

Grupo Temma

BIBLIOTECA  
DO  
IBICT



# Análise documentária:

a análise da síntese

4.05

BRASILERO  
MACAO  
IA E TECNOLOGIA

# 1. LINGUAGENS DE INDEXAÇÃO

*Eunides A. do Vale*

A escolha de uma linguagem de indexação é fator essencial para a eficácia de um sistema de recuperação da informação. Deve-se levar em consideração os objetivos do sistema, o tipo de usuário e a abrangência ou especificidade do assunto a ser tratado.

As linguagens se dividem em dois grupos:

- Pré-coordenadas, que combinam ou coordenam os termos no momento da indexação. São utilizadas principalmente em sistemas manuais: catálogos de bibliotecas, bibliografias e índices impressos etc.
- Pós-coordenadas, que combinam ou coordenam os termos no momento da busca. São utilizadas principalmente em sistemas automatizados: bases/bancos de dados, bibliotecas ou centros de documentação que operam com computador.

## 1.1. LINGUAGENS DE INDEXAÇÃO PRÉ-COORDENADAS

### 1.1.1. Sistemas tradicionais de classificação

São as mais antigas linguagens documentárias e abrangem todas as áreas do conhecimento. Atualmente, têm como objetivo principal a arrumação dos livros nas estantes por assunto. Representam a indexação através de uma notação, basicamente composta por números, letras ou a mistura de ambos.

Ex.: 324.63 = Sistemas eleitorais

Os sistemas mais utilizados no Brasil são:

- Classificação Decimal de Dewey (CDD);
- Classificação Decimal Universal (CDU);
- Classificação da Library of Congress (LC).

Esses sistemas são estruturados hierarquicamente, dividindo o conhecimento em dez classes principais, subdivididas sucessivamente.

É através dessas subdivisões que muitas vezes se forja a inserção de novos assuntos, os quais em outras estruturas teriam um tipo de relação não-hierárquica.

No exemplo abaixo, retirado da CDD, percebe-se a relação hierárquica existente entre a Sociologia e a Antropologia Cultural e Social.

Ex.: CDD:	18ª ed.	301	Sociology
		301.2	Culture and cultural processes
			Class here primitive races (formerly 57217), cultural (formerly 390) and social anthropology, comprehensive works on anthropology.
	19ª ed.	301/307	Sociology
		306	Culture and institutions
			Class here cultural and social anthropology.

Outro aspecto a ser considerado refere-se à diferença de conceituação de determinados assuntos, como pode ser observado no exemplo a seguir:

Ex.: CDD:	19ª ed.	572	Human races
			Class here Ethnology
	19ª ed.	306	Culture and institutions

Class here cultural and social anthropology.

Tesauro de Sociologia (VIET, 1971):

Social and cultural Anthropology/Ethnology

UF Ethnology

Enquanto a CDD atribui conceitos diferentes e os classifica como assuntos diversos, em classes completamente diferentes, o Tesauro de Sociologia (VIET, 1971) considera-os como termos equivalentes em língua inglesa e francesa. O *Dicionário Aurélio* define Etnologia como: 1. Ramo de Antropologia que estuda a cultura dos chamados povos naturais. 2. Estudo e conhecimento, sob o aspecto cultural, das populações primitivas. 3. Estudo comparativo de todos os povos. (Sinônimo geral: Antropologia cultural).

Referindo-se especialmente à CDU, lembramos que ela apresenta dispositivos de síntese, que dão maior flexibilidade ao sistema, embora, em muitos casos, isso leve à formação de notações muito grandes, dificultando sua leitura e compreensão, principalmente por parte dos usuários. A CDU é também utilizada para a elaboração do catálogo sistemático.

#### 1.1.1.2 Library of Congress (LC)

Foi elaborada para ser aplicada especialmente ao acervo da Library of Congress dos EUA. Também está estruturada hierarquicamente e suas classes são criadas de acordo com o perfil e desenvolvimento do acervo. É suplementada por lista de cabeçalho de assunto.

