

Avaliação de Descritores Relativos às Ciências da Informação: relato de pesquisa

Regina Helena van der Laan
Glória Isabel Sattamini Ferreira
Martha E. K. Kling Bonotto
Iara Conceição Bitencourt Neves
Inês M. de Gasperin

RESUMO

Relato de pesquisa que apresenta a proposta do uso da Teoria Comunicativa da Terminologia na metodologia do controle do vocabulário de indexação. A avaliação qualitativa deste vocabulário teve como base a seleção de descritores em uma lista constituída de 2.129 descritores. O corpus da coleta de descritores constituiu-se de artigos de periódicos brasileiros em Ciência da Informação, publicados a partir de 1995, sob a responsabilidade de programas de Pós-graduação. Evidenciou-se nos resultados o uso de descritores com diferenças de sentido entre as unidades de indexação, a falta de uma rede de remissivas e a necessidade de emprego de qualificadores para especificar a área do assunto do termo.

PALAVRAS-CHAVE: Ciências da Informação. Terminologia. Indexação. Descritores.

1 INTRODUÇÃO

O processo de indexação reveste-se de grande complexidade e envolve várias etapas. A etapa que consiste na representação alfabética dos conceitos/ idéias expressas nos textos (atribuição de descritores) é de grande importância, pois os termos são a via de acesso mais utilizada pelo usuário para a busca de informações. Assim, se não houver sintonia entre os termos utilizados pelo indexador para representar e os termos utilizados pelo usuário para buscar, as informações que esses termos representam ficarão irremediavelmente perdidas.

Nessa perspectiva, estão sendo avaliados os descritores empregados no processo de representação de informações na área de Ciências da Informação (CI) do Sistema de Automação de Bibliotecas (SABi) da UFRGS. Essa avaliação baseia-se em observações informais nas quais foram percebidas divergências entre as unidades de indexação e as expressões de busca empregadas pelos usuários, geralmente resultando no insucesso dessas pesquisas. Essas divergências terminológicas refletem-se no processo de busca de informações uma vez que o usuário irá empregar os termos de seu acervo lingüístico, nem sempre coincidente com o vocabulário empregado pelo Sistema de Recuperação da Informação (SRI).

A avaliação que está sendo realizada baseia-se na compreensão de que as unidades indexadoras têm valor de termo. Uma outra problemática refere-se ao surgimento de novos conceitos na área de CI cujos termos se encontram, ainda, em fase de consolidação. Conseqüentemente, não foram ainda integralmente apropriados pelos especialistas que, ora utilizam uma terminologia, ora outra para referir-se ao mesmo conceito.

Outra questão que pode dificultar o processo de comunicação entre os dois interlocutores, ou seja, o usuário e o SRI, refere-se à idéia de que cada conceito de um domínio do conhecimento é representado por um único termo indexador. Isso ocorre uma vez que os atuais descritores estão sendo determinados em conformidade com o princípio de univocidade, por isso caracterizados pela monorreferencialidade, ou seja, apenas um código é empregado para representar cada conceito. Dessa forma, perde-se toda a riqueza vocabular que existe, mesmo se tratando de léxico especializado. Deve-se ainda observar que, mesmo ao tratar de uma linguagem de especialidade própria de um grupo de especialistas, há uma dinâmica permanente de criação de novos termos, ressignificação de termos consagrados e variações denominativas de conceitos.

Como atualmente apenas a unidade de indexação autorizada é registrada na base, torna-se a única possibilidade de acesso às informações; se algum usuário verbalizar o mesmo conteúdo utilizando-se de uma outra expressão, diferente da autorizada pelo sistema, não obterá sucesso em suas buscas. Isso ocorre porque a base SABi, ainda não disponibiliza um sistema de remissivas de assuntos.

O exposto evidencia a problemática do vocabulário utilizado para representar as informações em um SRI e a necessidade de uma nova abordagem para o seu estabelecimento, de modo a tornar o sistema mais eficiente e eficaz

2 CAMINHADA TEÓRICA

O propósito da elaboração de índices é possibilitar a recuperação das informações contidas no documento a ser indexado. A indexação temática pode ser definida como o processo pelo qual é determinado o tema principal ou assunto e os assuntos secundários tratados no documento e que devem ser, posteriormente, traduzidos para uma linguagem documental.

