

2 Considerações a respeito da D.D.I.B.

A Divisão de Documentação e Informação Bibliográfica D.D.I.B. - é uma unidade técnica, subordinada ao Departamento Econômico da Confederação Nacional da Indústria.

O sistema de informação da D.D.I.B. é um sistema operado manualmente. Ele foi planejado e implantado em 1969, com o objetivo de dar apoio bibliográfico aos técnicos do Departamento Econômico. Desde então, esse sistema vem desempenhando atividades relativas à seleção, aquisição, tratamento, armazenamento e recuperação da informação de interesse para seus usuários - atualmente cerca de 50 técnicos, entre economistas graduados e pós-graduados e estagiários de Economia.

A coleção bibliográfica da D.D.I.B. é formada por publicações afeitas às áreas econômica e afins. Especificamente, ela se compõe de monografias, folhetos, anais de congressos e de conferências, publicações seriadas sobre legislação econômica, obras de referência, anuários, relatórios e outras publicações periódicas.

Esse material encontra-se catalogado de acordo com as normas do Código de Catalogação Anglo-Americano-C.C.A.A., 2. ed., classificado conforme a Classificação Decimal Universal-C.D.U., Edição Média em Língua Portuguesa, 1976, e indexado por assunto segundo vocabulário desenvolvido pelo próprio sistema da Divisão.

Atualmente, o acervo da D.D.I.B. constitui-se de 2811 monografias (inclusive folhetos, publicações seriadas e anais de congressos) e 438 títulos de periódicos, nacionais e estrangeiros (inclusive relatórios anuais de empresas e anuários).

No que se refere a artigos de periódicos, anualmente cerca de 1000 documentos são indexados, resumidos e classificados temática e cronologicamente.

O tipo de informação tratada pela D.D.I.B. e a necessidade de informação de seus usuários foram determinantes na opção de uma linguagem de indexação mais flexível do que a contida nos instrumentos de controle de vocabulário até então conhecidos.

Assim, na época da implantação do Sistema, optou-se pela adaptação do índice temático da publicação "Índice de periódicos brasileiros de Economia", entendendo-se que o mesmo satisfaria as necessidades do Sistema, uma vez que os assuntos nele arrolados refletiam a literatura produzida por dois dos periódicos mais expressivos da área econômica: Conjuntura Econômica (editado pela Fundação Getúlio Vargas) e Desenvolvimento & Conjuntura (editado pela Confederação Nacional da Indústria).¹

Com a continuidade do trabalho de indexação, houve necessidade

¹ D. Furtado & L. de L. Saide. *Índice de periódicos brasileiros de Economia*, Rio de Janeiro, FIRJAN, 1968. 266p.

de se ampliar o vocabulário inicialmente usado, o que foi feito mediante o acréscimo de novos termos selecionados da linguagem utilizada *correntemente na literatura da área, ligeiramente controlados, através de sinônimos e de termos relacionados.*

A proporção que a coleção bibliográfica aumentou e se diversificou, também aumentaram as dificuldades de indexação e, conseqüentemente, de recuperação da informação. Isso fez com que se pensasse em uma linguagem mais apropriada ao Sistema.

O reconhecimento de melhores resultados no tratamento da informação com controle terminológico conduziu à decisão da reestruturação da linguagem de indexação que vinha sendo usada. Soergel (1974:29), por exemplo, afirma que "é altamente reconhecido, hoje em dia, que as linguagens de indexação que usam controle terminológico dão melhores resultados, e a maioria das linguagens de indexação usam controle da terminologia...".

Por outro lado, a inexistência de uma linguagem documentária específica da área econômica, nacional ou estrangeira, também contribuiu para aquela decisão.

Uma análise da literatura pertinente permitiu verificar que as linguagens para tratamento da informação econômica encontram-se diluídas nas linguagens de Ciências Sociais. Os resultados da pesquisa feita por Meyriat (1980), por exemplo, revelaram haver sete linguagens documentárias de Economia, porém todas englobadas nas linguagens de Ciências Sociais, algumas enciclopédias, tentando cobrir todo o universo do conhecimento; outras, tentando definir o campo de forma muito abrangente, incluindo áreas limítrofes e tangentes de outras áreas especiais.

