunos ejemplos:

Braga, por ejemplo, realizó un estudio relativo a los títulos de artículos científicos de Química. Para ello, aplicó la técnica de índices KWIC para contraponer las palabras de los títulos y de los resúmenes con tablas de frecuencia. Y explicó que los títulos, a pesar de ser puntos de acceso al contenido de los documentos, resultaban insuficientes para una perfecta indización y recuperación, ya que la proporción de palabras significativas del resumen era del orden de doce por cada una del título.

En otro estudio efectuado en Euratom, mencionado en [Dijk y Slype, 1972, p. 105], se defendía que los títulos, de promedio, dotaban de bastantes menos términos de indización que los resúmenes.

Weinberg [1981] estudió sesenta y cinco artículos y sus resúmenes sobre Ingeniería civil. Comprobó que el 23% de los términos de indización no aparecía en los resúmenes sino en el texto; que el 44% sólo se manifestaba una vez en los

resúmenes; y que el 34% de los términos solamente se materializaba en los resúmenes.

D.B. Cleveland, A.D. Cleveland y Wise [1984]* indizaron de forma automática los títulos, los resúmenes y los textos para cerciorarse de las fuentes o combinaciones que ofrecían mejores resultados en la recuperación. Tras diversos ensayos con las fuentes combinadas concluyeron que la indización automática con los títulos o los resúmenes proporcionaba índices de exhaustividad y precisión semejantes a los obtenidos por la indización del texto completo.

Con los trabajos anteriores se ha brindado un número considerable de experimentos en defensa de una o varias fuentes. Ahora resta desplegar algunas de las propuestas para la automatización de la indización desarrolladas para mostrar las opciones preferidas por los investigadores:

Entre los sistemas que se nutren a partir de los títulos y de los resúmenes destacan: Klingbiel y Rinker [1976] ; Meulen y Janssen [1977] ; Barnes, Costantini y Perschke [1978] ; Dillon [1983] ; Martínez, Lucey y Linder [1987] ; Biebricher [1988] ; Evans et al., [1991] ; Robredo [1991] ; Hersh et al. [1992] ; Silvester, Genuardi y Klingbiel [1994].

La segunda opción elegida es la utilización de los resúmenes [Salton, 1972 ; Britvin, 1981 ; Seo, 1993 ; Hmeidi, 1995 ; Wan, 1995].

En menor medida, otros prototipos emplearon sólo los títulos [Valle Bracero y Fernández García, 1983 ; Scheele, 1983], mientras que otros los textos [Haller, 1982 ; Andreewsky y Ruas, 1982 ; Haller, 1983].

3.2.2 Ensayo en las Bases de datos del CSIC.

Ante este panorama de diversidad, parece obligado afrontar un ensayo propio a partir del cual, tomar una decisión adecuada a nuestra propuesta. Para ello, analizamos referencias de artículos de distintas áreas del conocimiento:

- Humanidades (Biblioteconomía y Documentación y Psicología).
- Áreas con un grado mayor de experimentación (Biología y Medicina).
- Áreas más experimentales (Química y Física).

Todas estas disciplinas y ciencias están presentes en las Bases de datos ISOC, IME o ICYT del Consejo Superior de Investigaciones Científicas.

3.2.2.1 Material y métodos.

El método seguido para el ensayo fue el siguiente. De cada una de las áreas seleccionadas se localizaron de una a tres revistas científicas que publicaran trabajos en estos ámbitos¹. A continuación, se obtuvieron de las diferentes Bases de datos mencionadas setenta y cinco registros que cumplieran alguna de estas dos condiciones: contener el campo Resumen -imprescindible para los análisis que se deseaban acometer- o bien que se tuviera acceso a la fuente en papel, pero a la vez que contuvieran el resumen del artículo.

Este proceso nos proporcionó para cada área del conocimiento setenta y cinco referencias de artículos con sus respectivos títulos, resúmenes y descriptores asignados por los profesionales que analizan esta información. En definitiva, se obtuvieron cuatrocientos cincuenta registros con 2077 descriptores.

A partir de aquí se comenzó a buscar manualmente en cada uno de los registros los descriptores asignados tanto en los títulos como en los resúmenes. Sólo se seleccionaron aquellos descriptores idénticos en esas zonas. Con esto se pretende expresar que cuando se encontraba por ejemplo, «análisis bibliométrico» y como descriptor «estudio bibliométrico», «préstamo entre bibliotecas» y como descriptor «préstamo interbibliotecario», o «recuperación de información» y como descriptor «recuperación de la información» no se contabilizaban.

* Estos trabajos también vienen referenciados en Bastos Vieira [1988].

¹ Área de Biblioteconomía y Documentación: *Revista Española de Documentación Científica*, *Jornadas Españolas de Documentación Automatizada* (celebradas en 1994) y el *Boletín de la Asociación Andaluza de Bibliotecarios*.

Área de Medicina: *Actas Urológicas Españolas*, *Oncología* y *Endocrinología*.

Área de Química: *Anales de Química*.

Área de Biología: *Anales de Biología* y *Monografías de Flora y Vegetación Béticas*.

Área de Psicología: *Investigaciones Psicológicas*, *Anales de Psicología* y *Anuario de Psicología*.

3.2.2.2 Resultados.

Los resultados de estos análisis se muestran en la siguiente tabla:

Áreas	Total descrip. asignados	Descrip en Título	%	Descrip. en Resumen	%	Descrip. en Título y Resumen	Total descrip. en Título o Resumen	%	Descrip. no presentes en Título o Resumen
Bib-Doc.	396	39	9,8	62	15,6	86	187	47,2	209
Medicina	186	47	25,2	10	5,3	47	104	55,9	82
Química	410	19	4,6	58	14,1	76	153	37,3	257
Biología	330	33	10	49	14,8	43	125	37,8	205
Psicología	395	13	3,2	72	18,2	46	131	33,1	264
Física	360	17	4,7	41	11,3	34	92	25,5	268
Total	2077	168	8	292	14	332	792	39,4	1285

Del análisis de los resultados obtenidos podemos expresar que:

1. Los títulos de los artículos de Medicina son los que albergan más cantidad de descriptores, pero es necesario considerar que también posee el número más bajo de descriptores asignados a los setenta y cinco registros analizados. Por tanto, si reparamos en esta circunstancia, el área que presenta los títulos más significativos, de cara a la indización, es la de Biblioteconomía y Documentación. Por el contrario, la que menos es Psicología seguida de Química.
2. El área con los resúmenes más significativos es Psicología seguida de Biblioteconomía y Documentación. En cambio, Medicina y Física son las que menos descriptores han presentado en los resúmenes.
3. Los títulos y los resúmenes de los artículos de Medicina son los que más descriptores acogen. Hay que observar de nuevo, que esta área es la que menos descriptores tiene asignados en los registros revisados. A continuación, aparecen las áreas de Biblioteconomía y Documentación y Biología.
4. Los resúmenes de los documentos, siempre y cuando estén bien elaborados, son fuentes que proporcionan más conceptos con posibilidad de convertirse en

descriptores que los títulos. Si bien es cierto que hemos tenido la oportunidad de leer resúmenes de muy mala calidad.

Una vez concluido el ensayo nos propusimos averiguar qué número de términos aportaban los textos de los documentos. Para tal objetivo, leímos, minuciosamente, los setenta y cinco artículos ya analizados de Biblioteconomía y Documentación. Esta lectura procuró otros 98 términos de indización no presentes en el título ni en el resumen.

Veamos de nuevo la tabla con los datos conseguidos:

Área	Descrip. asignados	Descrip. en Título o Resumen	%	Descrip. solamente en texto	%	Descrip. en Título, Resumen y Texto	%
Bib-Doc.	396	187	47,2	98	24,7	285	71,9

Se observa que el análisis de los textos aporta el 24,7% de los términos asignados como descriptores. Por tanto, según estos datos, la explotación de los textos para labores de indización no parece ser una postura caprichosa, sino más bien ineludible. En consecuencia, trasladada esta conclusión a la intención de automatizar la indización, parece razonable que los textos deben ser una fuente más, pero siempre y cuando la metodología utilizada no sea excesivamente compleja y no consuma demasiados recursos y tiempo.

3.2.2.3 Conclusiones.

En definitiva, la utilización de los títulos, los resúmenes y los textos se deben considerar para la automatización de la indización por las siguientes razones:

1. Según los datos precedentes, parece justificado la incorporación de los textos para fines de la indización porque aporta el 24,7% de los descriptores asignados a los artículos analizados. Y los títulos, los resúmenes y los textos han contribuido con el 71,9% de los descriptores asignados. Además, recordemos que, durante el ensayo, consideramos como términos diferentes por ejemplo: «análisis bibliométrico» y «estudio bibliométrico», «préstamo entre bibliotecas» y «préstamo interbibliotecario» o «recuperación de información» y «recuperación de

la información». Esto significa que con simples relaciones de equivalencia el número de descriptores localizados en estas fuentes será superior.

2. Para tomar como únicas fuentes de análisis los títulos y los resúmenes hay que considerar que los títulos sean lo bastante precisos, y que los resúmenes están bien elaborados, cuestión difícil de predecir. Además, hay revistas españolas de Biblioteconomía y Documentación que no parece que obliguen a los autores a la presentación de un resumen del trabajo que van a publicar. Publicaciones como la *Revista General de Información y Documentación*, *AABADOM*, *Information World* o *Métodos de Información* no ofrecen los resúmenes de los artículos en la mayor parte de las ocasiones. Por otro lado, la presentación y metodología de los resúmenes varía enormemente. Tras este estudio, podemos afirmar que la revista que más cuida la forma y contenido de los resúmenes es la *Revista Española de Documentación Científica*.

3.3 LA ELECCIÓN DE UN VOCABULARIO CONTROLADO.

A lo largo del segundo capítulo hemos constatado que los métodos estadísticos aplicados en la automatización de la indización de los documentos no solucionan problemas importantes, tales como el reconocimiento de sinonimias y la selección de términos compuestos, principalmente.

Por otro lado, la utilización del PLN supone una serie de inconvenientes como: a) la carencia de una biblioteca básica de programas informáticos de libre distribución para el español como, etiquetadores morfológicos o sintácticos, por lo que resulta necesario construirlos para tal fin; b) en el caso de disponer de este tipo de herramientas, a los datos resultantes tras los procesamientos del lenguaje natural hay que aplicar bien cálculos estadísticos o bien procesarlos contra un vocabulario controlado; c) incluso si se adopta esta metodología es necesario asumir que el PLN requiere una importante cantidad de recursos y de tiempo para ejecutar los procesos.

Por estas limitaciones se propone una metodología para la automatización de la indización fundamentada en la comparación de patrones admitidos desde un vocabulario controlado. Los patrones admitidos son todos los términos incluidos en el vocabulario controlado, ya sean términos preferentes o no preferentes, con la

salvedad de que los términos propuestos finalmente por el sistema para convertirse en descriptores, serán sólo términos preferentes.

3.3.1 La justificación de la elección de un vocabulario controlado.

A continuación, se justifica la utilización de un vocabulario controlado para la automatización de la indización.

1. El valor de la terminología en los textos científico-técnicos.

“Empecemos con una pregunta ingenua, pero también páfida: ¿De qué trata el siguiente texto?”:

Se hace un planteamiento general de la problemática de la ##### en relación con la ##### contenida en los #####. Se exponen los principales problemas que presenta la ##### de las entradas en los ##### y se analizan los nuevos ##### elaborados por la #####, la ##### y la #####. Finalmente, se describen los intentos llevados a cabo por dos organismos internacionales, ##### e #####, para lograr un mayor grado de ##### en las ##### y ##### teniendo en cuenta la utilización de ##### en el ##### y el intercambio internacional de los mismos.

Con las palabras precedentes entrecomilladas comenzaba Picht [1996, p. 10], un apartado de su reciente artículo sobre Terminología. En nuestro caso hemos cambiado el texto, pero sin embargo, el efecto es el mismo. Tras su lectura no se intuye nada del tema que trata ni del contenido del mismo. A continuación se reproduce el texto sin ‘censuras’:

“Se hace un planteamiento general de la problemática de la ordenación de los asientos bibliográficos en relación con la recuperación de la información contenida en los catálogos. Se exponen los principales problemas que presenta la ordenación de las entradas en los catálogos automatizados y se analizan los nuevos códigos elaborados por la Library of Congress, la British Library y la Library Association. Finalmente, se describen los intentos llevados a cabo por dos organismos internacionales, IFLA e ISO, para lograr un mayor

grado de uniformidad en las reglas de ordenación de catálogos y bibliografías teniendo en cuenta la utilización de ordenadores en el tratamiento de los datos bibliográficos y el intercambio internacional de los mismos”².

Este mismo ensayo sobre textos de otra área puede provocar resultados aún más desconcertantes. Veamos dos ejemplos:

- Ya sabemos que la ##### o ##### (#### y #####) en la ##### da lugar al #####, primera ##### del nuevo ser con #####. A partir de ahí comienza su ##### (por #####) que origina un número creciente de ##### mientras que el ##### realiza su descenso por la #####.

“Ya sabemos que la unión de los gametos o células sexuales (óvulo y espermatozoide) en la fecundación da lugar al cigoto, primera célula del nuevo ser con dotación cromosómica diploide. A partir de ahí comienza su Segmentación (por división mitótica) que origina un número creciente de células mientras que el óvulo fecundado realiza su descenso por la trompa de Falopio”.

- Las ##### no pueden explicarse por un único ##### o #####. La ##### probables es que la ##### da lugar a #####, y la ##### da lugar a #####. A su vez, la ##### y la ##### producen un incremento de la ##### y de la #####, con aumento de la ##### y de la #####.

“Las manifestaciones clínicas no pueden explicarse por un único factor fisiológico o bioquímico. La secuencia patogénica probable es que la hiperglucemia materna da lugar a hiperglucemia fetal, y la respuesta pancreática del feto da lugar a hiperinsulinemia fetal. A su vez, la hiperglucemia y la hiperinsulinemia fetal producen un incremento de la tasa de

² Este texto corresponde al resumen de un artículo publicado en la Revista Española de Documentación Científica, 18, 3, 1995, p. 283.

glucosa hepática y de la síntesis de glucógeno, con aumento de la lipogénesis y de la síntesis proteica”.³

Según se desprende de estos textos la terminología:

Constituye un requisito primordial para la existencia y el funcionamiento de un lenguaje específico.

Representa una condición esencial para la comunicación entre los profesionales.

Cumple una función representativa y cognitiva.

Proporciona los fundamentos conceptuales de los textos científico-técnicos.

Por tanto, si capturamos la terminología presente en los textos tenemos acceso a los conceptos comunicados, lo que nos permite conocer los temas y los contenidos tratados en los documentos.

2. Un vocabulario controlado que incluya relaciones de equivalencia es una buena herramienta para detectar los sinónimos en las fuentes. Esta posibilidad ya ha sido explotada tanto para la recuperación de la información como en la automatización de la indización en los sistemas SMART [Salton, 1980] o SHAPIRE [Hersh, 1992].

3. Como se comentó en el capítulo anterior, a partir de ensayos con el sistema SMART, se comprobó que los índices de precisión y exhaustividad en la recuperación son similares a la indización de profesionales si el sistema de indización automática emplea un tesoro para reconocer términos sinónimos.

4. En el informe de Hodge [1992], sobre el estado de la automatización de la indización en instituciones públicas y privadas (Bibliotecas especializadas, institutos de investigación, Centros de Documentación y empresas productoras de bases de

³ Se trata de dos fragmentos de los apuntes de una asignatura de la Facultad de Biología de la Universidad de Murcia.

datos o de servicios de indización o resúmenes), se ha constatado que las veintitrés entidades analizadas utilizan un vocabulario controlado, independientemente del grado de automatización de la indización conseguido.

Por todas estas razones consideramos que un sistema de indización semiautomático o automático debe fundamentarse en un vocabulario controlado. Además, la elección de esta posibilidad viene respaldada por otros investigadores que ya la adoptaron para sus propuestas como: Strode [1977], Dillon [1982], Scheele [1983], Valle Bracero y Fernández García [1983], Martínez, Lucey y Linder [1987], Biebricher, et al. [1988], Lovtsov [1990], Coret, Ducloy y Menillet [1991], Ciganik [1990], los sistemas PASSAT, SINTEX y ALEXDOC [Slupe, 1991], Evans [1991], Hersh et al. [1992], Silvester, Genuardi y Klingbiel [1994] o Wan [1995].

3.3.2 La elaboración de la lista de términos autorizados.

Como hemos referido, resulta relevante el papel jugado por la terminología para la indización. Por tanto, en el empeño por construir una metodología para acometer esta tarea de modo automático, se ha tropezado con la ausencia de un amplio vocabulario controlado (instrumento terminológico) sobre Biblioteconomía y Documentación. Al inicio de esta investigación solamente conocíamos los siguientes tesauros para el español:

ROZO RÍOS, D.S., GARZÓN HERRÁN, M.E. (1980). Tesauro en Documentación e información. Bogotá: FID.

TESAURO de la Unesco. (1982). París: Unesco.

PENICHE DE SÁNCHEZ MACGREGOR, S. (1992). Vocabulario controlado en Bibliotecología, Ciencia de la Información y temas afines. México: Universidad Nacional Autónoma de México.

La primera obra recoge 560 términos. La terminología presente en el tesauro está más cercana a la concepción y uso del lenguaje de América latina que de España. Por otro lado, la segunda referencia fue concebida para la indización y la recuperación de documentos en el Servicio Informatizado de Documentación de la Unesco. Las áreas abarcadas son los ámbitos de actuación de esta organización: desde aspectos generales a Educación, Cultura y Humanidades, Ciencia y

Tecnología, Comunicación y Ciencias sociales. Pero dedicaron una parte a la Información, Bibliotecas y Archivos. Y por último, la tercera obra contiene 1400 términos entre descriptores y no descriptores.

Ante esta falta de herramientas útiles para nuestros objetivos nos propusimos la elaboración de un vocabulario controlado (lista de términos autorizados) sobre Biblioteconomía y Documentación. En primer lugar, contactamos con los indizadores del área de Biblioteconomía y Documentación de la Base de Datos ISOC del Consejo Superior de Investigaciones Científicas. Y nos reiteraron la carencia de este tipo de herramientas para nuestro campo. A raíz de esta visita nos facilitaron un listado con los descriptores utilizados en la indización de dicha Base de datos.

Como quiera que las obras arriba mencionadas resultaban insuficientes para proveernos de terminología, consultamos otras relacionadas como diccionarios terminológicos, bilingües y de acrónimos. Los trabajos en cuestión fueron:

BUONOCORE, D. (1976). Diccionario de Bibliotecología: términos relativos a la bibliología, bibliografía, bibliofilia, biblioteconomía, archivología, documentología, tipografía y materias afines. Buenos Aires: Marymar.

DICCIONARIO de Información y Documentación (1989). Madrid: ICYT.

MARTÍNEZ DE SOUSA, J. (1993). Diccionario de Bibliología y Ciencias afines. Madrid: Fundación Germán Sánchez Ruiperez.

LOZANO PALACIOS, A. (1993). Vocabulario inglés-español español-inglés para los estudios de Biblio-Documentación. Granada: Universidad.

DICCIONARIO de acrónimos en información y documentación (1994). Madrid: CINDOC.

En definitiva, con estas obras, el material manejado y la terminología propia, elaboramos una lista de términos autorizados del área de Biblioteconomía y Documentación compuesto por 1995 términos, de los cuales 1362 son descriptores y 633 no descriptores. Este vocabulario constituye la base del proyecto que

proponemos para la constitución de un sistema para la automatización de la indización. En el anexo 8 se reproduce la lista íntegra de términos autorizados.

A continuación se muestran los veinte primeros términos de la lista de términos autorizados como aparecen en el anexo y posteriormente, como deben disponerse para su consulta por el algoritmo.

1. 3W USE WORLD WIDE WEB
2. AACR USE REGLAS DE CATALOGACION
3. ABREVIATURAS
4. ABSTRACTS USE RESUMENES
5. ACCESIBILIDAD DE LA INFORMACION

6. ACCESIBILIDAD UNIVERSAL A LA INFORMACION USE DISPONIBILIDAD UNIVERSAL DE LAS PUBLICACIONES

7. ACCESO A BASES DE DATOS
8. ACCESO A LA INFORMACION
9. ACCESO A LOS LIBROS
10. ACCESO AL DOCUMENTO
11. ACCESO REMOTO
12. ACCESO UNIVERSAL A LAS PUBLICACIONES USE DISPONIBILIDAD UNIVERSAL DE LAS PUBLICACIONES

13. ACERVO USE COLECCIONES
14. ACERVO BIBLIOGRAFICO USE FONDOS BIBLIOGRAFICOS
15. ACERVO DOCUMENTAL USE FONDOS DOCUMENTALES
16. ACM USE ASOCIACIONES PROFESIONALES
17. ACREDITACION PROFESIONAL
18. ACRODABA USE BASE DE DATOS DE ACRONIMOS
19. ACRONIMOS
20. ACTAS DE CONGRESOS

- 3W USE WORLD WIDE WEB
- AACR USE REGLAS DE CATALOGACION
- ABREVIATURAS

ABSTRACTS USE RESUMENES

ACCESIBILIDAD

INFORMACION

UNIVERSAL

INFORMACION USE DISPONIBILIDAD UNIVERSAL DE LAS
PUBLICACIONES ACCESO

BASES

DATOS

DOCUMENTO

INFORMACION

LIBROS

ACCESO

REMOTO

UNIVERSAL

PUBLICACIONES USE DISPONIBILIDAD UNIVERSAL DE LAS
PUBLICACIONES

ACERVO USE COLECCIONES

BIBLIOGRAFICO USE FONDOS BIBLIOGRAFICOS

DOCUMENTAL USE FONDOS DOCUMENTALES

ACM USE ASOCIACIONES PROFESIONALES

ACREDITACION

PROFESIONAL

ACRODABA USE BASE DE DATOS DE ACRONIMOS

ACRONIMOS

ACTAS

CONGRESOS

3.4 PROPUESTA PARA LA AUTOMATIZACIÓN DE LA INDIZACIÓN.

La automatización de la indización que proponemos se fundamenta en la selección desde las fuentes (títulos, resúmenes y textos) de los patrones admitidos incluidos en el vocabulario controlado. El proceso se lleva a cabo en tres módulos diferentes:

En el primer módulo se prepara el texto para los análisis posteriores por medio de:

La señalización de las fuentes.

La horizontalización de las frases y oraciones comprendidas entre signos de puntuación.

La eliminación de las palabras vacías de los textos.

La contabilización de las palabras de las fuentes.

En el segundo módulo se seleccionan los términos candidatos a descriptores a través de un mecanismo de comparación de patrones admitidos. Se compara cada uno de los términos preferentes y no preferentes, contenidos en el vocabulario controlado, con las fuentes. También hay un proceso para la selección de palabras que no estén contenidas en la lista de términos autorizados ni en el fichero de palabras vacías.

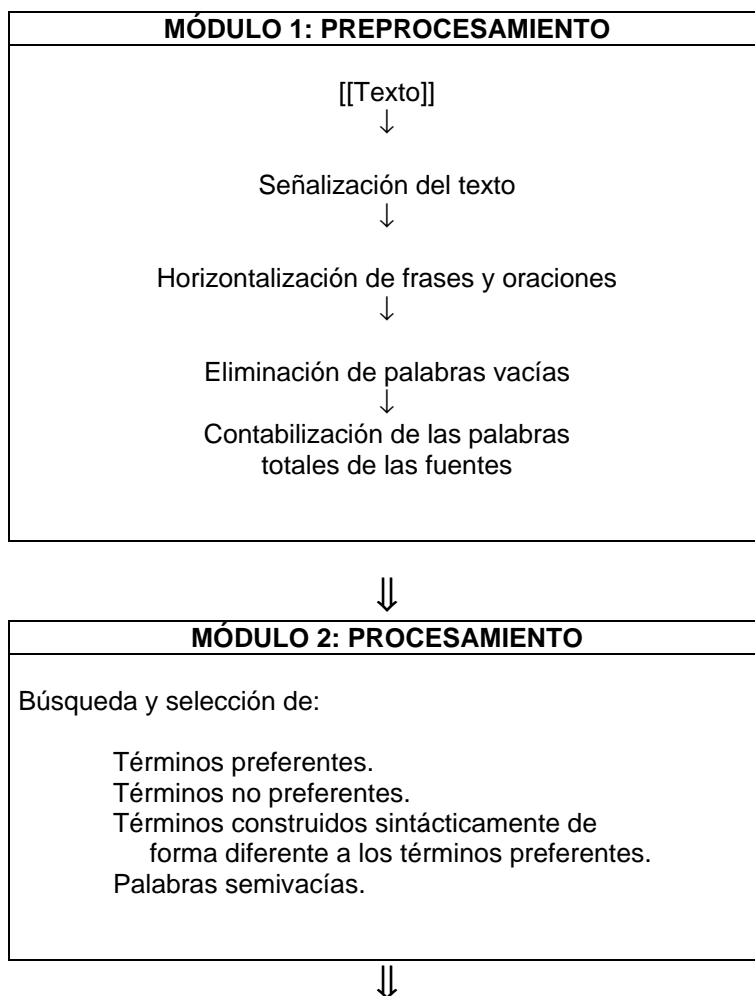
El tercer módulo acoge los términos seleccionados en el módulo anterior para valorar y ponderar los términos candidatos.

El sistema se compone de una Base de conocimientos en donde se almacenan los objetos y las relaciones entre ellos. Esta Base de conocimientos la concebimos como una red conceptual que presenta vínculos entre unos conceptos en lenguaje natural y otros en lenguaje controlado. Los objetos son los conceptos representados por el vocabulario controlado del área de Biblioteconomía y Documentación. Las relaciones entre los objetos establecen la correspondencia entre los términos, tanto preferentes como no preferentes.

Por último, el programa utiliza el análisis inferencial para localizar similitudes entre las fuentes y la Base de conocimientos. El programa, a través de los procesos de

búsqueda y comparación de patrones, genera una relación de términos preferentes con indicación de dónde se localizan (T= título, R= resumen y T= texto), que permanece opaca al usuario. Posteriormente, en virtud de unos principios de valoración y ponderación establecidos, el programa origina dos listados: uno con los términos de indización propuestos y otro con las palabras que no aparecen en la lista de términos autorizados ni en el fichero de palabras vacías, para que el usuario decida si se convierten en descriptores. Al final, se muestran los resultados al usuario y se inicia la fase de validación y edición de la indización propuesta por el sistema.

A continuación, se muestra de forma esquematizada cada uno de los pasos presentados:





MÓDULO 3: VALORACIÓN
<p>Aplicación de los principios de valoración:</p> <p>A. Si un término autorizado aparece en la fuente título y en la fuente resumen se convierte en término de indización.</p> <p>B. Si un término autorizado aparece en la fuente título y en la fuente texto se convierte en término de indización.</p> <p>C. Si un término autorizado aparece en la fuente resumen y en la fuente texto se convierte en término de indización.</p> <p>D. Si una palabra semivacia aparece en el título, el resumen y el texto se presenta al usuario para su posible incorporación como término de indización.</p> <p>E. Si una palabra semivacia aparece en el texto diez veces o más y además, en ocho párrafos o más diferentes y no está incluida en ninguno de los términos propuestos se presenta como palabra semivacia.</p>



PRESENTACIÓN DE LOS RESULTADOS PARA SU VALIDACIÓN Y EDICIÓN		
<p>Título: #####</p> <p>Resumen: ##### ##### #####.</p>		
TÉRMINOS PROPUESTOS		PALABRAS SEMIVACÍAS PROPUESTAS
	<p>AÑADIR</p> <p>SUPRIMIR</p> <p>←</p>	

3.4.1 Los módulos del sistema.

A continuación explicamos los módulos que componen el sistema, y se detallan los diferentes procesos acometidos en cada uno de ellos.

3.4.1.1 Módulo 1: Preprocesamiento.

Esta primera etapa tiene por objeto la preparación de las fuentes que van a ser utilizadas para la selección de términos, es decir, se trata de la preparación de las fuentes para su procesamiento. Como ya hemos señalado, las fuentes para el análisis son los títulos, los resúmenes y los textos. El preprocesamiento se lleva a cabo en diferentes fases, una vez que las fuentes están disponibles en formato legible por ordenador. Veamos estas fases:

1 Señalización del texto.

El primer estadio es la señalización de las siguientes partes de las fuentes:

- Comienzo y fin de título.
- Comienzo y fin de resumen.
- Comienzo de texto.
- Fin de texto.
- Comienzo y fin de párrafos.

Para ello, utilizaremos las siguientes marcas:

#CTi# : Comienzo de título.

#FTi# : Fin de título.

#CR# : Comienzo de resumen.

#FR# : Fin de resumen.

#CTe# : Comienzo de texto.

#FTe# : Fin de texto.

#CP# : Comienzo de párrafo.

#FP# : Fin de párrafo.

El objeto de marcar cada una de las partes mencionadas es el siguiente. La señalización del título, del resumen, del comienzo del texto y de los párrafos se lleva a cabo porque en el momento de la valoración y ponderación de los términos candidatos a descriptores es necesario conocer el lugar de donde se extrajeron. Las marcas **#Cti#**, **#Fti#**, **#CR#**, **#FR#** y **#Cte#** se colocan de forma manual⁴ mientras que las de **#Fte#**, **#CP#** y **#FP#** se ponen de modo automático.

2. Horizontalización de las frases.

Este proceso consiste en colocar de forma horizontal aquellas frases u oraciones comprendidas entre estos signos de puntuación: puntos, comas, puntos y comas, y dos puntos. De este modo, se persigue facilitar los procesos acometidos sobre los textos. Este método de horizontalización de frases u oraciones se ha empleado para fines semejantes en el sistema de indización semiautomático de la NASA, en SAPHIRE, en CLARIT y en la propuesta de Simón Granda y Lema Garzón.

Veamos como queda horizontalizado el siguiente fragmento:

El PLN sigue siendo una disciplina desconocida para investigadores y, sobre todo, para profesionales del área de Biblioteconomía y Documentación.

1. El PLN sigue siendo una disciplina desconocida para investigadores y,
2. sobre todo,
3. para profesionales del área de Biblioteconomía y Documentación.

3 Eliminación de las palabras vacías.

En esta fase se suprimen aquellas palabras que en ninguna circunstancia llegarán a ser candidatas para convertirse en términos de indización. Nos referimos a palabras como artículos, adverbios, preposiciones, conjunciones o pronombres. En

⁴ Si observamos la estructura que presentan los artículos de revista comprobamos que por lo general, los títulos aparecen en mayúsculas, con letra grande y en negrita. En cambio, los resúmenes suelen ir precedidos de la palabra «Resumen» y el tamaño de la letra es más pequeño que la del texto. Por tanto, con esta información se podría señalar el comienzo y fin de cada fuente de forma automática.

consecuencia, la supresión de lo que se conoce en el ámbito de la Biblioteconomía y Documentación como «palabras vacías» queda justificado por tres motivos:

1. Se busca economizar en los procesos acometidos por el sistema. En este sentido, el descarte de las palabras vacías provoca que disminuya el número de palabras a procesar.
2. Este primer aspecto, irremediablemente, se traduce en un menor tiempo de análisis. Uno de los objetivos generales que se persigue con la automatización de la indización es que el tiempo empleado por el programa sea similar o inferior al de un profesional.
3. Resulta ventajoso no contar con este tipo de palabras en la etapa de búsqueda de conceptos contruidos de forma diferente respecto a los términos autorizados.
- 4 Para los textos de diferentes áreas temáticas e idiomas se ha comprobado que, aproximadamente, el cincuenta por ciento de las palabras manejadas son palabras de este tipo [Silvester, Genuardi y Klingbiel, 1994, p. 632 ; Simón Granda y Lema Garzón, 1983, p. 1260].

Por estas razones, utilizamos el denominado fichero de palabras vacías. Estos ficheros han sido empleados en otras propuestas:

En SMART, se empleaba un listado de palabras clasificadas en orden decreciente a su valor de discriminación para excluir todas las de frecuencia muy alta [Salton, 1980, p. 29].

En el proyecto SIMPR definieron diferentes ficheros de palabras vacías, en concreto tres: palabras vacías como “among”, “been”, “keep” o “nothing”; palabras referidas a la estructura de un documento como “appendix”, “annex”, “equation”, “footnote” o “reference”; y aquellas que muestran la posición o estado de un objeto como “inner”, “outer” o “upper” [Karetnyk, Karlsson y Smart, 1991, p. 22].

En la propuesta de Robredo [1991, p. 131], se utilizaba un fichero con palabras vacías del tipo: “abaixo”, “cada”, “comparada”, “demais”, “duzentos”, “enquanto”, “fora”, “enfrente”, entre otras.

En el sistema de indización semiautomático de la NASA [Silvester, Genuardi, Klingbiel, 1994, p. 637] constituyeron un fichero con doscientas cincuenta palabras vacías tras el análisis estadístico de textos formados por cuatro millones de palabras. Las palabras vacías son del tipo “about”, “are”, “efficient”, “investigated”, “other” o “within”.

El fichero de palabras vacías que hemos construido está compuesto por 274 palabras obtenidas de la descomposición en palabras simples de:

- Conjunciones como: “a pesar de”, “de tal manera que”, “no obstante”.
- Adverbios como: “a posteriori”, “después de”, “cada vez más”.
- Preposiciones como: “a cambio de”, “con motivo de”, “gracias a”.

Las conjunciones, los adverbios y las preposiciones se han conseguido de varios trabajos dedicados a la morfosintaxis del español [Carratalá, 1980 ; Almela, 1986]. En el anexo 5 se presenta el listado completo de las palabras vacías.

La eliminación de estas palabras provoca que términos preferentes como ‘BASES DE DATOS DE TEXTO COMPLETO’ o ‘DISEÑO ASISTIDO POR ORDENADOR’, queden reducidos a “bases datos texto completo” y “diseño asistido ordenador”. Y una vez finalizada la indización por el programa, si a los términos propuestos les fueron suprimidas algunas palabras, se restablecen automáticamente.

Por tanto, las frases y oraciones horizontalizadas en el ejemplo anterior:

1. El PLN sigue siendo una disciplina desconocida para investigadores y, sobre todo,
2. para profesionales del área de Biblioteconomía y Documentación. quedan del siguiente modo:

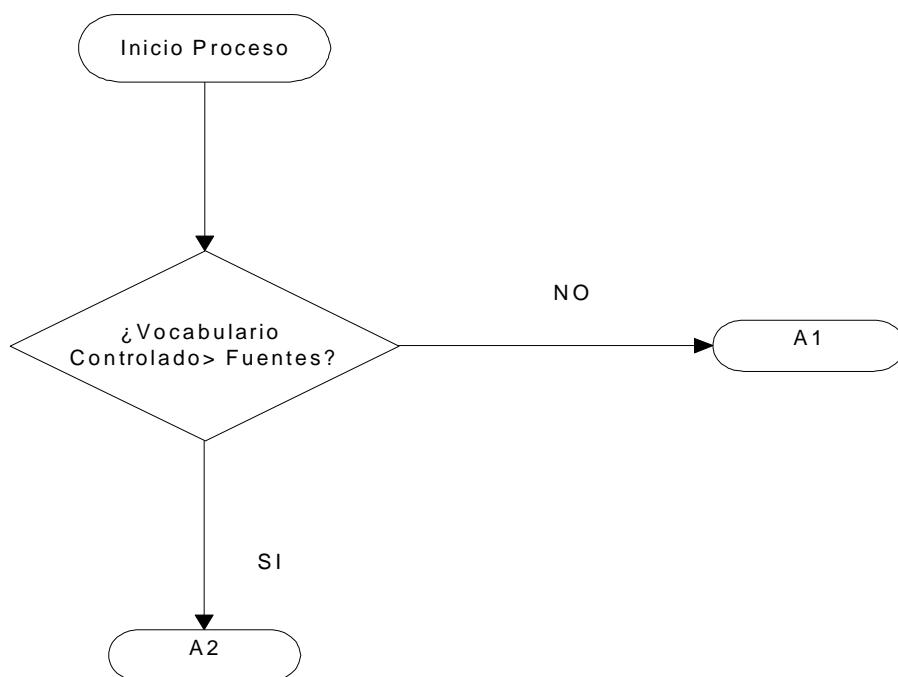
1. PLN sigue siendo disciplina desconocida investigadores,
2. profesionales área Biblioteconomía Documentación.

Como se constata, la eliminación de las palabras vacías reduce el número de palabras, y en ocasiones, de frases u oraciones a procesar. Esta circunstancia repercute positivamente en los procesos de búsqueda y de selección acometidos.

La decisión de elegir este orden, es decir, primero la horizontalización de las frases y oraciones y después, la eliminación de las palabras vacías se debe a que, en el proceso de selección de candidatos es preciso presentar las palabras semivacías en su contexto, para que el usuario compruebe su validez para asignarlas como términos de indización.

4. Contabilización del número de palabras de las fuentes.

Una vez calculado el número de palabras de las fuentes se decide cual de los dos algoritmos de procesamiento se utiliza. El algoritmo 1 se emplea cuando el vocabulario controlado es menor que las fuentes, mientras que el algoritmo 2 se utiliza cuando el vocabulario controlado es mayor que las fuentes. De este modo, se optimiza el proceso y el tiempo de ejecución.



3.4.1.2 Módulo 2: Procesamiento

En la automatización de la indización, tradicionalmente, se han seleccionado los candidatos a términos de indización de distintos modos:

1. Los sistemas basados íntegramente en métodos estadísticos eliminaban las palabras vacías y calculaban la frecuencia de aparición de éstas en los documentos y en la colección. Después, ponderaban las palabras en razón de los umbrales establecidos, su frecuencia inversa de aparición y, su valor de discriminación.
2. Posteriormente, se incorporaron técnicas del PLN que han consistido en la selección de estructuras sintagmáticas tras el análisis morfológico y el sintáctico de los textos del tipo: sustantivos, sustantivo + Sintagma Adjetivo, o sustantivo + Sintagma Preposicional. A estas estructuras candidatas se aplicaban cálculos estadísticos para distinguir qué términos se convertían en términos de indización.
3. También se comparaba el texto con un vocabulario controlado y a los términos presentes en ambos lugares se le aplicaban cálculos estadísticos para seleccionar los términos de indización.

En nuestra propuesta la idea central se dirige directamente a la selección desde las fuentes (títulos, resúmenes y textos) de los términos candidatos a descriptores. Para ello, contamos con mecanismos para buscar, seleccionar, valorar y ponderar los patrones admitidos incluidos en la lista de términos autorizados y hallados en las fuentes. Los términos candidatos a descriptores se extraen a través de un proceso de comparación de patrones admitidos. Un algoritmo busca y selecciona términos preferentes, términos no preferentes y términos contruidos sintácticamente de forma diferente a los términos preferentes.

En cuanto a la posibilidad de que una palabra aparezca en minúscula o mayúscula ('lenguajes de indización' o 'LENGUAJES DE INDIZACIÓN'), con o sin tilde ('gestión de la información' o 'gestion de la informacion'), con o sin diéresis ('lingüística' o 'linguística'), o incluso palabras en singular o plural ('publicación electrónica' o 'publicaciones electrónicas'), tendrán que tenerse en cuenta con el fin de que se consideren iguales.

Vamos a comprobar cómo se lleva a cabo cada una de las búsquedas y selecciones, pero en primer lugar, mostraremos el esquema general de los algoritmos 1 y 2 respectivamente.