Aqui cabe ressaltar as diferentes representações que os autores empregam para referir o conceito de linguagem documental, evidenciando dessa forma que, mesmo entre os especialistas da área de Ciências da Informação, o princípio de univocidade, caracterizado pela monorreferencialidade não se confirma, uma vez que termos diferentes como: Linguagem Documental, Linguagem Documentária e Linguagem de Indexação são encontrados com uma única acepção (VAN DER LAAN, 2004).

Pode-se definir esse conceito, conforme Van Slype (1991), como um sistema de signos que possibilitam representar o conteúdo dos documentos, tendo como finalidade sua recuperação.

Entre as diversas linguagens documentárias destaca-se o tesouro que, segundo Affonso (1987, p. 1):

[. . .] em termos de função, é um instrumento de controle terminológico usado para traduzir a linguagem natural dos documentos, dos indexadores e dos usuários em linguagem mais restrita que é denominada linguagem de sistema, linguagem documentária ou linguagem de indexação, bem como para verter essa linguagem em linguagem natural.

O tesouro tem por finalidade padronizar os termos usados na indexação e na busca das informações, de forma consistente e com bom nível de especificidade. Na construção de um tesouro é estabelecido um controle do vocabulário, visando que cada conceito seja expresso por um descritor.

Com o propósito de estabelecer o descritor preferido para representar cada conceito é necessário que os termos sejam coletados no discurso dos especialistas em textos da área. A seguir, é recomendável que se faça um cotejo com outras fontes, tais como tesouros da mesma área ou área afim, dicionários, vocabulários, esquemas de classificação e índices de publicações periódicas.

Na elaboração de tesouros ainda são estabelecidas relações entre os conceitos e entre os termos. As relações entre os termos estabelecem-se por equivalência e determinam a relação entre palavras sinônimas, quase sinônimas ou indicam a preferência de grafia. As relações entre os conceitos podem ser de tipo hierárquico, que indicam relações gênero/espécie ou todo/

parte e as associativas que estabelecem uma associação entre os conceitos. O estabelecimento dessas relações associativas tem sua importância evidenciada, uma vez que irá indicar ao usuário e ao bibliotecário a existência de assuntos relacionados (VAN DER LAAN, 2004).

A relação que se estabelece entre os usuários e o SRI é uma relação de comunicação. Essa comunicação muitas vezes é prejudicada devido à utilização de termos que não possuem o mesmo significado entre os interlocutores; por outro lado, existe também a utilização de termos diferentes para significados iguais – variações terminológicas.

A comunicação se faz através da linguagem, escrita, falada ou simbólica. Os signos utilizados devem fazer sentido, tanto para o emissor como para o receptor da mensagem. Na área científica é fundamental que seja utilizada uma terminologia compreensível entre os especialistas da área.

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.”

O reconhecimento da importância da Terminologia está relacionado à preocupação dos cientistas, dos séculos XVIII e XIX, com a proliferação de termos que dificultavam a comunicação entre os especialistas. O grande desenvolvimento científico e tecnológico ocorrido, principalmente a partir do final do século XIX, provocou o surgimento de novos conceitos e novas áreas de especialização e, conseqüentemente, de novos termos. Nesse contexto, emerge a problemática das negociações comerciais, dificultadas pela utilização de nomenclaturas diferentes pelos diversos países. Dessa forma, a necessidade de entendimentos comerciais em nível internacional determina o surgimento de padrões industriais e, conseqüentemente, de suas nomenclaturas (CABRÉ, 1993).

Essa problemática se amplia nos dias atuais com o surgimento dos grandes blocos econômicos, que vêm alterando as relações comerciais internacionais e mesmo toda uma sorte de intercâmbios, forçando a compatibilização de termos técnicos, comerciais e científicos entre as diversas línguas.

Nessa perspectiva, o processo de normalização e padronização dos termos propostos pela Terminologia Clássica, em princípio, propiciaria uma comunicação mais viável entre os especialistas, uma vez que todos utilizariam o mesmo signo lingüístico para transmitir suas descobertas e idéias em uma determinada área do conhecimento. O que, de acordo com os seguido-

res da Teoria Geral da Terminologia (TGT), propiciaria um melhor entendimento dos especialistas entre si e desses com o seu meio, além de uma considerável melhora nas comunicações comerciais.

Um dos aspectos postulados pela TGT é a normalização dos termos, fixando o uso de um termo e descartando a utilização de outros termos para o mesmo conceito. Os defensores da TGT afirmam que a padronização do vocabulário técnico e científico contribui para univocidade da comunicação entre os especialistas de um domínio de conhecimento.