A nível nacional, resultado semelhante foi observado a partir de levantamento feito para identificar as Instituições que desenvolvem linguagens documentárias próprias, adaptam ou traduzem tesouros ou outros vocabulários controlados (IBICT-1984).

Quanto à possibilidade de se traduzir e/ou adaptar linguagens já existentes, ainda que não específicas da área, as seguintes restrições são colocadas:

- a) cada sistema possui clientela própria com necessidades de *informação próprias, e só por ela o sistema pode ser legitimado;*
- b) a linguagem é fenômeno cultural. Portanto, a terminologia usada na área econômica nacional refletirá, necessariamente, as peculiaridades da nossa cultura;
- c) a experiência que o próprio sistema de informação da D.D.I.B. teve com adaptação de vocabulário (conforme exposto anteriormente) não se provou satisfatória, e
- d) a estruturação do conceito depende do sistema de conceitos ao qual ele pertence, o que, por sua vez, determina sua definição.

Some-se a esse fatores, os já abordados em 1("a" a "g").

3 Tesouro

3.1 Considerações gerais

Os instrumentos de indexação/recuperação destinam-se a descrever informação, de forma a que ela seja processada, armazenada, recuperada e comunicada. Eles são linguagens de informação, ou linguagens documentárias que, segundo Soergel (1967) e Gardin (1969) são metalinguagens, a um só tempo dependentes e determinantes da forma e do conteúdo das representações documentárias.

Há várias classes gerais de instrumentos destinados a controlar os vocabulários usados em sistemas de informação: as tabelas de classificação, as listas de cabeçalhos de assunto e os tesouros. Somente estes últimos serão aqui focalizados, por serem a preocupação central deste estudo.

Para Spark Jones (1972), o tesouro é a mais antiga e persistente forma de classificação de vocabulário, pretendido como contribuição para descrição da linguagem e para ajuda no seu uso.

O tesouro apresenta certas peculiaridades, se comparado com outras linguagens documentárias. Ele difere dos sistemas de classificação hierárquica por assunto porque, enquanto estes têm estrutura hierárquica, o tesouro apresenta, ainda, relações associativas. Difere, também, das listas de cabeçalho de assunto, porque estas baseiam-se em palavras, e não em conceitos e apresentam estrutura ambígua.

A palavra "tesouro" provém do vocábulo latino "thesaurus", o qual deriva do grego "thesaurós" e que foi usado durante vários séculos com o significado de "tesouro ou armazém/repositório de palavras".

Segundo afirma Mikhailov (1973:398), esse termo foi utilizado pela primeira vez por Brunetto Latini (1220-1294), para designar uma enciclopédia sistematizada, à qual deu o nome de "Os livros do tesouro". Em 1532, Thierry e R. Etinne publicaram seu "Dicionário ou tesouro do idioma latino", dicionário da língua latina, com arranjo alfabético. Em 1572, A. Etienne publicou um dicionário: "Thesaurus linguae Graecae", cuja confecção havia sido iniciada por seu pai R. Etienne.

Em 1736, o Shorter Oxford Dictionary registrou pela primeira vez o uso da expressão inglesa "treasury or storehouse of knowledge", definido-a como "tesouro ou armazém de conhecimento, similar a um dicionário ou a uma enciclopédia.

Já o dicionário Webster's define o termo como "um livro de palavras, ou de informação sobre um determinado campo, ou um conjunto de conceitos, especificamente um dicionário de sinônimos".

O mais conhecido entre os dicionários ideológicos da antiguidade é o "Onomástikon", de Polux, cuja confecção fundamentou-se em

trabalhos mais antigos, como os dicionários de Aristófanes de Bizancio, Pánfilo e Dídimos el Chalkentros. Nesse dicionário, as palavras do idioma grego (incluindo todos os sinônimos) estão agrupadas em dez classes: Deuses, Homem (partes do corpo), Genealogias, Ciência e Arte, Caça, Alimentos, Ofícios, Leis, Organização urbana, Utensílios (Mikhailov, 1973:400).

Atualmente, o tesouro que goza de maior popularidade é o "Thesaurus of English Words and Phrases", concebido pelo inglês Peter Roget em 1852. O Tesouro é definido como "uma coleção... arranjada, não em ordem alfabética como num dicionário, mas segundo as idéias que elas representam..." (Roget, 1962).