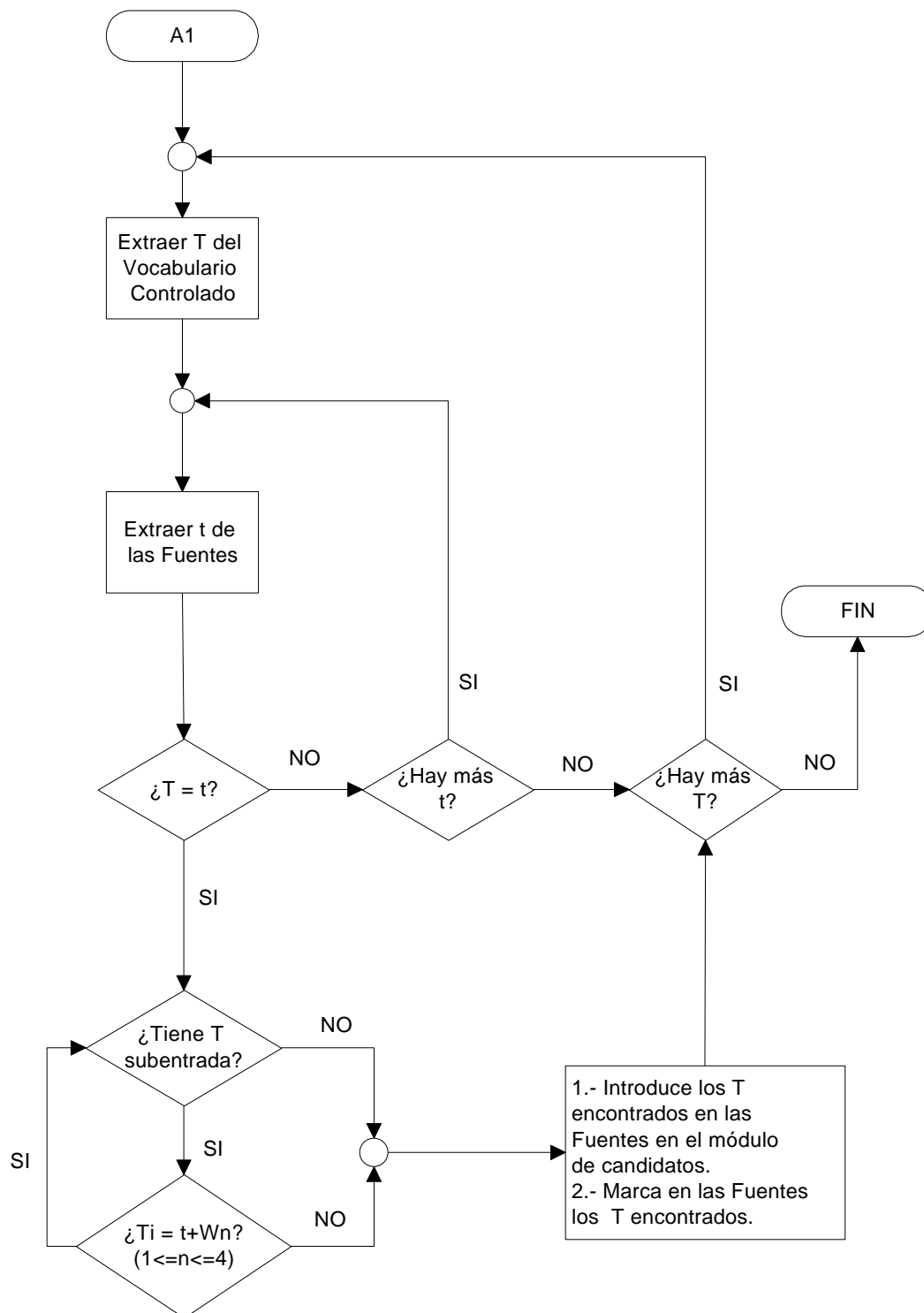
El algoritmo A1 se utiliza cuando el tamaño de las fuentes es mayor que el del vocabulario controlado⁵. El algoritmo A1 trabaja del siguiente modo:

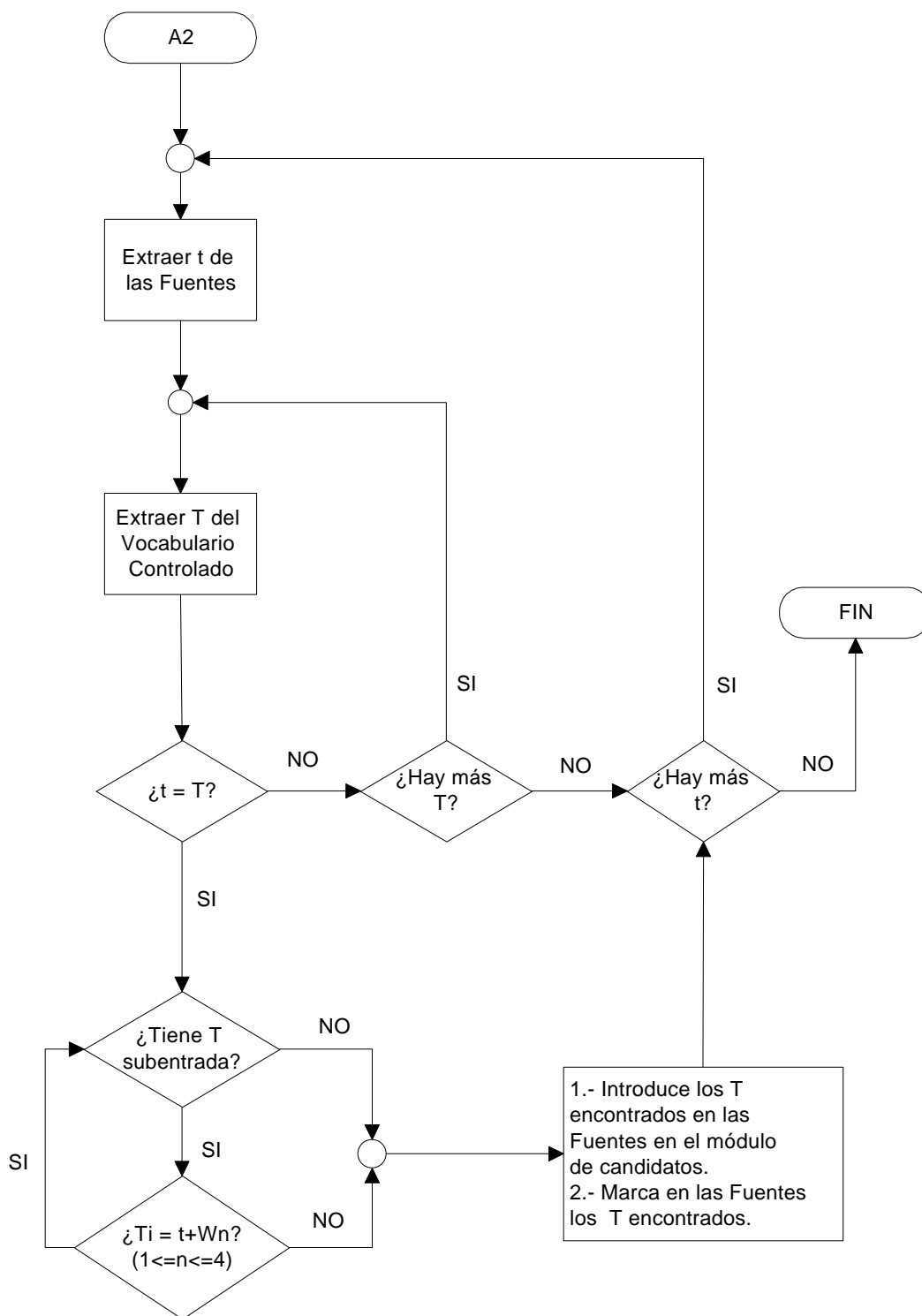
- 1.- Extrae el término de la primera entrada de la lista de términos autorizados.
2. Extrae la primera palabra de la fuente.
3. Comprueba si las dos palabras extraídas son iguales.
4. Si no son iguales se comprueba si hay más palabras (t) en las fuentes.

Si no hay más palabras, se comprueba si hay más términos en la lista de términos autorizados (T). Si no hay más términos finaliza el proceso.

En el caso de que hubiera más palabras (t) en la fuente, se extrae la siguiente, y se vuelve a comparar la palabra de la fuente (t) con el término de la lista de términos autorizados (T). Y se repite el mismo proceso hasta que sean iguales o no haya más palabras (t) en las fuentes.

⁵ El procedimiento de trabajo en el A2 es semejante al A1 a excepción de que la primera palabra extraída es de la fuente en lugar del vocabulario controlado, por tanto, el mecanismo de extracción y comparación es inverso al de A1.





5. Si la palabra de la fuente (t) y el término de la lista de términos autorizados (T) son iguales, se comprueba si la entrada del vocabulario controlado utilizada tiene subentradas.

- Si no tiene subentrada, se comprueba si existe relación de equivalencia del término (T) de la entrada principal, y el término preferente se envía al módulo de candidatos. También se marca para que no se tenga en cuenta si se vuelve a localizar.
- Si tiene subentrada, se comprueba si coincide con alguna de las cuatro palabras siguientes de las que hay a la derecha de la última palabra (t) localizada en la fuente.

6. Si alguna de las cuatro palabras (t) coincide con el término (T) del vocabulario controlado, se comprueba si T tiene más subentradas y se repite el mismo proceso. En el caso de que no tenga más subentradas o ninguna de las cuatro palabras (t) de la fuente coincide con T, se comprueba si existen relaciones de equivalencia y se traslada/n el/los término/s preferente/s al módulo de candidatos y se marca/n el/los localizado/s.

7. Se comprueba si hay más T en el vocabulario controlado:

- Si no hay más palabras (T) finaliza el proceso
- Si hay más palabras (T) continúa el proceso desde 2.

A continuación, se explica detalladamente cada uno de los procesos que es capaz de llevar a cabo este algoritmo:

1 Búsqueda y selección de términos preferentes.

Las fuentes han sufrido diversos procesos como hemos visto anteriormente. En uno de ellos, las frases y oraciones comprendidas entre los signos de puntuación han sido horizontalizadas, y en otro, se han eliminado las palabras vacías. Ahora interviene el algoritmo de procesamiento para la búsqueda y la selección de los patrones admitidos en las fuentes.

En la columna de la izquierda de la siguiente tabla se encuentra la fuente que se va a procesar, y en la parte derecha se reproducen unos fragmentos de la lista de términos autorizados.

<u>FUENTE: Texto</u>	<u>LISTA DE TÉRMINOS AUTORIZADOS</u>
1. recogidos actas notariales 1856.	ACTAS CONGRESOS
2. adquisición gestiona ordenador,	... ADQUISICION ORDENADOR
3. mejores servicios son alerta informativa.	... SERVICIOS ALERTA INFORMATIVA

Para descubrir en las oraciones o frases horizontalizadas términos preferentes o sinónimos ('camuflados') con diferentes estructuras sintácticas se procede de la siguiente manera:

1. Se extrae la primera palabra de la lista de términos autorizados. En este caso, 'ACTAS'.
2. Posteriormente, se extrae la primera palabra de la fuente (t). En este caso 'recogidos'.
3. Se comprueba si son palabras iguales.

Como no son iguales, se comprueba si hay más t. En este caso se extrae t 'actas'. Se comprueba de nuevo si el término extraído de la fuente ('actas') es igual al T ('ACTAS') de la lista de términos autorizados, t y el T son iguales.

4. Se comprueba si T ('ACTAS') tiene subentrada. T ('ACTAS') tiene la subentrada ('CONGRESOS').

5 Se comprueba si la subentrada ('CONGRESOS') aparece en alguna de las cuatro palabras⁶ siguientes a la t ('actas'), es decir, ('notariales') y ('1856'). En este caso no ha aparecido.

6 Se introduce el T encontrado en la fuente ('ACTAS') en el módulo de candidatos.

7 Se marca en la fuente el T ('ACTAS') encontrado.

8 Se comprueba si hay más T en la lista de términos autorizados. Si no los hubiera finaliza este proceso, y si los hubiera se extrae el siguiente T y se extrae la primera t de la fuente con lo que se inicia de nuevo el proceso.

2 Búsqueda y selección de términos no preferentes.

El proceso de búsqueda y selección de términos no preferentes es idéntico al descrito anteriormente pero esta vez con los términos sinónimos. Esto significa que si en una fuente se localiza alguno de estos términos no preferentes, el término introducido en el módulo de candidatos no es el localizado sino el preferente.

Veamos otro ejemplo:

<u>LISTA DE TÉRMINOS AUTORIZADOS</u>	<u>FUENTE: Texto</u>
...BIBLIOTECOLOGIA USE BIBLIOTECONOMIA	procesamiento lenguaje natural
Y	trata analizar textos programas
DOCUMENTACION	diseñados objetivo,
...PROCESAMIENTO	tareas Bibliotecología formación
LENGUAJE	específica.
NATURAL USE ANALISIS	
AUTOMATICO	
DE TEXTOS	
...	

⁶ Un método similar para la comparación de palabras próximas se utiliza en el sistema de indización semiautomático de la NASA. Allí comprobaron empíricamente que el límite de proximidad más conveniente era de cinco palabras con el fin de no obtener cadenas de palabras inapropiadas. Pero en nuestro caso el límite está establecido en cuatro porque con el fichero de palabras vacías utilizado eliminamos palabras como "a", "de", "desde", "el", "para", "por", "y" que son conservadas en el programa norteamericano.

Según lo expresado, al localizar en la fuente 'procesamiento lenguaje natural' en el módulo a candidatos se introduce 'ANALISIS AUTOMATICO TEXTO' y lo mismo ocurre con 'Bibliotecología', que queda depositado 'BIBLIOTECONOMIA Y DOCUMENTACION'. Este mismo proceso se produce con términos no preferentes como 'diseminación información' o 'préstamo (entre) bibliotecas' que pasarían a estar presentes en el módulo de candidatos como 'DIFUSION INFORMACION' y 'PRESTAMO INTERBIBLIOTECARIO' respectivamente.

De modo que el módulo a candidatos mostraría en estos momentos la siguiente situación:

BIBLIOTECONOMIA DOCUMENTACION-Re
ANALISIS AUTOMATICO TEXTOS-Te
BIBLIOTECONOMIA DOCUMENTACION-Te

3 Búsqueda y selección de términos contruidos sintácticamente de modo diferente a los términos del vocabulario controlado.

A los sistemas basados en los métodos estadísticos y a los fundamentados en el uso de vocabularios controlados se ha achacado que cuando un término no aparece escrito del mismo modo en el texto, no son capaces de detectarlos. Esto se debe a la riqueza del lenguaje natural para representar un mismo concepto con palabras diferentes, y del mismo modo, también posibilita pequeñas variaciones desde el punto de vista sintáctico, aunque persista el mismo significado.

En un texto podemos encontrar las oraciones 'la adquisición se efectúa por ordenador' o 'el material es cartográfico' que encierran los términos preferentes incluidos en la lista de términos autorizados 'ADQUISICION POR ORDENADOR' y 'MATERIAL CARTOGRAFICO'. En las líneas donde aparecen estas oraciones se hace referencia a los conceptos representados por dichos términos del vocabulario controlado, pero con estructuras sintácticas diferentes.

Para paliar esta situación, cuando en una fuente se encuentra un t igual a un T, se comprueba si la subentrada del T coincide con alguna de las cuatro palabras siguientes a la t de la fuente.

4 Búsqueda y selección de términos no incluidos en el vocabulario controlado ni en el fichero de palabras vacías.

Como se ha explicado, el sistema permite buscar términos preferentes, no preferentes y términos con estructura sintáctica diferente a los preferentes. Todos estos términos quedaron marcados cuando pasaron a ser candidatos. De este modo, en las frases y oraciones horizontalizadas, podríamos ver las palabras no recogidas en el fichero de palabras vacías ni en la lista de términos autorizados. A estas palabras las hemos denominado 'palabras semivacías'.

El objetivo perseguido con esta fase es que no quede una palabra simple o compuesta sin tener la oportunidad de convertirse en descriptor, por el hecho de que no conste en el vocabulario controlado.

Las palabras semivacías pertenecientes al título pasan automáticamente al módulo de palabras semivacías. Cada una de estas palabras actuará de 'patrón admitido' con respecto a las otras, es decir, se procesarán confrontándolas con los resúmenes y los textos para comprobar si se han utilizado también en estas fuentes. En este módulo se llevará a cabo su procesamiento y su valoración de acuerdo con los principios establecidos para este grupo de palabras.

Vamos a comprobar cómo se procede con el siguiente título y resumen⁷:

Título: Errores ortográficos en el ingreso en bases de datos.

Resumen: Se estudian los problemas de la corrección ortográfica en el ingreso de registros en bases de datos en idioma español. Se evalúan los pros y contras de cuatro métodos de control: doble entrada, hapax legomena, trigramas y uso de diccionarios, con vistas a determinar cuales de estos procedimientos ofrecen mejor relación de costo/resultado. El trabajo está enfocado a los procesos de ingreso por digitación, y no se analizan los errores ortográficos de los ingresos por lectura óptica.

A continuación mostramos las palabras semivacías de las fuentes título y resumen en cada una de sus frases y oraciones horizontalizadas:

Título:

1. errores ortográficos ingreso

Resumen:

1. estudian problemas corrección ortográfica ingreso registros idioma español.
2. evalúan pros contras control:
3. doble entrada
4. hapax,
5. legomena,
6. trigramas
7. uso,
8. vistas determinar procedimientos ofrecen relación resultado.
9. trabajo enfocado procesos ingreso digitación, analizan errores ortográficos ingresos lectura óptica.

A estas palabras semivacías se les aplican unos criterios de valoración que se explican en el siguiente apartado.

3.4.1.3 Módulo 3: Valoración y ponderación.

Esta tercera etapa tiene por objeto la selección de los términos que el sistema propone como términos de indización para representar el contenido del documento analizado. Para ello es preciso la aplicación de un método para la valoración de los mismos. Este mecanismo se requiere porque de lo contrario, el programa designa como términos de indización todos aquellos que, contenidos en el vocabulario controlado, aparecen en las fuentes procesadas.

El párrafo mostrado a continuación pertenece a un artículo indizado en la Base de datos ISOC. Los términos en **negrita** forman parte de los descriptores asignados al documento. En cambio, los términos subrayados no han sido seleccionados como descriptores.

⁷ El título y el resumen pertenece a un artículo publicado en la Revista Española de Documentación Científica, 18, 3, 1995, p. 307-319

“Con esta nueva tendencia de la instrucción bibliográfica, que en español se ha dado en llamar “**formación de usuarios**”, aparece la polémica entre los dos posibles objetivos de un servicio de documentación: la función de referencia, en cuyo caso el **bibliotecario** asume un papel de intermediario, frente a la función de formación, que mediante la enseñanza de patrones de trabajo, dotaría al usuario de los recursos necesarios para encontrar por sí mismo la información requerida.”

El ejemplo que sigue es aún más claro. Ninguno de los términos subrayados fueron seleccionados como descriptores para representar el contenido del documento al que pertenece este párrafo:

“En un servicio de documentación podemos encontrar dos tipos de usuarios claramente diferenciables. De una parte iniciado en informática y telecomunicaciones que se desenvuelve con soltura en el manejo de bases de datos y sólo precisa de una formación propiamente documental.”

Por ello, es necesario un mecanismo de valoración y/o ponderación porque de lo contrario, se propondrían todos los términos del vocabulario controlado que fueran encontrados en las fuentes.

Antes de iniciar este proceso de valoración se eliminan los términos que puedan estar repetidos en las mismas fuentes. La selección de los términos que son propuestos como descriptores se efectúa en virtud de los siguientes criterios de valoración:

- A) Si un término autorizado aparece en la fuente título y en la fuente resumen se convierte en término de indización.
- B) Si un término autorizado aparece en la fuente título y en la fuente texto se convierte en término de indización.
- C) Si un término autorizado aparece en la fuente resumen y en la fuente texto se convierte en término de indización.

D) Si una palabra semivacía aparece en el título, resumen y texto se presenta al usuario para su posible incorporación como término de indización.

E) Si una palabra semivacía aparece en el texto diez veces o más y además, en ocho párrafos⁸ diferentes o más, y no está incluida en ninguno de los términos propuestos, se presenta como palabra semivacía.

La elección del criterio A y B es debida a la importancia de los títulos y los resúmenes en la indización de los documentos. El criterio C es para dar importancia a términos que aparezcan en el texto. Y por último, los criterios D y E se aplican a las palabras semivacías y posibilita que un término que no aparece en el vocabulario controlado ni en el fichero de palabras vacías tenga la opción de convertirse en término de indización.

Si recordamos, uno de los pasos iniciales en la primera etapa (módulo de preprocesamiento) fue la eliminación de las palabras vacías para agilizar los procesos. Esto propició que durante el tratamiento de las fuentes se trabajara con los términos tal como aparecen aquí:

Términos durante el proceso

ADQUISICION ORDENADOR
CONTROL VOCABULARIO
SERVICIOS ALERTA INFORMATIVA

Presentación definitiva

Antes de que el sistema muestre los términos de indización propuestos para representar el contenido del documento analizado, se han de normalizar de nuevo

⁸ La contabilización de los párrafos en los que aparece una palabra es para evitar que, en uno o varios párrafos se trate un tema determinado y se utilice una palabra en bastantes ocasiones, pero en el resto del texto no se haga referencia a ella.

con la incorporación de las palabras eliminadas. En este ejemplo quedaría del siguiente modo:

ADQUISICION POR ORDENADOR
CONTROL DEL VOCABULARIO
SERVICIOS DE ALERTA INFORMATIVA

3.4.2 Análisis de un documento en función de la propuesta precedente.

A continuación, mostramos las fuentes utilizadas para este análisis. Estas fuentes deben estar en formato legible por ordenador, y deben eliminarse los gráficos, las tablas, y demás información no textual, en el caso de que fuera necesario.

Título

Análisis de los descriptores de diferentes áreas del conocimiento indizadas en bases de datos del CSIC. Aplicación a la indización automática

Resumen

Se estudia el valor de los títulos y resúmenes de los artículos científicos como fuentes suministradoras de términos para la indización de los documentos en seis áreas del conocimiento indizadas en las Bases de datos ISOC, IME e ICYT del CSIC. Asimismo, se examina la estructura sintagmática de los términos de indización hallados en el campo "Descriptores", y la posible relación entre el número de descriptores de un documento con la cantidad de páginas del mismo. Para tales fines se seleccionaron las áreas del conocimiento de Biblioteconomía y Documentación, Medicina, Química, Biología, Psicología y Física. Y se realizaron seis búsquedas en estas Bases de datos de las que seleccionamos 450 referencias bibliográficas (75 por área) proporcionando un total de 2077 descriptores.

Texto

1 Introducción. La indización es una operación compleja pero esta dificultad se torna doble cuando se intenta obtener de forma automática. Mediante la indización automática se pretende que sea un algoritmo el que proponga todos los términos de indización tras el análisis de un documento o algunas de sus partes. Tradicionalmente, en la indización automática se han venido utilizando dos métodos distintos pero a veces convergentes en algunos ensayos, esto es, medios no lingüísticos, iniciados a finales de los cincuenta, y lingüísticos incorporados posteriormente.

Cuando se pretende diseñar un sistema de indización automática basado en la extracción de conceptos uno de los planteamientos inmediatos es decidir si las fuentes de las que lograr los términos candidatos a descriptores, serán los documentos completos o los títulos y resúmenes de los mismos.

MÓDULO 1: PREPROCESAMIENTO

1 Señalización de las fuentes.

#CTi#Análisis de los descriptores de diferentes áreas del conocimiento indizadas en bases de datos del CSIC. Aplicación a la indización automática**#FTi#**

#CR#Se estudia el valor de los títulos y resúmenes de los artículos científicos como fuentes suministradoras de términos para la indización de los documentos en seis áreas del conocimiento indizadas en las Bases de datos ISOC, IME e ICYT del CSIC. Asimismo, se examina la estructura sintagmática de los términos de indización hallados en el campo "Descriptores", y la posible relación entre el número de descriptores de un documento con la cantidad de páginas del mismo. Para tales fines se seleccionaron las áreas del conocimiento de Biblioteconomía y Documentación, Medicina, Química, Biología, Psicología y Física, y se realizaron seis búsquedas en estas Bases de datos de las que seleccionamos 450 referencias bibliográficas (75 por área) proporcionando un total de 2077 descriptores.**#FR#**

#CTe##CP#1 Introducción. La indización es una operación compleja pero esta dificultad se torna doble cuando se intenta obtener de forma automática. Mediante la indización automática se pretende que sea un algoritmo el que proponga todos los términos de indización tras el análisis de un documento o algunas de sus partes. Tradicionalmente, en la indización automática se han venido utilizando dos métodos distintos pero a veces convergentes en algunos ensayos, esto es, medios no lingüísticos, iniciados a finales de los cincuenta, y lingüísticos incorporados posteriormente.**#FP#**

#CP#Cuando se pretende diseñar un sistema de indización automática basado en la extracción de conceptos uno de los planteamientos inmediatos es decidir si las fuentes de las que lograr los términos

candidatos a descriptores serán los documentos completos o los títulos y resúmenes de los mismos. **#FP##FTe#**

2 Horizontalización de frases y oraciones.

El siguiente paso es la horizontalización de cada una de las frases y oraciones comprendidas entre los signos de puntuación (coma, punto, dos puntos o punto y coma). En este caso las fuentes quedan de esta forma:

#Cti#

1 Análisis de los descriptores de diferentes áreas del conocimiento indizadas en bases datos CSIC.

2 Aplicación a la indización automática.

#Fti#

#CR#

1 Se estudia el valor de los títulos y resúmenes de los artículos científicos como fuentes suministradoras de términos para la indización de los documentos en seis áreas del conocimiento indizadas en las Bases datos ISOC,

2 IME e ICYT del CSIC.

3 Asimismo,

4 se examina la estructura sintagmática de los términos de indización hallados en el campo "Descriptores",

5 y la posible relación entre el número de descriptores de un documento con la cantidad de páginas mismo.

6 Para tales fines se seleccionaron las áreas del conocimiento de Biblioteconomía y Documentación,

7 Medicina,

8 Química,

9 Biología,

10 Psicología y Física,

11 y se realizaron seis búsquedas en estas Bases datos de las que seleccionamos 450 referencias bibliográficas (75 por área) proporcionando un total de 2077 descriptores. **#FR#**

#CTe#

#CP#

1 Introducción.

2 La indización es una operación compleja pero esta dificultad se torna doble cuando se intenta obtener de forma automática.

3 Mediante la indización automática se pretende que sea un algoritmo el que proponga todos los términos de indización tras el análisis de un documento o algunas de sus partes.

4 Tradicionalmente,

5 en la indización automática se han venido utilizando dos métodos distintos pero a veces convergentes en algunos ensayos,

6 esto es, medios no lingüísticos,

7 iniciados a finales de los cincuenta,

8 y lingüísticos incorporados posteriormente.

#FP#

#CP#

9 Cuando se pretende diseñar un sistema de indización automática basado en la extracción de conceptos uno de los planteamientos inmediatos es decidir si las fuentes de las que lograr los términos candidatos a descriptores serán los documentos completos o los títulos y resúmenes de los mismos.

#FP#

#Fte#

3 Eliminación de las palabras vacías.

Una vez horizontalizado el texto se procede a la supresión de las palabras vacías con la ayuda del fichero disponible para tal fin. El texto quedaría del siguiente modo:

#Cti#

1 Análisis descriptores diferentes áreas conocimiento indizadas bases datos CSIC

2 Aplicación indización automática.

#Fti#

#CR#

3 estudia valor títulos resúmenes artículos científicos fuentes suministradoras términos indización documentos áreas del conocimiento indizadas Bases datos ISOC,

4 IME ICYT CSIC

5 examina estructura sintagmática términos indización hallados campo "Descriptores",

6 posible relación número descriptores documento cantidad páginas.

7 fines seleccionaron áreas conocimiento Biblioteconomía Documentación,

8 Medicina,

9 Química,

10 Biología,

11 Psicología

12 Física,

13 realizaron búsquedas Bases datos seleccionamos 450 referencias bibliográficas (75 área) proporcionando total 2077 descriptores.

#FR#

#Cte#

#CP#

1 Introducción.

2 indización operación compleja dificultad torna intenta obtener forma automática.

3 indización automática pretende sea algoritmo proponga términos indización análisis documento partes.

4 Tradicionalmente,

5 indización automática han venido utilizando métodos convergentes ensayos,

6 medios no lingüísticos,

7 iniciados finales cincuenta,

8 lingüísticos incorporados posteriormente.

#FP#

#CP#

1 pretende diseñar sistema indización automática basado extracción conceptos planteamientos inmediatos decidir fuentes lograr términos candidatos descriptores serán documentos completos títulos resúmenes.

#FP##Fte#

De las 272 palabras que componen las tres fuentes han sido eliminadas 138 palabras vacías, es decir, el 50,7 % del total, lo que está acorde con los datos proporcionados con anterioridad, cuando se señaló que, aproximadamente, el 50% de las palabras de los textos se consideran vacías para los fines de la indización. En definitiva, con las 134 restantes serán sobre las que se lleven a cabo los procesos.

MÓDULO 2: PROCESAMIENTO

1 Búsqueda y selección de términos preferentes.

Una vez horizontalizado el texto y eliminadas las palabras vacías de las fuentes, comienza el procesamiento de los patrones admitidos con las fuentes. El resultado es el siguiente:

Patrones admitidos obtenidos del título

1. DESCRIPTORES-Ti
2. BASES DATOS-Ti
3. CSIC-Ti
4. INDIZACION AUTOMATICA-Ti

Patrones admitidos obtenidos del resumen

1. TITULOS-Re
2. RESUMENES-Re
3. ARTICULOS CIENTIFICOS-Re
4. TERMINOS INDIZACION-Re
5. DOCUMENTOS-Re
6. BASES DATOS-Re
7. ISOC-Re
8. IME-Re
9. ICYTC-Re
10. CSIC-Re
11. DESCRIPTORES-Re
12. BIBLIOTECONOMIA DOCUMENTACION-Re
13. MEDICINA-Re
14. QUIMICA-Re
15. BIOLOGIA-Re
16. PSICOLOGIA-Re
17. FISICA-Re
18. REFERENCIAS BIBLIOGRAFICAS-Re

Patrones admitidos obtenidos del texto

1. INDIZACION-Te
2. INDIZACION AUTOMATICA-Te
3. ALGORITMOS-Te
4. TERMINOS INDIZACION-Te
5. ANALISIS DOCUMENTOS-Te
6. CONCEPTO-Te
7. DESCRIPTORES-Te
8. DOCUMENTOS-Te
9. TITULOS-Te
10. RESUMENES-Te

2 Búsqueda y selección de términos no preferentes.

En estas fuentes no hay ninguno caso.

3 Búsqueda y selección de términos contruidos sintácticamente diferentes a los términos preferentes.

En estas fuentes no hay ninguno caso.

4 Búsqueda y selección de palabras semivacías.

Palabras semivacías obtenidas del título

análisis diferentes áreas conocimiento indizadas.
Aplicación

Palabras semivacías obtenidas del resumen

áreas conocimiento indizadas

Palabras semivacías obtenidas del Texto

análisis diferentes áreas conocimiento indizadas

MÓDULO 3: VALORACIÓN Y PONDERACIÓN.

A continuación, hay que valorar los términos candidatos para clarificar cuales pasan a convertirse en descriptores.

Los procesos descritos precedentemente han proporcionado los siguientes candidatos seleccionados de las distintas fuentes:

<u>Título</u>	<u>Resumen</u>	<u>Texto</u>
DESCRIPTORES	TITULOS	INDIZACION
BASES DATOS	RESUMENES	INDIZACION
CSIC	ARTICULOS CIENTIFICOS	AUTOMATICA
INDIZACION	TERMINOS INDIZACIÓN	ALGORITMOS
AUTOMATICA	DOCUMENTOS	TERMINOS INDIZACION
	BASES DATOS	ANALISIS
	ISOC	DOCUMENTOS
	IME	CONCEPTO
	ICYTC	DESCRIPTORES
	CSIC	DOCUMENTOS
	DESCRIPTORES	TITULOS
	MEDICINA	RESUMENES
	QUIMICA	
	BIOLOGIA	
	PSICOLOGIA	
	FISICA	
	BIBLIOTECONOMIA	
	DOCUMENTACION	
	REFERENCIAS	
	BIBLIOGRAFICAS	

Ya hemos señalado que el método de valoración se fundamenta en principios posicionales (A, B, C y D) y frecuenciales (E). Apliquemos los principios posicionales⁹:

⁹ La aplicación del principio frecuencial (E), no se ha llevado a cabo porque supone calcular la frecuencia de aparición de cada palabra en el texto íntegro y para este ejemplo sólo se han utilizado los dos primeros párrafos del mismo.

A) Si un término autorizado aparece en la fuente título y en la fuente resumen se convierte en término de indización.

DESCRIPTORES / BASES DATOS / CSIC

B) Si un término autorizado aparece en la fuente título y en la fuente texto se convierte en término de indización.

DESCRIPTORES / BASES DATOS / CSIC / INDIZACION AUTOMATICA

C) Si un término autorizado aparece en la fuente resumen y en la fuente texto se convierte en término de indización.

TERMINOS INDIZACION / INDIZACION AUTOMATICA
/ DESCRIPTORES / DOCUMENTOS / TITULOS / RESUMENES

D) Si una palabra semivacía aparece en el título, resumen y texto se presenta al usuario para su posible incorporación como término de indización.

AREAS CONOCIMIENTO INDIZADAS

Una vez eliminados los términos repetidos, los candidatos seleccionados son los siguientes:

Términos propuestos

BASES DATOS / CSIC / TERMINOS INDIZACION / DESCRIPTORES /
INDIZACION AUTOMATICA / DOCUMENTOS / TITULOS / RESUMENES

Palabras semivacías propuestas:

AREAS
CONOCIMIENTO
INDIZADAS

El último proceso consiste en la devolución de las palabras vacías suprimidas a los candidatos en la etapa del preprocesamiento. Para ello, se acude a un fichero inverso que contiene todos los términos incluidos en la lista de términos autorizados de este modo:

Términos sin palabras vacías

...
 RECONOCIMIENTO OPTICO CARACTERES

 RECORTES PRENSA
 RECUPERACION DATOS
 RECUPERACION DOCUMENTOS
 ...
 ...

Términos definitivos

...
 RECONOCIMIENTO OPTICO DE
 CARACTERES
 RECORTES DE PRENSA
 RECUPERACION DE LA INFORMACION
 RECUPERACION DE LA INFORMACION
 ...
 ...

La presentación definitiva de los términos y palabras semivacías propuestas por el programa aparecerán en una pantalla de este modo:

<p>Título Análisis de los descriptores de diferentes áreas del conocimiento indizadas en bases de datos del CSIC. Aplicación a la indización automática.</p> <p>Resumen Se estudia el valor de los títulos y resúmenes de los artículos científicos como fuentes suministradoras de términos para la indización de los documentos en seis áreas del conocimiento indizadas en las Bases de datos ISOC, IME e ICYT del CSIC. Asimismo, se examina la estructura sintagmática de los términos de indización hallados en el campo "Descriptores", y la posible relación entre el número de descriptores de un documento con la cantidad de páginas del mismo. Para tales fines se seleccionaron las áreas del conocimiento de Biblioteconomía y Documentación, Medicina, Química, Biología, Psicología y Física. Y se realizaron seis búsquedas en estas Bases de datos de las que seleccionamos 450 referencias bibliográficas (75 por área) proporcionando un total de 2077 descriptores.</p>		
TÉRMINOS PROPUESTOS		PALABRAS SEMIVACÍAS PROPUESTAS
BASES DE DATOS CSIC TÉRMINOS DE INDIZACIÓN DESCRIPTORES INDIZACIÓN AUTOMÁTICA TÍTULOS RESÚMENES	← AÑADIR SUPRIMIR	<u>ANALISIS ÁREAS</u> <u>CONOCIMIENTO INDIZADAS</u>

Se presentan el título y el resumen del documento analizado junto a los términos y las palabras semivacías propuestas para que el usuario tenga la mayor información posible en el momento de la validación y edición de las propuestas efectuadas. Por esta circunstancia, se han elegido estas dos fuentes que son de suma importancia en la indización.

En la columna de la izquierda se presentan los términos propuestos, mientras que en la columna del centro hay tres botones de función:

El botón con la flecha hacia la izquierda (\leftarrow) incorpora una palabra semivacía como término de indización una vez marcada al colocar el cursor sobre ella.

Si se pulsa el botón AÑADIR se coloca el cursor a continuación del último término propuesto por el sistema, para que el usuario introduzca el término o los términos que considere oportunos para completar la indización de ese documento.

Y con el botón SUPRIMIR se elimina un término propuesto una vez marcado.

En la columna de la derecha se muestran las palabras semivacías propuestas. El programa permite la visualización de las frases u oraciones horizontalizadas donde han sido localizadas. Esta posibilidad se lleva a cabo pulsar el ratón, una vez situados sobre la palabra. El objetivo de esta opción es tener información sobre el contexto donde han aparecido estas palabras con el fin de decidir si se incorpora como término de indización. En este ejemplo, si se pulsara sobre 'conocimiento' la respuesta será como aparece en el siguiente recuadro. Como se puede observar, el sistema presenta cada frase u oración horizontalizada donde ha aparecido la palabra "conocimiento" con especificación de la fuente y la palabra en cuestión, resaltada para su fácil localización.

Frases u oraciones horizontalizadas donde aparece la palabra semivacia seleccionada: "CONOCIMIENTO"

Título:

1 Análisis de los descriptores de diferentes áreas del **conocimiento** indizadas en bases datos del CSIC.

Resumen:

3 Se estudia el valor de los títulos y resúmenes de los artículos científicos como fuentes suministradoras de términos para la indización de los documentos en seis áreas del **conocimiento** indizadas en las Bases datos ISOC,

8 Para tales fines se seleccionaron las áreas del **conocimiento** de Biblioteconomía y Documentación,

Texto:

11 Otra cuestión a clarificar es si los títulos y resúmenes de documentos científicos -en este caso artículos de revistas- de diferentes áreas del **conocimiento** proporcionan similar número de conceptos útiles para la indización.

21 se ha abordado el análisis de referencias de artículos de distintas áreas del **conocimiento** que van desde las humanidades,

...
...

A continuación, se ofrece el análisis finalizado de un documento que requiere edición para comprobar el funcionamiento del sistema:

<p>Título: Errores ortográficos en el ingreso en bases de datos.</p> <p>Resumen: Se estudian los problemas de la corrección ortográfica en el ingreso de registros en bases de datos en idioma español. Se evalúan los pros y contras de cuatro métodos de control: doble entrada, hapax legomena, trigramas y uso de diccionarios, con vistas a determinar cuales de estos procedimientos ofrecen mejor relación de costo/resultado. El trabajo está enfocado a los procesos de ingreso por digitación, y no se analizan los errores ortográficos de los ingresos por lectura óptica.</p>		
TÉRMINOS PROPUESTOS		PALABRAS SEMIVACÍAS PROPUESTAS
BASES DE DATOS DICIONARIOS	← AÑADIR SUPRIMIR	<u>ERRORES ORTOGRAFICOS</u>

En la fase de validación y edición de este documento vamos a tener que utilizar las tres opciones que ofrece el sistema:

Eliminar un término de indización asignado.

Incorporar uno desde las palabras semivacías.

Introducir otros que el usuario considere oportunos.

Comencemos, en primer lugar, por los errores de asignación. La asignación del término ‘DICIONARIOS’ se ha producido porque este término está presente en la lista de términos autorizados y además, ha cumplido uno de los principios de valoración. Pero esta asignación parece injustificada según el contenido del artículo, por lo que es conveniente eliminarla. La supresión de este término se lleva a cabo de forma manual, es decir, a través del icono SUPRIMIR .

En segundo lugar, el módulo de palabras semivacías sugiere las palabras ‘ERRORES’ y ‘ORTOGRÁFICOS’. Si se revisa el título y el resumen del documento

se comprueba que queda justificada la inclusión de 'ERRORES ORTOGRAFICOS' como término de indización. Y para ello utilizaremos la flecha (←).

En tercer lugar, está justificado incorporar un concepto relativo al control de la indización para la obtención de la máxima calidad en las Bases de datos, según el contenido de este documento. En este caso 'CONTROL DE CALIDAD' que es añadido también de forma manual, por medio del icono AÑADIR.

En definitiva, este ejemplo clarifica en qué consiste la validación y la edición por parte del usuario.

Este artículo queda indizado del siguiente modo:

<p>Título: Errores ortográficos en el ingreso en bases de datos.</p> <p>Resumen: Se estudian los problemas de la corrección ortográfica en el ingreso de registros en bases de datos en idioma español. Se evalúan los pros y contras de cuatro métodos de control: doble entrada, hapax legomena, trigramas y uso de diccionarios, con vistas a determinar cuales de estos procedimientos ofrecen mejor relación de costo/resultado. El trabajo está enfocado a los procesos de ingreso por digitación, y no se analizan los errores ortográficos de los ingresos por lectura óptica.</p>		
TÉRMINOS PROPUESTOS		PALABRAS SEMIVACÍAS PROPUESTAS
BASES DE DATOS ERRORES ORTOGRAFICOS CONTROL DE CALIDAD	← AÑADIR SUPRIMIR	

3.4.3 La evaluación de la propuesta.

Tradicionalmente, la evaluación de la indización, ya sea de forma manual o automática, se ha llevado a cabo de dos modos diferentes.