#### 1.1.1.3 Hierarquização

A interdisciplinaridade das ciências mostra que a hierarquização do conhecimento não é suficiente para evidenciar os vários tipos

de relações possíveis entre assuntos, num sistema de recuperação de informação; daí a necessidade de maior reflexão sobre as linguagens documentárias que utilizam esse tipo de estruturação. O exemplo abaixo mostra o assunto **Eleição** estruturado por um sistema de classificação tradicional e por um tesouro.

Ex.: CDD	- 19ª ed.	Tesouro de Sociologia (VIET,
32	Political Science	1971)
324	The Political process	
324.6	Elections	Elections/Elections
324.63	Electoral systems	RT Electoral system/Système électoral
	Relação hierárquica	Relação Associativa

### 1.1.2 Listas de Cabeçalhos de Assunto

Representam os assuntos sob a forma de cabeçalhos já estruturados. Geralmente são listas gerais que arrolam termos de todas as áreas do conhecimento. No Brasil, as mais conhecidas e usadas são:

Library of Congress of Subject Headings (LCSH) e Sears List of Subject Headings (Sears), sendo a primeira de maior porte que a segunda. Essas listas apresentam-se alfabeticamente, utilizando alguns sinais como o traço, a vírgula e o parêntese para estruturar os cabeçalhos indiretos.

No exemplo a seguir, retirado da LCSH e citado por Foskett (1973), observa-se o uso do singular e plural e o uso do cabeçalho direto e indireto, o que interfere na recuperação de assuntos similares.

Ex.:	Libraries, naval
porém	Libraries, military see military libraries
	Library administration (não Libraries-Administration)
porém	Acquisitions (Libraries)
	Cataloging of moving-pictures
porém	Classification - Moving-pictures

Contudo, o singular e o plural tornam-se importantes quando diferenciam o sentido do cabeçalho.

Ex.:           Alemão - língua  
              Alemães - povo

As listas tradicionais também apresentam sinais de relação em função dos cabeçalhos, como por exemplo:

X (see) cabeçalhos sinônimos

XX cabeçalhos mais amplos ou genéricos e cabeçalhos relacionados

see also - cabeçalhos específicos e relacionados

em negrito - cabeçalhos eleitos para entrada

Com a utilização apenas desses sinais, as listas não fazem uma distinção clara entre os vários tipos de relações, como pode-se observar no exemplo a seguir, em que o mesmo assunto "Plantas" aparece estruturado por uma lista de cabeçalho de assunto e um tesouro. Na lista de cabeçalho de assunto os termos específicos e relacionados aparecem num mesmo grupo, precedido pelo termo see also (relação de associação "livre"), enquanto que no tesouro cada termo aparece especificado na sua relação com os demais (relação de associação estruturada).

Ex.:   (Sears, 11ª ed.)	Transformação para uma estrutura de tesouro
Plants	Plants
see also	TE alpine plants
Fertilization of Plants	TR Fertilization of plants
Alpine Plants, Flowers,	TR Flowers
Gardening	
X Flora	TR Gardening
XX Botany	UP Flora
	TG Botany

### 1.1.2.1 Remissivas (LCSH)

Foskett (1973) cita um estudo de J. Daily sobre a LCSH, em que foi detectado que, de 36.468 cabeçalhos principais, cerca de 7.033 não possuem remissivas, o que demonstra que só poderiam ser localizados pelas entradas previstas na lista, exigindo, portanto, um profundo conhecimento do instrumento pelo indexador.

### 1.1.3 Classificações Facetadas

Ranganathan é considerado o criador das classificações facetadas. Denominou sua classificação de Colon Classification, a qual ficou conhecida como Classificação dos Dois Pontos.

Revolucionou a estrutura dos sistemas tradicionais de classificação, introduzindo o princípio da divisão de assuntos em categorias ou facetas.

Suas categorias ou facetas principais são conhecidas pela sigla PMEST (FOSKETT, 1973):

ST - Espaço/Tempo

E - Energia (como força compulsora), ex.: Exportações em economia, currículo em educação.