O tesouro de Roget tem duas características: seu objetivo e sua forma. Seu objetivo é ajudar o usuário a se mover de uma idéia para a palavra que ele pode usar, para expressar aquela idéia, num texto escrito. As idéias são simbolizadas por palavras. O tesouro de Roget gruba palavras, tipicamente sinônimos, além de termos genéricos e relacionados e antônimos, mas sua essência não é determinada por um objetivo específico.

Segundo Spark Jones (1972), as linhas entre os tesouros e as classificações não são sempre claras. A partir dessa observação, a autora baseia-se em trabalhos que exibem algum agrupamento de palavras da linguagem natural sob rótulos conceituais explícitos ou implícitos, com uma indicação de relações entre as classes e identifica várias linhas na história do tesouro:

- a) classificação de vocabulários por assunto, agrupando palavras relacionadas, como, por exemplo, o clássico "Onomástikon" de Polux (acima referido).
- b) dicionários enfatizando relações verbais, como o "Synonymous Etymological, and Pronouncing English Dictionary" (Perry, 1805).
- c) sinonímia, como forma particular de relação verbal: "Justesse de la Langue Française" (Girard, 1718).
- d) adoção de ordem conceitual, muito mais do que alfabética, como no "English Synonyms Explained in Alphabetical Order" (Crabb, 1826).

Ex: a) Conhecimento, Ciência, Aprendizagem, Erudição.

b) Cultivo, Cultura, Civilização, Refinamento.

O trabalho de Crabb pode ser considerado como predecessor do tesouro de Roget, embora Roget não pensasse em seu trabalho como um dicionário, mas sim como uma classificação de idéias, como pode ser observado na introdução do seu tesouro, onde ele diz que as palavras estão arranjadas "de acordo com as idéias que elas expressam... estritamente de acordo com seu significado" (Roget, 1962).

Schulte-Albert (1979) afirma que, dentre todas essas iniciativas, dois trabalhos realizados durante o Século XVII - Dalgarno (1661) e Wilkins (1668) - são considerados contribuições fundamentais à história do tesouro. Dalgarno freqüentemente grupou termos opostos e forneceu,

às vezes, um ponto mediano num "continuum", entre extremos; Wilkins incluiu sinônimos, antônimos, termos que mostravam relações hierárquicas e, também, relações de outros tipos.

Roget viu o trabalho de Wilkins como o pioneiro na construção de tesouros para a classificação de idéias.

De um modo geral, o tesouro começou a ser usado na década de 50, como instrumento para conversão dos conceitos e suas relações expressos na linguagem natural dos documentos em uma linguagem de informação, com controle de sinônimos, quasi-sinônimos, etc. e estruturas sintáticas simplificadas, isto é, privilegiando palavras, e não expressões.

Existe muita contradição, e até incompreensão, a respeito da natureza e das funções de um tesouro.

Essa confusão é atribuída, em grande parte, à ambigüidade da terminologia usada no campo da Ciência da Informação.

Segundo Neufeld (1972), "freqüentemente os usos e objetivos de um tesouro não são compreendidos. Há confusão sobre se o tesouro deva ser um sistema para representação ou para recuperação de documentos, e sobre como medir sua eficácia em qualquer das duas abordagens", embora algumas medidas tenham sido sugeridas, tais como rapidez e consistência do indexador, facilidade de recuperação, revocação e relevância. Mesmo Spark Jones (1971), que apresenta opinião definitiva sobre as funções de recuperação do tesouro, reconhece explicitamente a confusão generalizada em torno de sua função.

Já no Século XVII, Wilkins sugeriu dois usos do seu esquema de classificação: armazenamento e recuperação. A partir de então, as controvérsias a respeito do significado do tesouro tomaram vulto.

Roberts (1984) faz um resumo significativo dessa situação, quando diz que Eugene Wall, confrontado pela falta de acordo a respeito do significado do tesouro, tentou eliminar esse problema, afirmando que os usuários de "thesaurus" não deveriam esperar ter o mesmo referente em mente. Mas, de qualquer forma, tesouros literários e tesouros para recuperação tinham o mesmo objetivo; Clapp pleiteou que a lista de cabeçalhos de assunto da Library of Congress deveria ser tida como o primeiro tesouro, evidenciando assim as incertezas causadas pela imprecisão terminológica; Korotkin decidiu que a única forma de acabar com o caos terminológico era chamar tudo de tesouro: todos os tipos de listas de cabeçalho de assunto, listas alfabéticas, listas de conceitos, quer arranjadas alfabeticamente, por relações ou por conceitos.