Un primer método ha sido hallar los índices de consistencia de la indización entre dos indizadores profesionales o entre un profesional y un sistema automático. Para estos fines se ha empleado la siguiente fórmula:

$$C_i = T_c / (T_m + T_a) - T_c$$

donde:

C_i = índice de consistencia de indización entre dos indizadores o sistemas.

T_c = número de términos comunes asignados.

T_m = número de términos asignados por el indizador uno de forma manual.

T_a = número de términos asignados por el indizador dos o de modo automático.

Un segundo método para la evaluación de sistemas de indización es a través del cálculo de los índices de exhaustividad y precisión en la recuperación. La fórmula de la exhaustividad en la recuperación representa el cociente entre el número de documentos pertinentes recuperados y el número total de documentos pertinentes en la colección. En cambio, la precisión representa el cociente entre los documentos pertinentes recuperados y el número total de documentos recuperados¹⁰.

Para la evaluación vamos a aplicar el sistema propuesto, manualmente, a cincuenta artículos que hayan sido analizados por indizadores profesionales de la Base de datos ISOC. Después, utilizaremos la fórmula para calcular los índices de consistencia entre la indización de nuestro sistema y la de esta Base de datos. Para ello, hemos elegido los artículos publicados durante 1995, 1994, 1993 y 1992 en la Revista Española de Documentación Científica. La elección de esta publicación se debe a que está considerada la de mayor calidad de las publicadas en España en Biblioteconomía y Documentación¹¹. Esta revista alberga cada año dieciséis artículos, por tanto en cuatro años han sido publicados sesenta y cuatro trabajos.

¹⁰ Las fórmulas para hallar estos valores se expusieron en el epígrafe 1.4.1

¹¹ Los trabajos de la Revista Española de Documentación Científica se recogen, periódicamente, en la Base de datos bibliográfica LISA, que es la más importante a nivel internacional en el área de Biblioteconomía y Documentación. Además, también se indiza en las Bases de datos INSPEC y PASCAL.

Empezamos con el análisis de todos los publicados en 1995, 1994, 1993 y 1992 respectivamente. Cuando encontrábamos un trabajo sin resumen (en concreto dos) o en un idioma diferente al español (tres) pasábamos al siguiente artículo.

La fórmula aplicada ha sido referenciada y utilizada para comparar la consistencia entre sistemas automáticos y manuales y entre varios indizadores en trabajos de Salton y McGill [1983] ; Lustig y Knorz [1986] ; Lancaster [1991] o Silvester, Genuardi y Klingbiel [1994].

A continuación, se muestran los términos propuestos por el sistema tras la aplicación manual de los módulos: preprocesamiento, procesamiento y valoración (sólo los principios A, B, y C)¹².

ARTÍCULO 1

TÉRMINOS ANTES DE SU VALORACIÓN

MERCADO DE TRABAJO-Ti
DIPLOMADOS-Ti
BIBLIOTECONOMIA DOCUMENTACION-Ti

MERCADO DE TRABAJO-Re
DIPLOMADOS-Re
BIBLIOTECONOMIA DOCUMENTACION-Re
CUESTIONARIOS-Re
INDICES-Re
ARTICULO-Re

DIPLOMADOS-Te
BIBLIOTECONOMIA DOCUMENTACION-Te
UNIVERSIDAD-Te
ESCUELAS BIBLIOTECONOMIA DOCUMENTACION-Te
ESCUELAS UNIVERSITARIAS-Te
DIPLOMATURAS-Te
MERCADO LABORAL-Te
PLANES DE ESTUDIO-Te
INDICE-Te
CUESTIONARIOS-Te
...

TÉRMINOS PROPUESTOS TRAS SU VALORACIÓN

MERCADO DE TRABAJO
DIPLOMADOS
BIBLIOTECONOMIA Y
DOCUMENTACION
INDICES
ENCUESTAS

Apliquemos los criterios de valoración A, B, y C:

A) Si un término autorizado aparece en la fuente título y en la fuente resumen se convierte en término de indización.

MERCADO DE TRABAJO / DIPLOMADOS /
BIBLIOTECONOMIA DOCUMENTACION

B) Si un término autorizado aparece en la fuente título y en la fuente texto se convierte en término de indización.

DIPLOMADOS / BIBLIOTECONOMIA DOCUMENTACION

C) Si un término autorizado aparece en la fuente resumen y en la fuente texto se convierte en término de indización.

DIPLOMADOS / BIBLIOTECONOMIA DOCUMENTACION /
CUESTIONARIOS / INDICE

TÉRMINOS ASIGNADOS DE MODO AUTOMÁTICO

1. MERCADO DE TRABAJO
2. DIPLOMADOS
3. BIBLIOTECONOMIA Y DOCUMENTACION
4. INDICES
5. ENCUESTAS

TÉRMINOS ASIGNADOS DE MODO MANUAL

1. ESCUELAS UNIVERSITARIAS DE BIBLIOTECONOMIA Y DOCUMENTACION
2. CONCEPTO DE SI MISMO
3. PRESTIGIO PROFESIONAL
4. CATEGORIAS PROFESIONALES
5. MERCADO DE TRABAJO
6. STATUS SOCIAL

De acuerdo a la fórmula expresada más arriba hallamos el índice de consistencia entre los dos sistemas de indización para el artículo anterior:

$$C_i = \frac{1}{6 + 5 - 1} = 0,1 = 10\% \text{ de consistencia}$$

¹² La aplicación de los principios D y E no se han llevado a cabo porque se trata, en un caso de un 'diálogo' entre el sistema y el usuario (D), y en otro, de calcular la frecuencia de aparición de las palabras (E).

ARTÍCULO 2

TÉRMINOS ANTES DE SU VALORACIÓN	TÉRMINOS PROPUESTOS
CD-ROM-Ti	ENCICLOPEDIAS
ENCICLOPEDIAS ELECTRONICAS-Ti	ENCICLOPEDIAS ELECTRONICAS
	HIPERTEXTO
ENCICLOPEDIAS-Re	NUEVAS TECNOLOGIAS
SISTEMAS DE RECUPERACION-Re	MULTIMEDIA
HIPERTEXTO-Re	HIPERMEDIA
NUEVAS TECNOLOGIAS-Re	
MULTIMEDIA-Re	
HIPERMEDIA-Re	
ENCICLOPEDIAS-Te	
MULTIMEDIA-Te	
HIPERTEXTO-Te	
NUEVAS TECNOLOGIAS-Te	
ENCICLOPEDIAS ELECTRONICAS-Te	
HIPERMEDIA-Te	

Apliquemos de nuevo, los criterios de valoración A, B, y C:

A) Si un término autorizado aparece en la fuente título y en la fuente resumen se convierte en término de indización.

B) Si un término autorizado aparece en la fuente título y en la fuente texto se convierte en término de indización.

ENCICLOPEDIAS ELECTRÓNICAS

C) Si un término autorizado aparece en la fuente resumen y en la fuente texto se convierte en término de indización.

ENCICLOPEDIAS / HIPERTEXTO /NUEVAS TECNOLOGÍAS /
MULTIMEDIA / HIPERMEDIA

TÉRMINOS ASIGNADOS DE MODO AUTOMÁTICO

- 1. ENCICLOPEDIAS
- 2. ENCICLOPEDIAS ELECTRONICAS
- 3. HIPERTEXTO
- 4. NUEVAS TECNOLOGIAS
- 5. MULTIMEDIA
- 6. HIPERMEDIA

TÉRMINOS ASIGNADOS DE MODO MANUAL

- 1. EDICION ELECTRONICA
- 2. ENCICLOPEDIAS
- 3. HIPERMEDIA
- 4. HIPERTEXTO
- 5. MULTIMEDIA

$$C_i = \frac{4}{5 + 6 - 4} = 0,5714 = 57,14\% \text{ de consistencia}$$

El resto de las tablas de comparación e índices de consistencia se pueden observar en el anexo 7. A continuación se muestran los índices de consistencia resultantes:

ARTÍCULO 1 → 10 %	ARTÍCULO 22 → 10 %
ARTÍCULO 2 → 57,14 %	ARTÍCULO 23 → 6,45 %
ARTÍCULO 3 → 25 %	ARTÍCULO 24 → 7,14 %
ARTÍCULO 4 → 25 %	ARTÍCULO 25 → 21,42%
ARTÍCULO 5 → 0 %	ARTÍCULO 26 → 22,22 %
ARTÍCULO 6 → 14,28 %	ARTÍCULO 27 → 12,5 %
ARTÍCULO 7 → 20 %	ARTÍCULO 28 → 33,33 %
ARTÍCULO 8 → 18,75 %	ARTÍCULO 29 → 22,22 %
ARTÍCULO 9 → 30,76 %	ARTÍCULO 30 → 0 %
ARTÍCULO 10 → 60 %	ARTÍCULO 31 → 30 %
ARTÍCULO 11 → 26,66 %	ARTÍCULO 32 → 20 %
ARTÍCULO 12 → 14,28 %	ARTÍCULO 33 → 33,33 %
ARTÍCULO 13 → 37,5 %	ARTÍCULO 34 → 27,27 %
ARTÍCULO 14 → 66,66 %	ARTÍCULO 35 → 9,09 %
ARTÍCULO 15 → 25 %	ARTÍCULO 36 → 22,22 %
ARTÍCULO 16 → 14,28 %	ARTÍCULO 37 → 28,57 %
ARTÍCULO 17 → 11,11 %	ARTÍCULO 38 → 33,33 %
ARTÍCULO 18 → 12,5 %	ARTÍCULO 39 → 74,42 %
ARTÍCULO 19 → 10 %	ARTÍCULO 40 → 37,5 %
ARTÍCULO 20 → 16,66 %	ARTÍCULO 41 → 50 %
ARTÍCULO 21 → 60 %	ARTÍCULO 42 → 25 %

ARTÍCULO 43 → 44,44 %	ARTÍCULO 48 → 38,46 %
ARTÍCULO 44 → 13,33 %	ARTÍCULO 49 → 0 %
ARTÍCULO 45 → 40 %	ARTÍCULO 50 → 20 %
ARTÍCULO 46 → 30 %	
ARTÍCULO 47 → 14,28 %	MEDIA ⇒⇒⇒ 25,93 %

De forma manual se han asignado 309 descriptores lo que supone una media de 6,1 por artículo, y de forma automática se asignaron 312, lo que supone 6,2. La consistencia oscila entre el 0 % de tres artículos al 74,42% de uno. La media resultante total es de 25,93 %.

A continuación presentamos diferentes trabajos sobre la consistencia en la indización entre indizadores y entre sistemas automáticos y profesionales.

Ensayos entre indizadores	Consistencia
Rodgers [1961]	24 %
Jacoby [1962]	10 %
Slamecka y Jacoby [1962]	12,9 %
Slamecka y Jacoby [1963]	40 %
Korotkin y Oliver [1964]	47,5 %
Sievert y Andrews [1991]	47,2 %
MEDIA	30,26 %

Ensayos entre sistemas semiautomáticos y profesionales	Consistencia
Silvester, Genuardi y Klingbiel [1994]	1987* ⇒20,8 % 1988' ⇒ 23,4 % 1989' ⇒26 % 1992' ⇒ 29 % 1993' ⇒ 33,3 % MEDIA 26.5 %

* Sobre 2500 documentos

' Sobre 100 documentos

3.4.4 Los problemas detectados.

A pesar de que el número de documentos analizados ha sido pequeño para establecer conclusiones definitivas, sí se han observado unos problemas cuya modificación permitirá conseguir mejores resultados:

—En ocasiones en la indización manual han sido asignados términos no específicos del ámbito de Biblioteconomía y Documentación del tipo: PRESTIGIO PROFESIONAL / STATUS SOCIAL / TRABAJO EN GRUPO / PARTICIPACIÓN EN GRUPO / PLAN ESTRATÉGICO / PROGRAMAS EUROPEOS DE COOPERACIÓN / DISTRIBUCIÓN ESPACIAL / COOPERACIÓN INTERNACIONAL. Estos y otros términos, al no estar recogidos en la lista de términos autorizados, ha sido la causa de que en algunos documentos los índices de consistencia hayan sido inferiores de lo esperado. En consecuencia, consideramos necesario la ampliación del vocabulario con un mayor número de términos del área, con sus sinónimos, así como de otros términos más generales, que no tienen por qué pertenecer a esta área en cuestión.

Esta ampliación no solamente viene justificada por las deficiencias encontradas sino que parece una característica generalizada que los sistemas para la automatización de la indización empleen vocabularios controlados muy amplios como se comprueba en los siguientes ejemplos. En el sistema de la NASA la base del conocimiento cuenta con más de once mil entradas con posibilidad de transformarse en descriptores. En el Centro Internacional de Información Científica y Técnica de Moscú se utilizó un tesoro con más de dieciocho mil descriptores y más de cuatro mil no descriptores. Y por último, el tesoro utilizado en el sistema SAPHIRE consta de más de veintiocho mil términos y más de setenta y ocho mil sinónimos.

—Cuando los resúmenes son excesivamente largos aumenta el número de términos propuestos por el sistema, lo que provoca, en ocasiones, una distorsión en la indización propuesta.

4.- CONCLUSIONES

La parte dedicada a las conclusiones tiene por objeto el compendio de las principales aportaciones efectuadas en la investigación acometida. Si se nos pidiera una conclusión global de este trabajo nos veríamos en la obligación de precisar que hemos analizado y concretado el desarrollo diacrónico acaecido en la automatización de la indización desde finales de los años cincuenta hasta el presente. También hemos puesto de relieve los inconvenientes más relevantes de las principales metodologías adoptadas en las distintas propuestas. Hemos diseñado a nivel teórico un mecanismo de indización semiautomático para el español y en concreto, para el área de Biblioteconomía y Documentación extrapolable a otras áreas. Pero principalmente, hemos abierto, a nuestro juicio, un rico y variado camino de investigación que permanecía intacto en nuestro país.

No obstante, al inicio y durante esta investigación se plantearon múltiples interrogantes a los que no se puede responder de forma tan general y concisa como se ha planteado en el párrafo precedente. Por tanto, como exige la costumbre académica se muestran a continuación las principales conclusiones alcanzadas en este trabajo:

1. Se observan diferencias importantes a la hora de definir la indización y de establecer las etapas constitutivas del proceso por parte de los investigadores del área de Biblioteconomía y Documentación. En la mayoría de las definiciones examinadas se echa en falta la mención de que no solamente es materia de indización el documento, sino también la pregunta planteada por el usuario o el cliente. Además, por lo general, tampoco se menciona la posibilidad de asignar como término de indización un concepto implícito. Por otro lado, no parece haber consenso en el establecimiento de las etapas que componen el proceso de la indización.
- 2 La indización de los documentos tiene por objeto permitir el almacenamiento de los mismos a través de un conjunto de términos que representan los conceptos

esenciales tratados en dichos documentos. El fin perseguido con la indización de las preguntas es encaminar la recuperación de los documentos previamente almacenados. Por tanto, podemos expresar que el objetivo general de la indización es el almacenamiento y la recuperación de la información.

3 El establecimiento del tiempo dedicado o que debería dedicarse a la indización de un documento no es una tarea fácil. El tiempo empleado obedece, en la mayoría de las ocasiones, a las directrices marcadas por la institución en donde se lleva a cabo esta tarea. Y esta política depende de los tipos y las necesidades de sus usuarios, del personal disponible para la realización de esta operación, de la clase de documentos analizados y del tipo de información contenida o del tamaño de los documentos.

4 La exhaustividad en la indización tiene que ver con la selección y/o asignación de todos los conceptos esenciales explícitos o implícitos en el documento o la pregunta planteada por el usuario o el cliente del sistema, y no con el número de términos de indización asignados a un documento. En cambio, los factores que pueden determinar el grado de exhaustividad de la indización son el nivel de precoordinación utilizado, la corrección de la indización y las directrices marcadas.

5 La obtención de índices semejantes de consistencia entre un mismo indizador en diferentes momentos o entre dos indizadores es difícil. Esta circunstancia nos lleva a considerar la inconsistencia en la indización como un elemento inherente a esta tarea más que como una anomalía producida esporádicamente. Según los distintos ensayos analizados, la consistencia se sitúa entre el veinticinco y sesenta por ciento, como mínimo y máximo respectivamente.

6 Para referirse a la automatización de la indización existe una amplia variedad de expresiones. En concreto, se han localizado una veintena que alude a tres conceptos diferentes. En primer lugar, a programas que asisten en el proceso de almacenamiento de los términos de indización una vez obtenidos de modo manual.

En segundo lugar, a sistemas que analizan los documentos de modo automático pero los términos de indización propuestos los valida y edita (si es necesario) un profesional. Y en tercer lugar, a los que no emplean métodos de validación y edición ya que los términos propuestos se almacenan directamente como descriptores de dicho documento.

7 Con las incipientes iniciativas para la automatización de la indización a finales de los años cincuenta, pero principalmente al inicio de los sesenta, surgieron las primeras manifestaciones en contra de automatizar esta tarea. Esta oposición todavía persiste por parte de profesionales e investigadores del campo de la Biblioteconomía y Documentación, basada en la idea fundamental de que un programa informático no puede simular convenientemente el trabajo de un profesional para analizar, detectar, relacionar y seleccionar los conceptos esenciales explícitos e implícitos tratados en los documentos.

8 Por contra, los investigadores que han trabajado en el diseño y el desarrollo de herramientas para la automatización de la indización han ido incorporando de forma progresiva principios y metodologías de otras ciencias y disciplinas. Este empeño está guiado por la idea de que con la automatización de esta operación se pueden solucionar problemas como la subjetividad, la reducción del tiempo dedicado a la indización de cada documento, y la disminución del coste económico que acarrea este proceso.

9 La interdisciplinariedad es una característica inherente a la automatización de la indización debido a la complejidad que supone su ejecución. Por tanto, estas dificultades necesitan resolverse desde un prisma interdisciplinar y por tanto, con soluciones multidisciplinares. Para ello, en algunas propuestas intervienen ciencias y disciplinas como la Lingüística, la Lingüística computacional, la Estadística, la Terminología, la Informática y la Inteligencia Artificial representada en los Sistemas Expertos.

- 10 En la bibliografía analizada para la realización de este trabajo, no se ha observado que la constitución de grupos de investigación interdisciplinarios para acometer la automatización de la indización sea una práctica generalizada. Si bien es cierto que, ha habido excepciones donde investigadores de diferentes áreas han aportado sus conocimientos teóricos y prácticos para conseguir este objetivo. No obstante, no concebimos un modo para la automatización de la indización que no se fundamente en un trabajo que recoja soluciones multidisciplinarios, de ahí la necesidad de constituir grupos de investigación formados, al menos, por lingüistas, informáticos y documentalistas.
- 11 Los primeros intentos en la automatización de la indización tomaron como fundamentos los métodos estadísticos. Las propuestas iniciales presentadas a finales de los cincuenta se asentaron sobre sencillos métodos estadísticos para el cálculo de aparición de las palabras tanto en documentos concretos como en colecciones específicas. A partir de aquí, surgieron fórmulas más complejas como la ponderación de la frecuencia inversa o el cálculo del poder de discriminación de los términos. Los métodos estadísticos fueron empleados casi de modo exclusivo hasta finales de los setenta y principios de los ochenta. Estos métodos se han mostrado incapaces para el reconocimiento de relaciones semánticas simples como la sinonimia, y ofrecen dificultades para el reconocimiento y el cómputo de términos compuestos.
- 12 La otra metodología adoptada ha sido el uso de métodos lingüísticos. A partir de los ochenta estos métodos adquirieron cada vez más relevancia, debido a los progresivos avances producidos en el procesamiento del lenguaje natural. Las propuestas para la automatización de la indización han utilizado, fundamentalmente, el nivel morfológico (la descomposición de las palabras en sus raíces y la asignación de categorías gramaticales) y el nivel sintáctico (la desambiguación lexical y la búsqueda en los textos de estructuras sintagmáticas preestablecidas). El principal inconveniente de la utilización del PLN (procesamiento del lenguaje natural) es la cantidad de recursos, proceso y tiempo que consume, para después tener que contrastar los resultados con un vocabulario controlado, y ponderar los términos

candidatos. Además, habría que añadir la carencia de una biblioteca de programas informáticos básica de libre distribución para el español, con los que emprender investigaciones apoyadas en esta metodología.

13 Habitualmente, las dos metodologías mencionadas con anterioridad se han combinado en las diferentes propuestas. Asimismo, en ocasiones, los sistemas para la automatización de la indización han utilizado un vocabulario controlado, lo que hace que una misma propuesta utilice técnicas del PLN para analizar los textos, un vocabulario controlado con el que validar los términos candidatos, y los métodos estadísticos para ponderar dichos términos. Esta circunstancia ha dado lugar a lo que hemos denominado “metodologías o sistemas híbridos”.

14 Las investigaciones para la automatización de la indización difundidas a través de publicaciones periódicas, actas de congresos o tesis doctorales han sido un hecho constante durante las tres últimas décadas en países como Francia y Brasil, influenciado este último por Estados Unidos que ha sido el que ha llevado la iniciativa en este tipo de investigaciones desde finales de los años cincuenta.

15 En España, en cambio, apenas ha sido tratado este tema por profesionales e investigadores del área de Biblioteconomía y Documentación. Solamente ha habido dos propuestas para la automatización de la indización presentadas en sendos artículos. La primera en 1983, se basó en seleccionar desde los títulos de los documentos, términos contenidos en un vocabulario controlado de Metalurgia y con una estructura sintáctica preestablecida. La segunda propuesta presentada en 1990, se fundamentó en la localización en los textos de un conjunto de estructuras sintácticas igualmente preestablecidas.

16 En el proceso de indización, los resúmenes son fuentes que proporcionan más términos con posibilidad de convertirse en descriptores que los títulos. En cuanto al área de Biblioteconomía y Documentación cabe mencionar que los títulos y los resúmenes de los artículos analizados proporcionaron el 47,2% de los descriptores asignados a esos documentos, mientras que los textos el 24,7%. Por

tanto, debido al importante número de términos proporcionados por los textos parece ineludible el uso de esta fuente, junto a los títulos y los resúmenes, para cualquier intento de automatizar la indización.

17 Los vocabularios controlados con sus relaciones de equivalencia o los tesauros deben constituir los ejes centrales de las propuestas dirigidas a la automatización de la indización . Esta idea está justificada por el valor que adquiere la terminología en los textos científico-técnicos y por la posibilidad que brindan para la detección de sinónimos. A pesar de todo, los sistemas fundamentados en estas herramientas tienen que disponer de métodos para que un término no incluido en el vocabulario controlado o en el tesauro tenga la opción de ser propuesto como término de indización.

18 No existe un vocabulario controlado amplio y actualizado sobre Biblioteconomía y Documentación en español. Esta circunstancia nos llevó a la elaboración de un vocabulario controlado sobre esta materia por ser de suma importancia para desarrollar nuestra propuesta. El vocabulario controlado está constituido por un total de 1995 términos, de los cuales 1362 son términos preferentes y 633 no preferentes.

19 Cada vez que se ha implantado un sistema automático para la indización de los documentos se ha llevado a cabo una comparación con los métodos manuales para comprobar su grado de eficiencia. En la mayor parte de los ensayos realizados en esta dirección, los resultados obtenidos de forma automática suelen ser similares a los logrados por los indizadores profesionales. No obstante, en algunas de estas pruebas, se ha concluido que la exhaustividad y precisión en la recuperación han sido inferiores con los sistemas automáticos.

20 A finales de los ochenta afloraron nuevos y variados caminos de investigación encaminados a la automatización de la indización de información no textual (imágenes y sonido). Cuando todavía no hay una metodología reconocida y consolidada para la automatización de la indización, y cuando aún se cuestiona

por parte de investigadores y profesionales del área de Biblioteconomía y Documentación la eficacia de los sistemas automáticos, han aparecido iniciativas dirigidas a crear instrumentos para la automatización de la indización de información multimedia.

21 La implantación de sistemas para la automatización de la indización es prácticamente nula en los Centros de Documentación, Bibliotecas especializadas, empresas de servicios de indización y resumen y en productores de bases de datos. Si bien es cierto que hay excepciones, y los productos utilizados en estos lugares oscilan entre programas que asisten en el momento de la indización, medios automáticos que validan los términos asignados por los indizadores para asegurar un control de calidad, y ya en menor medida, sistemas de indización automática sin ningún tipo de validación y edición del resultado propuesto.

22 Para la automatización de la indización se utilizan una serie de herramientas que no siempre coinciden en todos los desarrollos propuestos. Las más comunes son métodos para la necesaria ponderación de los términos (aplicación de cálculos estadísticos), la utilización de vocabularios controlados, ficheros de palabras vacías, analizadores morfológicos y en menor grado sintácticos.

23 Las primeras propuestas para automatizar la indización en los años sesenta, por lo general, tomaron como fuentes de análisis los títulos. Posteriormente, se fue generalizando el uso de los títulos y los resúmenes de los documentos. La adopción de los títulos, los resúmenes y los textos como fuentes de análisis viene justificada por las siguientes razones. En primer lugar, por los resultados hallados en el estudio llevado a cabo en las Bases de datos del CSIC. En segundo lugar, porque hay revistas españolas de Biblioteconomía y Documentación en las que los artículos publicados no siempre van acompañados de sus respectivos resúmenes. Y tercero, porque hemos adoptado una metodología sencilla que propicia que el análisis de las fuentes sea rápido.

- 24 Ante la imposibilidad de disponer de todas las posibles formas por las cuales se puede representar sintácticamente un concepto se cuenta en el vocabulario controlado con 633 sinónimos de términos preferentes. Además, el algoritmo para el procesamiento tiene la capacidad de detectar términos preferentes pero con diferente estructura sintagmática a los incluidos en el vocabulario controlado.
- 25 En la evaluación del sistema hemos detectado la necesidad de ampliar el vocabulario controlado en dos direcciones. En primer lugar, con términos preferentes y no preferentes del área de Biblioteconomía y Documentación, y en segundo lugar, con términos más generales o incluso que pertenezcan a otras áreas. Consideramos que esta ampliación redundará positivamente en la obtención de mejores índices de consistencia.
- 26 El índice medio de consistencia obtenido en la evaluación manual ha sido del 25,93%. Este porcentaje se mejorará cuando se aumente el vocabulario controlado. No obstante, este resultado no está muy alejado del índice de consistencia medio (26,5%) obtenido en las cinco evaluaciones del sistema de indización semiautomático operativo en el Centro de Información Aeroespacial de la NASA desde mediados de los ochenta.
- 27 Por último, juzgamos que con este trabajo hemos contribuido a abrir el camino de investigación hacia la automatización de la indización. Hasta el momento, no había sido abordado por los profesionales e investigadores del área de Biblioteconomía y Documentación españoles. Esperemos que esta brecha abierta se amplíe en numerosas líneas de investigación.

5.- BIBLIOGRAFÍA

AENOR. (1994). Documentación. Normas fundamentales. Madrid: AENOR.

ALBERICO, R., MICCO, M. (1990). Expert systems for reference and information retrieval. Westport, CT: Meckler.

ALBRECHTSEN, H. (1993). "Subject analysis and indexing: from automated indexing to domain analysis". *The Indexer*, 18, 4, p. 219-224.

ALMELA, R. (1986). Materiales para el estudio del español: (curso universitario de morfémica y lexémica. Murcia: Universidad.

AMAT, N. (1989). Documentación y nuevas tecnologías de la información. Madrid: Pirámide.

AMAT, N. (1994). La documentación y sus tecnologías. Madrid: Pirámide.

ANDERSON, J.D.(1994). "Standards for indexing: revising the American National Standard Guidelines Z39.4". *Journal of the American Society for Information Science*, 45, 8, p. 628-636.

ANDERSON, P.F. (1988). "Expert systems, expertise and the library and information professions". *Library and Information Science Research*, 10, p. 369-370.

ANDREEWSKY, A., RUAS, V. (1982). Indexação automática baseada em métodos lingüísticos e estatísticos e su aplicabilidade à lingua portuguesa. Rio de Janeiro, PUC-DI.

ARNTZ, R., PICHT, H. (1995). *Introducción a la terminología*. Madrid: Fundación Germán Sánchez Ruipérez.

ARTANDI, S. (1971). "Coordinate indexing". En A. Kent y H. Lancour (eds). *Encyclopedia of Library and Information Science*, vol. 5, p. 679-682.

ARTANDI, S. (1976). "Machine indexing: linguistic and semiotic implications". *Journal of the American Society for Information Science*, July-August, p. 235-239.

ATHERTON, P. (1978). *Manual para sistemas y servicios de información*. París: UNESCO.

BARANOW, U.G. (1983). "Perspectivas na contribuição da lingüística e de áreas afins à ciência da informação". *Ciencia da Informação, Brasília*, 12, 1, p. 23-35.

BARNES, C.I, COSTANTINI, L., PERSCHKE, S. (1978). "Automatic indexing using the SLC II System". *Information Processing & Management*, 14, 2, p. 107-119.

BASILIO, M. de P. (1979). "Interface lingüística e ciência da informação: potencialidades na análise de estruturas lexicais". *Reuniao Brasileira de Ciência da Informação*, 2. Rio de Janeiro: IBICT, 1, 6, p. 233-246.

BASTOS VIEIRA, S. (1988). "Indexação automática e manual: revisao de literatura". *Ciencia da Informao, Brasília*, 17, 1, p. 43-57.

BELKIN, N.J. (1988). "The nature and function of explanation in intelligent information retrieval". *11 th International Conference on Research and Development in Information Retrieval*, Grenoble, France, juin 13-15, Yves Chiaramella (ed.). France: Presses Universitaires de Grenoble.

BERTRAND-GASTALDY, S., PAGOLA, G. (1992). "L'analyse du contenu textuel en vue de la construction de thésaurus et de l'indexation assistées par ordinateur; applications possibles avec SATO". *Documentation et bibliothèques*, 38, 2, p. 75-89.

BIEBRICHER, P., et al. (1988). "The automatic indexing system AIR/PHYS - From research to application". *Proceedings of the 1988 ACM Conference on Research and Development in Information Retrieval*, p. 333-342.

BLOSSEVILLE, M.J., et al. (1992). "Automatic document classification: natural language processing, statistical analysis, and expert system techniques used together". *Proceedings of the Fifteenth Annual International ACM/SIGIR Conference on Research and Development in Information Retrieval*, N. Belkin, et al. (eds.), Copenhagen: ACM Press, p. 51-59.

BONURA, L.S. (1994). *The art of indexing*. New York: John Wiley.

BOOKSTEIN, A, SWANSON, D.R. (1975). "A decision theoretic foundation for indexing". *Journal of the American Society for Information Science*, 26, 1, p. 45-50.

BORDOGNA, G., et al. (1990). "Pictorial indexing for an integrated pictorial and textual IR environment". *Journal of the Information Science*, 16, p. 165-173.

BORKO, H., BERNIER, C.L. (1978). *Indexing concepts and methods*. New York: Academic Press.

BOYCE, B., LOCKARD, M. (1975). "Automatic and manual indexing performance in a small file of medical literature". *Bulletin of Medical Library Association*, 63, 4, p. 378-385.

BRAGA, L.M. (1982). *Palavras de títulos e resumos como acesso ao conteúdo do documento: uma análise numérica*. Rio de Janeiro, URFJ/IBICT.

BRANHORST, W.T., ECKERT, P.F. (1966) Guide to the processing, storage, and retrieval of bibliographic information an the NASA scientific and technical information facility. Document NASA, CR-62.033, june.

BRITVIN, V.G. (1981). Structural and semantic analysis of abstracts and document search pattern information in automatic indexing in the neft-3 petroleum industry information system. En Development and Operation of an information system for the oil industry.

CABRÉ, M.T. (1993). La terminología. Teoría, metodología, aplicaciones. Barcelona: Antártida.

CARRATALÁ, E. (1980). Morfosintaxis del castellano actual. Barcelona: Labor.

CARRERAS PANCHÓN, A. (ed.), (1994). Guía práctica para la elaboración de un trabajo científico. Bilbao: CITA.

CARRETERO, J., RODRÍGUEZ, S. (1997). "COES: herramienta lingüística de libre distribución para la Lengua Española". *NOVATICA*, 126, p. 39-44.

CARROLL, J., ROELOFFS, R. (1969). "Computer selection of keywords using word-frequency analysis". *American Documentation*, 20, 3, p. 227-233.

CAVANAGH, J.M.A. (1989). "Library applications of knowledge-based systems". *The Reference Librarian*, 8, 23, p. 7-8.

CERDÁ MASSÓ, R. (cord.), (1986). Diccionario de Lingüística. Madrid: Anaya.

CIGANIK, M. (1990). "Automation or subject catloguing and automated thesaurus indexing: future prospects". *Kniznice a Vedeckke Informacie*, 22, 2, p. 49-57.

CLEVELAND, D.B., CLEVELAND, A.D. (1990). Introduction to indexing and abstracting. Englewood: Libraries unlimited, Inc.

CLEVELAND, D.B., CLEVELAND, A.D., WISE, O.B. (1984). "Less than fullest indexing using a non-boolean searching model". *Journal of the American Society for Information Science*, 35, 1, p. 19-28.

CLEVERDON, C.W. (1962). Aslib Cranfield Research Project: report on the testing and analysis of an investigation into the comparative efficiency of indexing systems, Cranfield, England.

CLEVERDON, C.W. (1984). "Optimizing convenient on-line access to bibliographic databases". *Information Services & Use*, 4, p. 37-47.

COLL-VINENT, R. (1988). Información y poder. Barcelona: Herder.

COLL-VINENT, R. (1982). Profesionales de la documentación. Barcelona: A.T.E.

COLL-VINENT, R., BERNAL CRUZ, F.J. (1990). Curso de Documentación. Madrid: Dossat.

COOPER, W.S. (1969). "Is interindexer consistency a hobgoblin?". *American Documentation*, 20, p. 268-278.

COOPER, W.S. (1976). Automatic indexing and indexing for complex systems: an appendix to "Indexing documents by Gedanken experiments", Xeroxed.

COOPER, W.S. (1978). "Indexing documents by Gedanken experimentation". *Journal of the American Society for Information Science*, may, p. 107-119.

COOPER, W.S., MARON, M.E. (1978). "Foundations of probabilistic and utility-theoretic indexing". *Journal of the Association for Computing Machinery*, 25, 1, p. 67-80.

CORET, A., DUCLOY, J., MENILLET, D. (1991). Indexing and the information specialist's workstation. *INIST Info* 8, 2-3.

COUTURE de TROISMONTTS, R. (1975). *Manual de técnicas en documentación*. Buenos Aires: Marymar.

COYAUD, M. (1972). *Linguistique et Documentation*. Paris: Librairie Larousse.

CROFT, W.B., TURTLE, H.R., LEWIS, D.D. (1991). "The use of phrases and structured queries in information retrieval". *Proceedings of the Fourteenth Annual International ACM/SIGIR Conference on Research and Development in Information Retrieval*, A. Bookstein, Y. Chiaramella, G. Salton y V.V. Raghavan (eds.), Chicago, p. 32-45.

CRYSTAL, D. (1984). "Linguistics and indexing". *The Indexer*, 14, 1, p. 3-7.

CHAUMIER, J. (1988). *Le traitement linguistique de l'information*. Paris: Entreprise Moderne d'Édition.

DAMERAU, F.J. (1965). "An experiment in automatic indexing". *American Documentation*, 16, 4, p. 283-289.

DAY, R.A. (1995). *How to write publish & a scientific paper*. Cambridge: University.

DÍEZ CARRERA, C. (1994). *Las industrias de la lengua: panorámica para los gestores de información*. Madrid: Biblioteca Nacional.

DIJK, M., VAN SLYPE, G. (1972). El servicio de documentación frente a la explosión de la información. Buenos Aires: Consejo Superior de Investigaciones Científicas y Técnicas.

DILLON, M. (1982). "Thesaurus-based automatic book indexing". *Information Processing & Management*, 8, 4, p. 167-178.

DOCUMENTACIÓN. (1996). "Presentación de Tesis y documentos similares. ISO 7144, UNE 50-136". *Revista Española de Documentación Científica*, 19, 2, p. 203-218.

DRISCOLL, J., et al. (1991). "The operation and performance of an artificially intelligent keywording system". *Information Processing & Management*, 27, 1, p. 43-54.

ECO, U. (1994). Cómo se hace una tesis. Técnicas y procedimientos de investigación, estudio y escritura. Barcelona: Editorial Gedisa.

ELLIS, D., et al. (1994). "On the creation of hypertext links in full-text documents: measurement of inter-linker consistency". *Journal of Documentation*, 50, 2, p. 67-98.

ESPELT, C. (1995). "Bases teóricas en la enseñanza de lenguajes documentales". *Actas del I Encuentro ISKO-España*, p. 126-134.

ESTEBAN NAVARRO, M.A. (1995). "Aplicaciones de la Terminología para la docencia de la gestión de lenguajes documentales". *Actas del I Encuentro de ISKO-España*, p. 205-212.

EVANS, D.A, et al. (1991). "Automatic indexing of abstracts via natural-language processing using a simple thesaurus". *Med Decis Making*, 11, p. 108-115.

FAGAN, J.L. (1987). Experiments in automatic phrase indexing for document retrieval: a comparison of syntactic and non-syntactic methods. New York: Cornell University.

FAGAN, J.L. (1989). "The effectiveness of a non-syntactic approach to automatic phrase indexing for document retrieval". *Journal of the American Society for Information Science*, 40, 2, p. 115-132.

FARAJ, N., et al. (1996). "Analyse d'une méthode d'indexation automatique basée sur une analyse syntaxique de texte". *Canadian Journal of Information and Library Science*, 21, 1, p. 1-21.

FARROW, J. (1994). Indexing as a cognitive process. En A. Kent, H. Lancour y J.E. Daily (eds). *Encyclopedia of Library and Information Science*, vol. 53, p. 155-171.

FEINBERG, H. (1973). Title derivative indexing techniques: a comparative study. Metuchen: Scarecrow Press.

FEITEN, B., GUNZEL, S. (1994). "Automatic indexing of a sound database using self-organizing neural nets". *Computer Music Journal*, 18, 3, p. 53-65.

FIDEL, R. (1994). "User centered indexing". *Journal of the American Society for Information Science*, 45, 8, p. 572-576.

FOLTZ, P., DUMAIS, S.T. (1992). "Personalized information delivery: an analysis of information filtering methods". *Communications of the ACM*, 35, 12, p. 51-60.

FRENZEL, L.E. (1989). A fondo: Sistemas expertos. Madrid: Anaya Multimedia.

FUGMANN, R. (1990). "Unused possibilities in indexing and classification". Interface. *Proceedings 1st International ISKO Conference*, Darmstadt, 14-17 aug., 1990. Frankfurt: INDEKS Verlagn.

FUNK, M., REID, C.A., McGOOGAR, L.S. (1983). "Indexing consistency in MEDLINE". *Bulletin of the Medical Library Association*, 71, p. 176-183.

GARCÍA GUTIÉRREZ, A. (1984). *Lingüística documental*. Barcelona: Mitre.

GARCÍA GUTIÉRREZ, A. (1989a). "Normalización general y documental: concepto, historia e instituciones". En López Yepes, J. (comp.). *Fundamentos de información y documentación*, Madrid: Eudema, p. 227-260

GARCÍA GUTIÉRREZ, A. (1989b). "Nuevos parámetros para una teoría de la indización de documentos". En José López Yepes (comp.). *Fundamentos de información y documentación*, Madrid: Eudema, p. 381-383

GARCÍA GUTIÉRREZ, A. (1990). *Estructura lingüística de la Documentación: teoría y método*. Murcia: Universidad.

GARCÍA GUTIÉRREZ, A. (1992). *Análisis documental del discurso periodístico*. Madrid: Centro de Tratamiento de la Documentación.

GARCÍA GUTIÉRREZ, A. (1995). "La investigación documentológica: hacia un ajuste interdisciplinar". *Scire*, 1, 2, p.11-20.

GARDIN, J.C. (1973). "Document analysis and linguistics theory". *Journal of Documentation*, 29, p. 137-168.

GIBB, F., ed. (1986). *Experts systems in libraries*. London: Taylor Graham.

GIL LEIVA, I., RODRÍGUEZ MUÑOZ, J.V. (1996a). "Tendencias en los sistemas de indización automática. Estudio evolutivo". *Revista Española de Documentación científica*, 19, 3, p. 273-291.

GIL LEIVA, I., RODRÍGUEZ MUÑOZ, J.V. (1996b). "El procesamiento del lenguaje natural aplicado al análisis del contenido de los documentos". *Revista General de Información y Documentación*, 6, 2, p. 205-218.

