

Dilza Fonseca da Motta

**Método Relacional como Nova Abordagem
para a Construção de Tesouros**

Rio de Janeiro, 1987

Apresentação

Embora este trabalho não esteja relacionado diretamente com a Educação, a sua amplitude, dada a natureza da preocupação que o motivou, o faz importante para aplicação em qualquer área do conhecimento humano, inclusive para a Formação Profissional.

A sua proposta metodológica para o desenvolvimento de tesouros ajuda a suprir uma lacuna existente em nosso acervo bibliográfico e a sua divulgação ocorre num momento bastante oportuno por coincidir com os esforços ora realizados para a construção de um tesouro da Formação Profissional no País.

Julgamos assim que a divulgação desta obra resulte numa contribuição junto aos técnicos de Formação Profissional, além daqueles de outras áreas, que estejam buscando soluções alternativas para a construção de tesouros.

O trabalho foi originalmente apresentado por Dilza Fonseca da Motta, em 1986, ao Instituto Brasileiro de Informação em Ciência e Tecnologia, como requisito parcial para a obtenção do Grau de Mestre em Ciência da Informação. Sua orientadora foi a Professora Hagar Espanha Gomes, sendo seu co-orientador o Professor Valdir Ramalho de Melo.

Lauro Pio de Miranda
Diretor Técnico

Resumo

O estabelecimento das relações genéricas, partitivas e outras em tesouros não tem tido sua metodologia norteada adequadamente. Propõe-se o uso de uma nova abordagem - "Método Relacional" - como meio de estabelecer aquelas relações em sistemas conceituais, de forma mais objetiva. Tal método fundamenta-se na Teoria do Conceito, que, por sua vez, tem na definição do conceito sua ferramenta básica. O Método Relacional parece ser útil em várias circunstâncias, tais como: determinação de sinônimos, atualização do sistema, inclusão de termos em uma só categoria e mapeamento de áreas de assunto. Sugere-se a aplicação do método a outros sistemas de informação e seu teste, a fim de avaliar suas vantagens e desvantagens.

Abstract

The methodology for the establishment of generic, partitive and other relations in thesauri has not been adequate. This dissertation suggests the use of a new approach - "Relational Method" - as a means to establish those relations in conceptual systems more objectively. Such method is based on the Conceptual Theory, which, on its turn, has the definition of the concept as its basic tool. The Relational Method may have a wide range of uses, namely: the setting up of synonyms, the updating of the system, the inclusion of terms in just one category, and the mapping out of subject areas. It is suggested that the Relational Method be applied to information systems and that it be tested for further evaluation of its advantages and disadvantages.

SUMÁRIO

1	Introdução	15
2	Considerações a respeito da D.D.I.B.	19
3	Tesouro	21
3.1	Considerações gerais	21
3.2	Métodos de agrupamento de termos para a construção de tesouros	24
4	Conceito, Teoria do Conceito e Interrelações Conceituais - Revisão da Literatura	29
5	Método Relacional	39
6	Material e Método	41
7	Resultados	47
8	Conclusões	61
9	Anexos	63
9.1	Ocorrência de termos na indexação 1982-1983	63
9.2	Termos usados na indexação por ordem decrescente de ocorrência 1982-1983	70
9.3	Distribuição de freqüência dos termos usados na indexação 1982-1983	77
9.4	Termos selecionados segundo o princípio da garantia literária	79
9.5	Definições dos conceitos	79
9.6	Mapeamento da área de política fiscal	84
9.7	Mapeamento da área de política econômica	85
9.8	Mapeamento da área de desenvolvimento econômico	86
10	Bibliografia consultada	87

1 Introdução

O problema da linguagem como elemento simbólico exato para a descrição e comunicação das descobertas e do desenvolvimento científicos é, há muito tempo, preocupação mundial.

Já no Século XVII, com o desenvolvimento das ciências naturais, estudiosos estavam interessados em estabelecer uma linguagem universal que resolvesse o problema da barreira lingüística, quer de forma geral, como meio de comunicação entre os povos, quer especificamente no campo da ciência.

Esse problema persiste até hoje, agravado basicamente por três fatores. Um deles é a valorização da informação, por ser um dos produtos mais demandados pela sociedade moderna. O outro, a expansão contínua das áreas científicas, o que acarreta grande volume de informação bibliográfica e interdisciplinaridade das áreas. Por isso, a informação registrada está cada vez mais complexa e multidimensional em sua essência. Geralmente ela envolve uma interrelação de produtos, mecanismos, procedimentos e suas respectivas características. Os conceitos emergentes e a conseqüente alteração no significado dos conceitos científicos existentes são o terceiro fator agravante dessa problemática.