M - Matéria, ex.: Ouro como material de dinheiro dentro da Economia. Instrumento em música. Marfim na Escultura.

P - Personalidade - corresponde às coisas, tipos de coisas, tipos de ação. Ex.: Pessoas em Sociologia/Psicologia.

A classificação de Ranganathan em si não tem aplicação no mundo ocidental. É uma classificação geral, com 42 classes principais, sendo aplicada, a cada uma dessas classes, a análise em facetas.

#### 1.1.3.1 Classificação facetada pós-Ranganathan

Desde 1952, na Inglaterra, o Classification Research Group (CRG)

vem se dedicando aos estudos das classificações facetadas (Foskett, Shera, Vickery, etc.). No Brasil, os primeiros estudos datam de 1970 - curso de pós-graduação do IBICT (BARBOSA, 1972).

A classificação facetada define-se como a aplicação com rigor da análise em facetas à sua estruturação. Através da síntese de mais de uma faceta representa-se um assunto composto.

Ex.: Vegetação própria de terra roxa (VICKERY, 1980):

9ir/7n

tipo de solo 9

terra roxa ir

relação entre as facetas

constituintes 7

vegetação n

O quadro apresentado mostra parcialmente as facetas e suas subdivisões aplicadas à Ciência do Solo, que deram origem ao exemplo (VICKERY, 1980):

#### Facetas

9 tipos de solo

8 estrutura

7 constituintes

6 propriedades

5 processos no solo

4 operações

3 técnicas de laboratório

1 geral

#### Desdobramento das facetas

9 tipos de solo

b orto-elúvio

:

:

i subtropical e tropical

ir terra roxa

:

:

8 estrutura

b simples-grão

:

:

7 constituintes

b químicos

c inorgânicos

:

:

n vegetação

:

6 propriedades

As classificações facetadas normalmente se aplicam a uma área do conhecimento, tendo, portanto, um caráter de especialização. Essa especialização ou assunto núcleo é considerada o Universo ou campo que se quer classificar, por ex.: Sociologia, Petróleo, Transporte etc.

Faceta é a diferença ou característica de uma parte do conhecimento analisado (CURSO..., 1978).

Subfaceta é a divisão de uma faceta pelas diferenças apresentadas (CURSO..., 1978).

Além de Ranganathan, outros estudiosos como, por exemplo, Vickery, estabeleceram novas categorias para divisão do Universo a ser classificado. Acredita-se que, ao elaborar uma classificação facetada, primeiro deve-se verificar se as categorias já existentes podem ser aplicadas ou se é necessária uma adaptação ou até mesmo a elaboração de novas categorias.

Segue abaixo um exemplo de uma categorização "livre" aplicada ao campo: Alimentação de grupos indígenas, baseado no texto de Barbosa (1972).

Ex.:	Facetas	Subfacetas
	obtenção	caça pesca coleta roça
	preparo	cru cozido assado fermentado
	consistência	sólido líquido

## Facetas

## Subfacetas

ocasião de consumo

cotidiano

festa

funeral

propriedade nutriente

vitaminas

A

B

C

proteínas

sais minerais

As classificações facetadas serviram de base para elaboração de tesouros.

### 1.1.4 Ordem de citação em sistemas pré-coordenados

Normalmente, cada sistema sugere uma ordem de citação. Essas ordens não devem ser encaradas como inflexíveis, uma vez que podem ser modificadas de acordo com as necessidades e objetivos do sistema de informação.

Ex.: PMEST - ordem de citação de Ranganathan;  
lugar, tempo, forma, língua - ordem de citação da CDU.

## 1.2 LINGUAGENS DE INDEXAÇÃO PÓS-COORDENADAS

### 1.2.1 Unitermo

Criado e empregado por Mortimer Taube em 1953. Pode ser definido como a indexação por palavras únicas retiradas do contexto, na maioria das vezes, sem controle de vocabulário e destinado aos sistemas pós-coordenados.