GIL LEIVA, I., RODRÍGUEZ MUÑOZ, J.V. (1996c). "Los orígenes del almacenamiento y recuperación de información". *Boletín de la Asociación Andaluza de Bibliotecarios*, 42, p. 9-18.

GIL LEIVA, I., RODRÍGUEZ MUÑOZ, J.V. (1997). "Análisis de los descriptores de diferentes áreas del conocimiento indizadas en bases de datos del CSIC. Aplicación a la indización automática". *Revista Española de Documentación Científica*, 20, 2, p. 150-161.

GIL LEIVA, I. (1997a). "La indización automática". *IBERSID, II Encuentro sobre Sistemas de Información y Documentación*, Zaragoza 17-21 febrero.

GIL LEIVA, I. (1997b). "La investigación en indización y resumen en España". *Educación y Biblioteca*, 83, octubre.

GIL URDICIAIN, B. (1992). "Función de los lenguajes documentales en el tratamiento de la información en las organizaciones". *Revista General de Información y Documentación*, 2, 2, p. 195-200.

GIL URDICIAIN, B. (1994b). "Niveles de análisis documental de contenido". *Documentación de las Ciencias de la Información*, 17, p. 77-84.

GIL URDICIAIN, B. (1996a). *Manual de lenguajes documentales*. Madrid: Noesis.

GIL URDICIAIN, B. (1996b). Comunicación personal, noviembre.

GÓMEZ GUINOVART, J. (1996). "Aportaciones a la metodología de evaluación de los sistemas de verificación automática de la sintaxis". *Revista de la Sociedad Española para el Procesamiento del Lenguaje Natural*, 19, p. 7-13.

GONZÁLEZ, J.W. (ed.), (1990). Aspectos metodológicos de la investigación científica. Un enfoque multidisciplinar. Murcia: Universidad.

GRISHMAN, R. (1991). Introducción a la lingüística computacional. Madrid: Visor distribuciones.

GUINCHAT C., MENOUE, M. (1983). Introducción general a las ciencias y técnicas de la información y de la documentación. París: Unesco.

HALLER, J. (1983). "Análise automática de textos em sistemas de informação". *Revista de Biblioteconomia de Brasília*, 11, 1, p. 105-113.

HALLER, J. (1982). "Processamento de textos em linguagem natural." *Congreso Nacional de Informática*, 15, Rio de Janeiro.

HARTER, S.P. (1975a). "Probabilistic approach to automatic keyword indexing. Parts 1". *Journal of the American Society for Information Science*, 26, 4, p. 197-206.

HARTER, S.P. (1975b). "Probabilistic approach to automatic keyword indexing. Parts 2". *Journal of the American Society for Information Science*, 26, 5, p. 280-289.

HERSH, W.R., et al. (1992). "A comparison of retrieval effectiveness for three methods of indexing medical literature". *The American Journal of the Medical Sciences*, 30, 5, p. 292-300.

HMEIDI, I.I. (1995). Design and implementation of automatic word and phrase indexing for information retrieval with arabic documents (indexing, automatic indexing). Illinois Institute of Technology.

HODGE, G.M. (1992). Automated support to indexing. Philadelphia: National Federation of Abstracting and Information Services, NFAIS Report Series 3.

HODGE, G.M. (1993). "Computer-assisted database indexing: the state-of-the-art". Indexing, providing access to information: looking back, looking ahead. *The Proceedings of the 25th Annual Meeting of the American Society of Indexers* (Alexandria, VA), may 1993, Nancy C. Mulnavy (ed.), Port Aransas: American Society of Indexers, p. 33-44.

HODGE, G.M. (1994). "Computer-assisted database indexing: the state-of-the-art". *The Indexer*, 19, 1, p. 23-27.

HOOPER, R.S. (1965). Indexer consistency test: origin, measurements, results and utilization. Bethesda, MD: IBM Corp.

HUMPHREY, S. (1989). "MedIndex system: medical indexing expert system". *Information Processing & Management*, 25, 1, p. 73-78.

HUMPHREY, S.M. (1994). "Knowledge-based systems for indexing". En R. Fidel (ed.). *Challenges in indexing electronic text and images*. Medford ; New York: ASIS, 1994, p. 161-175

ISO 5963-1985. (1985). Documentation. Méthodes pour l'analyse des documents. la détermination de leur contenu et la sélection des termes d'indexation. Genève: ISO.

ISO. Documentación. (1989). "Directrices para el establecimiento y desarrollo de Tesoros monolingües (Parte II)". PNE - 50 106 ISO 2788-1986 (E). *Revista Española de Documentación Científica*, 12, 4, p. 601-629.

IZQUIERDO ARROYO, J.M., MORENO FERNÁNDEZ, L.M. (1994). "Listas de encabezamientos de materia y thesauri en perspectiva comparada". *Documentación de las Ciencias de la Información*, 17, p. 287-310.

IZQUIERDO ARROYO, J.M., MORENO FERNÁNDEZ, LM. (1995). "Problemas de terminología metalingüística en los lenguajes documentales". *Actas I Encuentro ISKO-España*, 4-5 noviembre 1993, F.J. García Marco (ed.). Zaragoza, p. 51-63.

JACOBY, J. (1962). Methodology for indexer reliability aids on the reliability of indexers. radc-tdr-63-116, Documentation, Inc., Bethesda, Maryland, march.

JONES, L.P., et al. (1990). "INDEX: The statistical basis for an automatic conceptual phrase-indexing system". *Journal of the American Society for Information Science*, 41, 2, p. 87-97.

KARETNYK, D., KARLSSON, F., SMART, G. (1991). "Knowledge-based indexing of morpho-syntactically analysed language". *Expert Systems for Information Management*, 4, 1, p. 1-29.

KLINGBIEL, P.H., RINKER, C. (1976). "Evaluation of Machine-Aided Indexing". *Information Processing & Management*, 12, 6, p. 351-366.

KOROTKIN, A.L., OLIVER, L.H. (1964). The effect of subject matter familiarity and the use of an indexing aid upon inter-indexer consistency, General Electric Company, Bethesda, Maryland, february.

KORYCINSKI, C., NEWELL, A.F. (1990). "Natural-language processing and automatic indexing". *The Indexer*, 17, 1, p. 21-29.

KRAFT, D.A. (1964). "A comparison of keyword in context (KWIC) indexing of titles with a subject heading classification system". *American Documentation*, 15, 1, p. 48-52.

KUNTZ, R. (1991). "The application of Expert Systems to indexing". *Current Studies in Librarianship*, spring/fall, p. 3-26.

KWOK, K.L. (1974). "Cited titles: a new source of keyword extraction for automatic document classification and retrieval". *Proceedings ASIS Annual Meeting*, 37. Atlanta, 13-17, october, Washinton: ASIS, v.11, p. 56-57

LANCASTER, F.W. (1969). "MEDLARS: report on the evaluation of its operating efficiency". *American Documentation*, april, p. 119-142.

LANCASTER, F.W. (1978). "Precision and recall". En A. Kent, H. Lancour y J.E. Daily (eds). *Encyclopedia of Library and Information Science*, vol. 23, p. 170-180.

LANCASTER, F.W. (1991). *Indexing and abstracting in theory and practice*. London: The Library Association.

LEONARD, L.E. (1975). *Inter-indexer consistency and retrieval effectiveness: Measurement of relationships*. Unpublished disserttion, University of Illinois, Urbana-Champaign.

LEONARD, L.E. (1977). *Inter-indexer consistency studies, 1954-1975: a review of the literature and summary of study results*. University of Illinois Graduate School of Library Science Occasional Papers, 131.

LEWIS, D.D. (1992). "An evaluation of phrasal and clustered representations on a text categorization task". *Proceedings of the Fiteenth Annual International*

ACM/SIGIR Conference on Research and Development in Information Retrieval, N. Belkin, et al. (ed.), Copenhagen: ACM Press, p. 37-50.

LIVONEN, M. (1990). "Interindexer consistency and the indexing environment". *International Forum on Information and Documentation*, 15,1, p. 16-21.

LONG, B. (1980). "Linguistique et indexation". *Documentaliste-Sciences de l'information*, 17, 3, p. 99-106.

LOPES GÍNEZ DE LARA, M. (1993). "Algumas contribuições da semiologia e da semiótica para a análise das linguagens documentárias". *Ciencia da Informação, Brasília*, 22, 2, p. 223-226.

LÓPEZ YEPES, J. (1995). *La aventura de la investigación científica. Guía del investigador y del director de investigación*. Madrid: Síntesis.

LOVTSOV, A. (1990). *Automatic indexing in reports and dissertations database*. ICSTI, Internal Report.

LUHN, H.P. (1957). "A statistical approach to mechanized encoding and searching of literary information". *IBM Journal of Research and Development*, 1, 4, p. 309-317.

LUHN, H.P. (1958). "The automatic creation of literature abstracts". *IBM Journal of Research and Development*, 2, 2, p. 159-165.

LUSTIG, G., KNORZ, G. (1986). *AIR/PHYS pilot application project: pilot application of automatic indexing and improved retrieval methods using the PHYS data base (1-30)*. Karlsruhe: Frachinformationszentrum, Energie Physik Mathematik GmbH.

MAHAPATRA, M., CHANDRA BISWAS, S. (1986). "Interdependence of PRECIS role operators: a quantitative analysis of their association". *Journal of the American Society for Information Science*, 37, 1, p. 20-25.

MAIZELL, R. (1960). "Value of titles for indexing purposes". *Revue de la Documentation*, 27, p. 126-127.

MANIEZ, J. (1992). Los lenguajes documentales y de clasificación. Concepción, construcción y utilización en los sistemas documentales. Madrid: FGSR.

MARCETICH, J., SCHUYLER, P. (1981). "The use of AID to promote indexing consistency at the National Library of Medicine". *Eighty-first Annual Meeting of the Medical Library Association*, Montreal, June.

MARKEY, K. (1984). "Interindexer consistency tests: a literature review and report of a test of consistency in indexing visual materials". *Library and Information Science Research*, 6, p. 155-177.

MARON, M.E. (1979). "Depth of indexing". *Journal of the American Society for Information Science*, July, p.224-228.

MARQUES CINTRA, A.M. (1983). "Elementos de lingüística para estudos de indexação". *Ciencia da Informação, Brasília*, 12, 1, p. 5-22.

MARTÍNEZ, C., LUCEY, J., LINDER, E. (1987). "An expert system for machine-aided-indexing". *Journal of Chemical Information and Computer Sciences*, 27, 4, p. 158-162.

MEDLARS manual indexing. (1967). National Library of Medicine.

MEULEN, W.A., JANSSEN, P.J. (1977). "Automatic versus manual indexing". *Information Processing & Management*, 13, 1, p. 13-21.

MICHELL, G. (1979). "The natural language foundations of indexing language relations". *The Canadian Journal of Information Science*, 4, p. 99-104.

MILSTEAD, J.L. (1994). "Needs for research in indexing". *Journal of the American Society for Information Science*, 45, 8, p. 577-582.

MOLINER, M. (1988). *Diccionario de uso del español*. Madrid: Gredos.

MONTGOMERY, C., SWANSON, D.R. (1962). "Title indexing". *American Documentation*, 13, p. 359-364.

MONTGOMERY, C. (1972). "Linguistics and Information Science". *Journal of the American Society for Information Science*, 23, p. 195-219.

MOOERS, C.N. (1972). "Descriptors". En A. Kent y H. Lancour (eds). *Encyclopedia of Library and Information Science*, vol. 7, p. 31-45.

MOOERS, C.N. (1985). "The indexing language of an information retrieval system". En L. Mai Chan (ed.), *Theory of subject analysis*. Littleton: Libraries unlimited, p. 247-261.

MOREIRO GONZÁLEZ, J.A (1989a). "El resumen científico en el contexto de la teoría de la documentación. Texto y descripción sustancial". *Documentación de las Ciencias de la Información*, 12, p. 147-170.

MOREIRO GONZÁLEZ, J.A. (1989b). "Método y tipología del resumen científico". En J. López Yepes (comp.). *Fundamentos de información y documentación*, Madrid: Eudema, p. 414-430.

MOREIRO GONZÁLEZ, J.A. (1990). Introducción bibliográfica y conceptual al estudio evolutivo de la Documentación. Barcelona: PPU.

MOREIRO GONZÁLEZ, J.A. (1993a). Aplicación de las ciencias del texto al resumen documental. Madrid: Universidad Carlos III.

MOREIRO GONZÁLEZ, J.A. (1993b). "Implicaciones documentales en el procesamiento del lenguaje natural". *Ciencias de la Información*, 24, 3, p. 48-54.

MOREIRO GONZÁLEZ, J.A. (1993c). "Perspectiva documental del procesamiento del lenguaje natural". *Sociedad Española para el Procesamiento del Lenguaje Natural*, 13, p. 41-45.

NATALI, J.W. (1978). "Documentação e Lingüística: inter-relação e campos de pesquisa". *Revista Brasileira de Biblioteconomía y Documentação*, 11, 1-2, p. 33-42.

NEET, H.E. (1981). L'analyse documentaire, Ginebra: Institut d'Etudes Sociales.

NOCETI, M.A, FIGUEIREDO, R.C. (1978). "Línguas naturais e linguagens documentárias: traços inerentes e ocorrências de interação". *Revista Brasileira de Biblioteconomía y Documentação*, 6, 1, p. 23-27.

ORTEGA CAVERO, D. (1991). Thesaurus gran Sopena de sinónimos y asociación de ideas. Barcelona: Ramón Sopena.

PAINTER, A.F. (1963). An analysis of duplication and consistency of subject indexing involved in report handling at the Office of Technical Services. U.S, Department of Commerce, Office of Technical Services, Washington, D.C., march.

PALMA VILLALÓN, M^a del Valle. (1995). "Técnicas y métodos para mejorar la calidad de la indización y su recuperación en bases de datos documentales de

ciencias sociales y humanidades". 5es Jornades Catalanes de Documentació, p. 223-239.

PÉREZ ÁLVAREZ-OSSORIO, J.R. (1988). Introducción a la información y documentación científica. Madrid: Alhambra.

PICHT, H. (1996). "La terminología como factor de desarrollo". *Revista Interamericana de Bibliotecología*, 19, 1, p. 7-29.

PINTO MOLINA, M. (1989). El análisis de contenido: la indización de documentos. En J. López Yepes (comp.). *Fundamentos de información y documentación*, Madrid: Eudema, p. 347-374.

PINTO MOLINA, M. (1993). *Análisis documental: fundamentos y procedimientos*. Madrid: Eudema.

POLLITT, A.S. (1981). "An expert system as an on-line search intermediary". *Proceedings of the 5th International On-line Information Meeting*. Oxford: Learned Information, p. 25-32.

QUINN, B. (1994). "Recent theoretical approaches in classification and indexing". *Knowledge Organization*, 21, 3, p. 140-147.

RABITTI, F., SAVINO, P. (1992). "Automatic image indexation to support content-based retrieval". *Information Processing & Management*, 28, 5, p. 547-565.

REAL Academia de la Lengua Española. (1992). *Diccionario de la lengua española* [21ª ed.]. Madrid: Espasa-Calpe.

REICH, P., BIEVER, E.J. (1991). "Indexing consistency: the input/output function of thesaurus". *College and Research Libraries*, 52, 4, p. 336-342.

RIBEIRO PINHEIRO, L.V., MATHEUS LOUREIRO, J.M. (1995). "Traçados e limites da ciência da informação". *Ciência da Informação, Brasília*, 24, 1, p. 42-53.

RICHTER, N. (1988). *Grammaire de l'indexation alphabétique*. Le Mans: Bibliothèque de l'Université du Maine.

ROBERTSON, S.E. (1972). "Term specificity". *Journal of Documentation*, 28, p. 164-165.

ROBERTSON, S.E. (1974). "Specificity and weighted retrieval". *Journal of Documentation*, 30, 1974, p. 41-46.

ROBERTSON, S.E., HARDING, P. (1984). "Probabilistic automatic indexing by learning from human indexers". *Journal of Documentation*, 40, 4, p. 264-270.

ROBREDO, J.A. (1991). "Indexação automática de textos: uma abordagem otimizada e simples". *Ciência da Informação, Brasília*, 20, 2, p. 130-136.

RODGERS, D.J. (1961). A study of intra-indexer consistency, General Electric company, Washington, D.C., september.

RODRÍGUEZ MUÑOZ, J.V., et al. (1992). Reglas de producción para una base del conocimiento en la construcción de tesauros. En *Proceedings of the Forty-Sixth FID Conference and Congress*, october, p. 341-352.

RODRÍGUEZ MUÑOZ, J.V. (1994). *Construcción del esquema conceptual del tesoro mediante un modelo de datos*. Murcia: Universidad.

ROLLAND-THOMAS, P. (1995). "Essai sur la contribution de l'anthropologie culturelle aux fondements de la classification documentaire". *Documentation et bibliothèques*, janvier-mars, p. 7-18.

ROLLING, L. (1981). "Indexing consistency, quality and efficiency". *Information Processing & Management*, 17, p. 69-76.

ROMANO, D. (1987). Elementos y técnica del trabajo científico. Barcelona: Teide.

ROSENBERG, V. (1971). "A study of statistical measures for predicting terms used to index documents". *Journal of the American Society for Information Science*, 22, 1, p. 41-50.

ROWLEY, J. (1988). Abstracting and indexing. London: Clive Bingley.

RUHL, M.J. (1964). "Chemical documents and their titles: human concept indexing vs KWIC- machine indexing". *American Documentation*, 15, 2, p. 136-141.

SALTON, G. (1972). A new comparison between conventional indexing (MEDLARS) and automatic text processing (SMART). *Journal of the American Society for Information Science*, march-april, p. 75-84.

SALTON, G., YANG, C.S. (1973). "On the specification of term values in automatic indexing". *Journal of Documentation*, 29, 4, p. 351-372.

SALTON, G., YANG, C.S., YU, C.T. (1975). "A theory of term importance in automatic text analysis". *Journal of the American Society for Information Science*, 26, 1, p. 33-44.

SALTON, G. (1980). "The SMART system 1961-1976: Experiments in dynamic document processing". En A. Kent, H. Lancour y J.E. Daily (eds). *Encyclopedia of Library and Information Science*, vol. 28, p. 1-28.

SALTON, G., MCGILL, M.J. (1983). Introduction to modern information retrieval. New York: McGraw-Hill.

SALTON, G., BUCKLEY, C. (1988). "Term weighting approaches in automatic text retrieval". *Information Processing & Management*, 24, 5, p. 513-523.

SARACEVIC, T. (1995). "Interdisciplinary nature of information science". *Ciencia da Informaçao, Brasília*, 24, 1, p. 36-41.

SCHEELE, M. (1983). "Automatic indexing of titles and keywords on the bases of a model for an overall thesaurus of knowledge". *International Classification*, 10, 3, p. 135-137.

SCHUEGRAF, E.J., BOMMEL, M.F. (1993). "An automatic document indexing system based on cooperating expert systems: design and development." *Canadian Journal of Information and Library Science*, 18, 2, p. 32-50.

SEO, E. (1993). An experiment in automatic indexing with korean texts: a comparison of syntactic-statistical and manual methods (indexing, manual indexing). University of Illinois at Urbana-Champaign.

SHOVAL, P. (1985). "Principles, procedures and rules in an expert system for information retrieval". *Information Processing & Management*, 21, p. 375.

SIERRA BRAVO, R. (1994). Tesis doctorales y trabajos de investigación científica. Madrid: Editorial Paraninfo.

SIEVERT, M.E., ANDREWS, M.J. (1991). "Indexing consistency in Information Science Abstracts". *Journal of the American Society for Information Science*, 42, 1, p. 1-6.

SIEVERT, M.E., VERBECK, A. (1987). "The indexing of the literature of online searching: a comparison of ERIC and LISA". *Online Review*, 11, p. 95-104.

SILVA, B. (1972). Origen e evolução dos descritores. Río de Janeiro: Fundação Getúlio Vargas.

SILVESTER, J.P., GENUARDI, M.T., KLINGBIEL, P.H. (1994). "Machine-aided indexing at NASA". *Information Processing & Management*, 30, 5, p. 631-645.

SIMÓN GRANDA, J., LEMA GARZÓN, E. (1990). "Primeras experiencias sobre el análisis de textos en castellano aplicado a la indexación automática de información". *Terceras Jornadas Españolas de Documentación Automatizada*, p. 1255-1270.

SLAMECKA, V., JACOBY, J. (1962). Indexer consistency under minimal conditions. RADC-TDR-62-426, Documentation, In., Bethesda, Maryland, november.

SLAMECKA, V., JACOBY, J. (1963). Effect of indexing aids on the reliability of indexers. RADC-TDR-63-116, Documentation, Inc., Bethesda, Maryland, june.

SLYPE, G. (1991). Los lenguajes de indización: Concepción, construcción y utilización en los sistemas documentales. Madrid: Fundación Germán Sánchez Ruiperez.

SMIT, J.W. (1974). "Análise semântica e análise documentária". *Revista Brasileira de Semiótica*, 1, p. 168-176.

SOERGEL, D. (1994). "Indexing and retrieval performance: the logical evidence". *Journal of the American Society for Information Science*, 45, 8, p. 589-599.

SPARCK JONES, K. (1972). "A statistical interpretation of term specificity and its application in retrieval". *Journal of Documentation*, 28, p. 11-21.

SPARCK JONES, K. (1973). "Does indexing exhaustivity matter". *Journal of the American Society for Information Science*, 24, 5, p. 313-316.

SPARCK JONES, K., KAY, M. (1973). *Linguistics and information science*. New York: Academic Press.

SPARCK JONES, K. (1986). "Getting started in computerized indexing". *The Indexer*, 15, 1, p. 12.

SPEIGHT, F.Y. (1967). *Guide for source indexing and abstracting of the engineering literature*. New York: EJC.

STEVENS, M.E. (1965). *Automatic indexing: a state of the art report*, Monograph 91, National Bureau of Standards, Washington, D.C.

STRODE, M. S.. (1977). *Automatic indexing using a thesaurus*. The University of North Carolina at Chapel.

SVENONIUS, E. (1972). "An experiment in index term frequency". *Journal of the American Society for Information Science*, 23, 2, p. 109-121.

TESAURO EUROVOC. (1995). Luxemburgo: Oficina de Publicaciones Oficiales de las Comunidades Europeas.

TESAURO SPINES. (1988). MADRID: Instituto de Información y Documentación en Ciencia y Tecnología.

THEORY of subject analysis: a sourcebook. (1985). Mai Chan, P.A. Richmond y E. Svenonius (eds.). Littleton: Libraries Unlimited.

TONTA, Y. (1991). "A study of indexing consistency between Library of Congress and British Library cataloguers". *Library Resources and Technical Services*, 35, 2, p. 177-185.

UNESCO: Principes d'indexation. (1975). Paris: Unesco.

VALLE BRACERO, A., FERNÁNDEZ GARCÍA, J.A. (1983). "Automatización de la indización y coordinación de descriptores". *Revista Española de Documentación Científica*, 6, 1, p. 9-16.

VERA LUJÁN, A. Fundamentos de análisis sintáctico: De la palabra al texto. Murcia: Universidad, 1994.

VERDEJO MAILLO, M.F. (1994). Comprensión del lenguaje natural: Avances, aplicaciones y tendencias. Procesamiento del lenguaje natural: Fundamentos y aplicaciones, p. 5-29.

WALKER, D.E. (1981). "The organization and use of information: contributions of Information Science, Computational Linguistics and Artificial Intelligence". *Journal of the American Society for Information Science*, september, p. 347-363.

WAN, T. (1995). Experiments with automatic indexing and a relational thesaurus in a chinese information retrieval system. Illinois Institute of Technology.

WARD, M. "The future of the human indexer". *Journal of Librarianship and Information Science*, 28, 4, 1996, p. 217-225.

WARNER, A.J. (1991). "Quantitative and Qualitative Assessments of the impact of Linguistic Theory on Information Science". *Journal of the American Society for Information Science*, 42, 1, p. 64-71.

WEINBERG, B.H. (1981). Word frequency and automatic indexing. Columbia University.

WELLISCH, H.H. (1992). The art of indexing and some fallacies of its automation. *Logos*, 3, 2, p. 69-76.

WONG, S.K.M., YAO, Y.Y. (1992). "An information-theoretic measure of term specificity". *Journal of the American Society for Information Science*, 43, 1, p. 54-61.

WOODRUFF, A. C. PLAUNT. (1994). "GIPSY: Automated geographic of text documents". *Journal of the American Society for Information Science*, 45, 9, p. 645-655.

WU, H., SALTON, G. (1981). "A comparison of search term weighting: term relevance vs. inverse document frequency". *Proceedings of the Fourth ACM SIGIR Conference on Research and Development in Information Retrieval*, ACM SIGIR Forum, V XI, p. 30-39.

XIVRY, O. (1993). "Le traitement de l'information textuelle utilisation du systeme "SPIRIT": (Système Probabiliste d'indexation et de Recherche d'Informations Textuelles)". *Cahiers de la Documentacion*, 1, p. 15-23.

YU, C.T., SALTON, G. (1976). "Precision weighting an effective automatic indexing method". *Journal of the Association for Computing Machinery*, 23, p. 76-88.

ZIPF, G.Z. (1949). *Human behavior and the principle of least effort*. Massachussets.

ZUNDE, P., DEXTER, M.E. (1969). "Indexing consistency and quality". *American Documentation*, july, p. 259-267.

Anexo 1: Glosario*

Álgebra booleana: Sistema matemático de funciones lógicas utilizado en la recuperación de la información que relaciona entre sí los términos por medio de los operadores AND (Y), OR (O) y NOT (NO).

Algoritmo: Proceso definido o conjunto de reglas secuenciales y preestablecidas para la resolución de un problema, específicamente en un ordenador.

Almacenamiento de la información: Proceso de introducir datos en un fichero para su retención temporal o permanente y su posterior recuperación.

Análisis de contenido [documental]: Conjunto de procedimientos encaminados a analizar el contenido de un documento para obtener y representar su estructura cognitiva.

Análisis documental: Fase del proceso documental que tiene por objeto el control y representación abreviada de los datos formales y de contenido de un documento.

Catálogo: Lista ordenada de los asientos bibliográficos de una colección.

Centro de documentación: Organismo que asume las funciones de selección, tratamiento y difusión de la información a distintos niveles.

Clasificación: Lenguaje documental basado en la representación estructurada de uno o varios dominios del conocimiento en clases en las que las nociones y sus relaciones son representadas por una notación.

Coefficiente de exhaustividad en la recuperación: Coeficiente resultante entre el número de documentos relevantes recuperados y el número total de documentos relevantes en la colección.

Coefficiente de precisión en la recuperación: Coeficiente entre los documentos relevantes recuperados de un sistema, según los criterios de un usuario o cliente, y el número total de documentos recuperados.

Consistencia: grado de coincidencia en la representación de la información esencial de un documento por medio de términos de indización entre varios indizadores o entre un indizador y un sistema semiautomático o automático.

Cuasisinónimo: Aquellos términos en los que la significación es generalmente diferente en el uso normal de dicho término, pero que se puede considerar como sinónimos para las necesidades particulares de la indización.

* Para la definición de algunos términos se ha utilizado Guinchat y Menou [1983], Richter [1988], Pinto Molina [1993], Amat [1994] y Gil Urdiciain [1996].

Descriptor: Palabra o frase en lenguaje natural o controlado que representa el contenido de un documento o una pregunta utilizada en el almacenamiento y la recuperación de la información.

Descriptor principal: Refleja los conceptos esenciales del documento.

Descriptor secundario: Refleja conceptos accesorios o vagos del documento. Son útiles en combinación con los descriptores principales.

Diacrónico: Inscrito en el tiempo y de manera general unido a una evolución o a una historia.

Difusión de la información: Transmitir al usuario o al cliente la información que necesita o darle la opción de que la obtenga.

Documento: Unidad material o virtual de conservación y transmisión de información.

Encabezamiento de materia: Una o varias palabras que representan conceptos. Condensa el tema sobre el que trata un documento.

Encabezamiento compuesto: Uso de dos o más términos para expresar el tema de un documento.

Entropía documental: Principio de economía de lenguaje en virtud del cual debe expresarse la mayor cantidad de información con el mínimo número de palabras.

Identificadores: Palabras o frases clave muy específicas que se sacan del texto como nombres de universidades, acrónimos, nombres de empresas, hospitales así como ríos, ciudades o países.

Índice: Serie ordenada de puntos de acceso o entradas que conducen desde la información conocida por el usuario a otra desconocida.

Índice KWIC: Índice generado por ordenador, en el que han sido permutadas todas las palabras clave del título de un documento, con un número limitado de caracteres, para presentarlas ordenadas alfabéticamente de manera que cada una de ellas aparezca como un punto de acceso.

Indización: Operación documental dirigida a representar por medio de un lenguaje documental o natural los datos resultantes del análisis del contenido de un documento o de una demanda de información.

Indización postcoordinada: La combinación de términos de indización representativos del contenido de los documentos, se lleva a cabo en el momento de la búsqueda.

Indización precoordinada: Sistema que combina los diferentes conceptos (por tanto, términos de indización) que contiene un documento en el momento de la indización, esto es, antes de su almacenamiento.

Información: Mensaje que es transmitido por el emisor al receptor con ayuda de señales durante la comunicación.

Lenguaje controlado: Léxico construido con la ayuda de un conjunto de reglas, que puede unívoca y limitadamente, representar el contenido de documentos y demandas de información.

Lenguaje documental: Conjunto de términos normalizados y controlados vinculados entre sí a través de una sintaxis específica para expresar cómo deben utilizarse los términos en la fase de análisis o para explicitar sus relaciones de sinonimia, jerarquía, proximidad temática o antonimia. Son herramientas empleadas en la indización de los documentos (almacenamiento) y para la indización de las preguntas (recuperación).

Lenguaje jerárquico: Sigue un orden lógico para agrupar y aproximar los conceptos más específicos dentro de los conceptos más generales.

Lenguaje libre: Léxico en lenguaje natural, entresacado del documento y no manipulado por el documentalista.

Nota de aplicación o de alcance: Precisan el sentido de un descriptor y determinan su uso en las labores de indización.

No descriptor: Palabra o frase en lenguaje natural o controlado no utilizada para representar el contenido de un documento o una pregunta que remite al descriptor.

Palabra clave: Palabra o grupo de palabras seleccionadas bien del título, del resumen o del texto de un documento, o bien de una demanda documental, para caracterizar el contenido de dicho documento o demanda.

Palabra vacía: Palabra que no transmite información respecto al contenido de un documento.

PRECIS: Sistema de indización desarrollado por la Bibliografía Nacional Británica, en el que una serie de descriptores asignados por un indizador es manipulada por un ordenador en diversas combinaciones según un sistema de operadores de relación.

Procesamiento de la información: Paradigma de la psicología cognitiva cuyo rasgo más característico es la consideración de la mente humana como un sistema activo capaz de manipular símbolos.

Proceso documental: Conjunto de operaciones sucesivas de recogida, tratamiento y difusión de documentos e informaciones.

Recuperación de la información: Procedimiento para localizar o recuperar y facilitar información que ha sido solicitada y que trata de un tema específico.

Referencia bibliográfica: Conjunto de elementos aparentes y convencionales que posibilitan la identificación precisa y formal de un documento.

Resumen: Producto secundario resultante del complejo operativo de resumir, y consiste en la representación abreviada de la macroestructura global del texto original.

Ruido documental: Datos obtenidos en la recuperación que sobrepasan en profundidad, superficialidad o extensión los estrictamente solicitados.

Servicios de indización y resumen: Servicios secundarios encargados de la elaboración y difusión de boletines de índices y revistas de resúmenes.

Servicios secundarios: Servicios de actualización bibliográfica que suministran acceso rápido a la literatura primaria publicada y analizada.

Silencio documental: Datos solicitados al sistema en el módulo de búsqueda pero no obtenidos, aún existiendo, debido a distorsión en el proceso.

Sinónimo: Término que tiene el mismo sentido que otro, siendo prácticamente intercambiables.

Subencabezamiento de materia: Palabra o palabras que siguen a un encabezamiento para formar un encabezamiento compuesto.

Terminología: Conjunto organizado de términos en un dominio especializado, en el que los significados han sido explicados o definidos.

Término de indización: Palabra, frase o cualquier otra representación alfabética o numérica que indique el contenido de los documentos.

Término no preferente: palabra clave no utilizada para representar el contenido de un documento o pregunta que remite al descriptor.

Término preferente: Cada una de las palabras clave significativas que expresan y representan el contenido de un documento.

Tesauro: Lenguaje documental controlado que basa las relaciones entre sus términos en los operadores de jerarquía, definitorios, preferenciales y asociativos.

Unitérmino: Descriptor constituido por una única palabra.

Anexo 2: Definiciones sobre Indización

DEFINICIONES	AUTOR/ES
“Caracterizar determinado tema de información mediante el recurso de vincularlo coherentemente con una palabra clave o una frase clave”.	Silva [1972, p. 18]*
“Extraer de un documento o contexto conceptos, palabras o temas”.	Couture de Troismonts [1975, p. 58]*
“Describir y caracterizar un documento con la ayuda de representaciones de los conceptos contenidos en dicho documento para permitir una búsqueda eficaz de las informaciones contenidas en un fondo documental”.	UNESCO [1975, p. 2]**
“Analizar el contenido informacional de los registros del conocimiento y expresar el contenido informacional en el lenguaje del sistema de indización”.	Borko y Bernier [1978, p. 8]
“Analizar los documentos y aislar, en la riqueza del lenguaje natural empleado por los autores, todos los conceptos esenciales o pertinentes que deben ser retenidos en previsión de búsquedas posteriores”.	Neet [1981, p. 9]**
“Extraer de un documento original o de su resumen unos vocablos especialmente expresivos y con una enorme carga informativa”.	Coll-Vinent [1982, p. 88]
“Escoger los términos más apropiados para representar el contenido de un documento”.	Guinchat y Menou [1983, p. 177]
“Describir el contenido de documentos o	García Gutiérrez

* Citado por García Gutiérrez [1984, p. 104].

** Citado por Pinto Molina [1993, 208].

demandas documentales para posibilitar la elaboración de estrategias de recuperación mediante conceptos o materias”.	[1984, p. 105]
“Caracterizar determinado tema de información mediante el recurso de vincularlo coherentemente con una palabra-clave o una frase-clave”.	Coll-Vinent [1988, p. 99]
“Extraer de un documento original o de su resumen unos vocablos especialmente expresivos y con enorme carga informativa -las palabras clave-, muy indicativos del contenido esencial del documento indizado”.	Coll-Vinent [1990, p. 133]
“Representar el resultado del análisis de un documento por medio de un lenguaje de indización natural o controlado”.	Wellisch [1991] ¹
“El AD que tiene por objetivo la recuperación del documento mediante la representación de los trazos descriptivos de su contenido mediante términos, sintagmas u otras unidades discretas se denomina tradicionalmente indización”.	García Gutiérrez [1992, p. 31]

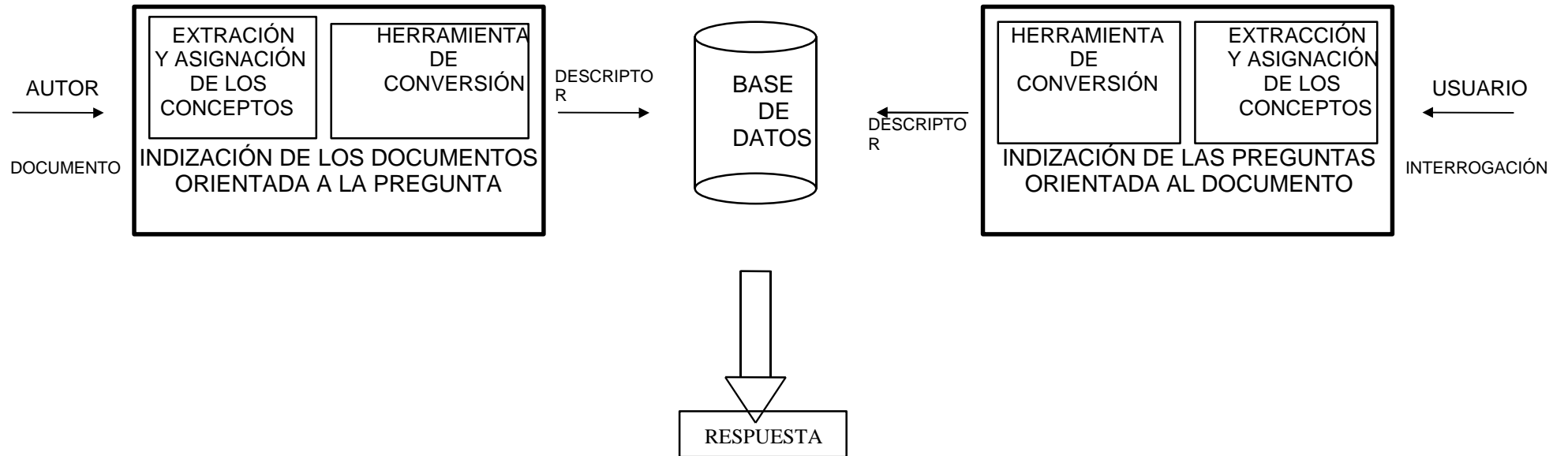
¹ Citado por Fidel [1994, p. 572].

