

ANÁLISE DA CAPACIDADE DE CODIFICAÇÃO DA INFORMAÇÃO DE VOCABULÁRIO CONTROLADO NA ÁREA DA SAÚDE¹

Cibele Araújo Camargo Marques dos Santos²

RESUMO

Análise da linguagem documentária da Base AdSAÚDE, que contém informação bibliográfica sobre temas vinculados à administração em saúde. A linguagem foi analisada do ponto de vista estrutural, do vocabulário e do seu uso na indexação e recuperação. Os dados obtidos demonstraram que as linguagens requerem administração sistemática para terem garantida a sua capacidade de codificação de informação, devendo-se estabelecer parâmetros que operem em dois níveis: 1) no nível da gestão da linguagem, que requer a elaboração de diretrizes para avaliação e revisão, baseadas em estudos de ocorrência de termos e comunicação das atualizações para o sistema; 2) no nível do controle terminológico, que requer a análise da adequação e pertinência da terminologia utilizada.

Palavras-chave: Linguagens documentárias, Vocabulário Controlado; Gerenciamento, Revisão, Administração em Saúde.

1. INTRODUÇÃO

Os Sistemas de Informação atuais operam de forma cooperativa e em rede estabelecendo parcerias entre diversas bibliotecas e centros de informação para desenvolver bases de dados, tanto bibliográficas como de texto completo. Procura-se,

¹ Baseado na dissertação de mestrado defendida em 16/09/2002 na ECA/USP, intitulada 'Linguagens documentárias e codificação da informação: estudo de vocabulário da área da saúde' orientada pela Professora Doutora Nair Yumiko Kobashi.

² Professora da Faculdade de Biblioteconomia e Ciência da Informação da Fundação Escola de Sociologia e Política e bibliotecária da Faculdade de Saúde Pública da USP. E.mail: cibelear@usp.br.

com essas iniciativas, solucionar simultaneamente a questão da cobertura da literatura científica produzida e a demanda crescente de informação.

Um exemplo de sistema cooperativo é o Sistema Latino-Americano e do Caribe de Informação em Ciências da Saúde - Sistema BIREME/OPAS/ OMS³. Nos sistemas de informação em geral, como também nos sistemas cooperativos, a utilização de uma linguagem documentária bem construída é fundamental para o tratamento e a recuperação de documentos relevantes.

A base AdSAÚDE, é uma base de dados na área de administração em saúde, integrada ao Sistema BIREME, desenvolvida pela Biblioteca/CIR da Faculdade de Saúde Pública da USP, em parceria com outras instituições. A base disponibiliza informação considerada necessária para a tomada de decisão nas instâncias políticas, sistêmicas e de serviços de saúde, visando à melhoria da qualidade da assistência à saúde e qualidade de vida da população brasileira. São usuários da AdSAÚDE os pesquisadores e técnicos dos serviços de saúde, hospitais, secretarias de saúde, ministérios e institutos de pesquisa.

O objeto de estudo desta pesquisa foi a linguagem documentária utilizada para a representação de conteúdo na área de saúde, especificamente na temática da administração em saúde, visando avaliar a sua capacidade de codificação de informação. Constatou-se que a terminologia do Vocabulário DeCS na área de Saúde Pública, apesar de todos os avanços alcançados, apresenta problemas que comprometem o tratamento e a recuperação de informação. Considerou-se que os problemas poderiam estar na estrutura hierárquica e na terminologia utilizada.

³ O Sistema BIREME desenvolve atualmente a Biblioteca Virtual em Saúde – BVS, integrada por diversas bibliotecas virtuais temáticas (Adolescência, Ambiente, Saúde Pública) e BVS regionais (Brasil, Colômbia, Costa Rica, Cuba entre outras) sendo que todas essas fontes utilizam o Vocabulário DeCS para indexação. Para a produção das bases de dados, o Sistema BIREME dispõe da Metodologia LILACS. A representação de conteúdo dos

Antes da apresentação dos resultados da pesquisa, serão caracterizados os conceitos fundamentais utilizados no trabalho, a saber: linguagem documentária e avaliação de linguagem documentária.