Como conseqüência, percebem-se esforços científicos, cada vez maiores, no sentido de padronizar a terminologia usada pelas várias ciências.

É sabido que cada área científica tem seus próprios padrões de comunicação, o que se faz sentir também através da linguagem usada por cada uma delas. As diferenças entre a terminologia empregada pelas Ciências Naturais e as Ciências Sociais foi percebida, segundo Bonzi (1984:247), por vários autores, como Adam, Stoan, Cleverdon e Riggs, entre outros.

As Ciências Físicas, por exemplo, tendem a ter conceitos relativamente definíveis, por lidarem com objetos e as relações físicas entre eles. O mesmo não acontece com as Ciências Sociais, que lidam com o comportamento humano e valores sociais que investigam, em grande parte, conceitos abstratos que raramente são percebidos por duas pessoas com um mesmo grau de igualdade ou semelhança.

Em suas comunicações, os cientistas sociais apoiam-se, quase que exclusivamente, em termos derivados do uso da linguagem corrente, resultando na proliferação de significados.

Ao contrário dos cientistas naturais que, via de regra, se utilizam de palavras cunhadas especificamente para representar novos conceitos, os cientistas sociais têm um padrão comportamental responsável pelo surgimento de homógrafos e de sinônimos, segundo opinião de Riggs (1978). Os homógrafos ocorrem quando surge um novo conceito e uma palavra familiar é adotada para expressá-lo, tendo essa palavra, porém,

um novo significado. Já os sinônimos decorrem do uso freqüente de diferentes palavras para designar o mesmo conceito.¹

Em meio a essa multiplicação dos significados de palavras familiares, o cientista social necessita explicitar os novos significados. Assim, eles começam a contextualizar os termos, a dar suas próprias definições, novas conceituações. Dessa forma, a ambigüidade de palavras usadas em sentidos particulares aumenta e a necessidade de novas explicações aumenta paralelamente, formando-se um círculo vicioso.

Percebe-se, então, que a mesma capacidade que as Ciências Naturais têm de moldar neologismos para novos conceitos não é facultada às Ciências Sociais, razão pela qual a necessidade de padronização terminológica faz-se sentir mais agudamente nessas áreas.

Em relação à indexação e recuperação de documentos, essa padronização é conseguida através de linguagens documentárias, construídas especificamente com o objetivo de controlar a linguagem usada correntemente.

Embora a literatura sobre indexação revele que as idéias sobre o uso de linguagem natural ou de linguagens documentárias sejam contraditórias e difíceis de se coordenar, a exposição feita até aqui a respeito da terminologia usada pelos cientistas em geral, e, particularmente, pelos cientistas sociais parece tornar clara a necessidade de se converter a linguagem natural, tal como usada nos textos científicos e técnicos, em linguagens controladas que facilitem a tarefa de indexação/recuperação.

Já no Século XVII, Bacon e outros estudiosos suspeitavam que a linguagem natural que descreveria a ciência apoiava-se mais nas palavras do que na natureza das coisas. Eles temiam que a linguagem natural tendesse a obscurecer a realidade, mais do que a descrevê-la com precisão.

Mais modernamente, muitos relatórios de implementação de controle de vocabulário para um determinado sistema exibem resultados benéficos, auferidos através de consistências na indexação ou recuperação de documentos.²

Até onde se sabe, mesmo a evidência mais forte em favor da linguagem natural, representada pelos resultados dos testes Cranfield, há que ser vista com restrições, já que o próprio relator dos testes afirma, entre outras coisas, que "seria absurdo qualquer organização abandonar

¹ Essa diferença de comportamentos leva Fred W. Riggs, em *Indexing glossary: a new rationale*, Frankfurt / Main, INDEKS - Verlag, 1982, p. 4, a distinguir duas formas de neologismos (termos usados para designar novos conceitos): "neoterismos", para as Ciências Naturais, e "neosemanticismos", para Ciências Sociais.

² A esse respeito, M. Neufeld menciona em seu artigo "Linguistic approaches to the construction and use of thesauri: a review", 1972, os trabalhos de Janning (1966), Pickford (1971), Tinker (1966) e Weinstein (1966).

a indexação convencional ou linguagens de indexação controladas com base nos resultados desses testes" (Cleverdon, 1967:191).

Além disso, entende-se que a linguagem controlada tenha algumas vantagens sobre a linguagem natural. Especificamente, em relação à indexação, pode-se citar:

- a) controle de sinônimos e quasi-sinônimos
- b) distinção de homógrafos
- c) minimização de esforço intelectual na atribuição de descritores e conseqüente racionalização do trabalho
- d) maior margem de consistência da indexação realizada por diversos indexadores
- e) aumento da consistência do vocabulário dentro de uma mesma área
- f) restrição do número de descritores usados pelo sistema, evitando sua sobrecarga.
- g) expressão da linguagem a nível conceitual (idéias), e não a nível verbal (palavras).