Anexo 3: Etapas en la Indización

	ETAPAS	AUTOR/ES
<u>Dos etapas</u> :	- a) reconocimiento y extracción de conceptos informativos, y b) traducción de esos conceptos al lenguaje documental.	Chaumier [1988, p.23]
	- a) análisis del contenido que resulta en la selección de los conceptos para representar el documento, y b) traducción de los conceptos seleccionados al lenguaje de indización utilizado por el sistema de información o base de datos.	Fidel [1994, p. 573]
<u>Tres etapas</u> :	- a) Examinar el documento y establecer los contenidos de la materia, b) identificar los conceptos principales del tema o de los documentos, y c) traducir los conceptos o términos seleccionados en términos del lenguaje de indización.	Amat [1989, p. 177]
	- a) examen del documento y determinación de su contenido, b) identificación y selección de los conceptos principales de su contenido, y c) selección de los términos de indización.	UNE 50-121-91 [p. 155]
<u>Cuatro etapas</u> :	-a) conocimiento del contenido conceptual del documento, b) extracción de los conceptos en lenguaje natural, c) traducción de esos conceptos al lenguaje documental, y d) búsqueda de otros conceptos pertinentes no expresados por el autor.	Dijk y Slype [1972, p. 105]
	- a) toma de contacto con el documento, b) identificación de los conceptos explícitos e implícitos del documento c) traducción de los conceptos expresados en lenguaje	[Slype, 1992, p. 116]

	natural a descriptores, y d) establecimiento de enlaces sintácticos entre los descriptores.	
<u>Cinco o más etapas</u> :	- a) revisión de los objetivos de esta operación, si se considera necesario, b) conocimiento previo del documento, c) determinación de su tema principal, d) identificación de los elementos del contenido que deben describirse y la extracción de los términos correspondientes, e) verificación de la pertinencia de los términos elegidos, f) traducción de los términos del lenguaje natural a los términos correspondientes del lenguaje documental, si procede, g) verificación de la pertinencia de esta descripción, y h) formalización de la descripción cuando el sistema prevé reglas especiales de presentación o de escritura.	Guinchat y Menou [1983, p. 179]
	- a) registros de los datos bibliográficos, b) análisis del contenido de los documentos a partir del título, resumen y texto completo, c) determinación de los temas, d) conversión de los conceptos extraídos al lenguaje de indización, y e) reexaminación de la indización.	Cleveland y Cleveland [1990, p. 104]

Anexo 4: Proceso Íntegro de la Indización



Anexo 5: Palabras vacías

- | | | |
|----------------|-----------------|---------------|
| 1. A | 50.CIERTA | 99.EFECTO |
| 2. ABAJO | 51.CIERTAMENTE | 100.EL |
| 3. ACASO | 52.CIERTAS | 101.ELLA |
| 4. ACUERDO | 53.CIERTO | 102.ELLO |
| 5. ADELANTE | 54.CIERTOS | 103.ELLOS |
| 6. ADEMÁS | 55.CINCO | 104.EMBARGO |
| 7. ADENTRO | 56.COMO | 105.EN |
| 8. ADÓNDE | 57.COMPAÑÍA | 106.ENCIMA |
| 9. AFUERA | 58.CON | 107.ENSEGUIDA |
| 10.AHÍ | 59.CONDICIÓN | 108.ENTONCES |
| 11.AHORA | 60.CONFORME | 109.ENTRE |
| 12.AL | 61.CONSIGUIENTE | 110.ES |
| 13.ALGO | 62.CONTRA | 111.ESA |
| 14.ALGÚN | 63.CONTRARIO | 112.ESAS |
| 15.ALGUNA | 64.CUAL | 113.ESE |
| 16.ALGUNAS | 65.CUALES | 114.ESO |
| 17.ALGUNO | 66.CUALESQUIERA | 115.ESOS |
| 18.ALGUNOS | 67.CUALQUIER | 116.ESTA |
| 19.ALREDEDOR | 68.CUALQUIERA | 117.ESTAN |
| 20.ALLÍ | 69.CUÁN | 118.ESTAS |
| 21.AMBAS | 70.CUANDO | 119.ESTE |
| 22.AMBOS | 71.CUANTA | 120.ESTO |
| 23.AMPLIAMENTE | 72.CUANTAS | 121.ESTOS |
| 24.ANTE | 73.CUANTO | 122.EXCEPTO |
| 25.ANTES | 74.CUANTOS | 123.FAVOR |
| 26.APENAS | 75.CUARTO | 124.FIN |
| 27.AQUEL | 76.CUATRO | 125.FRENTE |
| 28.AQUELLA | 77.CUYA | 126.FUERA |
| 29.AQUELLAS | 78.CUYAS | 127.GRACIAS |
| 30.AQUELLO | 79.CUYO | 128.HACIA |
| 31.AQUELLOS | 80.CUYOS | 129.HASTA |
| 32.AQUÍ | 81.DE | 130.HAY |
| 33.ARRIBA | 82.DEBAJO | 131.HOY |
| 34.ASÍ | 83.DEL | 132.INCLUSIVE |
| 35.ASIMISMO | 84.DELANTE | 133.INCLUSO |
| 36.ATRÁS | 85.DEMASIADA | 134.JAMÁS |
| 37.AUN | 86.DEMASIADAS | 135.JUNTO |
| 38.AUNQUE | 87.DEMASIADO | 136.LA |
| 39.BAJO | 88.DEMASIADOS | 137.LADO |
| 40.BASTANTE | 89.DENTRO | 138.LAS |
| 41.BASTANTES | 90.DESDE | 139.LE |
| 42.BIEN | 91.DESPUÉS | 140.LEJOS |
| 43.BUSCA | 92.DETRÁS | 141.LES |
| 44.CADA | 93.DIEZ | 142.LO |
| 45.CAMBIO | 94.DOBLE | 143.LOS |
| 46.CASI | 95.DÓNDE | 144.LUEGO |
| 47.CAUSA | 96.DOS | 145.LUGAR |
| 48.CERCA | 97.DURANTE | 146.MAL |
| 49.CERO | 98.E | 147.MANERA |

- | | | |
|-----------------|-------------------------|-------------|
| 148.MAÑANA | 201.POR | 253.TERCERO |
| 149.MAS | 202.PORQUE | 254.TERCIO |
| 150.MEDIANTE | 203.POSTERIORI | 255.TODA |
| 151.MEDIO | 204.PRETEXTO | 256.TODAS |
| 152.MEJOR | 205.PRIMER | 257.TODAVÍA |
| 153.MENOS | 206.PRIMERA | 258.TODO |
| 154.MENUDO | 207.PRIMERAS | 259.TODOS |
| 155.MERCED | 208.PRIMEROS | 260.TRAS |
| 156.MI | 209.PRINCIPAL-
MENTE | 261.TRES |
| 157.MIENTRAS | 210.PRIORI | 262.U |
| 158.MISMO | 211.PRONTO | 263.UN |
| 159.MITAD | 212.PUES | 264.UNA |
| 160.MODO | 213.PUESTO | 265.UNAS |
| 161.MOMENTO | 214.PUNTO | 266.UNO |
| 162.MOTIVO | 215.QUE | 267.UNOS |
| 163.MUCHA | 216.QUIEN | 268.VARIAS |
| 164.MUCHAS | 217.QUIENES | 269.VARIOS |
| 165.MUCHO | 218.QUINTO | 270.VECES |
| 166.MUCHOS | 219.QUIZÁ | 271.VEZ |
| 167.MUY | 220.REGULAR | 272.Y |
| 168.NADA | 221.RESPECTO | 273.YA |
| 169.NADIE | 222.SALVO | |
| 170.NI | 223.SE | |
| 171.NINGÚN | 224.SEA | |
| 172.NINGUNA | 225.SEGÚN | |
| 173.NINGUNAS | 226.SEGUNDO | |
| 174.NINGUNO | 227.SEIS | |
| 175.NINGUNOS | 228.SENDAS | |
| 176.NOS | 229.SENDOS | |
| 177.NOSOTROS | 230.SI | |
| 178.NUESTRO | 231.SIEMPRE | |
| 179.NUESTROS | 232.SIETE | |
| 180.NUEVAMEN-TE | 233.SIN | |
| 181.NUEVE | 234.SINO | |
| 182.NUNCA | 235.SO | |
| 183.O | 236.SOBRE | |
| 184.OBSTANTE | 237.SOLAMENTE | |
| 185.OCHO | 238.SOLO | |
| 186.OTRA | 239.SON | |
| 187.OTRAS | 240.SU | |
| 188.OTRO | 241.SUS | |
| 189.OTROS | 242.SUYO | |
| 190.PARA | 243.TAL | |
| 191.PARTE | 244.TALES | |
| 192.PASO | 245.TAMBIÉN | |
| 193.PENA | 246.TAMPOCO | |
| 194.PEOR | 247.TANTA | |
| 195.PERO | 248.TANTAS | |
| 196.PESAR | 249.TANTO | |
| 197.POCA | 250.TANTOS | |
| 198.POCAS | 251.TARDE | |
| 199.POCO | 252.TEMPRANO | |
| 200.POCOS | | |

ANEXO 7: Índices de consistencia resultantes

ARTÍCULO 3

TÉRMINOS ASIGNADOS DE MODO AUTOMÁTICO

1. SISTEMAS DE INFORMACIÓN
2. LIBROS ELECTRONICOS
3. CD-ROM
4. DISEÑO
5. METODOLOGIA
6. FASES
7. COSTES
8. AUTORES

TÉRMINOS ASIGNADOS DE MODO MANUAL

1. LIBROS ELECTRONICOS
2. CD-ROM
3. DISEÑO
4. PROPIEDAD INTELECTUAL
5. TECNOLOGIA
6. ANALISIS
7. PROBLEMATICA

$$C_i = \frac{3}{7 + 8 - 3} = 0,25 = 25\% \text{ de consistencia}$$

ARTÍCULO 4

TÉRMINOS ASIGNADOS DE MODO AUTOMÁTICO

1. SISTEMAS DE INFORMACION
2. USUARIOS

TÉRMINOS ASIGNADOS DE MODO MANUAL

1. ORGANIZACIONES
2. TRABAJO EN GRUPO
3. SISTEMA DE INFORMACION
4. PARTICIPACION EN GRUPO
5. USUARIOS
6. PRODUCTIVIDAD CIENTIFICA
7. NECESIDAD DE INFORMACION
8. DIFUSION DE LA INFORMACION

$$C_i = \frac{2}{8 + 2 - 2} = 0,25 = 25\% \text{ de consistencia}$$

ARTÍCULO 5

TÉRMINOS ASIGNADOS DE MODO AUTOMÁTICO

1. PROFESIONALES DE LA INFORMACION
2. SERVICIOS DE INFORMACION
3. CALIDAD
4. USUARIOS
5. ARTICULOS

TÉRMINOS ASIGNADOS DE MODO MANUAL

1. INFORMACION
2. DOCUMENTACION
3. PROFESIONALES
4. RESPONSABILIDAD PROFESIONAL
5. DEONTOLOGIA

$$C_i = \frac{0}{5 + 5 - 0} = 0 = 0\% \text{ de consistencia}$$

ARTÍCULO 6

TÉRMINOS ASIGNADOS DE MODO AUTOMÁTICO

- 1. BASES DE DATOS
- 2. DICCIONARIOS

TÉRMINOS ASIGNADOS DE MODO MANUAL

- 1. BASES DE DATOS
- 2. REGISTROS BIBLIOGRAFICOS
- 3. ORTOGRAFIA
- 4. GRABACION DE DATOS
- 5. CONTROL DE CALIDAD
- 6. EVALUACION

$$C_i = \frac{1}{6 + 2 - 1} = 0,1428 = 14,28 \% \text{ de consistencia}$$

ARTÍCULO 7

TÉRMINOS ASIGNADOS DE MODO AUTOMÁTICO

- 1. BIBLIOTECAS
- 2. DEMANDA DE INFORMACION
- 3. EVALUACION

TÉRMINOS ASIGNADOS DE MODO MANUAL

- 1. BIBLIOTECAS
- 2. INFORMATIZACION
- 3. SISTEMAS DE GESTION DOCUMENTAL
- 4. EVALUACION
- 5. CAMBIO TECNOLOGICO
- 6. AGENTES DEL CAMBIO
- 7. NECESIDADES
- 8. ESPECIFICACIONES
- 9. PLAN ESTRATEGICO

$$C_i = \frac{2}{9 + 3 - 2} = 0,20 = 20\% \text{ de consistencia}$$

ARTÍCULO 8

TÉRMINOS ASIGNADOS DE MODO AUTOMÁTICO

- 1. RECUPERACION DE INFORMACION
- 2. CATALOGOS
- 3. ORDENADORES
- 4. CATALOGOS AUTOMATIZADOS
- 5. ASIENTOS BIBLIOGRAFICOS
- 6. IFLA
- 7. ISO
- 8. LIBRARY OF CONGRESS
- 9. BRITISH LIBRARY
- 10. AMERICAN LIBRARY ASSOCIATION
- 11. BIBLIOGRAFIA

TÉRMINOS ASIGNADOS DE MODO MANUAL

- 1. CATALOGOS
- 2. REGISTROS BIBLIOGRAFICOS
- 3. NORMAS
- 4. CATALOGOS AUTOMATIZADOS
- 5. REGLAS DE CATALOGACION
- 6. ORDENACION
- 7. MANTENIMIENTO
- 8. LIBRARY ASSOCIATION

$$C_i = \frac{3}{8 + 11 - 3} = 0,1875 = 18,75 \% \text{ de consistencia}$$

ARTÍCULO 9

TÉRMINOS ASIGNADOS DE MODO AUTOMÁTICO

1. REDES DE COMUNICACIONES
2. SISTEMAS DE INFORMACION
3. PUBLICACIONES PERIODICAS
4. ARTICULOS
5. DEMANDA DE INFORMACION
6. SUMINISTRO DE DOCUMENTOS
7. ALERTA INFORMATIVA
8. PRODUCTOS

TÉRMINOS ASIGNADOS DE MODO MANUAL

1. SUMINISTRO DE DOCUMENTOS
2. ACCESO AL DOCUMENTO
3. SERVICIOS DE INFORMACION ELECTRONICA
4. SERVICIOS DE ALERTA
5. PUBLICACIONES PERIODICAS
6. INDUSTRIA EDITORIAL
7. INDUSTRIA DE LA INFORMACION
8. PROVEEDORES
9. REDES DE COMUNICACIONES

$$C_i = \frac{4}{9 + 8 - 4} = 0,3076 = 30,76 \% \text{ de consistencia}$$

ARTÍCULO 10

TÉRMINOS ASIGNADOS DE MODO AUTOMÁTICO

1. MERCADO DE TRABAJO
2. OFERTAS DE EMPLEO
3. DIARIOS
4. DOCUMENTACION
5. BIBLIOTECONOMIA
6. ARCHIVISTICA
7. DOCUMENTALISTAS
8. BIBLIOTECARIOS
9. ARCHIVEROS

TÉRMINOS ASIGNADOS DE MODO MANUAL

1. DOCUMENTALISTAS
2. BIBLIOTECARIOS
3. ARCHIVEROS
4. OFERTA DE EMPLEO
5. MERCADO DE TRABAJO
6. DIARIOS
7. REQUISITOS PROFESIONALES

$$C_i = \frac{6}{7 + 9 - 6} = 0,6 = 60\% \text{ de consistencia}$$

ARTÍCULO 11**TÉRMINOS ASIGNADOS DE MODO AUTOMÁTICO**

1. LENGUAJES DOCUMENTALES
2. ANALISIS DE CONTENIDO DOCUMENTALES
3. PUBLICACIONES PERIODICAS
4. INFORMACION Y DOCUMENTACION
5. PRODUCCION BIBLIOGRAFICA
6. ARTICULOS
7. AUTORES
8. CENTROS DE INVESTIGACION
9. CDU
10. TESAUROS
11. RESUMENES
12. CINDOC

TÉRMINOS ASIGNADOS DE MODO MANUAL

1. LENGUAJES DOCUMENTALES
2. ANALISIS DE CONTENIDO DOCUMENTAL
3. PUBLICACIONES PERIODICAS
4. INFORMACION
5. DOCUMENTACION
6. ANALISIS BIBLIOMETRICO
7. PRODUCCION BIBLIOGRAFICA

$$C_i = \frac{4}{7 + 12 - 4} = 0,2666 = 26,66 \% \text{ de consistencia}$$

ARTÍCULO 12**TÉRMINOS ASIGNADOS DE MODO AUTOMÁTICO**

1. BIBLIOTECAS
2. COOPERACION BIBLIOTECARIA

TÉRMINOS ASIGNADOS DE MODO MANUAL

1. BIBLIOTECAS
2. COOPERACION CIENTIFICA
3. PROGRAMAS EUROPEOS DE COOPERACION
4. TECNOLOGIAS DE LA INFORMACION
5. PROYECTOS DE INVESTIGACION
6. PARTICIPACION

$$C_i = \frac{1}{6 + 2 - 1} = 0,1428 = 14,28\% \text{ de consistencia}$$

ARTÍCULO 13**TÉRMINOS ASIGNADOS DE MODO AUTOMÁTICO**

1. BIBLIOTECAS
2. SUMINISTRO DE DOCUMENTOS
3. TARIFAS
4. COSTES

TÉRMINOS ASIGNADOS DE MODO MANUAL

1. ACCESO AL DOCUMENTO
2. TARIFAS
3. SUMINISTRO DE DOCUMENTOS
4. TRANSFERENCIA DE LA INFORMACION
5. PRECIOS
6. FOTOCOPIAS
7. DERECHOS DE AUTOR

$$C_i = \frac{3}{7 + 4 - 3} = 0,375 = 37,5 \% \text{ de consistencia}$$

ARTÍCULO 14**TÉRMINOS ASIGNADOS DE MODO AUTOMÁTICO**

1. BASES DE DATOS
2. DIRECTORIOS

TÉRMINOS ASIGNADOS DE MODO MANUAL

1. BASES DE DATOS
2. DIRECTORIOS
3. ESPECIALIZACION

$$C_i = \frac{2}{3 + 2 - 2} = 0,666 = 66,66 \% \text{ de consistencia}$$

ARTÍCULO 15**TÉRMINOS ASIGNADOS DE MODO AUTOMÁTICO**

1. CENTROS DE DOCUMENTACION
2. BIBLIOTECAS
3. GESTION DE LA CALIDAD

TÉRMINOS ASIGNADOS DE MODO MANUAL

1. BIBLIOTECAS
2. CENTROS DE DOCUMENTACION
3. GESTION
4. CALIDAD TOTAL
5. INDICADORES
6. PROBLEMATICA
7. MEJORA

$$C_i = \frac{2}{7 + 3 - 2} = 0,25 = 25 \% \text{ de consistencia}$$

ARTÍCULO 16**TÉRMINOS ASIGNADOS DE MODO AUTOMÁTICO**

1. SERVICIOS DE INFORMACION
2. CONTROL DE CALIDAD
3. COSTES
4. USUARIOS

TÉRMINOS ASIGNADOS DE MODO MANUAL

1. SERVICIOS DE INFORMACION
2. ANALISIS COSTE-BENEFICIO
3. CALIDAD
4. PRECIOS

$$C_i = \frac{1}{4 + 4 - 1} = 0,1428 = 14,28 \% \text{ de consistencia}$$

ARTÍCULO 17**TÉRMINOS ASIGNADOS DE MODO AUTOMÁTICO**

1. SISTEMA DE INFORMACION
2. NECESIDAD DE INFORMACION
3. DOCUMENTOS

TÉRMINOS ASIGNADOS DE MODO MANUAL

1. SISTEMA DE INFORMACION
2. CONCEPTUALIZACION
3. MODELO
4. PROPOSICIONES
5. DOCUMENTACION
6. TEORIA GENERAL DE SISTEMAS
7. RECUPERACION DE LA INFORMACION

$$C_i = \frac{1}{7 + 3 - 1} = 0,1111 = 11,11 \% \text{ de consistencia}$$

ARTÍCULO 18**TÉRMINOS ASIGNADOS DE MODO AUTOMÁTICO**

1. CATALOGO COLECTIVO
2. PUBLICACIONES SERIADA

TÉRMINOS ASIGNADOS DE MODO MANUAL

1. MEDICINA
2. PUBLICACIONES PERIODICAS
3. CATALOGOS COLECTIVOS
4. DEMANDA DE INFORMACION
5. EVALUACION
6. SERVICIOS DE INFORMACION
7. INDICADORES

$$C_i = \frac{1}{7 + 2 - 1} = 0,125 = 12,5 \% \text{ de consistencia}$$

ARTÍCULO 19**TÉRMINOS ASIGNADOS DE MODO AUTOMÁTICO**

1. INVESTIGACION HISTORICA
2. PUBLICACIONES PERIODICAS
3. ANALISIS DE DATOS
4. ISOC
5. AUTORES

TÉRMINOS ASIGNADOS DE MODO MANUAL

1. FRANQUISMO
2. PRODUCCION BIBLIOGRAFICA
3. PUBLICACIONES PERIODICAS
4. HISTORIOGRAFIA
5. BASES DE DATOS
6. ANALISIS BIBLIOMETRICO

$$C_i = \frac{1}{6 + 5 - 1} = 0,1 = 10 \% \text{ de consistencia}$$

ARTÍCULO 20**TÉRMINOS ASIGNADOS DE MODO AUTOMÁTICO**

1. CONCEPTOS
2. DOCUMENTACION

TÉRMINOS ASIGNADOS DE MODO MANUAL

1. INFORMACION
2. CONCEPTOS
3. PARADIGMA
4. COGNICION
5. TEORICA DE LA INFORMACION

$$C_i = \frac{1}{5 + 2 - 1} = 0,1666 = 16,66 \% \text{ de consistencia}$$

ARTÍCULO 21**TÉRMINOS ASIGNADOS DE MODO AUTOMÁTICO**

1. BIBLIOTECAS ESPECIALIZADAS
2. CENTROS DE DOCUMENTACION
3. COLECCIONES
4. PUBLICACIONES PERIODICAS
5. BIOMEDICINA
6. BASES DE DATOS
7. CD-ROM
8. METOLOGIA
9. TITULOS

TÉRMINOS ASIGNADOS DE MODO MANUAL

1. BIOMEDICINA
2. BIBLIOTECAS ESPECIALIZADAS
3. CENTROS DE DOCUMENTACION
4. PUBLICACIONES PERIODICAS
5. COLECCIONES
6. BASES DE DATOS
7. ANALISIS COMPARATIVO

$$C_i = \frac{6}{7 + 9 - 6} = 0,6 = 60 \% \text{ de consistencia}$$

ARTÍCULO 22**TÉRMINOS ASIGNADOS DE MODO AUTOMÁTICO**

1. PUBLICACIONES PERIODICAS
2. FISICA

TÉRMINOS ASIGNADOS DE MODO MANUAL

1. FISICA
2. PRODUCCION CIENTIFICA
3. ACTIVIDAD CIENTIFICA
4. ANALISIS BIBLIOMETRICO
5. ANALISIS DE CITAS
6. AUTORES
7. INSTITUCIONES
8. REVISTAS CIENTIFICAS
9. ESTUDIO DE CASO

$$C_i = \frac{1}{9 + 2 - 1} = 0,1 = 10 \% \text{ de consistencia}$$

ARTÍCULO 23**TÉRMINOS ASIGNADOS DE MODO AUTOMÁTICO**

1. SERVICIOS DE INFORMACION
2. UNIDADES DE INFORMACION
3. REDES DE COMUNICACION
4. INTERNET
5. TCP / IP
6. TELNET
7. FTP
8. ARCHIE
9. WAIS
10. GOPHER
11. VERONICA
12. WWW
13. BASES DE DATOS
14. OPAC
15. RECURSOS DE INFORMACION
16. PUBLICACIONES PERIODICAS
17. BIBLIOTECAS
18. BIBLIOTECARIOS
19. DOCUMENTALISTAS
20. CONTROL BIBLIOGRAFICO
21. COLECCIONES
22. DOCUMENTOS

TÉRMINOS ASIGNADOS DE MODO MANUAL

1. REDES DE COMUNICACIONES
2. TELEDOCUMENTACION
3. CORREO ELECTRONICO
4. REVISTAS ELECTRONICAS
5. CONFERENCIAS ELECTRONICAS
6. CLIENTE-SERVIDOR
7. BIBLIOTECAS
8. CENTROS DE DOCUMENTACION
9. ACCESO A LA INFORMACION
10. ACCESO AL DOCUMENTO

$$C_i = \frac{2}{10 + 23 - 1} = 0,064 = 6,45 \% \text{ de consistencia}$$

ARTÍCULO 24**TÉRMINOS ASIGNADOS DE MODO AUTOMÁTICO**

1. UNIVERSIDAD
2. BIBLIOTECAS
3. CATALOGACION RETROSPECTIVA
4. FASES
5. CLASIFICACION
6. INDIZACION
7. BIBLIOGRAFIA
8. MANUALES

TÉRMINOS ASIGNADOS DE MODO MANUAL

1. BIBLIOTECAS UNIVERSITARIAS
2. CONVERSION RETROSPECTIVA
3. PRECATALOGACION
4. CAPTURA DE REGISTROS
5. PROYECTO
6. FASES
7. EVALUACION

$$C_i = \frac{1}{7 + 8 - 1} = 0,071 = 7,14 \% \text{ de consistencia}$$

ARTÍCULO 25**TÉRMINOS ASIGNADOS DE MODO AUTOMÁTICO**

1. BIBLIOTECAS
2. CSIC
3. CONVERSION RETROSPECTIVA
4. CATALOGOS COLECTIVOS
5. METODOLOGIA
6. INFORMATICOS
7. FICHEROS
8. MANUALES

TÉRMINOS ASIGNADOS DE MODO MANUAL

1. BIBLIOTECAS
2. INFORMATIZACION
3. CONVERSION RETROSPECTIVA
4. REDES DE BIBLIOTECAS
5. EMPRESAS DE SERVICIOS
6. FASES
7. EVALUACION
8. METODOLOGIA
9. COSTES

$$C_i = \frac{3}{9 + 8 - 3} = 0,2142 = 21,42 \% \text{ de consistencia}$$

ARTÍCULO 26**TÉRMINOS ASIGNADOS DE MODO AUTOMÁTICO**

1. UNIVERSIDAD
2. BASES DE DATOS
3. ICYT
4. QUIMICA
5. PUBLICACIONES PERIODICAS

TÉRMINOS ASIGNADOS DE MODO MANUAL

1. UNIVERSIDAD
2. PRODUCCION CIENTIFICA
3. QUIMICA
4. PUBLICACIONES CIENTIFICAS
5. ANALISIS BIBLIOMETRICO
6. ESTUDIO COMPARATIVO

$$C_i = \frac{2}{6 + 5 - 2} = 0,2222 = 22,22 \% \text{ de consistencia}$$

ARTÍCULO 27**TÉRMINOS ASIGNADOS DE MODO AUTOMÁTICO**

1. INTERNET
2. USUARIOS
3. HIPERTEXTO
4. TELNET

TÉRMINOS ASIGNADOS DE MODO MANUAL

1. REDES DE INFORMACION
2. REDES DE COMUNICACIONES
3. PROGRAMAS DE ORDENADOR
4. HIPERTEXTO
5. FICHEROS INFORMATICOS

$$C_i = \frac{1}{5 + 4 - 1} = 0,125 = 12,5 \% \text{ de consistencia}$$

ARTÍCULO 28**TÉRMINOS ASIGNADOS DE MODO AUTOMÁTICO**

1. FUENTES DE INFORMACION
2. TIPOLOGIA DOCUMENTAL
3. PRODUCTOS

TÉRMINOS ASIGNADOS DE MODO MANUAL

1. BIOGRAFIA
2. BIBLIOGRAFIA
3. FUENTES DE INFORMACION
4. TIPOLOGIA DOCUMENTAL
5. EVALUACION

$$C_i = \frac{2}{5 + 3 - 2} = 0,3333 = 33,33 \% \text{ de consistencia}$$

ARTÍCULO 29

TÉRMINOS ASIGNADOS DE MODO AUTOMÁTICO

1. BASES DE DATOS
2. MERCADO
3. MARKETING
4. CD-ROM
5. PRODUCTOS
6. PRECIOS

TÉRMINOS ASIGNADOS DE MODO MANUAL

1. MATERIAL MEDICO QUIRURGICO
2. ANALISIS DE MERCADO
3. BASES DE DATOS
4. MARKETING
5. COMPETENCIA

$$C_i = \frac{2}{5 + 6 - 2} = 0,2222 = 22,22 \% \text{ de consistencia}$$

ARTÍCULO 30

TÉRMINOS ASIGNADOS DE MODO AUTOMÁTICO

1. DOCUMENTACION MEDICA
2. DOCUMENTOS

TÉRMINOS ASIGNADOS DE MODO MANUAL

1. MEDICINA
2. PUBLICACIONES PERIODICAS
3. MONOGRAFIAS
4. TESIS DOCTORALES
5. CONGRESOS
6. ANALISIS BIBLIOMETRICO
7. FUENTES DE INFORMACION

$$C_i = \frac{0}{7 + 2 - 0} = 0 \% \text{ de consistencia}$$

ARTÍCULO 31

TÉRMINOS ASIGNADOS DE MODO AUTOMÁTICO

1. ACCESO EN LINEA
2. BASES DE DATOS
3. COSTES
4. TARIFAS
5. PRECIOS
6. ANALISIS COMPARATIVO
7. TELECOMUNICACIONES

TÉRMINOS ASIGNADOS DE MODO MANUAL

1. BASES DE DATOS
2. REDES DE TRANSMISICON DE DATOS
3. VIDEOTEX
4. TARIFAS
5. ESTUDIO COMPARATIVO

8. ARTICULOS

$$C_i = \frac{3}{5 + 8 - 3} = 0,3 = 30 \% \text{ de consistencia}$$

ARTÍCULO 32**TÉRMINOS ASIGNADOS DE MODO AUTOMÁTICO**

1. SISTEMAS BIBLIOTECARIOS
2. COOPERACION BIBLIOTECARIA
3. CONTROL BIBLIOTECARIO
4. TELEDOCUMENTACION
5. INFORMÁTICA
6. PROFESIONALES DE LA DOCUMENTACION
7. INVESTIGACION Y DESARROLLO
8. BIBLIOTECAS UNIVERSITARIAS
9. SUMINISTRO DE DOCUMENTOS
10. USUARIOS

TÉRMINOS ASIGNADOS DE MODO MANUAL

1. BIBLIOTECAS UNIVERSITARIAS
2. COOPERACION BIBLIOTECARIA
3. TELECOMUNICACIONES
4. REDES DE BIBLIOTECAS
5. COOPERACION CIENTÍFICA
6. ACCESO A LA INFORMACION
7. REDES DE TRANSMISION DE DATOS
8. FINANCIACION

$$C_i = \frac{3}{8 + 10 - 3} = 0,2 = 20 \% \text{ de consistencia}$$

ARTÍCULO 33**TÉRMINOS ASIGNADOS DE MODO AUTOMÁTICO**

1. DISCOS OPTICOS
2. CD-ROM
3. EVOLUCION

TÉRMINOS ASIGNADOS DE MODO MANUAL

1. CD-ROM
2. EDICION ELECTRONICA
3. INDUSTRIA EDITORIAL
4. DATOS HISTORICOS
5. EVOLUCION

$$C_i = \frac{2}{5 + 3 - 2} = 0,3333 = 33,33 \% \text{ de consistencia}$$

ARTÍCULO 34**TÉRMINOS ASIGNADOS DE MODO AUTOMÁTICO**

1. UNIVERSIDAD
2. REDES DE AREA LOCAL
3. REDES DE CD-ROM
4. SOFTWARE
5. HARDWARE
6. PRODUCTOS
7. MERCADO
8. EVOLUCION
9. CD-ROM

TÉRMINOS ASIGNADOS DE MODO MANUAL

1. REDES DE CD-ROM
2. REDES DE AREA LOCAL
3. BASES DE DATOS
4. CD-ROM
5. ESTUDIO DE CASOS

$$C_i = \frac{3}{5 + 9 - 3} = 0,2727 = 27,27 \% \text{ de consistencia}$$

ARTÍCULO 35**TÉRMINOS ASIGNADOS DE MODO AUTOMÁTICO**

1. ALERTA INFORMATIVA
2. FORMACION CONTINUADA
3. CALIDAD
4. ENCUESTAS
5. LECTORES
6. DOCUMENTOS
7. PUBLICACIONES PERIODICAS
8. ARTICULOS

TÉRMINOS ASIGNADOS DE MODO MANUAL

1. SALUD
2. ARTICULOS DE REVISTA
3. ALERTA INFORMATIVA
4. REPERTORIOS BIBLIOGRAFICOS

$$C_i = \frac{1}{4 + 8 - 1} = 0,909 = 9,09 \% \text{ de consistencia}$$

ARTÍCULO 36**TÉRMINOS ASIGNADOS DE MODO AUTOMÁTICO**

1. EMPRESAS
2. NECESIDAD DE INFORMACION
3. CENTROS DE DOCUMENTACION
4. UNIVERSIDAD
5. USUARIOS
6. ENTREVISTAS
7. ENCUESTAS

TÉRMINOS ASIGNADOS DE MODO MANUAL

1. EMPRESAS
2. NECESIDAD DE INFORMACION
3. CANALES DE INFORMACION
4. FUENTES DE INFORMACION

$$C_i = \frac{2}{4 + 7 - 2} = 0,2222 = 22,22 \% \text{ de consistencia}$$

ARTÍCULO 37**TÉRMINOS ASIGNADOS DE MODO AUTOMÁTICO**

1. REVISTAS CIENTIFICAS
2. CIENCIA Y TECNOLOGIA
3. INVESTIGACION CIENTIFICA
4. INFORMACION CIENTIFICA

TÉRMINOS ASIGNADOS DE MODO MANUAL

1. CIENCIA Y TECNOLOGIA
2. REVISTAS CIENTIFICAS
3. DIFUSION
4. NORMALIZACION
5. EVALUACION

$$C_i = \frac{2}{5 + 4 - 2} = 0,2857 = 28,57 \% \text{ de consistencia}$$

ARTÍCULO 38**TÉRMINOS ASIGNADOS DE MODO AUTOMÁTICO**

1. BASES DE DATOS

TÉRMINOS ASIGNADOS DE MODO MANUAL

1. ARTE
2. BASES DE DATOS
3. LISTA

$$C_i = \frac{1}{3 + 1 - 1} = 0,3333 = 33,33 \% \text{ de consistencia}$$

ARTÍCULO 39**TÉRMINOS ASIGNADOS DE MODO AUTOMÁTICO**

1. BIBLIOTECAS UNIVERSITARIAS
2. PUBLICACIONES PERIODICAS
3. EVALUACION
4. TITULOS
5. SUSCRIPCIONES
6. CANCELACIONES

TÉRMINOS ASIGNADOS DE MODO MANUAL

1. BIBLIOTECAS UNIVERSITARIAS
2. PUBLICACIONES PERIODICAS
3. COLECCIONES
4. SUCRIPCION
5. CANCELACION
6. EVALUACION

$$C_i = \frac{5}{6 + 6 - 5} = 0,7442 = 74,42 \% \text{ de consistencia}$$

ARTÍCULO 40**TÉRMINOS ASIGNADOS DE MODO AUTOMÁTICO**

1. CENTROS DE DOCUMENTACION
2. LINGUISTICA
3. DOCUMENTACION
4. PRODUCTOS

TÉRMINOS ASIGNADOS DE MODO MANUAL

1. INDUSTRIAS DE LA LENGUA
2. LINGUISTICA
3. DOCUMENTACION
4. LENGUA
5. TRATAMIENTO AUTOMÁTICO
6. CENTROS DE DOCUMENTACION
7. BIBLIOTECAS

$$C_i = \frac{3}{7 + 4 - 3} = 0,375 = 37,5 \% \text{ de consistencia}$$

ARTÍCULO 41**TÉRMINOS ASIGNADOS DE MODO AUTOMÁTICO**

1. SISTEMAS DE GESTION DE BASES DE DATOS
2. BASES DE DATOS BIBLIOGRAFICAS
3. INFORMACION DE ACTUALIDAD
4. TESAUROS
5. ELABORACION DE TESAUROS
6. MANTENIMIENTO DE TESAUROS
7. MICROORDENADORES
8. FICHEROS
9. TERMINOS PREFERENTES
10. TERMINOS NO PREFERENTES

TÉRMINOS ASIGNADOS DE MODO MANUAL

1. SISTEMAS DE GESTION DE BASES DE DATOS
2. MICROORDENADORES
3. TESAUROS
4. ELABORACION
5. MANTENIMIENTO

$$C_i = \frac{5}{5 + 10 - 5} = 0,5 = 50 \% \text{ de consistencia}$$

ARTÍCULO 42**TÉRMINOS ASIGNADOS DE MODO AUTOMÁTICO**

1. CIENCIAS SOCIALES
2. CIENCIAS HUMANAS
3. EDUCACION
4. PSICOLOGIA
5. SOCIOLOGIA
6. BIBLIOMETRIA
7. PRODUCCION BIBLIOGRAFICA
8. PUBLICACIONES PERIODICAS
9. BASES DE DATOS
10. BUSQUEDA BIBLIOGRAFICAS
11. INFORMACION Y DOCUMENTACION
12. BIBLIOTECONOMIA
13. CENTROS DE DOCUMENTACION
14. BIBLIOTECAS
15. ESTADO DE LA CUESTION
16. EVOLUCION
17. AUTORES
18. DOCUMENTOS

TÉRMINOS ASIGNADOS DE MODO MANUAL

1. CIENCIAS SOCIALES
2. CIENCIAS HUMANAS
3. DOCUMENTACION
4. BIBLIOTECONOMIA
5. BASES DE DATOS
6. BIBLIOMETRIA
7. ANALISIS BIBLIOMETRICO

$$C_i = \frac{5}{7 + 18 - 5} = 0,25 = 25 \% \text{ de consistencia}$$

ARTÍCULO 43**TÉRMINOS ASIGNADOS DE MODO AUTOMÁTICO**

1. CATALOGOS SISTEMATICOS
2. BASES DE DATOS BIBLIOGRAFICOS
3. RECUPERACION DE LA INFORMACION
4. INFORMACION GEOGRAFICA
5. INDIZADORES
6. GEOGRAFIA
7. TOPONIMOS
8. DOCUMENTOS

TÉRMINOS ASIGNADOS DE MODO MANUAL

1. GEOGRAFIA
2. BASES DE DATOS
3. TOPONIMOS
4. CODIFICACION
5. RECUPERACION DE LA INFORMACION

$$C_i = \frac{4}{5 + 8 - 4} = 0,4444 = 44,44 \% \text{ de consistencia}$$

ARTÍCULO 44**TÉRMINOS ASIGNADOS DE MODO AUTOMÁTICO**

1. PAISES ARABES
2. INFORMACION CIENTIFICA
3. INFORMACION Y DOCUMENTACION
4. RECURSOS DE INFORMACION
5. DOCUMENTACION AUTOMATIZADA
6. POLITICA DE INFORMACION
7. NORMALIZACION
8. LENGUAJE
9. SISTEMAS INFORMATICOS
10. ARTICULOS

TÉRMINOS ASIGNADOS DE MODO MANUAL

1. PAISES EN DESARROLLO
2. INFORMACION
3. DOCUMENTACION
4. POLITICA DE INFORMACION
5. SISTEMAS DE INFORMACION
6. PAISES ARABES
7. FUENTES DE INFORMACION

$$C_i = \frac{2}{7 + 10 - 2} = 0,1333 = 13,33 \% \text{ de consistencia}$$

ARTÍCULO 45**TÉRMINOS ASIGNADOS DE MODO AUTOMÁTICO**

1. CIENCIA Y TECNOLOGIA
2. COOPERACION CIENTIFICA
3. REVISTAS CIENTIFICAS
4. ANALISIS BIBLIOMETRICO
5. CENTROS DE DOCUMENTACION
6. TITULOS
7. AUTORES

TÉRMINOS ASIGNADOS DE MODO MANUAL

1. COOPERACION CIENTIFICA
2. CIENCIA Y TECNOLOGIA
3. REVISTAS CIENTIFICAS
4. ANALISIS BIBLIOMETRICO
5. BASES DE DATOS
6. DISTRIBUCION ESPACIAL
7. AREAS TEMATICAS

$$C_i = \frac{4}{7 + 7 - 4} = 0,4 = 40 \% \text{ de consistencia}$$

ARTÍCULO 46**TÉRMINOS ASIGNADOS DE MODO AUTOMÁTICO**

1. COOPERACION CIENTIFICA
2. BASES DE DATOS
3. PUBLICACIONES PERIODICAS
4. ANALISIS BIBLIOMETRICO
5. DOCUMENTOS
6. AUTORES

TÉRMINOS ASIGNADOS DE MODO MANUAL

1. COOPERACION CIENTIFICA
2. CIENCIA Y TECNOLOGIA
3. REVISTAS CIENTIFICAS
4. BASES DE DATOS
5. ANALISIS BIBLIOMETRICO
6. AREAS TEMATICAS
7. DISTRIBUCION ESPACIAL

$$C_i = \frac{3}{7 + 6 - 3} = 0,3 = 30 \% \text{ de consistencia}$$

ARTÍCULO 47**TÉRMINOS ASIGNADOS DE MODO AUTOMÁTICO**

1. INFORMACION AUTOMATIZADA
2. BASES DE DATOS
3. INDUSTRIA
4. CALIDAD

TÉRMINOS ASIGNADOS DE MODO MANUAL

1. BASES DE DATOS
2. INDUSTRIA DE LA INFORMACION
3. DISTRIBUCION ESPACIAL
4. TEMATICA

$$C_i = \frac{1}{4 + 4 - 1} = 0,1428 = 14,28 \% \text{ de consistencia}$$

ARTÍCULO 48**TÉRMINOS ASIGNADOS DE MODO AUTOMÁTICO**

1. CENTROS DE DOCUMENTACION
2. UNIVERSIDAD
3. CIENCIAS SOCIALES
4. CIENCIAS HUMANAS
5. INFORMACION BIBLIOGRAFICA
6. DIFUSION DE LA INFORMACION
7. USUARIOS
8. PRODUCTOS
9. BASES DE DATOS
10. DISCOS OPTICOS
11. INDICES
12. DOCUMENTOS

TÉRMINOS ASIGNADOS DE MODO MANUAL

1. INFORMACION CIENTIFICA
2. BIBLIOGRAFIA
3. CENTROS DE DOCUMENTACION
4. CIENCIAS SOCIALES
5. HUMANIDADES
6. BASES DE DATOS

$$C_i = \frac{5}{6 + 12 - 5} = 0,3846 = 38,46 \% \text{ de consistencia}$$

ARTÍCULO 49**TÉRMINOS ASIGNADOS DE MODO AUTOMÁTICO**

1. INFORMACION Y DOCUMENTACION
2. ARTICULOS

TÉRMINOS ASIGNADOS DE MODO MANUAL

1. REDES DE INFORMACION
2. DOCUMENTACION
3. INFORMACION CIENTIFICA
4. COOPERACION INTERNACIONAL
5. DATOS HISTORICOS
6. PERSPECTIVAS

$$C_i = \frac{0}{6 + 2 - 0} = 0 \% \text{ de consistencia}$$

ARTÍCULO 50**TÉRMINOS ASIGNADOS DE MODO AUTOMÁTICO**

1. INFORMACION Y DOCUMENTACION
2. BIBLIOTECAS NACIONALES
3. BIBLIOTECAS
4. BIBLIOTECARIO
5. COOPERACION BIBLIOTECARIA
6. ELABORACION DE BASES DE DATOS
7. BASES DE DATOS BIBLIOGRAFICAS
8. OBRAS DE REFERENCIA