2. LINGUAGENS DOCUMENTÁRIAS: NATUREZA E FUNÇÕES

Uma linguagem documentária é um instrumento utilizado por um centro ou unidade de informação para a representação do conteúdo dos documentos (GUINCHAT E MENO, 1994). Além da função de representação, as linguagens documentárias atuam na recuperação da informação (TÁLAMO e col., 1992).

O uso das linguagens documentárias supõe, como qualquer linguagem, a observância de um conjunto de regras (CINTRA, 1983). Por ser linguagem, a linguagem documentária é composta de unidades signicas. O signo lingüístico, segundo a visão saussureana, é constituído de um conceito (significado) e de imagem acústica (significante). Nesta concepção, dois princípios básicos orientam a organização das línguas: o da *“arbitrariedade do signo e o do caráter linear do significante. A arbitrariedade do signo é responsável por problemas semânticos, como a polissemia, a sinonímia”* (CINTRA, 1983 p.8). A relação significado/significante é imotivada e se estabelece sob a forma de contrato social. Não pode, portanto, ser arbitrariamente alterada pelo indivíduo.

As linguagens naturais apresentam características que as tornam de difícil utilização para o tratamento da informação. Segundo CINTRA (1983), os signos lingüísticos, por representarem coisas através da relação arbitrária entre significante e significado, alimentam a polissemia da linguagem natural.

As linguagens documentárias, por outro lado, devem ter os signos com significados recortados e fixados para evitar a polissemia. Devem, além disso, ter os significados partilhados com a comunidade de usuários do sistema informacional. Com efeito, os descritores devem ter significado invariante para representar informações de maneira previsível. Estabelecer o significado nas linguagens documentárias é, desse modo, uma atividade complexa, pois reconceituar ou delimitar termos oriundos da linguagem natural, com sua pluralidade de significados, implica uma escolha entre várias possibilidades, que podem ser influenciadas por questões relacionadas a grupos sociais, regiões e países. Por ser uma linguagem constituída de parâmetros estáveis e predeterminados cultural e socialmente, vinculada a características institucionais, uma linguagem documentária interfere nas formas de organização e disseminação da informação.

A construção de uma linguagem documentária requer, como primeira operação, a determinação de hipóteses de segmentação da área de especialidade em categorias (KOBASHI e TÁLAMO, 2001).

Estes conjuntos são agrupamentos mutuamente excludentes. A identificação das categorias deve ser realizada considerando-se também a literatura da área, de modo a conferir garantia literária ao conjunto (LANCASTER, 1987), e responder às necessidades de grupos específicos de usuários.

O conhecimento categorizado neste tipo de linguagem pretende “*garantir de forma classificatória e preditiva, a organização do universo objeto de representação*” além de manter a “*contigüidade e semelhança entre o vocabulário a ser utilizado para a representação e o universo do conhecimento expresso nos textos*” (TÁLAMO e col., 1992 p.197). Semelhante categorização permite que, em face de um termo novo, este possa ser facilmente incluído na categoria adequada.

Desse modo, apenas um vocabulário construído a partir de uma rede de significações lógico-semânticas, baseadas em um dicionário terminológico, com descritores definidos de forma clara poderá proporcionar o rigor necessário para a melhoria dos procedimentos de indexação (TÁLAMO e col., 1992).

A atualização de uma linguagem documentária é necessária devido às mudanças constantes na linguagem natural e na área de especialidade. GUINCHAT E MENO (1994) sugerem controlar o emprego da linguagem registrando-se as dificuldades e soluções encontradas, bem como a ocorrência dos descritores. Deve-se, além disso, registrar os termos que não aparecem na linguagem para futura decisão sobre eventual inclusão.

Para isso, deve-se definir uma política de gerenciamento da linguagem que contemple a criação de novos descritores, o estabelecimento de novas relações, a substituição de termos, a eliminação de termos inúteis, a introdução de notas para descritores ambíguos, a modificação de termos, etc. (GUINCHAT E MENO, 1994).

3. AVALIAÇÃO DE LINGUAGENS DOCUMENTÁRIAS

A análise do Vocabulário DeCS foi precedida de pesquisa da literatura sobre avaliação, análise, revisão e gerenciamento de listas de cabeçalhos de assuntos, de vocabulários controlados e tesouros. O levantamento bibliográfico foi feito nas bases de dados LISA (Library and Information Science Abstracts), MEDLINE (da National Library of Medicine) e a LILACS (Literatura Latino-Americana e do Caribe em Ciências da Saúde).