Utiliza-se das chamadas Fichas Unitermo, como suporte para sua aplicação. Essas fichas são divididas em dez colunas numeradas de 0 a 9, deixando no alto um espaço para a palavra ou termo de

entrada.

Os documentos são numerados seqüencialmente e o número individual de cada um é anotado em todas as Fichas Unitermo que correspondem aos conceitos escolhidos pelo indexador como entrada. Por exemplo, o nº 298 de um documento sobre ensino de Biblioteconomia, é anotado nas fichas: Ensino coluna 8 e Biblioteconomia também na coluna 8. Nota-se que há uma correspondência entre o dígito final do número do documento com o número da coluna da Ficha Unitermo.

No momento da busca, comparam-se as fichas referentes aos conceitos que se está procurando; por exemplo: para encontrar documentos que tratam do Ensino de Biblioteconomia, conferem-se as fichas sobre Ensino e sobre Biblioteconomia, as de números comuns certamente corresponderão à pesquisa em questão.

É um método manual, sendo posteriormente também empregado o uso de fichas de coincidência óptica.

Para um sistema pequeno, de fácil manipulação, pode-se considerar um método satisfatório, mas à medida que o acervo cresce e conseqüentemente cresce também o número de fichas de assunto, torna-se difícil o seu manuseio. Muitos sistemas acabam se transformando em sistemas pré-coordenados ou então passam a se utilizar de computador.

### 1.2.2 Tesouros

Aplicados preferencialmente aos sistemas automatizados, são usados, por vezes, como base para indexação pré-coordenada em sistemas manuais, desvirtuando-se de seu objetivo principal.

#### Estrutura:

Sinais relacionados com a função do Descritor:



TG (BT) - termo geral  
TE (NT) - termo específico  
TR (RT) - termo relacionado  
NA (SN) - nota de alcance  
TGM (TT) - termo genérico maior  
UP (UF) - usado para  
USE (USE) - use

**Tipos de relações (AITCHISON & GILCHRIST, 1979):**

Relações de equivalência, incluindo sinônimos verdadeiros e quase sinônimos.

Ex.: Cosmonauta  
USE Astronauta

**Relações hierárquicas incluindo gênero/espécie e coisa/tipo:**

Ex.: Árvore frutífera  
TE macieira  
TG Árvore  
Máquina  
TE Máquina de escrever

A relação parte/todo é, em alguns casos, considerada como hierárquica.

Ex.: Brasil  
TE São Paulo  
TG América do Sul

**Poli-hierarquias - termo que pertence a mais de uma hierarquia:**

Ex.: Repressão sexual                      Repressão política  
TG Repressão                              TG Repressão

Relações Associativas ou não-hierárquicas, incluindo: coisa/parte; coisa/propriedade; coisa/processo etc.

Ex.: Fichário  
TE Gavetas

A apresentação dos tesouros pode ser alfabética ou sistemática, suplementada por um índice alfabético. Alguns incluem também a classificação facetada.

### 1.3 COORDENAÇÃO: LINGUAGENS PRÉ E PÓS-COORDENADAS

A coordenação dos termos é o que permite a precisão na recuperação da informação.

#### 1.3.1 Pré-coordenação

O primeiro termo é o que determina a recuperação, o que significa que, na pré-coordenação, é necessário o uso de muitas remissivas ou entradas múltiplas para explicitar todos os conceitos significativos, o que causa um aumento de custo para o sistema na fase de entrada. Podemos observar o número de entradas no exemplo abaixo:

Ex.: Entradas múltiplas

301.32:33 - Sociologia: C.Política: Economia

32:33:301 - C.Política: Economia: Sociologia

33:301:32 - Economia: Sociologia: C.Política

Remissivas

Demarcação de terras - Índios - Goiás

X Terras - Demarcação - Índios - Goiás

X Índios - Demarcação de terras - Goiás

X Goiás - Demarcação de terras - Índios

#### 1.3.1.1 Ordem de citação

A ordem de citação pré-fixada, na maioria das vezes, favorece a precisão na recuperação e dispensa o uso de artifícios de ligação

para evitar coordenações falsas. Mesmo assim, é necessário alertar que algumas vezes a prefixação dos termos acaba dispersando elementos de conceitos relacionados.

