

Base de Dados como Elemento Facilitador para o Tratamento dos Dados Terminológicos

Adriana Gonçalves Xavier
Ananda Feix Ribeiro
Laís Rosa dos Santos
Rafael Port da Rocha
Regina Helena van der Laan

RESUMO

Relato da elaboração e da implementação de uma base de dados desenvolvida como ferramenta auxiliar para o tratamento dos dados coletados para a pesquisa - Avaliação dos Descritores Relativos a Ciências da Informação, Empregados no Processo de Indexação Alfabética pela Biblioteca Setorial da Faculdade de Biblioteconomia e Comunicação/UFRGS. Essa avaliação está sendo desenvolvida baseada na Teoria Comunicativa da Terminologia. Nessa teoria o termo é analisado na perspectiva de uma unidade lexical terminológica, sendo necessário sua identificação em situação discursiva, ou seja, em funcionamento. A base de dados visa agilizar e facilitar o trabalho na consolidação dos termos e a elaboração de definições para a construção de um glossário. Este será construído com a finalidade de tornar-se critério avaliador dos descritores relativos à Área de Ciências da Informação. A base de dados foi desenvolvida segundo o modelo relacional que possibilita estabelecer com maior precisão as relações entre os campos constituintes da ficha de coleta dos dados terminológicos empregada na pesquisa. A base está estruturada nas seguintes entidades: compilador, fonte, termo, definição, contexto, marcas semânticas e definição para a elaboração do glossário. Destaca-se que essa ferramenta poderá ser empregada em outras pesquisas de cunho terminológico e para construção de tesouros.

PALAVRAS-CHAVE: Base de Dados. Terminologia. Indexação Alfabética.

1 INTRODUÇÃO

Este trabalho originou-se da necessidade de armazenar e tratar os dados coletados para a pesquisa Avaliação dos Descritores Relativos a Ciências da Informação, Empregados no Processo de Indexação Alfabética pela Biblioteca da Faculdade de Biblioteconomia e Comunicação (FABICO)/UFRGS.

Apesar de se ter conhecimento de programas para tratamento de dados terminológicos, optou-se pela elaboração de uma base de dados, possibilitando dessa forma o inter-relacionamento de uma pesquisa e de disciplinas do Curso de Biblioteconomia da UFRGS. Fato esse que foi considerado pelo grupo como uma experiência mais enriquecedora, valorizando, assim, não apenas o produto mas atribuindo valor também ao processo.

Com esse trabalho foi oportunizado aplicar os conhecimentos aprendidos nas disciplinas Linguagens Alfabéticas de Indexação, Introdução à Terminologia, Planejamento e Elaboração de Bases de Dados e Metodologia da Pesquisa.

A seguir será realizado um breve relato sobre a pesquisa de avaliação dos descritores, uma explanação sobre o processo de indexação e terminologia com o objetivo de contextualizar o presente trabalho. Após, será explicado como foi realizado a construção da base de dados, iniciando com a fundamentação teórica. Conclui-se com as vantagens observadas até o momento.

2 A PESQUISA

A pesquisa de avaliação dos descritores relativos à Área de Ciências da Informação, empregados no processo de indexação do Sistema de Automação de Biblio-tecas (SABi) da Biblioteca da FABICO/UFRGS se caracteriza pela interface entre a Ciências da Informação e a Terminologia.

A justificativa desse estudo teve base em observações informais de que os descritores relativos a Ciências da Informação, empregados no processo de indexação alfabética pela Biblioteca da Faculdade de Biblioteconomia e Comunicação, em geral, não obtinham respostas adequadas. Existiam divergências conceituais entre os descritores e a terminologia da Área, ocorriam falhas de comunicação entre o usuário e o sistema, além de existir silêncio ou ruído na recuperação de informações. Com base nessas observações tem-se como questão de pesquisa o questionamento se os descritores empregados estão representando adequadamente a terminologia dessa área.