Os estudos de maior interesse para esta proposta estão apresentados de forma resumida visando dar subsídios teóricos e metodológicos aos parâmetros de análise utilizados.

Em 1989, BERMEJO e col. avaliaram 47 tesouros de língua espanhola considerando as seguintes características: área de domínio, apresentação (incluindo introdução, relação entre as listas alfabéticas e sistemáticas, hierarquias e índices), consistência (relações recíprocas, forma dos descritores, termos traduzidos) e estrutura semântica (número de termos e campos semânticos). Os resultados deste estudo mostraram falhas na normalização dos tesouros, tanto na sua construção quanto na apresentação e indicaram a necessidade de cooperação entre instituições dos países de língua espanhola na construção e aprimoramento da qualidade dos tesouros.

Em 1979, HARRIS estudou as mudanças na terminologia em Ciência da Informação durante um período de 11 anos para determinar os efeitos das citadas mudanças em vocabulário da área. Chegou a uma taxa de mudança de 4% do total de termos excluídos e incluídos. Indicou neste estudo a importância de que sejam previstas mudanças e atualizações nos tesouros com a retirada dos termos que estão em desuso e a adição dos novos termos.

Em 1985, ante a possibilidade de disponibilização da base de dados CINAHL – Nursing and Allied Health online pelo DIALOG, foi realizada uma revisão da lista de cabeçalhos de assuntos utilizada na indexação da base visando melhorar a qualidade da recuperação pelo novo sistema de acesso. Para atender às necessidades dos usuários e as mudanças ocorridas na literatura em enfermagem foram revistas tanto a estrutura quanto os cabeçalhos (FISHEL e col., 1985).

Em 1987, JANIK e BRUNET analisaram a metodologia desenvolvida pelo Quebec Office for Handicapped People para a atualização e expansão do Quebec Thesaurus: Disabled children cuja primeira versão fora publicada em 1979. Neste novo tesouro intitulado Thesaurus: Handicapped People, aproveitou-se 50% da versão original com a expansão dos termos médicos e educacionais, com base nos novos

recursos políticos e sociais. Foram introduzidos também os termos da Classificação da Organização Mundial da Saúde para Deficientes. A estrutura original foi mantida, mas a apresentação gráfica foi melhorada. A terminologia e a estrutura foram analisadas detalhadamente para direcionar a revisão do tesouro, com a eliminação, por exemplo, de termos discriminatórios ou pejorativos.

Com relação ao MESH (Medical Subject Headings), GREEN e col. analisaram termos usados para indexar a literatura sobre Síndrome de Imunodeficiência Adquirida - AIDS na base de dados AIDSLINE da N. L. M. no período de 1983-1989. Observaram que entre 1987 a 1989 foram introduzidos muitos termos sobre o tema, resultado do aumento dos estudos no período. Além disso, indicaram que o alto índice de alteração de descritores nesses seis anos reflete a incerteza destes estudos sobre uma doença nova, as diversas abordagens e direções tomadas pelas investigações biomédicas em AIDS. Por outro lado, foi possível perceber a agilidade com que a N. L. M. gerenciou a atualização do vocabulário no período.

Alguns estudos relacionados ao DeCS foram realizados em Cuba por RODRIGUES CAMIÑO em 1998 e VALDÉS ABREU em 1996. Esses estudos enfatizaram a necessidade de serem comunicadas ao Sistema as mudanças ocorridas no vocabulário. Propôs-se a criação de um veículo de comunicação e uma lista de discussão dos bibliotecários que utilizam o DeCS, bem como a revisão sistematizada da referida linguagem incluindo notas de escopo e regras de uso para os qualificadores.

4. ANÁLISE DO VOCABULÁRIO DECS: CATEGORIAS SP1 E SP2

O Vocabulário DeCS – Descritores em Ciências da Saúde é um vocabulário controlado trilingüe que contém descritores em Inglês, Português e Espanhol. Foi desenvolvido a partir da tradução do MeSH – Medical Subject Headings da U.S.

National Library of Medicine com o objetivo de permitir o uso de terminologia comum para processamento e busca nos três idiomas.