TERMINOS ASIGNADOS DE MODO MANUAL

1. BIBLIOTECAS NACIONALES
2. COOPERACION BIBLIOTECARIA
3. CATALOGOS COLECTIVOS
4. COOPERACION INTERNACIONAL

$$C_i = \frac{2}{4 + 8 - 2} = 0,2 = 20 \% \text{ de consistencia}$$

ANEXO 8 VOCABULARIO CONTROLADO

1. 3W USE WORLD WIDE WEB
2. AACR USE REGLAS DE CATALOGACION
3. ABREVIATURAS
4. ABSTRACTS USE RESUMENES
5. ACCESIBILIDAD DE LA INFORMACION
6. ACCESIBILIDAD UNIVERSAL A LA INFORMACION USE
DISPONIBILIDAD UNIVERSAL DE LAS PUBLICACIONES
7. ACCESO A BASES DE DATOS
8. ACCESO A LA INFORMACION
9. ACCESO A LOS LIBROS
10. ACCESO AL DOCUMENTO
11. ACCESO REMOTO
12. ACCESO UNIVERSAL A LAS PUBLICACIONES USE
DISPONIBILIDAD UNIVERSAL DE LAS PUBLICACIONES
13. ACERVO USE COLECCIONES
14. ACERVO BIBLIOGRAFICO USE FONDOS BIBLIOGRAFICOS
15. ACERVO DOCUMENTAL USE FONDOS DOCUMENTALES
16. ACM USE ASOCIACIONES PROFESIONALES
17. ACREDITACION PROFESIONAL
18. ACRODABA USE BASE DE DATOS DE ACRONIMOS
19. ACRONIMOS
20. ACTAS DE CONGRESOS
21. ACTIVIDADES DE EXTENSION BIBLIOTECARIA
22. ACTIVIDADES DE EXTENSION CULTURAL
23. ACTUALIZACION DE LOS FICHEROS USE MANTENIMIENTO DE LOS
FICHEROS
24. ADAB USE ASOCIACIONES PROFESIONALES
25. ADBS USE ASOCIACIONES PROFESIONALES
26. ADMINISTRACION BIBLIOTECARIA USE ADMINISTRACION DE
BIBLIOTECAS
27. ADMINISTRACION DE BIBLIOTECAS
28. ADMINISTRACION PUBLICA
29. ADOLESCENTES
30. ADQUISICION
31. ADQUISICION COMPARTIDA USE ADQUISICION COOPERATIVA

32. ADQUISICION COOPERATIVA
33. ADQUISICION DE DOCUMENTOS
34. ADQUISICION DE FONDOS USE ADQUISICIÓN DE DOCUMENTOS
35. ADQUISICION DE LIBROS USE ADQUISICIÓN DE DOCUMENTOS
36. ADQUISICION GRATUITA
37. *ADQUISICION POR COMPRA*
38. ADQUISICION POR INTERCAMBIO
39. ADQUISICION POR ORDENADOR
40. AENOR
41. AFICHES USE CARTELES
42. AFNOR
43. AGENCIAS DE PRENSA
44. ALA USE ASOCIACIONES PROFESIONALES
45. ALERTA BIBLIOGRAFICA USE ALERTA INFORMATIVA
46. ALERTA INFORMATIVA
47. ALGORITMOS
48. ALMACENAMIENTO DE DOCUMENTOS
49. ALMACENAMIENTO DE IMAGENES
50. ALMACENAMIENTO DE INFORMACION USE ALMACENAMIENTO DE LA INFORMACION
51. ALMACENAMIENTO DE LA INFORMACION
52. ALMACENAMIENTO DE LIBROS USE ALMACENAMIENTO DE DOCUMENTOS
53. ALMACENAMIENTO OPTICO
54. ALMACENAMIENTO Y RECUPERACION DE LA INFORMACION USE PROCESAMIENTO DE LA INFORMACION
55. ALMANAQUES
56. ALTA DEFINICION
57. AMERICAN LIBRARY ASSOCIATION USE ASOCIACIONES PROFESIONALES
58. AMERICAN STANDARD CODE FOR INFORMATION INTERCHANGE USE ASCII
59. ANABAD USE ASOCIACIONES PROFESIONALES
60. ANALES
61. ANALISIS AUTOMATICO DE TEXTOS
62. ANALISIS BENEFICIO-COSTO USE ANALISIS COSTE-BENEFICIO
63. *ANALISIS BIBLIOGRAFICO*
64. ANALISIS BIBLIOMETRICO
65. ANALISIS COMPARATIVO
66. ANALISIS COSTE-BENEFICIO
67. ANALISIS CRITICO USE REVISION CRITICA
68. ANALISIS DE CONTENIDO USE ANALISIS DEL CONTENIDO DOCUMENTAL
69. ANALISIS DEL CONTENIDO USE ANALISIS DEL CONTENIDO DOCUMENTAL
70. ANALISIS DE CITAS

71. ANALISIS DE COSTOS Y BENEFICIOS USE ANALISIS DE COSTE-
BENEFICIO
72. ANALISIS DE DATOS
73. ANALISIS DE DOCUMENTOS USE ANALISIS DE LA INFORMACION
74. ANALISIS DE FACETAS USE ANALISIS POR FACETAS
75. ANALISIS DE FRECUENCIAS
76. ANALISIS DE INFORMACION USE ANALISIS DE LA INFORMACION
77. ANALISIS DE LA INFORMACION
78. ANALISIS DE MATERIAS USE ANALISIS DEL CONTENIDO DOCUMENTAL
79. ANALISIS DE RESULTADOS
80. ANALISIS DE SISTEMAS
81. ANALISIS DE TEXTOS
82. ANALISIS DE USUARIOS USE ESTUDIO DE USUARIOS
83. ANALISIS DEL CONTENIDO DOCUMENTAL
84. ANALISIS DOCUMENTAL
85. ANALISIS ESTADISTICO
86. ANALISIS FACETADO USE ANALISIS POR FACETAS
87. ANALISIS LINGÜISTICO
88. ANALISIS MATRICIAL USE LINGÜISTICA ESTADISTICA
89. ANALISIS MORFOLOGICO
90. ANALISIS POR ASOCIACION DE TERMINOS USE INDIZACION POR
ASOCIACION
91. ANALISIS POR CLUSTERS USE CLUSTERING
92. ANALISIS POR FACETAS
93. ANALISIS POR RACIMOS
94. ANALISIS PRAGMATICO
95. ANALISIS SEMANTICO
96. ANALISIS SINTACTICO
97. ANALISIS TEMATICO
98. ANALISIS TEXTUAL USE ANALISIS DE TEXTOS
99. ANALISTAS DE INFORMACION
100. ANALISTAS DE SISTEMAS
101. ANALIZADOR MORFOLOGICO USE ETIQUETADOR MORFOLOGICO
102. ANALIZADOR SINTACTICO USE ETIQUETADOR SINTACTICO
103. ANIMACION A LA LECTURA
104. ANIMACION CULTURAL
105. ANIMACION SOCIOCULTURAL
106. ANTIDICCIONARIOS USE DICCIONARIOS DE PALABRAS VACIAS
107. ANUARIOS
108. AÑO INTERNACIONAL DEL LIBRO
109. APLICACION DE ORDENADORES USE USO DE ORDENADORES
110. APLICACIONES INFORMATICAS
111. ARCHIE
112. ARCHIVEROS
113. ARCHIVEROS CIENTIFICOS
114. ARCHIVEROS TECNICOS
115. ARCHIVISTAS USE ARCHIVEROS
116. ARCHIVISTICA
117. ARCHIVOLOGIA USE ARCHIVISTICA
118. ARCHIVONOMIA USE ARCHIVISTICA

- 119. ARCHIVOS
- 120. ARCHIVOS ADMINISTRATIVOS
- 121. ARCHIVOS AMERICANOS
- 122. ARCHIVOS AUDIOVISUALES
- 123. ARCHIVOS CARTOGRAFICOS
- 124. ARCHIVOS CATEDRALICIOS
- 125. ARCHIVOS CENTRALES
- 126. ARCHIVOS COMARCALES
- 127. ARCHIVOS CORRIENTES
- 128. ARCHIVOS CRONOLOGICOS
- 129. ARCHIVOS DE DATOS USE ARCHIVOS LEGIBLES POR MAQUINA
- 130. ARCHIVOS DE DISTRITO
- 131. ARCHIVOS DE EMPRESAS
- 132. ARCHIVOS DE FUNDACIONES
- 133. ARCHIVOS DE GESTION
- 134. ARCHIVOS DE HOSPITALES
- 135. ARCHIVOS DE IMAGENES
- 136. ARCHIVOS DE IMPRESOS
- 137. ARCHIVOS DE LA ADMINISTRACION
- 138. ARCHIVOS DE LA ADMINISTRACION AUTONOMICA
- 139. ARCHIVOS DE LA ADMINISTRACION CENTRAL
- 140. ARCHIVOS DE LA ADMINISTRACION LOCAL
- 141. ARCHIVOS DE LA ADMINISTRACION PUBLICA
- 142. ARCHIVOS DE LOS GOBIERNOS REGIONALES USE ARCHIVOS GUBERNAMENTALES
- 143. ARCHIVOS DE OFICINA
- 144. ARCHIVOS DE PRENSA
- 145. ARCHIVOS DE PROGRAMAS
- 146. ARCHIVOS DE PROGRAMAS INFORMATICOS
- 147. ARCHIVOS DE PROTOCOLOS
- 148. ARCHIVOS DE RECORTES
- 149. ARCHIVOS DEL ESTADO USE ARCHIVOS GUBERNAMENTALES
- 150. ARCHIVOS DEL GOBIERNO LOCAL USE ARCHIVOS GUBERNAMENTALES
- 151. ARCHIVOS DEPARTAMENTALES
- 152. ARCHIVOS DIOCESANOS
- 153. ARCHIVOS DIPLOMATICOS USE ARCHIVOS GUBERNAMENTALES
- 154. ARCHIVOS ECLESIASTICOS
- 155. ARCHIVOS ELECTRONICOS
- 156. ARCHIVOS ESTATALES
- 157. ARCHIVOS FAMILIARES
- 158. ARCHIVOS FILMICOS
- 159. ARCHIVOS FOTOGRAFICOS
- 160. ARCHIVOS GENERALES
- 161. ARCHIVOS GUBERNAMENTALES
- 162. ARCHIVOS HISTORICOS
- 163. ARCHIVOS HISTORICOS GENERALES
- 164. ARCHIVOS HISTORICOS PROVINCIALES
- 165. ARCHIVOS ICONOGRAFICOS
- 166. ARCHIVOS INDUSTRIALES

167. ARCHIVOS INFORMATICOS
 168. ARCHIVOS INTERMEDIOS
 169. ARCHIVOS INTERNACIONALES
 170. ARCHIVOS JUDICIALES
 171. ARCHIVOS JUDICIALES MILITARES
 172. ARCHIVOS LEGIBLES POR MAQUINA
 173. ARCHIVOS LITERARIOS
 174. ARCHIVOS LOCALES
 175. ARCHIVOS MILITARES USE ARCHIVOS GUBERNAMENTALES
 176. ARCHIVOS MONASTICOS
 177. ARCHIVOS MUNICIPALES USE ARCHIVOS GUBERNAMENTALES
 178. ARCHIVOS MUSICALES
 179. ARCHIVOS NACIONALES
 180. ARCHIVOS NOBILIARIOS
 181. ARCHIVOS NOTARIALES
 182. ARCHIVOS OFICIALES USE ARCHIVOS GUBERNAMENTALES
 183. ARCHIVOS OPTICOS
 184. ARCHIVOS ORALES
 185. ARCHIVOS PARROQUIALES
 186. ARCHIVOS PERMANENTES
 187. ARCHIVOS PERSONALES
 188. ARCHIVOS PRIVADOS
 189. ARCHIVOS PROVINCIALES
 190. ARCHIVOS PUBLICOS
 191. ARCHIVOS REALES
 192. ARCHIVOS REGIONALES
 193. ARCHIVOS RELIGIOSOS
 194. ARCHIVOS SONOROS
 195. ARCHIVOS UNIVERSITARIOS
 196. ARCHIVOS VERTICALES
 197. ARL USE ASOCIACION DE BIBLIOTECAS DE INVESTIGACION
 198. ARQUITECTURA DE ARCHIVOS
 199. ARQUITECTURA DE BIBLIOTECAS
 200. ARTES GRAFICAS
 201. ARTICULOS
 202. ARTICULOS CIENTIFICOS
 203. ARTICULOS DE REVISTA
 204. ASCII
 205. ASIENTOS ANALITICOS
 206. ASIENTOS BIBLIOGRAFICOS
 207. ASIENTOS DE CATALOGOS USE ASIENTOS BIBLIOGRAFICOS
 208. ASIGNACION DE PESO USE PONDERACION
 209. ASISTENCIA A LOS USUARIOS USE ORIENTACION A LOS
 USUARIOS
 210. ASLIB
 211. ASOCIACION CANADIENSE DE LAS CIENCIAS DE LA INFORMACION USE
 ASOCIACIONES PROFESIONALES
 212. ASOCIACION DE ALUMNOS Y DIPLOMADOS EN BIBLIOTECONOMIA Y
 DOCUMENTACION USE ASOCIACIONES PROFESIONALES

213. ASOCIACION DE BIBLIOTECAS AMERICANAS USE ASOCIACIONES
PROFESIONALES
214. ASOCIACION DE BIBLIOTECAS DE INVESTIGACION
215. ASOCIACION ESPAÑOLA DE NORMALIZACION USE AENOR
216. ASOCIACION FRANCESA DE DOCUMENTALISTAS Y BIBLIOTECARIOS
ESPECIALIZADOS USE ASOCIACIONES PROFESIONALES
217. ASOCIACION FRANCESA DE NORMALIZACION USE AFNOR
218. ASOCIACION INTERNACIONAL DE NORMALIZACION USE ISO
219. ASOCIACION PARA LA INGENIERIA INFORMATICA USE ASOCIACIONES
PROFESIONALES
220. ASOCIACIONES PROFESIONALES
221. ASOCIACIONISMO
222. AUTOAPRENDIZAJE USE AUTOFORMACION
223. AUTOEDICION
224. AUTOFORMACION
225. AUTOMATIZACION
226. AUTOMATIZACION DE BIBLIOTECAS
227. AUTOMATIZACION DE CATALOGOS
228. AUTOPISTAS DE LA INFORMACION
229. AUTORES
230. AUTORIA USE AUTORES
231. AYUDA A LOS LECTORES USE ORIENTACION A LOS USUARIOS
232. BAAB
233. BANCOS DE DATOS
234. BANCOS DE DATOS TERMINOLOGICOS
235. BANCOS DE INFORMACION USE BASES DE DATOS
236. BARRERAS A LA INFORMACION
237. BARRERAS LINGÜISTICAS
238. BASE DE DATOS DE ACRONIMOS
239. BASES DE DATOS
240. BASES DE DATOS BIBLIOGRAFICAS
241. BASES DE DATOS BIBLIOGRAFICAS DE RESUMENES
242. BASES DE DATOS CATALOGRAFICAS
243. BASES DE DATOS DE TEXTO COMPLETO
244. BASES DE DATOS DISTRIBUIDAS
245. BASES DE DATOS DOCUMENTALES
246. BASES DE DATOS EN LINEA
247. BASES DE DATOS EN RED
248. BASES DE DATOS FACTUALES
249. BASES DE DATOS FEDERADAS
250. BASES DE DATOS FULL TEXT USE BASES DE DATOS DE TEXTO
COMPLETO
251. BASES DE DATOS ORIENTADAS A OBJETO
252. BASES DE DATOS RELACIONALES
253. BASES DE DATOS TEXTUALES
254. BASIC USE LENGUAJES DE PROGRAMACION
255. BEBETECAS
256. BIBLIOBUS USE BIBLIOTECAS MOVILES
257. BIBLIOFILIA

258. BIBLIOFILQS
 259. BIBLIOGRAFIA
 260. BIBLIOGRAFIAS
 261. BIBLIOGRAFIAS ALFABETICAS
 262. BIBLIOGRAFIAS ANALITICAS
 263. BIBLIOGRAFIAS ANOTADAS
 264. BIBLIOGRAFIAS CERRADAS
 265. BIBLIOGRAFIAS CORRIENTES
 266. BIBLIOGRAFIAS DE ALERTA
 267. BIBLIOGRAFIAS DE BIBLIOGRAFIAS
 268. BIBLIOGRAFIAS ESPECIALIZADAS
 269. BIBLIOGRAFIAS EXHAUSTIVAS
 270. BIBLIOGRAFIAS GENERALES
 271. BIBLIOGRAFIAS INTERNACIONALES
 272. BIBLIOGRAFIAS LOCALES
 273. BIBLIOGRAFIAS NACIONALES
 274. BIBLIOGRAFIAS NACIONALES CORRIENTES
 275. BIBLIOGRAFIAS NACIONALES RETROSPECTIVAS
 276. BIBLIOGRAFIAS PERIODICAS
 277. BIBLIOGRAFIAS PROSPECTIVAS
 278. BIBLIOGRAFIAS REGIONALES
 279. BIBLIOGRAFIAS RETROSPECTIVAS
 280. BIBLIOGRAFIAS SELECTIVAS
 281. BIBLIOGRAFIAS SISTEMATICAS
 282. BIBLIOGRAFIAS TEMATICAS USE BIBLIOGRAFIAS
 ESPECIALIZADAS
 283. BIBLIOGRAFIAS UNIVERSALES
 284. BIBLIOGRAFOS
 285. BIBLIOLANCHAS USE BIBLIOTECAS MOVILES
 286. BIBLIOLOGIA
 287. BIBLIOMANIA
 288. BIBLIOMETRIA
 289. BIBLIOTECA DE LA UNIVERSIDAD USE BIBLIOTECA UNIVERSITARIA
 290. BIBLIOTECARIOS
 291. BIBLIOTECARIOS DE CONSULTA USE BIBLIOTECARIOS DE
 REFERENCIA
 292. BIBLIOTECARIOS DE REFERENCIA
 293. BIBLIOTECARIOS REFERENCISTAS USE BIBLIOTECARIOS DE
 REFERENCIA
 294. BIBLIOTECAS
 295. BIBLIOTECAS ACADEMICAS
 296. BIBLIOTECAS ADMINISTRATIVAS
 297. BIBLIOTECAS AMBULANTES USE BIBLIOTECAS MOVILES
 298. BIBLIOTECAS BRAILLE
 299. BIBLIOTECAS CENTRALES
 300. BIBLIOTECAS CIENTIFICAS
 301. BIBLIOTECAS CIENTIFICAS Y TECNICAS
 302. BIBLIOTECAS CIRCULANTES
 303. BIBLIOTECAS DE AULA
 304. BIBLIOTECAS DE CARCELES USE BIBLIOTECAS DE PRISIONES

305. BIBLIOTECAS DE CONDADO
 306. BIBLIOTECAS DE DEPARTAMENTO
 307. BIBLIOTECAS DE DEPOSITO LEGAL USE BIBLIOTECAS DEPOSITARIAS
 308. BIBLIOTECAS DE DISTRITO
 309. BIBLIOTECAS DE EMPRESA
 310. BIBLIOTECAS DE HOSPITALES
 311. BIBLIOTECAS DE INSTITUCIONES RELIGIOSAS
 312. BIBLIOTECAS DE INVESTIGACION
 313. BIBLIOTECAS DE MEDICINA
 314. BIBLIOTECAS DE MONASTERIOS
 315. BIBLIOTECAS DE MUSICA
 316. BIBLIOTECAS DE PATENTES
 317. BIBLIOTECAS DE PRISIONES
 318. BIBLIOTECAS DE PROGRAMAS INFORMATICOS
 319. BIBLIOTECAS DE RECLUSOS USE BIBLIOTECAS DE PRISIONES
 320. BIBLIOTECAS DE SINDICATOS
 321. BIBLIOTECAS DE TITULARIDAD ESTATAL USE BIBLIOTECAS GUBERNAMENTALES
 322. BIBLIOTECAS DEL EJERCITO
 323. BIBLIOTECAS DEPOSITARIAS
 324. BIBLIOTECAS DEPOSITO
 325. BIBLIOTECAS ECLESIASTICAS
 326. BIBLIOTECAS ELECTRONICAS
 327. BIBLIOTECAS ESCOLARES
 328. BIBLIOTECAS ESPECIALES
 329. BIBLIOTECAS ESPECIALIZADAS
 330. BIBLIOTECAS ESTATALES USE BIBLIOTECAS GUBERNAMENTALES
 331. BIBLIOTECAS EXPERIMENTALES USE BIBLIOTECAS PILOTO
 332. BIBLIOTECAS FEDERALES USE BIBLIOTECAS GUBERNAMENTALES
 333. BIBLIOTECAS FILIALES USE BIBLIOTECAS SUCURSALES
 334. BIBLIOTECAS GENERALES
 335. BIBLIOTECAS GUBERNAMENTALES
 336. BIBLIOTECAS HISTORICAS
 337. BIBLIOTECAS INFANTILES
 338. BIBLIOTECAS INSTITUCIONALES
 339. BIBLIOTECAS INTERNACIONALES
 340. BIBLIOTECAS JURIDICAS
 341. BIBLIOTECAS JUVENILES
 342. BIBLIOTECAS MEDICAS USE BIBLIOTECAS DE MEDICINA
 343. BIBLIOTECAS MINISTERIALES
 344. BIBLIOTECAS MONASTICAS
 345. BIBLIOTECAS MOVILES
 346. BIBLIOTECAS MUNICIPALES
 347. BIBLIOTECAS NACIONALES
 348. BIBLIOTECAS NACIONALES DE PRESTAMO
 349. BIBLIOTECAS OFICIALES USE BIBLIOTECAS GUBERNAMENTALES
 350. BIBLIOTECAS PARA CIEGOS
 351. BIBLIOTECAS PARA INCAPACITADOS
 352. BIBLIOTECAS PARA INVIDENTES USE BIBLIOTECAS PARA CIEGOS
 353. BIBLIOTECAS PARA JOVENES USE BIBLIOTECAS JUVENILES

354. BIBLIOTECAS PARA NIÑOS USE BIBLIOTECAS INFANTILES
 355. BIBLIOTECAS PARLAMENTARIAS
 356. BIBLIOTECAS PARROQUIALES
 357. BIBLIOTECAS PARTICULARES USE BIBLIOTECAS PRIVADAS
 358. BIBLIOTECAS PERSONALES USE BIBLIOTECAS PRIVADAS
 359. BIBLIOTECAS PILOTO
 360. BIBLIOTECAS POPULARES
 361. BIBLIOTECAS PRIVADAS
 362. BIBLIOTECAS PROVINCIALES
 363. BIBLIOTECAS PUBLICAS
 364. BIBLIOTECAS PUBLICAS DEL ESTADO
 365. BIBLIOTECAS PUBLICAS INFANTILES
 366. BIBLIOTECAS PUBLICAS MUNICIPALES
 367. BIBLIOTECAS REGIONALES
 368. BIBLIOTECAS RURALES
 369. BIBLIOTECAS SUCURSALES
 370. BIBLIOTECAS TECNICAS USE BIBLIOTECAS CIENTIFICAS
 371. BIBLIOTECAS UNIVERSITARIAS
 372. BIBLIOTECAS URBANAS USE BIBLIOTECAS MUNICIPALES
 373. BIBLIOTECAS VIAJERAS
 374. BIBLIOTECAS VIRTUALES
 375. BIBLIOTECOGRAFIA
 376. BIBLIOTECOLOGIA USE BIBLIOTECONOMIA
 377. BIBLIOTECOLOGOS USE BIBLIOTECARIOS
 378. BIBLIOTECONOMIA
 379. BIBLIOTECONOMIA Y DOCUMENTACION
 380. BIBLIOTERAPIA
 381. BIOMEDICINA
 382. BLISS CLASSIFICATION USE CLASIFICACION DE BLISS
 383. BOE USE BOLETIN OFICIAL DEL ESTADO

 384. BOLETIN DE LA ASOCIACION ANDALUZA DE BIBLIOTECARIOS USE
 PUBLICACIONES PERIODICAS

 385. BOLETIN DE LA ASOCIACION NACIONAL DE ARCHIVEROS BIBLIOTECARIOS Y
 DOCUMENTALISTAS USE PUBLICACIONES PERIODICAS

 386. BOLETIN OFICIAL DEL ESTADO
 387. BOLETINES
 388. BOLETINES DE ALERTA BIBLIOGRAFICA USE BOLETINES DE ALERTA
 INFORMATIVA
 389. BOLETINES DE ALERTA INFORMATIVA
 390. BOLETINES DE INDICES
 391. BOLETINES DE NOTICIAS
 392. BOLETINES DE RESUMENES
 393. BOLETINES DE SUMARIOS USE BOLETINES DE RESUMENES
 394. BOLETINES DE TABLAS DE CONTENIDO
 395. BOLETINES DE TITULOS
 396. BOLETINES ELECTRONICOS
 397. BROW CLASSIFICATION USE CLASIFICACION DE BROWN

398. BUSQUEDAS AUTOMÁTICAS
 399. BUSQUEDAS AUTOMATIZADAS
 400. BUSQUEDAS BIBLIOGRÁFICAS
 401. BUSQUEDAS DE DATOS FACTUALES
 402. BUSQUEDAS DE INFORMACIÓN
 403. BUSQUEDAS EN EL TEXTO USE BUSQUEDAS EN LENGUAJE NATURAL
 404. BUSQUEDAS EN LENGUAJE NATURAL
 405. BUSQUEDAS EN LENGUAJE NO CONTROLADO USE BUSQUEDAS EN LENGUAJE NATURAL
 406. BUSQUEDAS EN LÍNEA
 407. BUSQUEDAS EN TEXTO LIBRE USE BUSQUEDAS EN LENGUAJE NATURAL
 408. BUSQUEDAS EXHAUSTIVAS
 409. BUSQUEDAS INTERACTIVAS
 410. BUSQUEDAS MANUALES
 411. BUSQUEDAS ON LINE USE BUSQUEDAS EN LÍNEA
 412. BUSQUEDAS POR MATERIAS
 413. BUSQUEDAS RETROSPECTIVAS
 414. BUSQUEDAS TEMÁTICAS
 415. CAD USE DISEÑO ASISTIDO POR ORDENADOR
 416. CADENA DOCUMENTAL USE PROCESO DOCUMENTAL
 417. CAI USE ENSEÑANZA ASISTIDA POR ORDENADOR
 418. CAIS USE ASOCIACIONES PROFESIONALES
 419. CAJAS VIAJERAS USE BIBLIOTECAS MÓVILES
 420. CALENDARIOS
 421. CALIDAD
 422. CALIDAD DE LA INFORMACIÓN
 423. CALIDAD DE LOS SERVICIOS
 424. CALIDAD TOTAL
 425. CAMBIOS TECNOLÓGICOS USE INNOVACIONES TECNOLÓGICAS
 426. CANAL DE COMUNICACIÓN
 427. CANJE USE CANJE DE PUBLICACIONES
 428. CANJE DE PUBLICACIONES
 429. CANJE INTERNACIONAL DE PUBLICACIONES
 430. CAPTACIÓN DE REGISTROS USE CAPTURA DE REGISTROS
 431. CAPTURA DE REGISTROS
 432. CARTELES
 433. CARTOGRAFÍA
 434. CARTOTECAS
 435. CATALOGACIÓN
 436. CATALOGACIÓN ANALÍTICA
 437. CATALOGACIÓN ASISTIDA POR ORDENADOR
 438. CATALOGACIÓN AUTOMÁTICA USE CATALOGACIÓN AUTOMATIZADA
 439. CATALOGACIÓN AUTOMATIZADA
 440. CATALOGACIÓN BIBLIOGRÁFICA
 441. CATALOGACIÓN CENTRALIZADA
 442. CATALOGACIÓN COMPARTIDA USE CATALOGACIÓN COOPERATIVA
 443. CATALOGACIÓN COOPERATIVA
 444. CATALOGACIÓN CORRIENTE
 445. CATALOGACIÓN DE MATERIAL NO BIBLIOGRÁFICO USE CATALOGACIÓN DE MATERIAL NO LIBRARIO

446. CATALOGACION DE MATERIAL NO LIBRARIO
447. CATALOGACION DE MATERIAS USE CATALOGACION POR MATERIAS
448. CATALOGACION DESCRIPTIVA
449. CATALOGACION EN LA FUENTE USE CATALOGACION EN LA PUBLICACION
450. CATALOGACION EN LA PUBLICACION
451. CATALOGACION EN ORIGEN USE CATALOGACION EN LA PUBLICACION
452. CATALOGACION EXHAUSTIVA
453. CATALOGACION LEGIBLE POR MAQUINA USE MARC
454. CATALOGACION POR MATERIAS
455. CATALOGACION RETROSPECTIVA
456. CATALOGACION SELECTIVA
457. CATALOGACION SIMPLIFICADA
458. CATALOGACION TEMATICA USE CATALOGACION POR MATERIAS
459. CATALOGADORES
460. CATALOGOS
461. CATALOGOS DE TITULOS
462. CATALOGOS ALFABETICOS DE AUTORES USE CATALOGOS DE AUTORES
463. CATALOGOS ALFABETICOS DE MATERIAS USE CATALOGOS DE MATERIAS
464. CATALOGOS ALFABETICOS DE TITULOS USE CATALOGOS DE TITULOS
465. CATALOGOS AUTOMATIZADOS
466. CATALOGOS BIBLIOGRAFICOS
467. CATALOGOS CENTRALIZADOS
468. CATALOGOS CLASIFICADOS USE CATALOGOS SISTEMATICOS
469. CATALOGOS COLECTIVOS
470. CATALOGOS COMERCIALES
471. CATALOGOS COMERCIALES DE LIBROS
472. CATALOGOS CRONOLOGICOS
473. CATALOGOS DE ACCESO
474. CATALOGOS DE ACCESO PUBLICO EN LINEA
475. CATALOGOS DE ADQUISICIONES
476. CATALOGOS DE AUTORES
477. CATALOGOS DE AUTORIDAD
478. CATALOGOS DE BIBLIOTECAS
479. CATALOGOS DE LIBRERIAS
480. CATALOGOS DE LIBROS DE OCASION
481. CATALOGOS DE MATERIAS
482. CATALOGOS DE PUBLICACIONES PERIODICAS
483. CATALOGOS DE PUBLICACIONES SERIADAS USE CATALOGOS DE PUBLICACIONES PERIODICAS
484. CATALOGOS DE TITULOS
485. CATALOGOS DICCIONARIOS
486. CATALOGOS DIVIDIDOS
487. CATALOGOS EDITORIALES
488. CATALOGOS EN LIBROS USE CATALOGOS IMPRESOS
489. CATALOGOS EN LINEA

490. CATALOGOS EN LINEA DE ACCESO PUBLICO USE CATALOGOS DE
ACCESO PUBLICO EN LINEA
491. CATALOGOS IMPRESOS
492. CATALOGOS LOCALES
493. CATALOGOS MANUALES
494. CATALOGOS MONUMENTALES
495. CATALOGOS POR MATERIAS
496. CATALOGOS SISTEMATICOS
497. CATALOGOS TOPOGRAFICOS
498. CATALOGOS UNIDOS USE CATALOGOS COLECTIVOS
499. CBU USE CONTROL BIBLIOGRAFICO UNIVERSAL
500. CD ROM
501. CD-ROM USE CD ROM
502. CD-ROM EN RED USE REDES DE CD ROM
503. CDU USE CLASIFICACION DECIMAL UNIVERSAL
504. CE USE UNION EUROPEA
505. CENTRO BIBLIOTECARIO EN LINEA USE OCLC
506. CENTRO NACIONAL DE INFORMACION Y DOCUMENTACION USE CINDOC
507. CENTROS DE ANALISIS DE INFORMACION USE CENTROS DE
DOCUMENTACION
508. CENTROS DE DOCUMENTACION
509. CENTROS DE DOCUMENTACION ESPECIALIZADOS
510. CENTROS DE DOCUMENTACION EUROPEA
511. CENTROS DE INFORMACION USE CENTROS DE DOCUMENTACION
512. CENTROS DE INFORMACION ESPECIALIZADOS
513. CENTROS DE INFORMACION REGIONAL
514. CENTROS DE INVESTIGACION
515. CENTROS DE PROCESOS DE DATOS
516. CENTROS UNIVERSITARIOS USE UNIVERSIDAD
517. CIENCIA BIBLIOGRAFICA
518. CIENCIA BIBLIOTECARIA
519. CIENCIA DE LA CIENCIA
520. CIENCIA DE LA COMPUTACION
521. CIENCIA DE LA DOCUMENTACION
522. CIENCIA DE LA INFORMACION
523. CIENCIA SISTEMICA
524. CIENCIA Y TECNOLOGIA
525. CIENCIAS BIOMEDICAS USE CIENCIAS DE LA SALUD
526. CIENCIAS DE LA COMUNICACION
527. CIENCIAS DE LA DOCUMENTACION
528. CIENCIAS DE LA SALUD
529. CIENCIAS HUMANAS
530. CIENCIAS SOCIALES
531. CIENCIOMETRIA
532. CINDOC
533. CINEMATECA USE FILMOTECAS
534. CINEMATOLOGIA
535. CINTAS DE VIDEO

536. CINTAS MAGNÉTICAS
537. CINTAS MAGNETOFONICAS USE CINTAS MAGNÉTICAS
538. CINTAS PERFORADAS
539. CIP USE CLASIFICACION INTERNACIONAL DE PATENTES
540. CIRCULACION USE PRESTAMO
541. CIRCULACION DE LIBROS
542. CIRCULACION Y PRESTAMO
543. CITACION USE CITAS
544. CITAS
545. CITAS BIBLIOGRAFICAS
546. CLASIFICACION
547. CLASIFICACION AUTOMÁTICA
548. CLASIFICACION COLONADA USE CLASIFICACIONES
BIBLIOGRAFICAS
549. CLASIFICACION DE BLISS USE CLASIFICACIONES BIBLIOGRAFICAS
550. CLASIFICACION DE BROWN USE CLASIFICACIONES
BIBLIOGRAFICAS
551. CLASIFICACION DE COLON USE CLASIFICACIONES
BIBLIOGRAFICAS
552. CLASIFICACION DE CUTTER USE CLASIFICACIONES
BIBLIOGRAFICAS
553. CLASIFICACION DE DEWEY USE CLASIFICACIONES
BIBLIOGRAFICAS
554. CLASIFICACION DE DOS PUNTOS USE CLASIFICACIONES
BIBLIOGRAFICAS
555. CLASIFICACION DE LA BIBLIOTECA DEL CONGRESO USE
CLASIFICACIONES BIBLIOGRAFICAS
556. CLASIFICACION DE LA LIBRARY OF CONGRESS USE CLASIFICACIONES
BIBLIOGRAFICAS
557. CLASIFICACION DE RANGANATHAN USE CLASIFICACIONES
BIBLIOGRAFICAS
558. CLASIFICACION DECIMAL DE DEWEY USE CLASIFICACIONES
BIBLIOGRAFICAS
559. CLASIFICACION DECIMAL UNIVERSAL
560. CLASIFICACION EXPANSIVA DE CUTTER USE CLASIFICACIONES
BIBLIOGRAFICAS
561. CLASIFICACION INTERNACIONAL DE PATENTES
562. CLASIFICACION POR FACETAS USE CLASIFICACIONES FACETADAS
563. CLASIFICACIONES ALFABÉTICAS POR MATERIAS USE
CLASIFICACIONES BIBLIOGRAFICAS

564. CLASIFICACIONES ANALITICAS BIBLIOGRAFICAS	USE	CLASIFICACIONES
565. CLASIFICACIONES ANALITICO-SINTETICAS CLASIFICACIONES BIBLIOGRAFICAS		USE
566. CLASIFICACIONES BIBLIOGRAFICAS		
567. CLASIFICACIONES COMBINATORIAS BIBLIOGRAFICAS	USE	CLASIFICACIONES
568. CLASIFICACIONES ENUMERATIVAS BIBLIOGRAFICAS	USE	CLASIFICACIONES
569. CLASIFICACIONES EXPANSIVAS BIBLIOGRAFICAS	USE	CLASIFICACIONES
570. CLASIFICACIONES JERARQUICAS BIBLIOGRAFICAS	USE	CLASIFICACIONES
571. CLASIFICACIONES DE PATENTES		
572. CLASIFICACIONES DECIMALES BIBLIOGRAFICAS	USE	CLASIFICACIONES
573. CLASIFICACIONES FACETADAS BIBLIOGRAFICAS	USE	CLASIFICACIONES
574. CLASIFICACIONES JERARQUICAS BIBLIOGRAFICAS	USE	CLASIFICACIONES
575. CLASIFICACIONES LINEALES BIBLIOGRAFICAS	USE	CLASIFICACIONES
576. CLASIFICACIONES PARA BIBLIOTECAS BIBLIOGRAFICAS	USE	CLASIFICACIONES
577. CLASIFICACIONES POR MATERIAS BIBLIOGRAFICAS	USE	CLASIFICACIONES
578. CLASIFICACIONES SELECTIVAS BIBLIOGRAFICAS	USE	CLASIFICACIONES
579. CLASIFICACIONES SISTEMATICAS BIBLIOGRAFICAS	USE	CLASIFICACIONES
580. CLAVES DE BUSQUEDAS		
581. CLUBES DE LECTORES		
582. CLUBES DE LECTURA		
583. CLUBES DEL LIBRO	USE	CLUBES DE LECTURA

584. CLUSTER USE CLUSTERING
 585. CLUSTERING
 586. COBOL USE LENGUAJES DE PROGRAMACION
 587. CODICES
 588. CODICOLOGIA
 589. CODIFICACION
 590. CODIGOS ALFABETICOS
 591. CODIGOS ALFANUMERICOS
 592. CODIGOS BINARIOS
 593. CODIGOS DE BUSQUEDAS
 594. CODIGOS DE CATALOGACION USE REGLAS DE CATALOGACION
 595. CODIGOS FUENTE
 596. CODIGOS NUMERICOS
 597. COEFICIENTES DE LLAMADA
 598. COEFICIENTES DE PRECISION
 599. COEFICIENTES DE RUIDO
 600. COEFICIENTES DE SILENCIO
 601. COLECCIONES
 602. COLECCIONES DE REFERENCIA
 603. COLECCIONES DOCUMENTALES
 604. COLECCIONES DE BIBLIOTECAS
 605. COLECCIONES DE LIBROS USE COLECCIONES DE BIBLIOTECAS
 606. COLECCIONES DE MUSEOS
 607. COLECCIONES DIPLOMATICAS
 608. COLECCIONES ESPECIALES
 609. COLECCIONES LOCALES
 610. COLECCIONES PRIVADAS
 611. COLEGIOS INVISIBLES
 612. COMPACT DISC USE DISCOS COMPACTOS
 613. COMPARTIR INFORMACION EN CD ROM USE REDES DE CD ROM
 614. COMPARTIR INFORMACION EN CD-ROM USE REDES DE CD ROM
 615. COMPATIBILIDAD DE SISTEMAS
 616. COMPRESION DE LA INFORMACION
 617. COMPRESORES
 618. COMPUTADORAS USE ORDENADORES
 619. COMPUTADORAS DIGITALES USE ORDENADORES DIGITALES
 620. COMPUTADORAS ANALOGICAS USE ORDENADORES ANALOGICOS
 621. COMUNICACION CIENTIFICA

 622. COMUNICACION DE DATOS CIENTIFICOS USE COMUNICACION
 CIENTIFICA

 623. COMUNICACION DE MASAS
 624. COMUNICACION HOMBRE/MAQUINA
 625. COMUNICACION TECNOLOGICA
 626. COMUNIDAD EUROPEA
 627. CONCEPTOS
 628. CONDICIONES DE EMPLEO USE CONDICIONES DE TRABAJO
 629. CONDICIONES DE TRABAJO