Um dos instrumentos de controle do vocabulário para fins de indexação/recuperação da informação é o tesauro.

A construção de tesauros envolve, basicamente, dois aspectos: seleção do vocabulário e estabelecimento de relações entre os termos. O primeiro tem sido fundamentado em dois princípios: o da garantia literária e o do endosso do usuário, derivados, respectivamente, dos métodos que, em geral, têm orientado a construção de sistemas de classificação: o indutivo e o dedutivo.³ Ao contrário desse, o segundo aspecto não tem sido norteado teoricamente. O interrelacionamento dos termos tem-se baseado, tão somente, nas normas que os tesauros exibem para tal fim, mas o embasamento teórico para as propostas que não esgotam as necessidades dos tesaurografos - é inexistente.

Ainda quanto ao estabelecimento de relações, sabe-se da necessidade de identificar outras, que não a genérica; é reconhecido que esse tipo de relação não é o único possível entre os conceitos. A introdução de tabelas auxiliares na C.D.D e do uso dos dois pontos na C.D.U. indicam a percepção de outros tipos de relações.

Sabe-se, também, que, por definição, um tesauro deve exibir relações genéricas e outras entre termos. Mas, como identificar "outras" relações, se as normas para construção de tesauros carecem de bases teóricas que orientem o seu estabelecimento?

O que se percebe num tesauro, na maioria das vezes é a fixação dessas relações de maneira um tanto arbitrária, isto é, de acordo com o conhecimento de cada pessoa e, não raras vezes, com base no "achismo", uma vez que as notas introdutórias dos tesauros, em geral, não estabelecem princípios que orientem o estabelecimento das relações.

³ Esses métodos e princípios estão referidos brevemente no item 3.2.

Dahlberg (1976:89) sugere que, com o auxílio de uma teoria analítica dos conceitos, uma terceira abordagem possa ser usada na construção de sistemas de classificação em geral - a abordagem relacional, que imprimiria maior objetividade aos sistemas. Paralelamente, a autora sugere pesquisas dirigidas à análise de conceitos, especificamente de combinações entre conceitos.

A partir dessas sugestões e à luz da Teoria do Conceito, resolveu-se fazer esta dissertação, que objetiva, de um modo geral, propor o "Método Relacional", enquanto instrumento básico para o estabelecimento das relações entre os termos de um tesouro em bases mais objetivas, e, portanto, mais sólidas do que aquelas que têm sido utilizadas até hoje. Especificamente, pretende-se que esse método seja aplicado ao sistema de vocabulário da Divisão de Documentação e Informação Bibliográfica do Departamento Econômico da Confederação Nacional da Indústria D.D.I.B., de forma a se desenvolver uma linguagem documentária mais adequada aos serviços por ela desempenhados.

A área de assunto escolhida para o desenvolvimento da dissertação foi Economia, o que se justifica por ser essa a especialidade da documentação afeita à D.D.I.B., material básico para a elaboração deste trabalho.

O presente estudo divide-se em oito partes, que compõem seu texto, além dos anexos e da bibliografia consultada. Após a Introdução, na qual foram expostos o problema que originou esta dissertação e seus objetivos, são feitas considerações a respeito da D.D.I.B.: sua criação, objetivos, constituição de seu acervo, serviços prestados e clientela à que se destina. O item 3 é dedicado a considerações gerais sobre o tesouro (história, natureza, função) e aos métodos de agrupamento de termos para sua construção. O item 4 aprecia a literatura dos temas centrais envolvidos neste trabalho, a saber: conceito, Teoria do Conceito e interrelações conceituais. O item 5 define, caracteriza e sugere o "Método Relacional", como método mais sólido para o estabelecimento de relações em tesouros. O item 6 descreve o material e a metodologia usados. Nos itens 7 e 8 são apresentados, respectivamente, os resultados e as conclusões do estudo.

As conclusões permitem observar que o "Método Relacional" parece ser um instrumento válido e seguro para a construção de tesouros, no que tange o estabelecimento de relações entre seus termos. Entretanto, é necessário que ele seja mais explorado, possivelmente com a colaboração de técnicos de outras áreas, como Psicologia (com o objetivo de melhor entender a natureza da formação das associações de idéias), e terminologia (para os problemas relativos à padronização conceitual). Além disso, faz-se necessário que o Método seja aplicado a outros sistemas de informação e que seja testado, para ter a sua validade legitimada.