Além dos termos originais do MeSH, foram desenvolvidas áreas específicas para as necessidades do sistema, tal como a de Saúde Pública. A categoria SP – Saúde Pública criada em 1986, foi acrescentada ao DeCS visando melhor representar a literatura latino-americana sobre o tema. Foram adicionados cerca de 3.000 termos entre vocabulário ativo (efetivamente usado, termos preferenciais) e vocabulário passivo (termos antigos, histórico e remissivos), inclusive os descritores do MeSH em Saúde Pública.

O Vocabulário DeCS por ser uma tradução do MESH, apresenta características híbridas de lista de cabeçalho de assunto e tesouro. Deve-se lembrar que o MeSH e o DeCS foram se transformando em vocabulário controlado sem abandonar as estruturas de origem. Mantiveram, portanto, a estrutura hierárquica fundamentada na divisão do conhecimento em classes e subclasses denominadas categorias, respeitando as ligações conceituais e semânticas sendo os termos apresentados em uma estrutura híbrida de conceitos e assuntos.

4.1. Estrutura hierárquica e léxico das Categorias SP do DeCS

Os descritores do Vocabulário DeCS estão subordinados e coordenados dentro da estrutura básica originária do sistema nocional do MeSH.

Como a estrutura do Vocabulário DeCS está baseada na tradução para o português e o espanhol dos descritores do MESH, foi mantida a organização original em categorias A, B, C e assim por diante.

Nas áreas de especificidade em que foi necessário fazer expansões, como Saúde Pública e Homeopatia, foram abertas as novas categorias SP e HP, respectivamente.

O vocabulário DeCS é polihierárquico, sendo que vários descritores da Categoria SP estão também em outras categorias, principalmente na categoria E (Técnicas e Equipamentos), categoria F (Psicologia e Psiquiatria), categoria G (Ciências Biológicas), categoria I (Antropologia, Educação, Sociologia e Fenômenos Sociais), categoria M (Pessoas) e os descritores da categoria N, que no MESH correspondem aos descritores de Assistência à Saúde.

A área da Saúde Pública está dividida nas 6 categorias abaixo:

SP1 – ADMINISTRAÇÃO E PLANEJAMENTO EM SAÚDE

SP2 – PRESTAÇÃO DE CUIDADOS DE SAÚDE

SP3 – DEMOGRAFIA

SP4 – SAÚDE AMBIENTAL

SP5 – EPIDEMIOLOGIA E BIOESTATÍSTICA

SP6 – NUTRIÇÃO

As categorias temporárias SP7 (REFORMA DO SETOR SAÚDE) e SP8 (DESASTRES) foram criadas para inserção do Vocabulário sobre Reforma do Setor Saúde da OPAS e do Vocabulário da Base Desastres, estando prevista a compatibilização destas terminologias com o DeCS. Este último deverá absorver, nas outras 6 categorias, os termos destas áreas após avaliação por especialistas. O vocabulário sobre Desastres já foi totalmente revisto e incorporado às categorias SP, principalmente na Categoria SP4 (Saúde Ambiental), que já continha descritores ligados ao tema.

Observam-se, no entanto, alguns problemas de ordem geral:

- a) A estrutura organizada em seis grandes categorias engessa o vocabulário, forçando a inserção de termos em uma ou mais das seis categorias. Tal inserção, muitas vezes, cria relações artificiais e pouco precisas. É o caso dos termos da área do Direito, por exemplo, que estão hierarquizados na categoria SP1 - ADMINISTRAÇÃO E PLANEJAMENTO EM SAÚDE;
- b) As seis opções de conceitos genéricos, representadas pelas categorias SP, além de serem insuficientes dentro de uma área de conhecimento multidisciplinar, não explicitam de forma clara a relação estabelecida entre descritor genérico e específico na elaboração da hierarquia.
- c) Com relação ao léxico os termos do DeCS mantêm a estrutura híbrida de pré e pós-coordenação.
- d) vocabulário utiliza termos simples e compostos, sendo que muitos descritores correspondem a frases substantivas como, por exemplo, o descritor da categoria SP2 GARANTIA DE QUALIDADE DOS CUIDADOS DE SAÚDE. A utilização de frases como descritores, prática originária das listas de cabeçalho de assunto, pode gerar dificuldades tanto para os indexadores quanto na recuperação de informação.
- e) Os descritores do DeCS utilizam modificadores indicados por parênteses para aumentar a especificidade e evitar ambigüidades, como no caso dos descritores TOMADA DE DECISÃO (ADMINISTRAÇÃO), DETERMINAÇÕES PRÉVIAS (ÉTICA MÉDICA), PLANOS DE IMPLEMENTAÇÃO ANUAL (SAÚDE PÚBLICA). Esse fato, a criação de frases e expressões, torna difícil a identificação do descritor a ser utilizado na pesquisa.