Conforme Wüster (1998), a diferença entre a TGT e a Linguística fundamenta-se na atitude que cada uma das áreas adota em relação ao estudo e à evolução da língua. Para o autor, três elementos básicos determinam essas diferenças. Primeiro para a Terminologia Clássica o ponto de partida são os conceitos e sua denominação. O significado do termo é denotativo, prescindindo das conotações. Os conceitos são representados por um único termo de forma unívoca e caracterizado pela monorreferencialidade – um único referente. Outro ponto refere-se ao fato de que na concepção wüsteriana, os terminólogos só se preocupam com o léxico, não há interesse na morfologia flexiva, nem na sintaxe, cujas regras se depreendem da língua geral. O último ponto assinalado pelo autor é que, para a investigação terminológica, se considera a língua do ponto de vista sincrônico. Mas esse ideal não se confirma na realidade, uma vez que o vocabulário utilizado pelo especialista não se caracteriza pela univocidade. Verifica-se que os próprios pesquisadores utilizam vocabulários de acordo com seus grupos de pesquisa, não existindo padronização nem mesmo em uma mesma área do conhecimento.

Essas considerações sobre a TGT nos permitem estabelecer uma aproximação com os tesouros, cujas metodologias de elaboração também partem dos mesmos princípios de padronização do vocabulário que será utilizado no processo de indexação. Considerações essas respaldadas por Gomes (1990) ao afirmar que os tesouros podem se beneficiar dos fundamentos da TGT.

As linguagens de indexação ou vocabulários controlados têm como finalidade ou razão de ser, recuperar rápida e facilmente os documentos que contêm informações pertinentes a um usuário.

A mediação entre quem deseja informação e quem a tem é um complexo processo comunicativo duplamente codificado. O bibliotecário codifica a informação utilizando uma linguagem de indexação. Nesse processo interferem:

- a) a política de indexação adotada pelo serviço;
- b) os conhecimentos do próprio indexador;
- c) os recursos disponíveis para as estratégias de busca nos sistemas

informatizados;
d) a linguagem de indexação.

O usuário utiliza seu código lingüístico e seu conhecimento da área para construir as questões de busca de informação. Nessa dupla codificação, é necessário que ambos os códigos utilizados sejam coincidentes para o sucesso nas buscas de informação.

Esse fato evidencia a importância do vocabulário utilizado no processo de representação/recuperação das informações nos diferentes SRI. Nessa perspectiva, está sendo avaliada a representatividade dos descritores relativos à área de Ciências da Informação, empregados no processo de indexação pela Biblioteca Setorial da Faculdade de Biblioteconomia e Comunicação (FABICO) da Universidade Federal do Rio Grande do Sul (UFRGS) e que alimentam a base SABI.

Um tesouro é um vocabulário controlado de um domínio específico do conhecimento, sendo que sua organização busca evidenciar as relações conceituais dessa área de especialidade. Dessa forma, entende-se que as unidades lexicais registradas nesses instrumentos de indexação deveriam ser constituídas em conformidade com a terminologia desse domínio específico do conhecimento. Além do que, essas unidades lexicais, mesmo pertencendo a uma linguagem de especialidade, comportam sinonímia e variação, evidenciando-se, assim, a importância do controle do vocabulário (VAN DER LAAN, 2004).

A padronização terminológica especializada não é garantia de uma relação comunicativa bem sucedida. Cabré (1999, p.118, tradução nossa) 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, portanto, abre a possibilidade de generalizar sua descrição através de uma teoria de base menos restrita.

Dessa forma, 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, p.119, tradução nossa) defende uma teoria que possa dar conta de toda essa complexidade.

A terminologia, vista à luz 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.

A mesma autora afirma 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, em conformidade com a dimensão textual e discursiva dos termos.

Dessa forma, este estudo fundamenta-se na interface entre as chamadas Ciências da Informação, mais especificamente a Arquivologia e a Biblioteconomia, e a Teoria Comunicativa da Terminologia (TCT).

3 NOSSO PERCURSO

Diante do exposto, evidenciou-se a importância de uma avaliação qualitativa com uma interface com a Teoria Comunicativa da Terminologia que, no nosso entender, melhor dá conta de toda a problemática da comunicação especializada.