Os objetivos da pesquisa são avaliar a pertinência dos descritores empregados na base do Sistema de Automação de Bibliotecas (SABi) da UFRGS, de acordo com a terminologia da área de Ciências da Informação, com finalidade de minimizar e/ou eliminar equívocos conceituais e avaliar os termos

utilizados pelo sistema, averiguando sua validade do ponto de vista terminológico.

O trabalho desenvolvido pelos bolsistas de iniciação científica constitui-se em coletar em corpus textual os termos expressos pelos especialistas da área, portanto em uma situação de funcionamento e não in vitro, ou seja, coletados de dicionários, glossários, sistemas de classificação entre outros.

Essa coleta é registrada em formulário dividido nos campos: termo; fonte de coleta do termo; contexto; definição; variantes; compilador e data da compilação.

O corpus textual é constituído por artigos em português publicados em periódicos nacionais vinculados a programas de pós-graduação em Ciências da Informação, no período de 1995 a 2003.

Os dados coletados ao longo dessa pesquisa serão registrados em um glossário de Ciências da Informação. Os glossários são ferramentas muito importantes para o indexador, pois vão auxiliá-lo na compreensão dos conceitos e também na escolha dos termos a serem indexados, qualificando o processo de indexação.

3 DO PROCESSO DE INDEXAÇÃO

Este estudo tem como objetivo maior possibilitar um melhor acesso às informações indexadas na base de dados SABI.

Este acesso é viabilizado através de um criterioso processo de organização e tratamento das informações registradas em algum tipo de suporte, ou seja, o processo de indexação. Isto significa afirmar que se entende este processo como um todo e estruturado em duas etapas, a representação descritiva e a representação temática. Na representação descritiva são registrados os dados como autoria, título e imprensa (local, editora e data); e na indexação temática são atribuídos os descritores extraído de um vocabulário controlado, identificando a categoria que mais adequadamente expressa o assunto genérico do texto (IRVING, 1997).

Considerou-se, para fins deste trabalho, a indexação apenas como o processo de determinação de assuntos de um documento e a sua tradução para uma linguagem documentária ou vocabulário controlado.

Partindo dessa concepção, entende-se que indexação é o processo pelo qual é determinado o tema principal, ou assunto, e os subtemas, ou assuntos secundários, tratados em um documento e, posteriormente, traduzidos para uma linguagem documentária. Entende-se que linguagem documentária é um sistema de signos estruturados, cuja finalidade é representar e recuperar as informações registradas nos documentos (VAN DER LAAN, 2002).

4 TERMINOLOGIA

Para Currás (1995, p.23) Terminologia é um “[. . .] processo ontológico pelo qual se chega à normalização dos sistemas formados pelas diferentes linguagens especializadas, de onde o conceito fixado pelo termo está em relação semântica com o resto dos termos desse sistema.” A autora conclui dizendo que: “A finalidade da terminologia se resume no melhor entendimento dos especialistas entre si e destes com o seu meio.”

Segundo Cabré (1993, p.43) “[. . .] a Terminologia nasce da necessidade manifestada pelos especialistas de organizar as denominações de seus sistemas de conceitos, com a finalidade de conseguir uma comunicação profissional mais confiável.” (tradução nossa).

Um dos aspectos postulados pela Terminologia clássica é a normalização dos termos, fixando o uso de um termo e descartando a utilização de outros para o mesmo conceito. A padronização do vocabulário técnico e científico, segundo a Teoria Geral da Terminologia melhoraria a comunicação entre os especialistas da área.

Esse tem sido um dos objetivos das Ciências da Informação, considerando que a linguagem é o veículo para a representação e recuperação de informações. Fato que não encontra respaldo nem mesmo no próprio discurso dos especialistas da área.

A falta de consolidação da Terminologia da área de Ciências da Informação, conforme Smit et al. (2004), tem influenciado a criação de muitos termos para referir as mesmas idéias ou ainda, mesmos termos são empregados para designar conceitos diversos.