- f) Por ser trilingüe o DeCS apresenta os problemas gerados por traduções, como é o caso dos termos ligados a cuidados de saúde. Existem vários termos, tais como CUIDADOS MÉDICOS, CUIDADOS PRIMÁRIOS DE SAÚDE, CUIDADOS DOMICILIARES DE SAÚDE, que vêm do termo *CARE*. A literatura no Brasil, no entanto, utiliza a terminologia Assistência Médica, Assistência Primária e Assistência domiciliar.

Portanto a estrutura hierárquica e o léxico utilizados nas categorias SP1 e SP2, apresentam problemas diversos que requerem as seguintes revisões:

- a) Verificar a possibilidade de expansão das noções básicas da área em novas categorias.
- b) Rever relações hierárquicas, associativas e coordenação entre os descritores.
- c) Revisar os conceitos de alguns descritores.
- d) Analisar descritores que correspondem a frases substantivas.
- e) Revisar utilização de qualificadores para modificar alguns descritores.
- f) Rever as traduções adequando-as à terminologia da área em português.
- g) Revisar relações de equivalência (sinônimos).

4.2. ANÁLISE DA CAPACIDADE DE CODIFICAÇÃO DA LINGUAGEM

Para avaliar a capacidade de codificação dos descritores do Vocabulário DeCS foi realizado um estudo de ocorrência de descritores da Categoria SP1 e SP2 na indexação dos registros da Base AdSAÚDE.

O estudo foi realizado com base em uma listagem que continha os descritores utilizados na base associados à incidência. Para a operacionalização da análise, foram

geradas três novas tabelas: uma tabela de ocorrência hierarquizada, uma de ocorrência em ordem decrescente e uma de termos não utilizados.

Foi possível através da Tabela de Hierarquia, avaliar o uso dos termos por conjuntos específicos de noções, como exemplificado a seguir:

- a) Nos assuntos ligados ao Direito e Legislação, os termos mais usados são: DIREITOS HUMANOS (total de 127 registros), DIREITO A MORRER (30 registros), JUSTIÇA SOCIAL (140 registros), JURISPRUDÊNCIA (21 registros), LEGISLAÇÃO AMBIENTAL (520 registros), LEGISLAÇÃO SANITÁRIA (28 registros), LEGISLAÇÃO SOBRE MEDICAMENTOS (24 registros).
- b) Nos assuntos ligados aos aspectos econômicos, os descritores mais usados são ANÁLISE ECONOMICA (62 registros), INDICADORES ECONÔMICOS (20 registros), PRODUTO INTERNO BRUTO (37 registros), TAXAS HOSPITALARES (98 REGISTROS), CUSTOS DIRETOS DE SERVIÇOS (55 registros), AUDITORIA FINANCEIRA (239 registros).

Pela Tabela de Ocorrência foi possível identificar os descritores mais usados na base como POLÍTICA DE SAÚDE (1135 registros), SISTEMA DE SAÚDE (810 registros), PARTICIPAÇÃO COMUNITÁRIA (672 registros), SERVIÇOS DE SAÚDE (665 registros), DESCENTRALIZAÇÃO (593 registros), POLÍTICA SOCIAL (578 registros), entre outros.

A Tabela de Ocorrência permite o mapeamento dos conceitos principais da área de Administração em Saúde, a análise da ocorrência de descritores por conjunto de

conceitos, bem como, a avaliação da tendência de uso de termos e a situação da terminologia na área.