Assim, nosso objeto de pesquisa é constituído pelos descritores relativos à área de Ciências da Informação, empregados no processo de indexação do sistema SAbi da Biblioteca Setorial da FABICO/UFRGS.

Conforme anteriormente mencionado, esta pesquisa encontra sua justificativa em observações informais, nas quais se evidenciou que, de modo geral, os descritores empregados como unidades de indexação não obtinham respostas adequadas; também foi detectada a existência de divergências conceituais entre os descritores empregados e a terminologia da área, que decorriam de falhas de comunicação entre o usuário e o sistema, resultando em um significativo grau de silêncio ou ruído no momento recuperação de informações.

Nosso objetivo geral é avaliar a representatividade dos descritores de Ciências da Informação empregados no processo de indexação na Biblioteca da FABICO em relação à terminologia desta área.

Inicialmente, solicitamos à Comissão Automação de Bibliotecas da UFRGS a elaboração de uma listagem dos descritores relativos à Biblioteca da Faculdade de Biblioteconomia e Comunicação. Cabe aqui um esclarecimento: o acervo dessa Biblioteca é composto por itens das áreas de Jornalis-

mo, Relações Públicas; Publicidade e Propaganda, Biblioteconomia e Arquivologia. Como nosso interesse restringia-se somente aos descritores da área de Ciências da Informação, foi organizada uma relação de termos que não deveriam constar nessa listagem, tais como: jornal; jornalismo; semiótica; relações públicas, entre outros. Dessa forma foi obtida uma relação de 4.503 descritores. Essa lista foi analisada pela equipe de pesquisa, resultando numa listagem de 2.129 descritores que foram considerados específicos da área de Ciências da Informação.

Ficou estabelecido que o corpus de coleta dos descritores seria constituído por artigos de periódicos em Ciências da Informação, brasileiros, publicados a partir de 1995, sob a responsabilidade de programas de pós-graduação, pertencentes ao acervo da Biblioteca da FABICO ou disponibilizados em rede.

Foi elaborado um instrumento de coleta de dados, estruturado em campos que possibilitasse o registro de informações pertinentes para a avaliação dos descritores.

Com a finalidade de verificar a representatividade dos descritores, está sendo realizado um levantamento em bases de dados de outras instituições. Além disso, existe uma interface com o sistema SABI, onde os descritores são constantemente testados.

4 ALGUNS DADOS PRELIMINARES

Com os dados obtidos até o presente momento, podemos perceber alguns problemas na determinação dos descritores, tais como diferenças de sentido entre as unidades de indexação e os termos expressos nos textos pelos especialistas. Devido à problemática da falta de consolidação da terminologia da área, referida por Smit, Tálamo e Kobashi (2004), há descritores que se referem ao mesmo conceito empregado com sentidos diferentes.

Outro problema observado foi a falta da rede de remissivas, o que impede o acesso às informações por variantes terminológicas. Essa problemática compromete a relação comunicativa que se estabelece entre o usuário e o sistema. Foi ainda evidenciada a necessidade do emprego de qualificadores, pois há uma extensa terminologia que é empregada em outras áreas, como por exemplo, o termo “classificação” que, além da área de Biblioteconomia, é encontrado também na área de Agronomia, Geologia, Medicina, Psicologia. Essa problemática ficou mais evidente pelo fato de o usuário, de forma geral, não selecionar a unidade de informação (biblioteca setorial) no momento de sua busca. Isso resulta em um número excessivo de documentos não pertinentes às necessidades do usuário.

Salientamos, ainda, que esta pesquisa estabelece uma inter-relação com o trabalho desenvolvido pelas professoras da área de Organização e Tratamento da Informação do Departamento de Ciências da Informação (DCI) da UFRGS que, no presente momento, estão estruturando um glossário da terminologia específica empregada nessa área, com a finalidade de publicação de um livro que servirá de referência em português para as disciplinas cujos conteúdos estão relacionados a esta temática. Além do fato de que pesquisa, devido a suas características, poderá tornar-se um recurso didático, articulando, dessa forma, a pesquisa com o ensino e a extensão em nível de graduação.

Por outro lado, acreditamos que a metodologia utilizada nesta pesquisa poderá ser utilizada como referência pelo o Sistema de Bibliotecas dessa Universidade em estudos similares.