Essa padronização do vocabulário contrapõe-se à dinâmica da língua. As normas são estáticas e a língua é dinâmica. O falante está constantemente criando novos termos, alterando os padrões. Portanto, a padronização terminológica especializada não é garantia de uma relação comunicativa bem sucedida. Isso ocorre, pois em situações profissionais reais há uma multiplicidade de registros que mantêm seu caráter especializado, mas apresentam características de unidades lexicais manifestadas em outras situações discursivas.

Cabré (1999, p.118) afirma que:

Esta observação nos leva a supor que as unidades terminológicas compartilham muitos elementos com outras unidades da linguagem natural e de outros sistemas simbólicos não lingüísticos e, para tanto, abre a possibilidade de generalizar sua descrição através de uma teoria de base menos restrita. (tradução nossa).

Nessa perspectiva, os termos podem fazer parte dos signos da linguagem natural, integrando-se ao acervo lingüístico do falante através de um aprendizado especializado.

Partindo dessas reflexões, Cabré (1999) defende uma teoria que possa dar conta de toda essa complexidade:

A Terminologia, vista a partir de uma teoria lingüística não redutiva que inclua a competência e a atuação dos falantes contemplados em sua heterogeneidade cognitiva e comunicativa, deve propor uma teoria que ao mesmo tempo dê conta dos fenômenos da linguagem geral, descreva as especificidades cognitivas, lingüísticas (gramaticais, pragmáticas, textuais e discursivas) e comunicativas das unidades terminológicas, e explique como o falante-especialista adquire estas especificidades e utiliza estas unidades. (CABRÉ, 1999, p.119, tradução nossa).

A autora continua afirmando que com esta teoria busca-se explicar os termos como unidades singulares, às vezes, similares a outras unidades de comunicação, admitindo variação conceitual e denominativa, considerando a dimensão textual e discursiva dos termos (CABRÉ, 1999).

Com esse enfoque integrador, objetiva-se obter elementos para uma nova abordagem metodológica para avaliação dos descritores. Nessa proposta, busca-se aproximar as necessidades dos destinatários da informação especializada e sua realidade lingüística com os princípios básicos necessários à organização e tratamento dessas informações.

5 A BASE DE DADOS

Na coleta de dados através de formulários em papel havia uma série de dificuldades, que exigiam um tempo excessivo. O volume muito grande de dados dificultava tanto na análise como na guarda organizada e no armazenamento deste material. Sendo assim foi iniciado o projeto para a elaboração e a implementação de uma base de dados. Dessa forma, a coleta de dados passou a ser feita através de formulários eletrônicos que alimentam a base de dados, e relatórios que permitem a visualização das informações da base sob diversos enfoques.

A base de dados visa, então, cooperar no trabalho de consolidação dos termos e na elaboração de definições para a construção de um glossário.

Para a elaboração da base de dados optou-se por desenvolvê-la segundo o Modelo Relacional, pois este modelo possibilita estabelecer com maior precisão as relações entre os campos constituintes da ficha de coleta dos

dados terminológicos empregada na pesquisa. A base de dados foi desenvolvida a partir da ficha construída originalmente para coleta de dados em papel.

A utilização do Modelo Relacional, com suas técnicas de modelagem, resulta na elaboração de uma base de dados íntegra e confiável, a medida que o desenvolvimento do modelo de dados passa por um processo de normalização, e inconsistências são evitadas com estabelecimento de restrições de integridade, que permitem o controle da redundância da informação, assim como mantêm consistentes os relacionamentos estabelecidos entre as entidades (DATE, 2000).

Esse modelo é baseado nas entidades e relacionamentos. As entidades são aqueles dados perceptíveis, ou seja, aqueles dados que serão inseridos na base; já os relacionamentos são as ligações entre as entidades, ficarão aparentes quando da manipulação dos dados (DATE, 2000). Ainda Date (2000, p. 10) afirma que “[...] o ponto significativo sobre [...] relacionamento [...] é que eles fazem parte dos dados tanto quanto as entidades básicas.”