Pela tabela de Termos Não Utilizados foi possível obter indicações para orientar a eliminação ou revisão de termos. De fato, descritores tanto da categoria SP1 e SP2 que não foram usados para indexação na Base AdSAÚDE durante mais de 10 anos de registros bibliográficos devem ter sua permanência avaliada. Deve-se considerar, igualmente, a possibilidade de os conceitos referentes a estes descritores serem inapropriados o que requer, de qualquer modo, a atualização da terminologia.

5. CONCLUSÕES

O estudo do Vocabulário Controlado DeCS na área de Administração em Saúde, indicou ser necessário dispor de parâmetros de gerenciamento das linguagens documentárias, desenvolver modelos para avaliação, bem como elaborar diretrizes para revisão e expansão. O processo de atualização e/ou revisão de uma linguagem documentária requer, em síntese, dois níveis de trabalho distintos: a gestão da linguagem e o controle terminológico.

A gestão da linguagem deverá contemplar:

- 1) A elaboração de diretrizes de avaliação e revisão que inclui prazos, verificação de falhas, além de comunicação das atualizações para os usuários do sistema.
- 2) Adoção dos seguintes mecanismos: consulta a especialistas e aos estudos de uso dos termos, assegurar a agilidade na atualização dos termos da linguagem; estabelecer mecanismos efetivos de cooperação entre as instituições que participam do sistema e estabelecer mecanismos de comunicação das mudanças.

3) O controle terminológico supõe a análise da adequação e pertinência da terminologia utilizada. Nesse sentido, merecem especial atenção as operações de estabelecimento de equivalências entre descritores de instrumentos multilíngues. É necessário considerar que as questões da linguagem, em sistemas de informação cooperativos que envolvem diferentes idiomas de trabalho, requerem abordagens que levem em conta os aspectos socioculturais, já que os instrumentos multilíngues são, antes de tudo, de natureza multicultural.

REFERÊNCIAS

1. BERMEJO, C. A; RUBIO, A. V; ROJO, A. S. Desarrollo de lenguagens documentales formalizados en lengua espanhola: 2. Evaluación de tesauros en lengua espanhola. **Revista espanhola de documentación científica**, v.12, n.3, p.283-305, 1989.
2. CINTRA, A. M. M. Elementos de lingüística para estudos de indexação. **Ciência da informação**, Brasília: v.12, n. 1, p.5-22, 1983.
3. FISHEL, C. C et al. List of subject headings: a nursing thesaurus revised. **Bulletin of the Medical Library Association**, v.73, n. 2, p. 153-159, Apr. 1985.
4. GREEN-BIERBAUM, E; BROOKS, T. A; BROOKS, R. M. Subject control of the literature of Acquired Immunodeficiency Syndrome (AIDS). **Information Processing and Management** , v.28 n. 1, p.89-98, 1992.
5. GUINCHAT, C; MENOU, M. Lenguagens documentárias. In: _____. **Introdução geral às ciências e técnicas da informação e documentação**. Brasília: IBICT, 1994. p.133-165.

6. HARRIS, J. Terminology change: effect on index vocabularies. **Information Processing and Management** v.15 n. 2, p.77-88, 1979.
7. JANIK, S; BRUNET, L. La mise a jour d'un thesaurus. **Documentaliste**, v. 24, n. 6, p. 215-229, Nov.-Dec. 1987.
8. KOBASHI, N. Y; TÁLAMO, M. F. G. M. **Aspectos metodológicos da construção de linguagens documentárias**. São Paulo: 2001. (Apresentado no Encontro Mercosul de Ensino em Ciência da Informação, Assunção, jul. 2001).
9. LANCASTER, F. W. **Construção e uso de tesouro: curso condensado**. Brasília: IBICT, 1987.
10. RODRÍGUEZ CAMIÑO, R. MeSH o DeCS: algunas consideraciones sobre la indización biomédica. **ACIMED**, v.6, n.3, 1998.
11. TÁLAMO, M. F. G. M.; KOBASHI, N. Y; LARA, M. L. G. Contribuição da terminologia para elaboração de tesouros. **Ciência da Informação**, Brasília: v.21, n.3, p.197-200, 1992.
12. VALDÉS ABREU, M. de la C. Necesidad de las notas de alcance de los calificadores en la Introducción del tesouro Descriptores en Ciencias de la Salud (DeCS). **ACIMED**, v.4, n.1, 1996.