Evaluation of Descriptors used for Information Science: research report

ABSTRACT

Research report that proposes an application of the communicative theory of terminology to a methodology for vocabulary control for indexing. The qualitative evaluation of this vocabulary was based on the selection of descriptors from a list of 2,129 descriptors. The corpus from which descriptors were collected consisted of articles published in Brazilian periodicals in Information Science, since 1995, sponsored by Post-graduate study programs. It became evident in the results that the same descriptors are being used with different meanings among indexing units; also that there is a lack of a net cross references and the need for qualifiers to specify the field of study of the term.

KEYWORDS: Information Science. Terminology. Indexing. Descriptors.

REFERÊNCIAS

AFFONSO, Lais da Boa Morte Ferreira. **Metodologia para Construção do Tesouro de Informática em Língua Portuguesa.** Rio de Janeiro: SERPRO, 1987.

CABRÉ, Maria Teresa. **La Terminología:** representación y comunicación. Elementos para una teoría de base comunicativa y otros artículos. Barcelona: IULA: Universitat Pompeu Fabra, 1999.

CABRÉ, Maria Teresa. **La Terminología:** teoría, metodología, aplicaciones. Barcelona: Antártida/Empúries, 1993.

CURRÁS, Emília. **Tesauros:** linguagens terminológicas. Brasília, DF: IBICT, 1995.

GOMES, Hagar Espanha (Coord.). **Manual de Elaboração de Tesouros Monolíngües**. Brasília, DF: Programa Nacional de Bibliotecas das Instituições de Ensino Superior, 1990.

SMIT, Johanna W. ; TÁLAMO, Maria de Fátima G. M.; KOBASHI, Nair Y. A Determinação do Campo Científico da Ciência da Informação: uma abordagem terminológica. **DataGramaZero**- Revista de Ciência da Informação, Rio de Janeiro, v. 5, n. 1, artigo 3. Disponível em <http://www.dgzero.org/fev04/F_I_art.htm>. Acesso em: 2 mar. 2004.

VAN DER LAAN, Regina Helena. **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. Orientada por: Maria da Graça Krieger. Disponível em: <<http://www.biblioteca.ufrgs.br/bibliotecadigital/2002-2/tese-bscsh-0339228.pdf>> Acesso em: 20 set. 2004.

VAN SLYPE, Georges. **Los Lenguajes de indezación: concepción, construcción y utilización en los sistemas documentales**. Madrid: Fundación Germán Sánchez Ruipérez, 1991.

WÜSTER, Eugen. **Introducción a la teoría general de la terminología y a la lexicografía terminológica**. Barcelona: IULA: Universitat Pompeu Fabra, 1998.

Regina Helena van der Laan, CRB-10/514

*Doutora em Letras pela Universidade Federal do Rio Grande do Sul.
Professora Adjunta do Departamento de Ciências da Informação
da Faculdade de Biblioteconomia e Comunicação da Universidade
Federal do Rio Grande do Sul.
E-mail: rhvdl@ufrgs.br*

Glória Isabel Sattamini Ferreira, CRB-10/176

*Mestre em Educação pela Pontifícia Universidade Católica do Rio
Grande do Sul.
Professora Assistente do Departamento de Ciências da Informação
da Faculdade de Biblioteconomia e Comunicação da Universidade
Federal do Rio Grande do Sul.
E-mail: gloria.ferreira@ufrgs.br*

Martha E. K. Kling Bonotto, CRB-10/755

*Mestre em Letras pela Universidade Federal do Rio Grande do Sul.
Professora Assistente do Departamento de Ciências da Informação
da Faculdade de Biblioteconomia e Comunicação da Universidade
Federal do Rio Grande do Sul.
E-mail: marthakbonotto@yahoo.com.br;
martha.bonotto@ufrgs.br*

Iara Conceição Bitencourt Neves, CRB-10/351

Doutora em Ciências da Comunicação pela Universidade de São Paulo.

Professora Adjunta do Departamento de Ciências da Informação da Faculdade de Biblioteconomia e Comunicação da Universidade Federal do Rio Grande do Sul.

E-mail: iara.neves@ufrgs.br

Inês M. de Gasperin, CRB-10/733

Especialista em Automação de Bibliotecas e Centros de Informações Documentárias pela Universidade Federal de Pernambuco.

Bibliotecária da Biblioteca da Faculdade de Biblioteconomia e Comunicação da Universidade Federal do Rio Grande do Sul, responsável pelo Setor de Processamento Técnico.

E-mail: inesmaria.degasperin@gmail.com