A base de dados está estruturada nas seguintes entidades: compilador, fonte, termo, definição, contexto, marcas semânticas e definição para a elaboração do glossário. Conforme representação da figura a seguir.

Na entidade compilador registra-se as informações relativas à pessoa que coletou os dados; na entidade fonte, a referência dos artigos constituintes do corpus textual; na entidade termo são registrados o termo identificado e as variantes; na entidade contexto, o contexto que aquele termo está inserido; na entidade definição, as definições dadas pelo autor para aquele termo; na entidade marcas semânticas, as marcas semânticas que são os traços característicos do conceito, que por sua vez serão a base para a elaboração da definição do glossário.

Como a base de dados está sendo utilizada como ferramenta auxiliar para a pesquisa Organização e Tratamento da Informação: uma avaliação crítica de sua terminologia, coordenada pela professora Martha Bonotto, verificou-se a necessidade da criação de um subcampo ao qual denominou-se de macrodescritores na entidade termo. Esses macrodescritores irão possibilitar a recuperação dos termos relativos especificamente a subárea temática ao qual se referem. Inicialmente os macrodescritores estão sendo definidos de acordo com as subÁreas do currículo do Curso de Biblioteconomia da UFRGS. Posteriormente, serão incluídos os macrodescritores relativos às demais Áreas constituintes das Ciências da Informação.

A base ainda está em fase de testagem, mas já foi possível perceber algumas vantagens tais como a facilidade de coleta de dados, devido ao expressi-

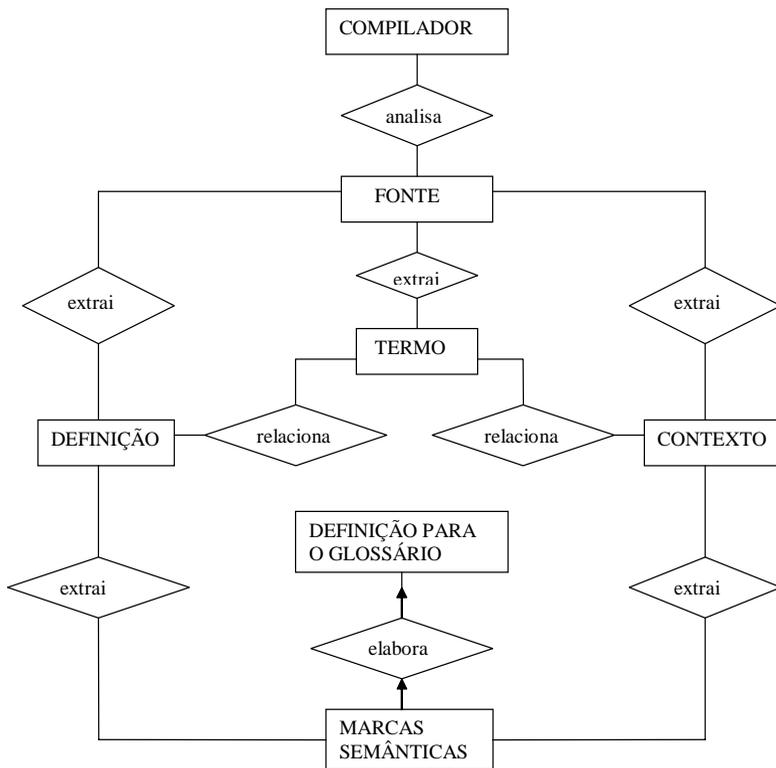


Figura - Diagrama das Entidades e Relacionamentos da Base
 Fonte: Elaborado pelos autores.

vo número de periódicos disponibilizados na Internet, eliminando o processo de digitação e, conseqüentemente, agilizando o tratamento terminológico.

Data-base as a Facilitador Element for the Treatment of Terminology Data.