Como se percebe, mesmo na década de 60, quando o tesouro teve seu desenvolvimento acelerado, muitas dúvidas persistiram quanto ao seu uso, a despeito da aceitação geral de seu nome. Roberts (1984:277) coloca, entre outras, as seguintes questões:

- ☐ "- era o tesouro uma ferramenta para indexadores?
- ☐ - era uma ajuda para os pesquisadores?
- ☐ - para ambos?

era necessário um tesouro indicar relações conceituais?
e se fosse, como?"

A definição dada pela Unesco (1973) parece não deixar dúvidas, pelo menos, quanto à finalidade do tesouro, já que esse é visto como um "instrumento para traduzir a linguagem natural de documentos, de indexadores ou de usuários, num sistema de linguagem documentária".

Como se vê, o tesouro é um sistema que serve, tanto à indexação, quanto à recuperação da informação; ele cumpre suas próprias funções, de acordo com o objetivo a que se destina.

Para fins deste trabalho, então, o tesouro será entendido da seguinte forma:

Sistema de vocabulário baseado em conceitos, incluindo termos preferidos (descritores), termos não preferidos (não descritores) e suas interrelações, que se aplica a um determinado ramo do conhecimento e que se destina a controlar a terminologia utilizada para a indexação/recuperação de documentos.

3.2 Métodos de agrupamento de termos para a construção de Tesouros

Um tesouro compreende os atributos usuais da linguagem:

- a) um léxico (vocabulário em si próprio)
- b) uma estrutura - sistema de regras pelo qual os termos no léxico são arranjados, através do interrelacionamento dos termos.

Portanto, pode-se dizer que a construção de tesouros tem, basicamente, dois aspectos:

- a) seleção de termos de determinado assunto, para o qual o tesouro é construído, cujo objetivo é a geração do léxico.

A geração de itens dá-se através de:

- análise pormenorizada da literatura primária, de livros - texto; de dicionários, de outros tesouros, de glossários, de sistemas de classificação, de enciclopédias, de listas de cabeçalhos de assunto.

- consulta a especialistas do assunto.

- experiência individual e conhecimento dos compiladores.

- realimentação do usuário.

- b) estabelecimento de relações entre os termos (sinonímia, relações hierárquicas e outras), cujo objetivo é a geração da estrutura.

O limite do conjunto de termos, como um todo, cujas relações semânticas serão demonstradas, tem sido traçado, até agora, com base em duas abordagens, comuns à construção de sistemas de classificação em geral: a dedutiva e a indutiva. Como os tesouros são instrumentos que se utilizam de processos classificatórios, esses métodos também têm sido adotados em sua elaboração e citados pelos autores de várias formas.

Assim é que o método dedutivo é mencionado por Lancaster (1972)

como "consensual" ou "committe approach"; por Wooster (1972) como "estalactite"; pela ANSI (1974), como "gestalt"; por Ghose & Dhawle (1977), como "método de categoria facetada", e por Fhromann (1983), como "a priori".

Em contraposição, esses mesmos autores citam o método indutivo como "empírico", "estalagmite", "analítico", "método estatístico a posteriori" e "a posteriori", respectivamente.

3.2.1 Método dedutivo

O método dedutivo é a primeira área da pesquisa em Classificação e a mais tradicional. Em Classificação, ele é melhor conhecido através do método clássico da divisão lógica.

Para a elaboração de sistemas de classificação, esse método consiste na subdivisão de um universo do conhecimento em disciplinas.

Nessa abordagem (global), o conjunto de termos é definido aprioristicamente, como uma rede de relações entre os significados dos termos, sendo essa rede construída via análise conceitual.

A terminologia da área a qual o tesouro se destina é obtida por consenso de especialistas que se reúnem em comitês e discutem sobre os termos que devem integrar o vocabulário, sem perder de vista os objetivos do sistema de informação ao qual o tesouro vai servir. A partir de então, os termos que efetivamente constituirão esse vocabulário são deduzidos, com base na definição prévia das classes gerais as quais os termos pertencem ou de suas facetas principais.