630. CONFECCION AUTOMATICA DE RESUMENES USE ELABORACION
AUTOMATICA DE RESUMENES
631. CONFECCION DE RESUMENES USE ELABORACION DE RESUMENES
632. CONFERENCIAS
633. CONGRESOS
634. CONSEJO SUPERIOR DE INVESTIGACIONES CIENTIFICAS USE CSIC
635. CONSERVACION
636. CONSERVACION DE DOCUMENTOS
637. CONSERVACION DE OBRAS DE ARTE
638. CONSERVADORES DE MUSEOS
639. CONSULTAS BIBLIOGRAFICAS
640. CONSULTAS DOCUMENTALES
641. CONTROL BIBLIOGRAFICO
642. CONTROL BIBLIOGRAFICO NACIONAL
643. CONTROL BIBLIOGRAFICO UNIVERSAL
644. CONTROL DE ADQUISICIONES
645. CONTROL DE AUTORIDADES
646. CONTROL DE CALIDAD
647. CONTROL DE VOCABULARIO USE CONTROL DEL VOCABULARIO
648. CONTROL DEL VOCABULARIO
649. CONTROL TERMINOLOGICO USE CONTROL DEL VOCABULARIO
650. CONVERSION
651. CONVERSION DE DATOS
652. CONVERSION DE FORMATOS
653. CONVERSION DE LOS FICHEROS MANUALES USE CONVERSION
RETROSPECTIVA
654. CONVERSION RETROSPECTIVA
655. COOPERACION BIBLIOTECARIA
656. COOPERACION CIENTIFICA
657. COOPERACION INTERBIBLIOTECARIA USE COOPERACION
BIBLIOTECARIA
658. COOPERACION INTERNACIONAL
659. COORDINACION BIBLIOTECARIA
660. COPIADORA USE REPROGRAFIA
661. COPISTERIA USE REPROGRAFIA
662. COPYRIGHT USE DERECHOS DE AUTOR
663. CORREO ELECTRONICO
664. CORTOMETRAJES
665. COSTES
666. COSTES DE LA INFORMACION
667. COSTOS USE COSTES
668. COTEJO
669. CPU USE UNIDAD CENTRAL DE PROCESAMIENTO
670. CREACION DE BASES DE DATOS USE PRODUCCION DE BASES DE
DATOS

671. CRIPTOGRAFIA
 672. CRITERIOS DE BUSQUEDA
 673. CRITERIOS DE RECUPERACION
 674. CRITERIOS DE RELEVANCIA
 675. CSIC
 676. CUADROS DE CLASIFICACION
 677. CUESTIONARIOS USE ENCUESTAS
 678. CURRÍCULA USE CURRICULUM
 679. CURRICULUM

 680. CURSILLOS DE FORMACION BIBLIOTECARIA USE ENSEÑANZA DE
 LA BIBLIOTECONOMIA

 681. CURSOS DE ACTUALIZACION
 682. CURSOS DE BIBLIOTECONOMIA
 683. CURSOS DE DOCTORADO
 684. CURSOS DE DOCUMENTACION
 685. CURSOS DE FORMACION
 686. CURSOS DE FORMACION BIBLIOTECARIA USE ENSEÑANZA DE LA
 BIBLIOTECONOMIA

 687. CURSOS DE POSTGRADO USE ESTUDIOS DE POSTGRADO
 688. CURSOS DE VERANO
 689. CHIP USE MICROPROCESADORES
 690. DACTILOGRAFIA
 691. DATOS BIBLIOGRAFICOS USE REFERENCIAS BIBLIOGRAFICAS
 692. DATOS ESTADISTICOS
 693. DATOS HISTORICOS
 694. DC USE CLASIFICACION DECIMAL DE DEWEY
 695. DDC USE CLASIFICACION DECIMAL DE DEWEY
 696. DDS USE SERVICIOS DE DIFUSION DE DOCUMENTOS
 697. DEMANDA DE INFORMACION
 698. DEMANDA DE LOS USUARIOS USE DEMANDA DE INFORMACION
 699. DEMANDA DOCUMENTAL USE DEMANDA DE INFORMACION
 700. DEMANDA INFORMATIVA USE DEMANDA DE INFORMACION
 701. DEPOSITO LEGAL
 702. DEPOSITOS DE MANUSCRITOS
 703. DERECHO
 704. DERECHO A LA INFORMACION
 705. DERECHOS DE AUTOR
 706. DESARROLLO CULTURAL
 707. DESARROLLO DE COLECCIONES
 708. DESCODIFICACION
 709. DESCRIPCION BIBLIOGRAFICA

 710. DESCRIPCION BIBLIOGRAFICA INTERNACIONAL NORMALIZADA USE
 ISBD

 711. DESCRIPCION DE DOCUMENTOS
 712. DESCRIPCION DE LIBROS USE DESCRIPCION DE DOCUMENTOS

713. DESCRIPTORES
714. DESTRUCCION DE EXPEDIENTES USE ELIMINACIÓN DE DOCUMENTOS
715. DETERIORO DE DOCUMENTOS
716. DETERIORO DEL PAPEL
717. DEVOLUCION DE DOCUMENTOS
718. DEVOLUCION DE LIBROS USE DEVOLUCION DE DOCUMENTOS
719. DEWEY DECIMAL CLASSIFICATION USE CLASIFICACIONES BIBLIOGRAFICAS
720. DIA DEL ARCHIVERO
721. DIA DEL BIBLIOTECARIO
722. DIA DEL LIBRO
723. DIA INTERNACIONAL DEL LIBRO
724. DIALOG
725. DIAPOSITIVAS
726. DIARIOS
727. DIAZOCOPIA USE REPROGRAFIA
728. DIAZOCOPIADORA USE REPROGRAFIA
729. DICCIONARIOS
730. DICCIONARIOS AUTOMATICOS
731. DICCIONARIOS BILINGUES
732. DICCIONARIOS CIENTIFICOS
733. DICCIONARIOS COMPUTADORIZADOS USE DICCIONARIOS AUTOMATICOS
734. DICCIONARIOS ELECTRONICOS
735. DICCIONARIOS ENCICLOPEDIICOS
736. DICCIONARIOS TECNICOS
737. DICCIONARIOS TERMINOLOGICOS
738. DIFUNDIR INFORMACION USE DIFUSION DE LA INFORMACION
739. DIFUSION
740. DIFUSION CULTURAL
741. DIFUSION DE CONOCIMIENTOS
742. DIFUSION DE LA INFORMACION
743. DIFUSION DEL LIBRO
744. DIFUSION SELECTIVA DE INFORMACION USE DIFUSION SELECTIVA DE LA INFORMACION
745. DIFUSION SELECTIVA DE LA INFORMACION
746. DIGITALIZACIÓN
747. DIGITALIZACION DE DOCUMENTOS
748. DIPLOMADOS
749. DIPLOMATURA USE ENSEÑANZA SUPERIOR
750. DIPLOMATURA DE BIBLIOTECONOMIA Y DOCUMENTACION
751. DIRECTORIOS
752. DIRECTORIOS ELECTRONICOS

- 753.DISCOGRAFIA
754.DISCOS
755.DISCOS COMPACTOS
756.DISCOS COMPACTOS INTERACTIVOS
757.DISCOS DE ALMACENAMIENTO OPTICO
758.DISCOS DE GRAMOFONO USE GRABACIONES EN DISCO
759.DISCOS FONOGRAFICOS USE GRABACIONES EN DISCO
760.DISCOS MAGNETICOS
761.DISCOS OPTICOS
762.DISCOS OPTICOS NUMERICOS
763.DISCOTECAS
764.DISEMINACION DE INFORMACION USE DIFUSION DE LA INFORMACION
765.DISEMINACION DE LA INFORMACION USE DIFUSION DE LA INFORMACION
766.DISEMINACION SELECTIVA DE INFORMACION USE DIFUSION SELECTIVA DE LA INFORMACION
767.DISEMINACION SELECTIVA DE LA INFORMACION USE DIFUSION SELECTIVA DE LA INFORMACION
768.DISEMINAR INFORMACION USE DIFUSION DE LA INFORMACION
769.DISEÑO
770.DISEÑO ASISTIDO POR ORDENADOR
771.DISEÑO DE SISTEMAS
772.DISPERSION DE LA INFORMACION
773.DISPERSION DE PUBLICACIONES USE DISPERSION DE LA INFORMACION
774.DISPONIBILIDAD DE PUBLICACIONES
775.DISPONIBILIDAD UNIVERSAL DE LAS PUBLICACIONES
776.DISPONIBILIDAD UNIVERSAL DE PUBLICACIONES USE DISPONIBILIDAD UNIVERSAL DE LAS PUBLICACIONES
777.DISPOSITIVOS DE ALMACENAMIENTO
778.DISPOSITIVOS DE ALMACENAMIENTO Y RECUPERACION
779.DISPOSITIVOS DE BUSQUEDA
780.DISPOSITIVOS DE LECTURA
781.DISQUETES
782.DISQUETTES USE DISQUETTES
783.DISTRIBUCION DE BASES DE DATOS
784.DISTRIBUCION DE LA INFORMACION
785.DISTRIBUCION DE LAS PALABRAS USE LINGÜISTICA ESTADISTICA
786.DISTRIBUCION SELECTIVA DE LA INFORMACION USE DIFUSION SELECTIVA DE LA INFORMACION

- 787.DISTRIBUIDORES DE BASES DE DATOS
 788.DIVULGACION CIENTIFICA Y TECNICA
 789.DOCENTES
 790.DOCUMAT USE CONGRESOS
 791.DOCUMENTACION
 792.DOCUMENTACION ADMINISTRATIVA
 793.DOCUMENTACIÓN APLICADA
 794.DOCUMENTACION AUDIOVISUAL
 795.DOCUMENTACION AUTOMATIZADA
 796.DOCUMENTACION CIENTIFICA
- 797.DOCUMENTACION CIENTIFICA-TECNICA USE DOCUMENTACION
 CIENTIFICA
- 798.DOCUMENTACION CINEMATOGRAFICA
 799.DOCUMENTACION CLINICA USE DOCUMENTACION MEDICA
 800.DOCUMENTACION DEPORTIVA
 801.DOCUMENTACION ECONOMICA
 802.DOCUMENTACION EDUCATIVA
 803.DOCUMENTACION EMPRESARIAL
 804.DOCUMENTACION ESCRITA
 805.DOCUMENTACION HISTORICA
 806.DOCUMENTACION JUDICIAL USE DOCUMENTACION JURIDICA
 807.DOCUMENTACION JURIDICA
 808.DOCUMENTACION LEGISLATIVA
- 809.DOCUMENTACION MECANIZADA USE DOCUMENTACION
 AUTOMATIZADA
- 810.DOCUMENTACION MEDICA
 811.DOCUMENTACION MUSICAL
 812.DOCUMENTACION NOTARIAL
 813.DOCUMENTACION PERIODISTICA
 814.DOCUMENTACION POLITICA
 815.DOCUMENTACION PUBLICA
 816.DOCUMENTACION VISUAL
 817.DOCUMENTALISTAS
 818.DOCUMENTOGRAFIA
 819.DOCUMENTOLOGIA
 820.DOCUMENTOLOGOS
 821.DOCUMENTOS
 822.DOCUMENTOS AUDIOVISUALES
- 823.DOCUMENTOS CENSURADOS USE DOCUMENTOS DE
 CIRCULACION RESTRINGIDA
- 824.DOCUMENTOS CIENTIFICOS
- 825.DOCUMENTOS CLASIFICADOS USE DOCUMENTOS DE CIRCULACION
 RESTRINGIDA

- 826.DOCUMENTOS CONFIDENCIALES USE DOCUMENTOS DE
CIRCULACION RESTRINGIDA
- 827.DOCUMENTOS CONTABLES
828.DOCUMENTOS DE ARCHIVO
829.DOCUMENTOS DE CIRCULACION RESTRINGIDA
830.DOCUMENTOS ECLESIASTICOS
831.DOCUMENTOS FUENTE
832.DOCUMENTOS GRAFICOS
- 833.DOCUMENTOS GUBERNAMENTALES USE PUBLICACIONES
OFICIALES
- 834.DOCUMENTOS INEDITOS
835.DOCUMENTOS JURIDICOS
836.DOCUMENTOS NO PUBLICADOS USE DOCUMENTOS INEDITOS
837.DOCUMENTOS OFICIALES
838.DOCUMENTOS PRIMARIOS
- 839.DOCUMENTOS SECRETOS USE DOCUMENTOS DE CIRCULACION
RESTRINGIDA
- 840.DOCUMENTOS SECUNDARIOS
841.DOCUMENTOS SONOROS
842.DONACIONES
843.DSI USE DIFUSION SELECTIVA DE LA INFORMACION
844.DUP USE DISPONIBILIDAD UNIVERSAL DE LAS PUBLICACIONES
- 845.ECONOMIA*
- 846.ECONOMIA DE LA INFORMACION
847.ECUACION DE BUSQUEDA USE ECUACIONES DE INTERROGACION
848.ECUACIONES DE INTERROGACION
849.ECHO
850.EDI USE INTERCAMBIO ELECTRONICO DE DATOS
851.EDICION ELECTRONICA
852.EDITORES
853.EDITORIALES
854.EDITORIALES OFICIALES
855.EDITORIALES PRIVADAS
856.EDUCACION A DISTANCIA
857.EDUCACION ABIERTA USE EDUCACION A DISTANCIA
858.EDUCACION CONTINUA USE FORMACION CONTINUADA
859.EDUCACION POR CORRESPONDENCIA USE EDUCACION A DISTANCIA
860.EDUCACION SUPERIOR USE ENSEÑANZA SUPERIOR
861.EDUCACION Y BIBLIOTECA USE PUBLICACIONES PERIODICAS
862.ELABORACION AUTOMATICA DE RESUMENES
- 863.ELABORACION AUTOMATICA DE TESAUROS

- 864.ELABORACION DE BASES DE DATOS USE PRODUCCION DE
BASES DE DATOS
- 865.ELIMINACION DE DOCUMENTOS
- 866.ELIMINACION DE LIBROS USE ELIMINACION DE DOCUMENTOS
- 867.EMPRESAS DE DOCUMENTACION
- 868.EMPRESAS DE INFORMACION USE EMPRESAS DE
DOCUMENTACION
- 869.EMPRESAS EDITORIALES
- 870.ENCABEZAMIENTOS DE MATERIA
- 871.ENCICLOPEDIAS
- 872.ENCICLOPEDIAS ELECTRONICAS
- 873.ENCUADERNACION
- 874.ENCUESTAS
- 875.ENSEÑANZA ASISTIDA POR ORDENADOR
- 876.ENSEÑANZA DE LA ARCHIVISTICA
- 877.ENSEÑANZA DE LA BIBLIOTECONOMIA
- 878.ENSEÑANZA DE LA DOCUMENTACION
- 879.ENSEÑANZA PRIMARIA
- 880.ENSEÑANZA SECUNDARIA
- 881.ENSEÑANZA SUPERIOR
- 882.ENSEÑANZA UNIVERSITARIA USE ENSEÑANZA SUPERIOR
- 883.ENSEÑANZA ABIERTA USE EDUCACION A DISTANCIA
- 884.ENTRADA A UN SISTEMA
- 885.ENTRADAS BIBLIOGRAFICAS USE REFERENCIAS
BIBLIOGRAFICAS
- 886.ENTRADAS CATALOGAFICAS USE FICHAS CATALOGAFICAS
- 887.ENTREVISTAS
- 888.ENTROPIA DOCUMENTAL
- 889.EQUIPOS ELECTRONICOS
- 890.EQUIPOS INFORMATICOS
- 891.EQUIPOS LECTORES DE MICROFORMAS USE LECTORES DE
MICROFORMAS
- 892.EQUIPOS PERIFERICOS
- 893.ERIC
- 894.ESCRITURA
- 895.ESCRITURA ELECTRONICA
- 896.ESCUELAS DE ARCHIVISTICA
- 897.ESCUELAS DE ARCHIVONOMIA USE ESCUELAS DE ARCHIVISTICA
- 898.ESCUELAS DE BIBLIOTECARIOS
- 899.ESCUELAS DE BIBLIOTECOLOGIA USE ESCUELAS UNIVERSITARIAS
DE BIBLIOTECONOMIA Y DOCUMENTACION
- 900.ESCUELAS DE BIBLIOTECONOMIA USE ESCUELAS
UNIVERSITARIAS DE BIBLIOTECONOMIA Y DOCUMENTACION

901.ESCUELAS DE CIENCIAS DE LA INFORMACION	USE	ESCUELAS UNIVERSITARIAS DE BIBLIOTECONOMIA Y DOCUMENTACION
902.ESCUELAS TECNICAS SUPERIORES	USE	UNIVERSIDAD
903.ESCUELAS UNIVERSITARIAS		
904.ESCUELAS UNIVERSITARIAS DE BIBLIOTECONOMIA Y DOCUMENTACION		
905.ESPECIALISTAS EN INFORMACION	USE	DOCUMENTALISTAS
906.ESPECIFICIDAD DE LA INDIZACION	USE	ESPECIFICIDAD EN LA INDIZACION
907.ESPECIFICIDAD EN LA INDIZACION		
908.ESPECIFICIDAD EN LA RECUPERACION		
909.ESPECIFICIDAD DE LA RECUPERACION	USE	ESPECIFICIDAD EN LA RECUPERACION
910.ESTADISTICA		
911.ESTADISTICA BIBLIOGRAFICA	USE	BIBLIOMETRIA
912.ESTADO DE LA CUESTION		
913.ESTADO DEL ARTE	USE	ESTADO DE LA CUESTION
914.ESTAMPAS		
915.ESTANDARIZACION	USE	NORMALIZACION
916.ESTANTERIAS		
917.ESTANTERIAS ABIERTAS	USE	ESTANTERIAS DE LIBRE ACCESO
918.ESTANTERIAS CERRADAS		
919.ESTANTERIAS DE ACCESO LIBRE	USE	ESTANTERIAS DE LIBRE ACCESO
920.ESTANTERIAS DE LIBRE ACCESO		
921.ESTRATEGIAS BOOLEANAS	USE	OPERADORES BOOLEANOS
922.ESTRATEGIAS DE BUSQUEDA		
923.ESTUDIANTES		
924.ESTUDIO DE CASOS		
925.ESTUDIO DE MERCADO		
926.ESTUDIO EVOLUTIVO	USE	EVOLUCION HISTORICA
927.ESTUDIO BIBLIOMETRICO	USE	ANALISIS BIBLIOMETRICOS
928.ESTUDIO COMPARATIVO	USE	ANALISIS COMPARATIVO
929.ESTUDIO DE CITAS	USE	ANALISIS DE CITAS
930.ESTUDIOS DE POSTGRADO		
931.ESTUDIO DE RESULTADOS	USE	ANALISIS DE RESULTADOS
932.ESTUDIOS DE TERCER CICLO		
933.ESTUDIO DE USUARIOS		
934.ESTUDIO ESTADISTICO	USE	ANALISIS ESTADISTICOS
935.ETAPAS	USE	FASES
936.ETICA		

937. ETIQUETADO DE LIBROS USE PREPARACION DE DOCUMENTOS
938. ETIQUETADOR MORFOLOGICO
939. ETIQUETADOR SEMANTICO
940. ETIQUETADOR SINTACTICO
941. EUDORA USE CORREO ELECTRONICO
942. EVALUACION
943. EVALUACION DE FONDOS
944. EVALUACION DE LA INFORMACION
945. EVALUACION DE LOS SERVICIOS
946. EVOLUCION HISTORICA
947. EXHAUSTIVIDAD EN LA INDIZACION
948. EXHAUSTIVIDAD EN LA RECUPERACION
949. EXISTENCIAS DE LIBROS USE FONDOS BIBLIOGRAFICOS
950. EXPORTACION DE REGISTROS BIBLIOGRAFICOS
951. EXPOSICIONES
952. EXPOSICIONES BIBLIOGRAFICAS
953. EXPOSICIONES DOCUMENTALES
954. EXPURGO
955. EXTENSION BIBLIOTECARIA
956. FACSIMIL

957. FACULTAD DE BIBLIOTECONOMIA Y DOCUMENTACION

958. FACULTAD DE DOCUMENTACION USE FACULTAD DE
BIBLIOTECONOMIA Y DOCUMENTACION

959. FASES

960. FEDERACION ESPAÑOLA DE SOCIEDADES DE ARCHIVISTICA
BIBLIOTECONOMIA Y DOCUMENTACION USE FESABID

961. FEDERACION INTERNACIONAL DE ASOCIACIONES DE BIBLIOTECAS USE
IFLA

962. FEDERACION INTERNACIONAL DE DOCUMENTACION USE FID

963. FEDERACION NACIONAL DE SERVICIOS DE INDIZACION Y RESUMEN USE
NFAIS

964. FEEDBACK USE RETROALIMENTACION
965. FERIA DEL LIBRO
966. FESABID
967. FICHAS CATALOGAFICAS
968. FICHAS DE CATALOGOS USE FICHAS CATALOGAFICAS
969. FICHAS PERFORADAS
970. FICHEROS
971. FICHEROS DE AUTORIDADES
972. FICHEROS INFORMATICOS
973. FID
974. FILE TRANSFERENCE PROTOCOL USE FTP

- 975.FILMOTECAS
- 976.FILMS USE PELICULAS
- 977.FILOLOGIA
- 978.FILOSOFIA
- 979.FLUJO DE DATOS

- 980.FLUJO DE INFORMACION USE TRANSFERENCIA DE LA INFORMACION

- 981.FOLLETOS
- 982.FOMENTO DE LA LECTURA
- 983.FONDOS
- 984.FONDOS ANTIGUOS
- 985.FONDOS ARCHIVISTICOS
- 986.FONDOS ARTISTICOS
- 987.FONDOS BIBLIOGRAFICOS
- 988.FONDOS DOCUMENTALES
- 989.FONDOS ESPECIALES
- 990.FONDOS ESPECIALIZADOS
- 991.FONDOS HISTORICOS
- 992.FONDOS MUSEISTICOS
- 993.FONOTECAS
- 994.FORMACION CONTINUA USE FORMACION CONTINUADA
- 995.FORMACION CONTINUADA

- 996.FORMACION DE ARCHIVEROS USE FORMACION DE PROFESIONALES DE LA INFORMACION

- 997.FORMACION DE BIBLIOTECARIOS USE FORMACION DE PROFESIONALES DE LA INFORMACION

- 998.FORMACION DE COLECCIONES USE DESARROLLO DE COLECCIONES

- 999.FORMACION DE DOCUMENTALISTAS USE FORMACION DE PROFESIONALES DE LA INFORMACION

- 1000.FORMACION DE MUSEOLOGOS USE FORMACION DE PROFESIONALES DE LA INFORMACION

- 1001.FORMACION DE PROFESIONALES USE FORMACION PROFESIONAL
- 1002.FORMACION DE PROFESIONALES DE LA INFORMACION
- 1003.FORMACION DE RACIMOS USE CLUSTERING
- 1004.FORMACION DE USUARIOS
- 1005.FORMACION PERMANENTE USE FORMACION CONTINUADA
- 1006.FORMACION PROFESIONAL
- 1007.FORMATO MARC USE MARC
- 1008.FORMATOS BIBLIOGRAFICOS
- 1009.FORMATOS DE REGISTROS
- 1010.FORMATOS GRAFICOS

1011. FORMULARIOS
 1012. FORTRAN USE LENGUAJES DE PROGRAMACION
 1013. FOTOCOMPOSICION
 1014. FOTOCOPIAS
 1015. FOTODOCUMENTACION
 1016. FOTOGRAFIA
 1017. FOTOGRAFIAS
 1018. FOTOS USE FOTOGRAFIAS
 1019. FOTOTECAS
 1020. FRECUENCIA DE CITACION
 1021. FRECUENCIA DE LAS PALABRAS
 1022. FRECUENCIA DE TERMINOS USE LINGÜISTICA ESTADÍSTICA
 1023. FTP
 1024. FUENTES ARCHIVISTICAS
 1025. FUENTES DE INFORMACION
 1026. FUENTES DE INFORMACION ELECTRONICAS
 1027. FUENTES DE REFERENCIA
 1028. FUENTES DOCUMENTALES
 1029. FUENTES INFORMATIVAS USE FUENTES DE INFORMACION
 1030. FUENTES PRIMARIAS USE DOCUMENTOS PRIMARIOS
 1031. FUENTES SECUNDARIAS USE DOCUMENTOS SECUNDARIOS
 1032. FUSION DE FICHEROS USE COTEJO
 1033. GABINETES DE LECTURA
 1034. GESTION BIBLIOTECARIA
 1035. GESTION DE ARCHIVOS
 1036. GESTION DE BASES DE DATOS
 1037. GESTION DE BIBLIOTECAS USE GESTION BIBLIOTECARIA
 1038. GESTION DE DOCUMENTOS
 1039. GESTION DE LA CALIDAD
 1040. GESTION DE LA DOCUMENTACION
 1041. GESTION DE LA INFORMACION
 1042. GESTION DE MUSEOS
 1043. GESTION DE RECURSOS DE INFORMACION
 1044. GESTION DE REGISTROS BIBLIOGRAFICOS
 1045. GESTION DOCUMENTAL
 1046. GESTION PRESUPUESTARIA
 1047. GESTOR DE BASES DE DATOS
 1048. GESTOR DE LA INFORMACION
 1049. GESTOR DOCUMENTAL
 1050. GLOSARIOS USE VOCABULARIOS
 1051. GOPHER
 1052. GRABACIONES EN DISCO

 1053. GRADO DE INDIZACION USE PROFUNDIDAD EN LA INDIZACION

 1054. GRUPOS DE DISCUSION
 1055. GRUPOS DE NOTICIAS USE GRUPOS DE DISCUSION
 1056. GUÍAS
 1057. GUÍAS DE FUENTES
 1058. GUÍAS DE LECTURA

- 1059.GUIAS DEL LECTOR
1060.HABITOS DE INFORMACION
1061.HABITOS DE LECTURA
1062.HABITOS DE LOS LECTORES USE HABITOS DE LECTURA
1063.HABITOS LECTORES USE HABITOS DE LECTURA
1064.HARDWARE
1065.HEMEROTECAS
1066.HIPERDOCUMENTOS
1067.HIPERMEDIA
1068.HIPERTEXTO
1069.HISTORIA DE LA BIBLIOGRAFIA
1070.HISTORIA DE LA BIBLIOTECONOMIA
1071.HISTORIA DE LA DOCUMENTACION
1072.HISTORIA DE LA ESCRITURA
1073.HISTORIA DE LA IMPRENTA
1074.HISTORIA DE LAS BIBLIOTECAS
1075.HISTORIA DE LOS ARCHIVOS
1076.HISTORIA DEL LIBRO
1077.HOLOGRAFIA
1078.HORA DEL CUENTO
1079.HORARIOS
1080.HTML USE LENGUAJE HTML
1081.HYPERMEDIA USE HIPERMEDIA
1082.IA USE INTELIGENCIA ARTIFICIAL
1083.ICONOTECAS
1084.ICYT
1085.IDENTIFICADORES
1086.IFLA
1087.ILUMINACION
1088.ILUSTRACIONES
1089.IMAGENES DIGITALES
1090.IMAGENES ELECTRONICAS
1091.IMAGENES MECANICAS
1092.IMPORTACION DE REGISTROS BIBLIOGRAFICOS
1093.IMPRENTAS
1094.IMPRESORES
1095.IMPRESORAS
1096.INCUNABLES
1097.INDEXACION USE INDIZACION

1098.INDEXACION ASISTIDA POR ORDENADOR USE INDIZACION ASISTIDA
POR ORDENADOR

1099.INDEXACION AUTOMATICA USE INDIZACION AUTOMATICA
1100.INDEXACION AUTOMATIZADA USE INDIZACION AUTOMATICA
1101.INDICACION DE REFERENCIAS USE ORIENTACION A LOS USUARIOS
1102.INDICADORES BIBLIOMETRICOS
1103.INDICADORES DE ROL
1104.INDICES
1105.INDICES ACUMULATIVOS

1106. INDICES ALFABETICOS
 1107. INDICES ALFABETICOS POR MATERIAS
 1108. INDICES ALFANUMERICOS
 1109. INDICES AUTOMATICOS
 1110. INDICES BIBLIOGRAFICOS
 1111. INDICES CLASIFICADOS
 1112. INDICES COORDINADOS
 1113. INDICES CORRELATIVOS
 1114. INDICES DE AUTORES
 1115. INDICES DE CITAS
 1116. INDICES DE CITAS BIBLIOGRAFICAS USE INDICES DE CITAS
 1117. INDICES DE CONTENIDOS
 1118. INDICES DE EXHAUSTIVIDAD EN LA RECUPERACION
 1119. INDICES DE LIBROS PROHIBIDOS
 1120. INDICES DE MATERIAS
 1121. INDICES DE PALABRAS CLAVE
 1122. INDICES DE PERTINENCIA
 1123. INDICES DE PRECISION
 1124. INDICES DE PUBLICACIONES DE RESUMENES
 1125. INDICES DE REFERENCIAS BIBLIOGRAFICAS USE INDICES DE
 CITAS
 1126. INDICES DE RELEVANCIA
 1127. INDICES DE RUIDO
 1128. INDICES DE SILENCIO
 1129. INDICES DE TERMINOS PERMUTADOS
 1130. INDICES DE TITULOS PERMUTADOS
 1131. INDICES GEOGRAFICOS
 1132. INDICES KWAC
 1133. INDICES KWIC
 1134. INDICES KWOC
 1135. INDICES KWOT
 1136. INDICES ONOMASTICOS
 1137. INDICES PERMUTADOS
 1138. INDICES PRECIS
 1139. INDICES PRECOORDINADOS
 1140. INDICES SISTEMATICOS
 1141. INDIZACION
 1142. INDIZACION ANALITICA
 1143. INDIZACION ASISTIDA POR ORDENADOR
 1144. INDIZACION AUTOMATICA
 1145. INDIZACION COMPUTADORIZADA USE INDIZACION AUTOMATICA
 1146. INDIZACION COORDINADA
 1147. INDIZACION DE TITULOS PERMUTADOS
 1148. INDIZACION EN CADENA
 1149. INDIZACION EXHAUSTIVA USE EXHAUSTIVIDAD EN LA INDIZACION
 1150. INDIZACION MECANIZADA USE INDIZACION AUTOMATICA
 1151. INDIZACION POR ASOCIACION
 1152. INDIZACION POR CITAS
 1153. INDIZACION POR COMPUTADORA USE INDIZACION AUTOMATICA
 1154. INDIZACION POR DESCRIPTORES

- 1155.INDIZACION POR ENCABEZAMIENTOS DE MATERIA
 1156.INDIZACION POR MAQUINA USE INDIZACION AUTOMATICA
 1157.INDIZACION POR MATERIAS
 1158.INDIZACION POR ORDENADOR USE INDIZACION AUTOMATICA
 1159.INDIZACION POR PALABRAS CLAVE
 1160.INDIZACION POR UNITERMINOS
 1161.INDIZACION POSTCOORDINADA
 1162.INDIZACION PRECOORDINADA
 1163.INDIZACION SEMIAUTOMATICA
 1164.INDIZADOS COORDINADOS
 1165.INDUSTRIA CULTURAL
 1166.INDUSTRIA DE LA COMUNICACION
 1167.INDUSTRIA DE LA INFORMACION
 1168.INDUSTRIA DEL LIBRO USE INDUSTRIA EDITORIAL
 1169.INDUSTRIA EDITORIAL
 1170.INDUSTRIA ELECTRONICA
 1171.INDUSTRIA INFORMATICA
 1172.INDUSTRIA VIDEOGRAFICA
 1173.INDUSTRIA DE LA LENGUA
 1174.INFOMETRIA USE BIBLIOMETRIA
 1175.INFORMACION
 1176.INFORMACION AUTOMATIZADA
 1177.INFORMACION BIBLIOGRAFICA
 1178.INFORMACION CIENTIFICA
 1179.INFORMACION CIUDADANA
 1180.INFORMACION CLASIFICADA
 1181.INFORMACION CLINICA USE DOCUMENTACION CLINICA
 1182.INFORMACION COMERCIAL
 1183.INFORMACION CORPORATIVA
 1184.INFORMACION DE ACTUALIDAD
 1185.INFORMACION DEPORTIVA
 1186.INFORMACION ECONOMICA
 1187.INFORMACION ELECTRONICA
 1188.INFORMACION EN HUMANIDADES
 1189.INFORMACION EN LA EMPRESA
 1190.INFORMACION EN LINEA
 1191.INFORMACION ESPECIALIZADA
 1192.INFORMACION ESTADISTICA USE DATOS ESTADISTICOS
 1193.INFORMACION ESTRATEGICA
 1194.INFORMACION FACTUAL
 1195.INFORMACION FORMAL
 1196.INFORMACION HISTORICA
 1197.INFORMACION HUMANISTICA
 1198.INFORMACION INFORMAL
 1199.INFORMACION INTERNA
 1200.INFORMACION JURIDICA
 1201.INFORMACION LOCAL
 1202.INFORMACION MEDICA USE DOCUMENTACION MEDICA
 1203.INFORMACION MULTIMEDIA
 1204.INFORMACION ON LINE USE INFORMACION EN LINEA

1205. INFORMACION PARLAMENTARIA
 1206. INFORMACION POLITICA
 1207. INFORMACION PRIMARIA
 1208. INFORMACION SECUNDARIA
 1209. INFORMACION TECNICA USE INFORMACION CIENTIFICA
 1210. INFORMACION TECNOLOGICA USE INFORMACION CIENTIFICA
 1211. INFORMACION TERRITORIAL
 1212. INFORMACION VISUAL
 1213. INFORMACION Y DOCUMENTACION
 1214. INFORMATICA
 1215. INFORMATICA DOCUMENTAL
 1216. INFORMATICA JURIDICA
 1217. INFORMATICOS
 1218. INFORMATION RETRIEVAL USE RECUPERACION DE LA INFORMACION
 1219. INFORMATIZACION

 1220. INFORMATIZACION BIBLIOTECARIA USE INFORMATIZACION DE
 BIBLIOTECAS

 1221. INFORMATIZACION DE ARCHIVOS
 1222. INFORMATIZACION DE BIBLIOTECAS
 1223. INFORMES
 1224. INFORMES ANUALES

 1225. INFORMES DE INVESTIGACIONES EN CURSO USE
 INVESTIGACIONES EN CURSO

 1226. INFORMES DE INVESTIGACIONES EN PROGRESO USE
 INVESTIGACIONES EN CURSO

 1227. INFORMES TECNICOS
 1228. INFORMETRIA
 1229. INGENIERIA DEL CONOCIMIENTO
 1230. INNOVACIONES BIBLIOTECARIAS
 1231. INNOVACIONES TECNOLOGICAS
 1232. INPUT USE ENTRADA A UN SISTEMA
 1233. INSTALACIONES BIBLIOTECARIAS

 1234. INSTITUCIONES DE INVESTIGACION USE CENTROS DE
 INVESTIGACION

 1235. INSTITUTO DE INFORMACION EN CIENCIAS SOCIALES Y HUMANIDADES
 USE ISOC

 1236. INSTITUTO DE INFORMACION Y DOCUMENTACION EN CIENCIA Y
 TECNOLOGIA USE ICYT

 1237. INTELIGENCIA ARTIFICIAL
 1238. INTERCAMBIO CULTURAL
 1239. INTERCAMBIO DE INFORMACION

1240. INTERCAMBIO DE INFORMACION BIBLIOGRAFICA
 1241. INTERCAMBIO DE PUBLICACIONES USE CANJE DE PUBLICACIONES
 1242. INTERCAMBIO DE REGISTROS BIBLIOGRAFICOS
 1243. INTERCAMBIO ELECTRONICO DE DATOS
 1244. INTERCONEXION DE SISTEMAS ABIERTOS
 1245. INTERFACE USE INTERFAZ
 1246. INTERFASES USE INTERFAZ
1247. INTERFASES DE ACCESO DIRECTO USE INTERFAZ DE ACCESO DIRECTO
1248. INTERFAZ
 1249. INTERFAZ DE ACCESO DIRECTO
 1250. INTERNATIONAL STANDARD BIBLIOGRAPHIC DESCRIPTION USE ISBD
 1251. INTERNATIONAL STANDARD BOOK NUMBERING USE ISBN
 1252. INTERNATIONAL STANDARD SERIAL NUMBERING USE ISSN
 1253. INTERNET
 1254. INTRANETS
 1255. INVENTARIOS
 1256. INVENTARIOS ARTISTICOS
 1257. INVESTIGACION APLICADA
 1258. INVESTIGACION BASICA
 1259. INVESTIGACION BILIOGRAFICA
 1260. INVESTIGACION CIENTIFICA
 1261. INVESTIGACION EN INFORMACION
 1262. INVESTIGACION HISTORICA
 1263. INVESTIGACION Y DESARROLLO
 1264. INVESTIGACIONES EN CURSO
1265. INVESTIGACIONES EN PROCESO USE INVESTIGACIONES EN CURSO
1266. IP USE PROTOCOLO DE INTERNET
 1267. IR USE RECUPERACION DE LA INFORMACION
 1268. ISBD
 1269. ISBD (B)
 1270. ISBD (M)
 1271. ISBN
 1272. ISBN (NBM)
 1273. ISKO USE ASOCIACIONES PROFESIONALES
 1274. ISO
 1275. ISOC
 1276. ISSN
 1277. ITEM
 1278. JASIS
1279. JORNADAS ESPAÑOLAS DE DOCUMENTACION AUTOMATIZADA USE CONGRESOS
1280. JORNADES CATALANES DE DOCUMENTACIO

1281. JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE USE
JASIS
1282. JUEGOS DE FICHAS USE FICHAS CATALOGRAFICAS
1283. KEYWORD AND CONTEX INDEX USE INDICES KWAC
1284. KEYWORD IN CONTEX INDEX USE INDICES KWIC
1285. KEYWORD OUT OF CONTEXT INDEX USE INDICES KWOC
1286. KEYWORD OUT OF TITLES INDEX USE INDICES KWOT
1287. LAN USE RED DE AREA LOCAL
1288. LARGOMETRAJES USE PELICULAS
1289. LASER
1290. LCC
1291. LECTORES
1292. LECTORES DE MICROFORMAS
1293. LECTURA
1294. LECTURA ASISTIDA POR ORDENADOR
1295. LECTURA INFANTIL
1296. LECTURA OPTICA
1297. LEGISLACION BIBLIOTECARIA
1298. LEGISLACION BIBLIOTECOLOGICA USE LEGISLACION BIBLIOTECARIA
1299. LEGISLACION SOBRE BIBLIOTECAS USE LEGISLACION BIBLIOTECARIA
1300. LEGISLACION SOBRE INFORMACION
1301. LENGUAJE HTML
1302. LENGUAJE LIBRE USE LENGUAJES NATURALES
1303. LENGUAJES
1304. LENGUAJES ARTIFICIALES
1305. LENGUAJES CONTROLADOS
1306. LENGUAJES DE CLASIFICACION
1307. LENGUAJES DE COMPUTACION USE LENGUAJES DE
PROGRAMACION
1308. LENGUAJES DE DESCRIPCION DE DATOS
1309. LENGUAJES DE DESCRIPTORES USE TESAUROS
1310. LENGUAJES DE INDIZACION
1311. LENGUAJES DE INTERROGACION
1312. LENGUAJES DE INTERROGACION ESTRUCTURADO USE SQL
1313. LENGUAJES DE MANIPULACION DE DATOS
1314. LENGUAJES DE MAQUINA
1315. LENGUAJES DE PROGRAMACION
1316. LENGUAJES DE RECUPERACION
1317. LENGUAJES DE USUARIO
1318. LENGUAJES DOCUMENTALES
1319. LENGUAJES DOCUMENTARIOS USE LENGUAJES DOCUMENTALES
1320. LENGUAJES ESPECIALIZADOS
1321. LENGUAJES FORMALES
1322. LENGUAJES NATURALES
1323. LENGUAJES PUENTE USE LENGUAJES DOCUMENTALES
1324. LEXICOGRAFIA
1325. LEXICOLOGIA
1326. LEXICOS USE DICCIONARIOS
1327. LEY DE BRADFORD

1328. LEY DE DISPERSION USE LEY DE BRADFORD
 1329. LEY DE LA PRODUCTIVIDAD CIENTIFICA USE LEY DE LOTKA
 1330. LEY DE LOTKA
 1331. LEY DE PRODUCTIVIDAD CIENTIFICA USE LEY DE LOTKA
 1332. LEYENDA BIBLIOGRAFICA
 1333. LIBERTAD DE INFORMACION USE DERECHO A LA INFORMACION
 1334. LIBRARY ASSOCIATION USE ASOCIACIONES PROFESIONALES