ABSTRACT

A report of the design and implementation of a database developed as an auxiliary tool for the treatment of the data collected for the research - Evaluation of the Relative Describers Sciences of the Information used in the Process of Alphabetical Indexation for the Biblioteca Setorial da Faculdade de Biblioteconomia e Comunicação/UFRGS. This evaluation is being developed based on the Teoría Comunicativa de la Terminología. In this theory the term is analyzed from the perspective of being a terminological lexical unit, its identification being necessary in a discourse situation, i.e., functioning. The database aims to speed up and to facilitate the work in the consolidation of the terms and the elaboration of definitions for the construction of a glossary. This will be constructed with the purpose to become a criterion appraiser of the relative describers in the area of Information Sciences. The database was developed according to a relational model that makes it possible to establish with bigger precision the relations between the constituent fields of the card for collecting the terminology data used in the research. The base is structured upon the following fields: compiler, source, term, definition, context, semantic marks and definition for the elaboration of the glossary. It points out that this tool can be used in other researches in terminology as well as for the construction of thesauri.

KEYWORDS: Database. Terminology. Alphabetical Indexing.

REFERÊNCIAS

- CABRÉ, M. T. **La Terminología:** teoría, metodología, aplicaciones. Barcelona: Antártida/Empúries, 1993.
- CABRÉ, M. T. **La Terminología:** representación y comunicación. Elementos para una teoría de base comunicativa y otros artículos. Barcelona: IULA: Universitat Pompeu Fabra, 1999.
- CURRÁS, E. **Tesauros:** linguagens terminológicas. Brasília, DF: IBICT, 1995.
- DATE, C. J. **Introdução a Sistemas de Banco de Dados.** Rio de Janeiro: Campus, 2000.
- IRVING, Holly Berry. Computer Assisted Indexing Training and Electronic Text

Conversion at NAL. **Knowledge Organization**, Frankfurt, v. 24, n. 1, p. 4-7, 1997.

SMIT, J. W.; TÁLAMO, M. F. G. M.; KOBASHI, N. Y. A Determinação do Campo Científico da Ciências da Informação: uma abordagem terminológica. **Data Grama Zero: Revista de Ciências da Informação**, Rio de Janeiro, v. 5, n. 1, fev.2004.

Disponível em: <<http://www.dgz.org.br>> Acesso em: 2 mar. 2004

VAN DER LAAN, R. H. **Tesouro e Terminologia**: uma inter-relação lógica. 2002. Tese (Doutorado)– Universidade Federal do Rio Grande do Sul. Instituto de Letras. Programa de Pós-Graduação em Letras. Porto Alegre, 2002. Orientador: Maria da Graça Krieger. Disponível em:

<<http://www.biblioteca.ufrgs.br/bibliotecadigital/2002-2/tese-bscsh-0339228.pdf>> Acesso em: 20 set. 2004.

Adriana Gonçalves Xavier

Graduanda do Curso de Biblioteconomia da Universidade Federal do Rio Grande do Sul.

Bolsista voluntária.

E-mail: adriana.xavier@ufrgs.br

Ananda Feix Ribeiro

Graduanda do Curso de Biblioteconomia da Universidade Federal do Rio Grande do Sul.

Bolsista voluntária.

E-mail: anafery@ig.com.br

Laís Rosa dos Santos

Graduanda do Curso de Biblioteconomia da Universidade Federal do Rio Grande do Sul.

Bolsista de Iniciação Científica.

E-mail: lais.rosa@click21.com.br

Rafael Port da Rocha

Doutor em Ciência da Computação pela Universidade Federal do Rio Grande do Sul.

Professor Adjunto do Curso de Biblioteconomia da Faculdade de Biblioteconomia e Comunicação da Universidade Federal do Rio Grande do Sul.

E-mail: rafael.rocha@ufrgs.br

Regina Helena van der Laan, CRB-10/514

Doutora em Letras pela Universidade Federal do Rio Grande do Sul. Professora Adjunta do Curso de Biblioteconomia da Faculdade de Biblioteconomia e Comunicação da Universidade Federal do Rio Grande do Sul.

E-mail: rhvdl@ufrgs.br