Três dos maiores sistemas de classificação são baseados no método dedutivo: CDD, CDU e a Classificação Bibliográfica de Bliss. Embora essas classificações sejam utilizadas até hoje por vários sistemas de informação, elas se têm mostrado, muitas vezes, deficientes, por sua impossibilidade de caracterizar um tópico por um único termo, já que um tópico pode ter várias facetas.

Ainda pelo método dedutivo (ou método de categoria facetada, segundo Ghose & Dhawle, 1977), o problema de várias facetas teve sua solução tentada por Ranganathan (1971), com suas cinco categorias fundamentais - PMEST; por Vickery (1968), que desenvolveu a noção de classificação facetada através de uma lista abrangente de categorias fundamentais; por Meincke e Atherton (1976), que dão nova versão da "abordagem categorial" para o universo do conhecimento, onde cada elemento possui características multidimensionais.

Em todas essas abordagens é assumido que as categorias não são ambíguas e são mutuamente exclusivas.

3.2.2 Método indutivo

Por esse método, o conjunto de termos é definido a partir da

investigação das redes que existem nas práticas humanas. Considera-se que as práticas humanas determinam o significado das palavras, antes de serem por ele determinadas.

A construção de sistemas de linguagens documentárias faz-se a partir de termos mais genéricos e mais específicos (abordagens de elementos), coletados na literatura corrente da área, até se chegar às classes gerais ou facetas principais. Assim, sistemas são construídos do particular para o geral, pela seleção de termos em grupos mutuamente exclusivos e de acordo com vários critérios, pela distinção de suas características e pela sua interrelação.

Richmond (1965) e Ghose & Dhawle (1977) discorrem sobre o método empírico e citam vários exemplos de sistemas e experiências onde o método foi usado na seleção de termos e no estabelecimento de suas interrelações.

3.2.3 Métodos dedutivo e indutivo e os princípios que regem a seleção dos termos.

Os métodos dedutivo e indutivo derivam dos dois princípios que regem a seleção de termos: "endosso do usuário" e "garantia literária", respectivamente.

A expressão "garantia literária" ("literary warrant"), usada por Hulme (1950), implica na determinação de classes, a partir de exame acurado e medição dessas, conforme identificadas na literatura, e não a partir da classificação do conhecimento. Hulme afirma "que um título de uma classe só é garantido quando a literatura provou que ele existe e que a unidade dos assuntos é expressa pelo encadeamento das classes, e não pela sua sistematização lógica". (Hulme 1950:22). Somente dessa forma a classificação é real, e não potencial.

Segundo Frommann (1983), o CRG-Classification Research Group, Vickery e Mills também assumem o princípio da "garantia literária", quando dizem que a seleção e o relacionamento de termos consiste na análise dos termos usados em grande número de textos sobre determinado assunto e enfatizam que um campo de assunto é definido pelo núcleo de problemas que dão origem à sua literatura. Para Dodebei (1979:13-4), Foskett e Goodman sustentam esse mesmo princípio, segundo o qual o Thesaurus of Eric Descriptors, por exemplo, foi desenvolvido.

Já o princípio do "endosso do usuário" ("user warrant" ou "personal warrant" para Lancaster, 1972) baseia-se na confirmação dos produtores e utilizadores de um sistema de informações, a respeito dos termos que devem integrar seu vocabulário. Tal confirmação tem como base a pressuposição de que os termos, por serem efetivamente usados pela comunidade à qual o sistema serve, serão os utilizados nas solicitações de busca bibliográfica.

Bush (1978) adere a esse princípio, quando propõe que, para a construção de tesouros de temas científicos e de engenharia, os termos sejam proporcionados através de seleção feita por especialista da área.

Várias foram as experiências em construção de tesouros em que os princípios da "garantia literária" e do "endosso do usuário" foram usados complementarmente. Dentre elas, pode-se citar os trabalhos de Dym, Pickford e Dodebei (1979), segundo informação dessa última autora.

Cabe ressaltar, entretanto, que esses princípios são referidos na literatura mais comumente para a seleção dos termos do que para seu interrelacionamento. Conforme dito anteriormente, existem nos tesouros normas que orientam o estabelecimento das relações entre os termos; o que não existe é o fundamento teórico das propostas.