 1335. LIBRARY OF CONGRESS CLASSIFICATION USE CLASIFICACION DE LA
 BIBLIOTECA DEL CONGRESO

 1336. LIBRE CIRCULACION DE LA INFORMACION
 1337. LIBRERIAS
 1338. LIBRERIAS ANTICUARIAS
 1339. LIBRERIAS DE OCASION
 1340. LIBRERJAS DE VIEJO
 1341. LIBREROS
 1342. LIBRO DE LAS HORAS
 1343. LIBROS
 1344. LIBROS ANTIGUOS
 1345. LIBROS DE ACTAS
 1346. LIBROS DE BOLSILLO
 1347. LIBROS DE CONTABILIDAD
 1348. LIBROS DE DEPOSITO
 1349. LIBROS DE IMAGENES
 1350. LIBROS DE OCASION
 1351. LIBROS DE REFERENCIA USE OBRAS DE REFERENCIA
 1352. LIBROS DE TEXTO
 1353. LIBROS ELECTRONICOS
 1354. LIBROS INFANTILES
 1355. LIBROS JUVENILES
 1356. LIBROS PARA CIEGOS
 1357. LIBROS PARA JOVENES USE LIBROS JUVENILES
 1358. LIBROS PARA NIÑOS USE LIBROS INFANTILES
 1359. LIBROS PROHIBIDOS
 1360. LIBROS RAROS
 1361. LICENCIATURA USE ENSEÑANZA SUPERIOR
 1362. LINGÜISTAS
 1363. LINGÜISTAS COMPUTACIONALES
 1364. LINGÜISTICA
 1365. LINGÜISTICA ESTADISTICA
 1366. LISA
 1367. LISTA AUTORIZADA DE TERMINOS
 1368. LISTA DE ENCABEZAMIENTOS DE MATERIA DEL CONGRESO USE LCSH
 1369. LISTA DE ENCABEZAMIENTOS DE MATERIAS
 1370. LISTA DE TERMINOS AUTORIZADA USE LISTA AUTORIZADA DE
 TERMINOS
 1371. LISTAS DE PALABRAS VACIAS
 1372. LISTAS DE TERMINOS
 1373. LITERATURA CIENTIFICA

- 1374.LITERATURA GRIS
 1375.LOCAL AREA NETWORK USE RED DE AREA LOCAL
 1376.LOGICA BOOLEANA USE OPERADORES BOOLEANOS
 1377.LOGICALES USE SOFTWARE
 1378.LOGICIALES USE SOFTWARE
 1379.LUDOTECAS
 1380.MACROORDENADORES
 1381.MACROTESAUROS
 1382.MACHINE READABLE CATALOG USE MARC
 1383.MACINTOSH
 1384.MANTENIMIENTO DE BASES DE DATOS

 1385.MANTENIMIENTO DE DOCUMENTOS USE PRESERVACION DE LOS
 DOCUMENTOS

 1386.MANTENIMIENTO DE FONDOS

 1387.MANTENIMIENTO DE LIBROS USE PRESERVACION DE LOS
 DOCUMENTOS

 1388.MANTENIMIENTO DE LOS FICHEROS

 1389.MANTENIMIENTO DE LOS REGISTROS DE ARCHIVO USE PRESERVACION
 DE LOS REGISTROS DE ARCHIVO

 1390.MANTENIMIENTO DE TESAUROS
 1391.MANUALES DE CALIDAD
 1392.MANUALES DE PROCEDIMIENTO
 1393.MANUSCRITOS
 1394.MANUSCRITOS ANTIGUOS USE MANUSCRITOS
 1395.MAPAS
 1396.MAPOTECA
 1397.MARC
 1398.MARKETING
 1399.MATERIAL AUDIOVISUAL
 1400.MATERIAL AUDITIVO
 1401.MATERIAL BIBLIOGRAFICO USE FONDOS BIBLIOGRAFICOS
 1402.MATERIAL CARTOGRAFICO
 1403.MATERIAL DE ARCHIVO USE REGISTROS DE ARCHIVO
 1404.MATERIAL DE REFERENCIA USE OBRAS DE REFERENCIA
 1405.MATERIAL ESPECIAL
 1406.MATERIAL IMPRESO USE PUBLICACIONES
 1407.MATERIAL INFORMatico
 1408.MATERIAL NO BIBLIOGRAFICO USE MATERIAL NO LIBRARIO
 1409.MATERIAL NO LIBRARIO
 1410.MATERIAL TRIDIMENSIONAL
 1411.MATERIAL VISUAL

 1412.MECANIZACION DE BIBLIOTECAS USE INFORMATIZACION DE
 BIBLIOTECAS

1413.MEDIATECAS		
1414.MEDICINA		
1415.MEDIOS AUDIOVISUALES		
1416.MEDIOS DE COMUNICACION SOCIAL	USE	MEDIOS DE COMUNICACION SOCIAL
1417.MEDIOS DE COMUNICACION DE MASAS SOCIAL	USE	MEDIOS DE COMUNICACION SOCIAL
1418.MEDIOS DE COMUNICACION ELECTRONICOS		
1419.MEDIOS DE COMUNICACION IMPRESOS	USE	MATERIAL IMPRESO
1420.MEDIOS DE COMUNICACION SOCIAL		
1421.MEDIOS DE INFORMACION SOCIAL	USE	MEDIOS DE COMUNICACION SOCIAL
1422.MEDLARS		
1423.MEDLINE		
1424.MEMORIA OPTICA		
1425.MENSAJERIA ELECTRONICA	USE	CORREO ELECTRONICO
1426.MERCADO		
1427.MERCADO DE LA EDICION	USE	INDUSTRIA EDITORIAL
1428.MERCADO DE LA INFORMACION		
1429.MERCADO DE TRABAJO	USE	MERCADO LABORAL
1430.MERCADO EDITORIAL	USE	INDUSTRIA EDITORIAL
1431.MERCADO LABORAL		
1432.MESH		
1433.METODO	USE	METODOLOGIA
1434.METODO CIENTIFICO		
1435.METODOLOGIA		
1436.METODOLOGIA CIENTIFICA	USE	METODO CIENTIFICO
1437.MICRÓFICHAS		
1438.MICROFILMACION		
1439.MICROFILMS		
1440.MICROFORMAS		
1441.MICROGRABACIONES	USE	MICROFORMAS
1442.MICROINFORMATICA		
1443.MICROORDENADORES		
1444.MICROPROCESADORES		
1445.MICROPROGRAMACION		
1446.MICROTESAUROS		
1447.MINIORDENADORES		
1448.MODELO DE DATOS		
1449.MODELO ENTIDAD RELACION		
1450.MODELO RELACIONAL		
1451.MODEM		
1452.MONOGRAFIAS	USE	LIBROS

- 1453.MOVIMIENTOS ASOCIATIVOS USE ASOCIACIONISMO
 1454.MUESTRAS USE MUESTREOS
 1455.MUESTREOS
 1456.MULTIMEDIA
 1457.MUSEISTICA
 1458.MUSEOGRAFIA
 1459.MUSEOLOGIA
 1460.MUSEOLOGOS
 1461.MUSEOS
 1462.MUSEOS ARQUEOLOGICOS
 1463.MUSEOS CIENTIFICOS
 1464.MUSEOS ESTATALES
 1465.MUSEOS LOCALES
 1466.MUSEOS PROVINCIALES
 1467.NECESIDAD DE INFORMACION
 1468.NECESIDAD DE LOS CLIENTES USE NECESIDAD DE INFORMACION
 1469.NECESIDAD DE LOS USUARIOS USE NECESIDAD DE INFORMACION
 1470.NECESIDAD INFORMATIVA USE NECESIDAD DE INFORMACION
 1471.NESTCAPE USE WORLD WIDE WEB
 1472.NEWS USE GRUPOS DE DISCUSION
 1473.NFAIS
 1474.NO DESCRIPTORES
 1475.NO-DESCRIPTORES USE NO DESCRIPTORES
 1476.NOMENCLATURA USE VOCABULARIOS
 1477.NORMALIZACION
 1478.NORMAS
 1479.NORMAS BIBLIOGRAFICAS
 1480.NORMAS DE ACCESO
 1481.NORMAS DE CATALOGACION

 1482.NORMAS INTERNACIONALES DE DESCRIPCION BIBLIOGRAFICA USE
 ISBD

 1483.NORMAS INTERNACIONALES PARA LA DESCRIPCION BIBLIOGRAFICA DE
 MATERIALES NO LIBRARIOS USE ISBN(NBM)

 1484.NORMAS INTERNACIONALES PARA LA DESCRIPCION BIBLIOGRAFICA DE
 PUBLICACIONES SERIADAS USE ISBD(B)

 1485.NORMAS INTERNACIONES PARA LA DESCRIPCION BIBLIOGRAFICA USE
 ISBD

 1486.NORMAS INTERNACIONES PARA LA DESCRIPCION BIBLIOGRAFICA DE
 MONOGRAFIAS USE ISBD(M)

 1487.NORMAS ISO
 1488.NOTAS BIBLIOGRAFICAS USE REFERENCIAS BIBLIOGRAFICAS
 1489.NOTAS DE ALCANCE
 1490.NOTAS DE APLICACION USE NOTAS DE ALCANCE
 1491.NUEVAS TECNOLOGIAS

1492. NUEVAS TECNOLOGIAS DE LA INFORMACION
 1493. NUEVOS SOPORTES
 1494. NUMERO INTERNACIONAL NORMALIZADO PARA LIBROS USE ISBN
 1495. NUMERO INTERNACIONAL NORMALIZADO PARA PUBLICACIONES SERIADAS
 USE ISSN
 1496. OBJETOS DE MUSEOS
 1497. OBRAS DE CONSULTA USE OBRAS DE REFERENCIA
 1498. OBRAS DE REFERENCIA
 1499. OBSTACULOS A LA INFORMACION USE BARRERAS A LA
 INFORMACION
 1500. OCLC
 1501. OCR
 1502. OFERTAS DE EMPLEO
 1503. OFICINAS DE INFORMACION USE SERVICIOS DE INFORMACION
 1504. OFIMATICA
 1505. ON LINE COMPUTER LIBRARY CENTER USE OCLC
 1506. ON LINE PUBLIC ACCESS CATALOGUES USE OPAC
 1507. ONLINE COMPUTER LIBRARY CENTER USE OCLC
 1508. ONLINE PUBLIC ACCESS CATALOGUES USE OPAC
 1509. OPAC USE CATALOGOS DE ACCESO PUBLICO EN LINEA
 1510. OPERADORES BOOLEANOS
 1511. OPERADORES LOGICOS USE OPERADORES BOOLEANOS
 1512. OPTICAL CHARACTER READER USE OCR
 1513. OPTICAL CHARACTER RECOGNITION USE OCR
 1514. ORACLE
 1515. ORDENACION ALFABETICA
 1516. ORDENACION CRONOLOGICA
 1517. ORDENACION GEOGRAFICA
 1518. ORDENACION SECUENCIAL
 1519. ORDENACION SISTEMATICA
 1520. ORDENADORES
 1521. ORDENADORES ANALOGICOS
 1522. ORDENADORES CONECTADOS USE REDES DE COMUNICACION
 1523. ORDENADORES DIGITALES
 1524. ORDENADORES GRAFICOS
 1525. ORDENADORES OPTICOS
 1526. ORDENADORES PARALELOS
 1527. ORDENADORES PERSONALES USE ORDENADORES
 1528. ORDENADORES PORTATILES
 1529. ORGANIZACION BIBLIOTECARIA
 1530. ORGANIZACION DE ARCHIVOS
 1531. ORGANIZACION DE FONDOS
 1532. ORGANIZACION DE MUSEOS
 1533. ORGANIZACIONES INTERNACIONALES
 1534. ORIENTACION A LOS USUARIOS
 1535. ORIENTACION BIBLIOGRAFICA
 1536. PAISES EN DESARROLLO USE PAISES EN VIAS DE DESARROLLO
 1537. PAISES EN VIAS DE DESARROLLO

- 1538.PALABRAS CLAVE
 1539.PALABRAS VACIAS
 1540.PARTITURAS MUSICALES
 1541.PATENTES
 1542.PATRIMONIO
 1543.PATRIMONIO ARQUEOLOGICO
 1544.PATRIMONIO ARQUITECTONICO
 1545.PATRIMONIO ARTISTICO
 1546.PATRIMONIO BIBLIOGRAFICO
 1547.PATRIMONIO CULTURAL
 1548.PATRIMONIO DOCUMENTAL
 1549.PATRIMONIO HISTORICO
 1550.PATRIMONIO HISTORICO-ARTISTICO
 1551.PATRIMONIO MUSICAL
 1552.PATRIMONIO URBANO
 1553.PC USE ORDENADORES
 1554.PELICULAS
 1555.PELICULAS CINEMATOGRAFICAS USE PELICULAS
 1556.PEQUEÑA Y MEDIANA EMPRESA USE PYME
 1557.PERDIDA DE DOCUMENTOS
 1558.PERDIDA DE LIBROS USE PERDIDA DE DOCUMENTOS

 1559.PERFECCIONAMIENTO PROFESIONAL USE FORMACION
 CONTINUADA

 1560.PERFILES DE BUSQUEDA USE ESTRATEGIAS DE BUSQUEDA
 1561.PERFILES DE INTERES USE PERFILES DE LOS USUARIOS
 1562.PERFILES DE LOS USUARIOS
 1563.PERGAMINOS
 1564.PERIODICOS USE DIARIOS
 1565.PERSONAL COMPUTER USE ORDENADORES
 1566.PERTINENCIA USE PRECISION
 1567.PESO USE PONDERACION
 1568.PETICION DE INFORMACION USE DEMANDA DE INFORMACION
 1569.PLANES DE ESTUDIOS
 1570.PLANIFICACION BIBLIOTECARIA
 1571.PLANIFICACION CULTURAL
 1572.PLANIFICACION DE LOS SERVICIOS BIBLIOTECARIOS
 1573.PLANIFICACION PRESUPUESTARIA
 1574.PLN USE PROCESAMIENTO AUTOMATICO DE TEXTOS
 1575.POLITICA
 1576.POLITICA ARCHIVISTICA
 1577.POLITICA BIBLIOTECARIA
 1578.POLITICA CULTURAL
 1579.POLITICA DE ADQUISICIONES
 1580.POLITICA DE ARCHIVO
 1581.POLITICA DE BIBLIOTECAS USE POLITICA BIBLIOTECARIA
 1582.POLITICA DE COMUNICACION
 1583.POLITICA DE INFORMACION
 1584.PONDERACION

- 1585.PONENCIAS
 1586.POSTCOORDINACION
 1587.PRECATALOGACION
 1588.PRECIOS USE TARIFAS
 1589.PRECIS
 1590.PRECISION
 1591.PRECOORDINACION
 1592.PRENSA DIARIA USE DIARIOS
 1593.PRENSA ESCRITA
 1594.PRENSA TIPOGRAFICA
 1595.PRESERVACION DE DOCUMENTOS
 1596.PRESERVACION DE LIBROS USE PRESERVACION DE DOCUMENTOS
 1597.PRESERVACION DE LOS REGISTROS DE ARCHIVO
 1598.PRESERVED CONTEXT INDEX SYSTEM USE INDICES PRECIS
 1599.PRESTAMO BIBLIOTECARIO
 1600.PRESTAMO DE BIBLIOTECA
 1601.PRESTAMO DE DOCUMENTOS
 1602.PRESTAMO DE LIBROS
 1603.PRESTAMO EN SALA
 1604.PRESTAMO ENTRE BIBLIOTECAS USE PRESTAMO
 INTERBIBLIOTECARIO
 1605.PRESTAMO INTERBIBLIOTECARIO
 1606.PRESTAMO INTERBIBLIOTECARIO INTERNACIONAL
 1607.PRINCIPIO DE ECONOMIA
 1608.PRINCIPIO DE ESPECIFICIDAD
 1609.PRINCIPIO DE USO
 1610.PROCEEDINGS USE ACTAS DE CONGRESOS
 1611.PROCESADORES DE TEXTOS
 1612.PROCESAMIENTO DE DATOS USE PROCESAMIENTO DE LA
 INFORMACION
 1613.PROCESAMIENTO DE INFORMACION USE PROCESAMIENTO DE
 LA INFORMACION
 1614.PROCESAMIENTO DE LA INFORMACION
 1615.PROCESAMIENTO DEL LENGUAJE NATURAL USE ANALISIS
 AUTOMATICO DE TEXTOS
 1616.PROCESO DOCUMENTAL
 1617.PRODUCCION BIBLIOGRAFICA
 1618.PRODUCCION CIENTIFICA USE PRODUCCION BIBLIOGRAFICA
 1619.PRODUCCION DE BASES DE DATOS
 1620.PRODUCCION DOCUMENTAL
 1621.PRODUCCION HISTORIOGRAFICA USE PRODUCCION BIBLIOGRAFICA
 1622.PRODUCTIVIDAD CIENTIFICA USE PRODUCCION BIBLIOGRAFICA
 1623.PRODUCTOR OS
 1624.PRODUCTOS

- 1625.PROFESIONALES DE LA DOCUMENTACION USE PROFESIONALES DE LA INFORMACION
- 1626.PROFESIONALES DE LA INFORMACION
- 1627.PROFESORES USE DOCENTES
- 1628.PROFUNDIDAD EN LA INDIZACION
- 1629.PROGRAMACION
- 1630.PROGRAMACION BIBLIOTECARIA USE PLANIFICACION BIBLIOTECARIA
- 1631.PROGRAMAS DE ASIGNATURAS
- 1632.PROGRAMAS DE COMPUTO USE SOFTWARE
- 1633.PROGRAMAS DE ESTUDIOS USE PLANES DE ESTUDIOS
- 1634.PROGRAMAS DE GESTION BIBLIOTECARIA
- 1635.PROGRAMAS DE INVESTIGACION USE PROYECTOS DE INVESTIGACION
- 1636.PROGRAMAS DE ORDENADOR USE SOFTWARE
- 1637.PROGRAMOTECA USE BIBLIOTECAS DE PROGRAMAS
- 1638.PROMOCION DE LA CULTURA
- 1639.PROMOCION DE LA LECTURA USE ANIMACION A LA LECTURA
- 1640.PROMOCION DE LAS BIBLIOTECAS
- 1641.PROPIEDAD INTELECTUAL USE DERECHOS DE AUTOR
- 1642.PROPIEDAD LITERARIA USE DERECHOS DE AUTOR
- 1643.PROTECCION CONTRA ROBOS USE SEGURIDAD EN BIBLIOTECAS
- 1644.PROTECCION DE DATOS
- 1645.PROTECCION DE DOCUMENTOS
- 1646.PROTOCOLO DE CONTROL DE TRANSMISION USE TCP
- 1647.PROTOCOLO DE INTERNET
- 1648.PROTOCOLO PARA LA TRANSFERENCIA DE FICHEROS USE FTP
- 1649.PROTOCOLOS DE COMUNICACION
- 1650.PROVEEDORES DE INTERNET
- 1651.PROYECTOS DE INVESTIGACION
- 1652.PROYECTOS FINAL DE CARRERA
- 1653.PUBLICACIONES
- 1654.PUBLICACIONES CIENTIFICAS
- 1655.PUBLICACIONES ELECTRONICAS
- 1656.PUBLICACIONES EN SERIE
- 1657.PUBLICACIONES ESPAÑOLAS
- 1658.PUBLICACIONES EXTRANJERAS
- 1659.PUBLICACIONES OFICIALES
- 1660.PUBLICACIONES PERIODICAS
- 1661.PUBLICACIONES RETROSPECTIVAS
- 1662.PUBLICACIONES SERIADAS
- 1663.PUNTOS DE INFORMACION
- 1664.PYME
- 1665.RDSI USE RED DIGITAL DE SERVICIOS INTEGRADOS
- 1666.REALIDAD VIRTUAL
- 1667.RECALL USE EXHAUSTIVIDAD EN LA RECUPERACION
- 1668.RECOGIDA DE DATOS

- 1669.RECONOCIMIENTO DE LA PALABRA HABLA USE RECONOCIMIENTO DEL HABLA
- 1670.RECONOCIMIENTO DEL HABLA
- 1671.RECONOCIMIENTO OPTICO DE CARACTERES
- 1672.RECORTES DE PRENSA
- 1673.RECUPERACION AUTOMATICA DE LA INFORMACION
- 1674.RECUPERACION AUTOMATIZADA
- 1675.RECUPERACION BIBLIOGRAFICA INFORMACION USE RECUPERACION DE LA INFORMACION
- 1676.RECUPERACION DE DATOS INFORMACION USE RECUPERACION DE LA INFORMACION
- 1677.RECUPERACION DE DOCUMENTOS INFORMACION USE RECUPERACION DE LA INFORMACION
- 1678.RECUPERACION DE INFORMACION EN LINEA INFORMACION EN LINEA USE RECUPERACION DE LA INFORMACION EN LINEA
- 1679.RECUPERACION DE LA INFORMACION
- 1680.RECUPERACION DE LA INFORMACION EN LINEA
- 1681.RECUPERACION DE LA INFORMACION ON LINE INFORMACION EN LINEA USE RECUPERACION DE LA INFORMACION EN LINEA
- 1682.RECUPERACION DE TEXTOS
- 1683.RECURSOS BIBLIOGRAFICOS
- 1684.RECURSOS COMPARTIDOS EN BIBLIOTECAS BIBLIOTECARIA USE COOPERACION
- 1685.RECURSOS COMPARTIDOS ENTRE BIBLIOTECAS BIBLIOTECARIA USE COOPERACION
- 1686.RECURSOS DE INFORMACION
- 1687.RECURSOS DOCUMENTALES USE FONDOS DOCUMENTALES
- 1688.RECURSOS HUMANOS
- 1689.RECURSOS INFORMATIVOS USE RECURSOS DE INFORMACION
- 1690.REDE DE AREA LOCAL
- 1691.REDE DE REDES USE REDES DE COMUNICACION
- 1692.REDE DIGITAL DE SERVICIOS INTEGRADOS
- 1693.REDE EUROPEA DE INFORMACION Y DOCUMENTACION SOBRE AMERICA LATINA USE REDIAL
- 1694.REDE LOCAL DE CD ROM USE REDES DE CD ROM
- 1695.REDE LOCAL DE CD-ROM USE REDES DE CD ROM
- 1696.REDE NACIONAL DE BIBLIOTECAS

- 1697.REDC
 1698.REDES COOPERATIVAS DE BIBLIOTECAS
 1699.REDES DE ARCHIVOS
 1700.REDES DE AREA EXTENSA
 1701.REDES DE BIBLIOTECAS
 1702.REDES DE CATALOGOS
 1703.REDES DE CD ROM
 1704.REDES DE CD-ROM USE REDES DE CD ROM
 1705.REDES DE CENTROS DE INFORMACION
 1706.REDES DE COMUNICACIONES
 1707.REDES DE INFORMACION
 1708.REDES DE ORDENADORES USE REDES DE COMUNICACION
 1709.REDES DE TRANSMISION DE DATOS
 1710.REDES DIGITALES
 1711.REDES INFORMATIZADAS

 1712.REDES TELEINFORMATICAS USE REDES DE TRANSMISION DE DATOS

 1713.REDES Y SISTEMAS DE BIBLIOTECAS USE REDES DE BIBLIOTECAS
 1714.REDC USE PUBLICACIONES PERIODICAS
 1715.REAL
 1716.REFERENCIAS BIBLIOGRAFICAS
 1717.REFERENCISTAS USE BIBLIOTECARIOS DE REFERENCIA
 1718.REGISTROS BIBLIOGRAFICOS
 1719.REGISTROS CAPTADOS USE CAPTURA DE REGISTROS
 1720.REGISTROS DE ARCHIVO
 1721.REGISTROS IMPORTADOS USE CAPTURA DE REGISTROS
 1722.REGISTROS PERMANENTES USE REGISTROS DE ARCHIVO
 1723.REGISTROS PUBLICOS USE ARCHIVOS PUBLICOS
 1724.REGLAS DE CATALOGACION

 1725.REGLAS DE CATALOGACION ANGLO-AMERICANAS USE REGLAS DE CATALOGACION

 1726.RELACIONES LOGICAS USE OPERADORES BOOLEANOS

 1727.REPARACION DE DOCUMENTOS USE RESTAURACION DE DOCUMENTOS

 1728.REPARACION DE LIBROS USE RESTAURACION DE DOCUMENTOS

 1729.REPERTORIOS BIBLIOGRAFICOS
 1730.REPERTORIOS BIBLIOGRAFICOS INTERNACIONALES
 1731.REPRODUCCION DE LA INFORMACION
 1732.REPRODUCCION FACSIMIL
 1733.REPROGRAFIA
 1734.RESEÑAS BIBLIOGRAFICAS USE REFERENCIAS BIBLIOGRAFICAS
 1735.RESEÑAS DE LIBROS USE REFERENCIAS BIBLIOGRAFICAS

- 1736.RESTAURACION DE DOCUMENTOS
 1737.RESTAURACION DE LIBROS USE RESTAURACION DE DOCUMENTOS
 1738.RESUMENES
 1739.RESUMENES CRITICOS USE RESUMENES
 1740.RESUMENES DE AUTOR USE RESUMENES
 1741.RESUMENES INDICATIVOS USE RESUMENES
 1742.RESUMENES INFORMATIVOS USE RESUMENES
 1743.RESUMENES ORIENTATIVOS USE RESUMENES
 1744.RETROALIMENTACION
 1745.REVISION BIBLIOGRAFICA
 1746.REVISION CRITICA
- 1747.REVISTA ESPAÑOLA DE DOCUMENTACION CIENTIFICA USE
 PUBLICACIONES PERIODICAS
- 1748.REVISTA GENERAL DE INFORMACION Y DOCUMENTACION
 1749.REVISTAS USE PUBLICACIONES PERIODICAS
 1750.REVISTAS BIOMEDICAS
 1751.REVISTAS CIENTIFICA Y TECNICA USE PUBLICACIONES PERIODICAS
 1752.REVISTAS CIENTIFICAS
 1753.REVISTAS CULTURALES
 1754.REVISTAS DE EXTRACTOS USE REVISTAS DE RESUMENES
 1755.REVISTAS DE RESUMENES
 1756.REVISTAS DE SUMARIOS USE REVISTAS DE RESUMENES
 1757.REVISTAS ELECTRONICAS
 1758.REVISTAS LATINOAMERICANAS
 1759.REVISTAS LITERARIAS
 1760.REVISTAS QUIMICAS
 1761.REVISTAS TECNICAS
 1762.ROBOS DE DOCUMENTOS
 1763.ROBOS DE LIBROS USE ROBOS DE DOCUMENTOS
 1764.RUIDO DOCUMENTAL
 1765.SALAS DE CONSULTA
 1766.SALAS DE LECTURA
 1767.SALAS GENERALES
 1768.SALAS INFANTILES
 1769.SALAS JUVENILES
 1770.SECTOR PRIVADO
 1771.SECTOR PUBLICO
 1772.SEDIC USE ASOCIACIONES PROFESIONALES
 1773.SEGURIDAD EN LAS BIBLIOTECAS
 1774.SEGURIDAD EN LOS ARCHIVOS
 1775.SELECCION
 1776.SELECCION DE DOCUMENTOS
 1777.SELECCION DE FONDOS USE SELECCION DE DOCUMENTOS
 1778.SELECCION DE LIBROS USE SELECCION DE DOCUMENTOS
 1779.SELECCION Y ADQUISICION
 1780.SELLOS
 1781.SEÑALIZACION EN BIBLIOTECAS
 1782.SEPLN USE ASOCIACIONES PROFESIONALES

- 1783.SERIE USE COLECCIONES
 1784.SERIES DOCUMENTALES
 1785.SERIES TEMPORALES
 1786.SERVICIOS ARCHIVISTICOS
 1787.SERVICIOS BIBLIOGRAFICOS
 1788.SERVICIOS BIBLIOTECARIOS
 1789.SERVICIOS CULTURALES
- 1790.SERVICIOS DE ALERTA BIBLIOGRAFICA USE SERVICIOS DE ALERTA
 INFORMATIVA
- 1791.SERVICIOS DE ALERTA INFORMATIVA
 1792.SERVICIOS DE BIBLIOGRAFIA USE SERVICIOS BIBLIOGRAFICOS
 1793.SERVICIOS DE DIFUSION DE DOCUMENTOS
 1794.SERVICIOS DE DOCUMENTACION USE CENTROS DE
 DOCUMENTACION
 1795.SERVICIOS DE EXTENSION BIBLIOTECARIA
 1796.SERVICIOS DE FOTODOCUMENTACION
- 1797.SERVICIOS DE INDIZACION USE SERVICIOS DE INDIZACION Y
 RESUMENES
- 1798.SERVICIOS DE INDIZACION Y RESUMENES
 1799.SERVICIOS DE INFORMACION
 1800.SERVICIOS DE INFORMACION BIBLIOGRAFICA
 1801.SERVICIOS DE INFORMACION ELECTRONICA
 1802.SERVICIOS DE INTERCAMBIO BIBLIOGRAFICO
 1803.SERVICIOS DE LECTURA
- 1804.SERVICIOS DE ORIENTACION AL LECTOR USE SERVICIOS DE
 REFERENCIA
- 1805.SERVICIOS DE PRESTAMO USE PRESTAMO BIBLIOTECARIO
 1806.SERVICIOS DE REFERENCIA
 1807.SERVICIOS DE REPROGRAFIA USE SERVICIOS DE
 FOTODOCUMENTACION
- 1808.SERVICIOS DE RESUMENES USE SERVICIOS DE INDIZACION Y
 RESUMENES
- 1809.SERVICIOS INFORMATIVOS USE SERVICIOS DE INFORMACION
 1810.SERVICIOS MUSEISTICOS
 1811.SERVIDOR DE INFORMACION DE GRAN ALCANCE USE WAIS
 1812.SGBS USE SISTEMAS DE GESTION DE BASES DE DATOS
 1813.SIDA
 1814.SIGILOGRAFIA
 1815.SIGLE
 1816.SIGNATURA TOPOGRAFICA
 1817.SIMPOSIO
 1818.SINDROME DE INMUNODEFICIENCIA ADQUIRIDA USE SIDA

- 1819.SISTEMA BIBLIOTECARIO
1820.SISTEMA BRAILLE
- 1821.SISTEMA DE INDIZACION CON CONSERVACION DEL CONTEXTO USE
 PRECIS
- 1822.SISTEMA DE INDIZACION PRECIS USE INDICES PRECIS
1823.SISTEMA DE INFORMACION DOCUMENTAL
1824.SISTEMA DE INFORMACION INTEGRADO
- 1825.SISTEMA DE INFORMACION SOBRE LITERATURA GRIS EN EUROPA
 USE SIGLE
- 1826.SISTEMA DE RECUPERACION Y ANALISIS DE LITERATURA MEDICA
 USE MEDLARS
- 1827.SISTEMA HOMBRE-MAQUINA
1828.SISTEMA NACIONAL DE INFORMACION
1829.SISTEMA NACIONAL DE INFORMACION EDUCATIVA
1830.SISTEMAS ARCHIVISTICOS
1831.SISTEMAS BIBLIOTECARIOS
1832.SISTEMAS DE CLASIFICACION
1833.SISTEMAS DE COINCIDENCIA OPTICA
1834.SISTEMAS DE GESTION
1835.SISTEMAS DE GESTION DE BASES DE DATOS
1836.SISTEMAS DE GESTION DOCUMENTAL
1837.SISTEMAS DE GESTION INTEGRADA DE LA DOCUMENTACION
1838.SISTEMAS DE INDIZACION
- 1839.SISTEMAS DE INDIZACION POR UNITERMINOS USE UNITERMINOS
- 1840.SISTEMAS DE INFORMACION
1841.SISTEMAS DE INFORMACION TERRITORIAL
- 1842.SISTEMAS DE MUSEOS
- 1843.SISTEMAS DE RECUPERACION USE SISTEMAS DE RECUPERACION DE
 INFORMACION
- 1844.SISTEMAS DE RECUPERACION DE INFORMACION
1845.SISTEMAS DE SEGURIDAD USE SEGURIDAD EN LAS BIBLIOTECAS
1846.SISTEMAS EXPERTOS
1847.SISTEMAS INFORMATICOS
1848.SISTEMAS INTEGRADOS
1849.SISTEMAS INTEGRADOS DE DOCUMENTACION
1850.SISTEMAS OPERATIVOS
1851.SMART
1852.SOCADI USE ASOCIACIONES PROFESIONALES

- 1853.SOCIEDAD CATALANA DE DOCUMENTACION E INFORMACION
USE ASOCIACIONES PROFESIONALES
- 1854.SOCIEDAD DE LA INFORMACION
- 1855.SOCIEDAD ESPAÑOLA DE DOCUMENTACION CIENTIFICA USE
ASOCIACIONES PROFESIONALES
- 1856.SOCIEDAD ESPAÑOLA PARA EL PROCESAMIENTO DEL LENGUAJE NATURAL
USE ASOCIACIONES PROFESIONALES
- 1857.SOCIEDAD INTERNACIONAL PARA LA ORGANIZACION DEL CONOCIMIENTO
USE ASOCIACIONES PROFESIONALES
- 1858.SOCIEDAD POSTINDUSTRIAL
- 1859.SOCIETAD CATALANA DE DOCUMENTACIO I INFORMACIO USE
ASOCIACIONES PROFESIONALES
- 1860.SOCIOLOGIA
- 1861.SOCIOLOGIA DE LA INFORMACION
- 1862.SOCIOLOGIA DE LA LECTURA
- 1863.SOCIOMETRIA
- 1864.SOFTWARE
- 1865.SOLICITUD DE INFORMACION USE DEMANDA DE INFORMACION
- 1866.SOLICITUD DE FOTOCOPIAS
- 1867.SOPORTE FISICO USE HARDWARE
- 1868.SOPORTE LOGICO USE SOFTWARE
- 1869.SOPORTES DE INFORMACION
- 1870.SOPORTES OPTICOS
- 1871.SPIRIT
- 1872.SQL
- 1873.STOPLIST USE LISTAS DE PALABRAS VACIAS
- 1874.SUBJECT INDEXING USE INDIZACION POR MATERIAS
- 1875.SUCURSALES DE BIBLIOTECAS USE BIBLIOTECAS SUCURSALES
- 1876.SUMINISTRO DE DOCUMENTOS
- 1877.SUPERORDENADORES
- 1878.SUSCRIPCIONES PARA PUBLICACIONES PERIODICAS USE
ADQUISICIONES
- 1879.SYMPOSIUM USE SIMPOSIO
- 1880.TARIFAS
- 1881.TARJETAS PERFORADAS
- 1882.TARJETAS POSTALES

1883.TASAS DE EXHAUSTIVIDAD EN LA RECUPERACION	USE	INDICES DE EXHAUSTIVIDAD
1884.TASAS DE LLAMADA RECUPERACION	USE	INDICES DE EXHAUSTIVIDAD EN LA
1885.TASAS DE PERTINENCIA	USE	INDICES DE PRECISION
1886.TASAS DE PRECISION	USE	INDICES DE PRECISION
1887.TASAS DE RUIDO	USE	INDICES DE RUIDO
1888.TASAS DE SILENCIO	USE	INDICES DE SILENCIO
1889.TASACION DE LIBROS		
1890.TCP/IP		
1891.TECNICAS DE INVESTIGACION	USE	METODO DE INVESTIGACION
1892.TECNOLOGIAS DE LA INFORMACION		
1893.TECNOLOGIAS DE LAS COMUNICACIONES		
1894.TECNOLOGIAS MULTIMEDIA		
1895.TECNOLOGIAS OPTICAS		
1896.TELARAÑA MUNDIAL	USE	WORLD WIDE WEB
1897.TELECOMUNICACIONES		
1898.TELECONFERENCIAS		
1899.TELEDESCARGA		
1900.TELEDOCUMENTACION		
1901.TELEMATICA		
1902.TELETEXTO		
1903.TELEVISION		
1904.TEOLOGIA		
1905.TEORIA DE LA COMUNICACION		
1906.TEORIA DE LA DOCUMENTACION		
1907.TEORIA DE LA INFORMACION		
1908.TEORIA DE SISTEMAS		
1909.TEORIAS ARCHIVISTICAS		
1910.TERCER CICLO		
1911.TERMINOLOGIA		
1912.TERMINOS COMBINADOS	USE	TERMINOS COMPUESTOS
1913.TERMINOS DE BUSQUEDA		
1914.TERMINOS DE INDIZACION		
1915.TERMINOS NO PREFERENTES	USE	NO DESCRIPTORES
1916.TERMINOS PRECOORDINADOS		
1917.TESAUROS		
1918.TESINAS	USE	TESINAS DE LICENCIATURA
1919.TESINAS DE LICENCIATURA		
1920.TESIS DE LICENCIATURA	USE	TESINAS DE LICENCIATURA
1921.TESIS DOCTORAL		
1922.THESAURI	USE	TESUROS
1923.THESAURO	USE	TESAUROS
1924.TIPOBIBLIOGRAFIAS		
1925.TIPOLOGIA DOCUMENTAL		
1926.TITULOS		

1927.TOMA DE DECISIONES
 1928.TRABAJOS DE GRADO
 1929.TRABAJOS FINAL DE CARRERA
 1930.TRADUCCION AUTOMATICA
 1931.TRADUCCION MECANICA USE TRADUCCION AUTOMATICA
 1932.TRADUCCION MECANIZADA USE TRADUCCION AUTOMATICA
 1933.TRANSCRIPCION DE DATOS
 1934.TRANSFERENCIA DE LA INFORMACION
 1935.TRANSFERENCIA ELECTRONICA DE DATOS
 1936.TRANSLITERACION
 1937.TRANSMISION DE DATOS

 1938.TRANSMISSION CONTROL PROTOCOL / INTERNET PROTOCOL USE
 TCP/IP

 1939.TRANSMISION DE INFORMACION ELECTRONICA USE
 TELECOMUNICACIONES

 1940.TRANSMISION ENTRE ORDENADORES USE REDES DE
 TRANSMISION DE DATOS

 1941.TRASLADO DE DOCUMENTOS
 1942.TRATAMIENTO AUTOMATICO DE DATOS

 1943.TRATAMIENTO AUTOMATICO DE TEXTOS USE ANALISIS
 AUTOMATICO DE TEXTOS

 1944.TRATAMIENTO DE DATOS
 1945.TRATAMIENTO DE FONDOS
 1946.TRATAMIENTO DE IMAGENES
 1947.TRATAMIENTO DE LA INFORMACION
 1948.TRATAMIENTO DOCUMENTAL
 1949.U.E USE UNION EUROPEA
 1950.UAP USE DISPONIBILIDAD UNIVERSAL DE LAS PUBLICACIONES
 1951.UBC USE CONTROL BIBLIOGRAFICO UNIVERSAL
 1952.UDC USE CLASIFICACION DECIMAL UNIVERSAL
 1953.UE USE UNION EUROPEA
 1954.UNED
 1955.UNIDAD CENTRAL DE PROCESAMIENTO
 1956.UNIDADES DE INFORMACION
 1957.UNION EUROPEA
 1958.UNITERMINOS
 1959.UNIVERSIDAD
 1960.UNIVERSIDAD NACIONAL DE EDUCACION A DISTANCIA USE UNED
 1961.USO DE LAS BIBLIOTECAS
 1962.USO DE LAS COLECCIONES
 1963.USO DE LOS FONDOS
 1964.USO DE ORDENADORES
 1965.USUARIOS

1966.USUARIOS DE LA DOCUMENTACION INFORMACION	USE	USUARIOS DE LA INFORMACION
1967.UTILIZACION DE LA DOCUMENTACION INFORMACION	USE	UTILIZACION DE LA INFORMACION
1968.VALOR DOCUMENTAL		
1969.VERONICA		
1970.VIDEOCASSETES		
1971.VIDEOCASSETTES	USE	VIDEOCASSETTES
1972.VIDEOCINTAS		
1973.VIDEOCONFERENCIAS		
1974.VIDEODISCOS		
1975.VIDEOLIBROS		
1976.VIDEOTECAS		
1977.VIDEOTEX		
1978.VISITAS GUIADAS		
1979.VOCABULARIOS		
1980.VOCABULARIOS CIENTIFICOS		
1981.VOCABULARIOS CONTROLADOS		
1982.VOCABULARIOS LIBRES	USE	LENGUAJE NATURAL
1983.VOCABULARIOS NO CONTROLADOS		
1984.VOCABULARIOS TECNICOS		
1985.WAIS		
1986.WAN	USE	RED DE AREA EXTENSA
1987.WEB	USE	WORLD WIDE WEB
1988.WIDE AREA NETWORK	USE	RED DE AREA EXTENSA
1989.WINDOWS		
1990.WORLD WIDE WEB		
1991.WORM		
1992.WWW	USE	WORLD WIDE WEB
1993.XEROCOPIA		
1994.XEROGRAFIA		
1995.XILOGRAFIA		