Ex.: Código penal na Nicarágua, classificado pela CDU, ficaria assim representado:

342.2 (728.5) (094.4)

342.2 - Direito Penal

(728.5) - Nicarágua

(094.4) - Código

O conceito código penal (342.2(094.4)) fica disperso, porque a ordem de citação da CDU recomenda o local antes da forma.

### 1.3.2 Pós-coordenação

Necessita de um suporte computacional adequado, visto que a saída manual é praticamente inviável em sistemas de médio ou grande porte. Dispensa a ordem de citação e possibilita múltiplas combinações no momento da busca. Opera preferencialmente com conceitos simples, sendo que esses conceitos podem ter uma ou mais palavras.

Ex.: uma só palavra: Tuberculose - para expressar doença no pulmão.

mais de uma palavra: Cientista político (diferente de Político).

Os manuais de construção de tesouros sugerem o uso da pré-coordenação nos casos em que a pós-coordenação pode alterar o sentido do conceito, por exemplo: Vida/padrão pode significar tanto padrão de Vida como Vida Padrão.

O uso de conceitos compostos, próprios de determinadas áreas, colaboram para a especificidade da linguagem, possibilitando

melhor precisão na recuperação.

1.3.2.1 Artificios de precisão utilizados para evitar associações falsas (AITCHISON & GILCHRIST, 1979)

Elos (links): Usados para mostrar quais os termos que devem ser relacionados num mesmo documento.

Ex.: Febre Amarela ataca turistas em uma comunidade de pescadores no Rio Amazonas.

Febre Amerela - 1000 A

Turistas - 1000 A

Comunidade - 1000 B

Pescadores - 1000 B

Rio Amazonas - 1000 A

Os elos de ligação evitam associações falsas, como por exemplo Febre amarela/Pescadores e Pescadores/Rio Amazonas, que não figuram no texto. No entanto, nem sempre os elos conseguem resolver todos os problemas de coordenação, como pode ser observado no exemplo acima em que a combinação correta Comunidade de pescadores (B)/Rio Amazonas (A), fica excluída.

Aitchison & Gilchrist (1979) afirmam que os "elos são valiosos em grandes coleções, mas devem ser usados com cautela".

Indicadores de função: Determinam a função que o termo apresenta no contexto: matéria-prima, produto final, componente etc.

Ex.: Utilização de madeira na construção de habitações populares na Amazônia      Habitações populares/  
Madeira/Amazônia (1)

Habitações populares em madeira na Amazônia      Habitações populares/  
Madeira/Amazônia (2)

## Indicadores de função:

- (1) matéria-prima
- (2) produto

As linguagens de indexação envolvem problemas gerais e específicos. Os primeiros enunciam-se como: conhecimento da área, interdisciplinaridade e ideologia; os segundos como: elaboração e adaptação - questões que serão abordadas nos capítulos posteriores.

## BIBLIOGRAFIA

- AITCHISON, J. & GILCHRIST, A. (1979). *Manual para construção de tesouros*. Rio de Janeiro, BNG/Brasilart.
- BARBOSA, Alice Príncipe (1972). Classificação facetada. *Ciência da Informação*. Rio de Janeiro, 1 (2):73-81.
- CURSO sobre lenguajes de indización de tesouros (1981), por el Grupo de Trabajo sobre Lenguajes de Indización. Paris, UNESCO. (Basado em el Seminario Regional sobre Lenguajes de Indización, Buenos Aires, UNESCO/CAICYT, 1978).
- FOSKETT, A.C. (1973). *A Abordagem temática da informação*. São Paulo, Ed. UnB/Polígono.
- VICKERY, B.C. (1980). *Classificação e indexação nas ciências*. Rio de Janeiro, BNG/Brasilart.
- VIET, Jean (1971). *Thesaurus pour le traitement de l'information en Sociologie*. Paris, Mouton.