

Bibliografia Sul-Riograndense de Ciências Biomédicas: elaboração de uma linguagem padronizada

MALVINA VIANNA ROSA *

Descreve linguagem padronizada, tipo «thesaurus» escolhida como meio de indexação. Apresenta problemas enfrentados por planejadores de «thesaurus». Salienta a estrutura classificatória básica, bem como a ordem de citação preferida. Expõe as etapas seguidas na construção do presente sistema de indexação.

1. LINGUAGEM

Em maio de 1973, criava-se em Porto Alegre o “Grupo de Trabalho em Documentação Biomédica” (GTDB), da Associação Rio Grandense de Bibliotecários (ARB), congregando profissionais ligados às bibliotecas especializadas no campo da biomedicina.

A “Bibliografia Sul-Rio-Grandense de Ciências Biomédicas” constitui uma das metas do primeiro plano de trabalho do grupo então formado. Estaria, assim, o GTDB colaborando com a documentação

* Bibliotecária-Chefe da Biblioteca Setorial da Faculdade de Odontologia, UFRGS.

regional e mesmo com a internacional, reunindo e divulgando a produção do Rio Grande do Sul.

Concomitantemente à sua elaboração, tal bibliografia tornaria possível a realização de exercício prático do grupo, para o qual seriam estudadas e revistas técnicas de indexação.

Cada participante do GTDB encarregou-se do levantamento do material dentro de sua área específica de ação. Foram reunidas 316 referências correspondentes à literatura biomédica produzida entre outubro de 1972 e outubro de 1973. Surgiu uma bibliografia corrente, com periodicidade anual e sinalética.

As referências são apresentadas segundo a ordem alfabética dos termos indexadores, descritores. Visando a maior eficiência na operação de recuperação da informação armazenada, estabeleceu-se uma linguagem padronizada, "thesaurus", elo entre indexadores e usuários da referida bibliografia.

2. LINGUAGEM PADRONIZADA DE INDEXAÇÃO E RECUPERAÇÃO

A linguagem padronizada é o agente principal da coincidência de vocabulários, respectivamente do indexador e do usuário, conduzindo as operações de indexação e de busca dentro de uma única terminologia, proporcionando melhores possibilidades de entendimento. Indexadores, ao representarem o assunto do documento por descritores, estão sendo portanto, intermediários dos autores, desejosos de terem suas informações divulgadas, e, dos usuários, buscando as informações. Através das linguagens padronizadas, identificam-se conceitos e suas relações nas palavras do texto-linguagem natural. Transformados em linguagem padronizada, onde os sinônimos são contro-

lados e as estruturas sintáticas simplificadas, serão instrumentos valiosos para a obtenção de informações. Muitas linguagens fornecem termos para a indexação pós-coordenada, principalmente, como os "thesaurus", originados do "thesaurus" da língua inglesa de Roget. O "Thesaurus" de Roget, ligando palavras-texto com idéias, auxilia o escritor a passar da idéia para o texto. Por sua vez, o "thesaurus", usado em indexação, ligando palavras-texto com idéias, auxilia o indexador a passar do texto à palavra-chave, apresentando uma transformação em direção reversa a do "thesaurus" original, isto é, "do texto" em lugar de "para o texto".

O planejador desses instrumentos deverá considerar vários itens para defini-los de acordo com suas finalidades. Um dos problemas com que se depara refere-se ao nível em que permitirá a inclusão de termos pré-coordenados, isto é, termos compostos correspondentes a conceitos únicos. Também serão determinados os controles de revocação (recall) e relevância (relevance) que caracterizarão a linguagem a ser construída. Estudará também espécie e número de usuários, "modus operandi" do sistema, se o sistema será manejado diretamente pelo usuário, ou se essa tarefa será entregue à pessoa qualificada. Finalmente, calculará a quantidade provável de informação armazenada, assim como a qualidade da mesma em relação à atualidade, isto é, se retrospectiva ou corrente. Essas etapas são as linhas mestras de uma linguagem e obviamente serão decididas "a priori".

3. DEFINIÇÕES ÚTEIS AO ESTABELECIMENTO DE UMA LINGUAGEM PADRONIZADA

Termo é a palavra genérica que representa a unidade semântica da linguagem de indexação. *Conceito* é o termo com significado específico.

Palavra-chave é o termo preferido utilizado na indexação. *Termo-guia* é o que não é usado, mas conduz à palavra-chave. *Descriptor* é o termo que engloba uma cadeia de conceitos reunidos segundo uma classificação ou mesmo de maneira aleatória, representando o assunto contido no documento. Assim, chamamos de **COROAS** o termo que se torna conceito quando, no contexto odontológico, lhe atribuímos o significado de prótese dentária parcial fixa e unitária. **JAQUETA** é o termo-guia que remete a palavra-chave **COROA**. Uma vez que **COROA** represente o assunto de um documento, será um descriptor.

Consideramos como revocação de um sistema a habilidade de recuperar documentos desejados, e relevância a qualidade do sistema de isolar documentos não desejados.

Supondo que os termos seguintes foram escolhidos como descritores — **BRÔNQUIOS—CARCINOMA**, **TRIPANOSOMÍASE—SUL - AMERICANA**, **VEIAS—VARICOSAS**, **VITAMINA-C** — o controle semântico aumentará a revocação, trazendo sinônimos que ampliarão o vocabulário. Portanto, considerando **BRÔNQUIOS—CARCINOMA** igual a câncer brônquico, **TRIPANOSOMÍASE—SUL - AMERICANA** o mesmo que Chagas, doença, **VEIAS—VARICOSAS** igual a varizes e **VITAMINA-C** a ácido ascórbico, teremos as seguintes entradas no vocabulário:

ácido ascórbico, veja — **VITAMINA-C**

BRÔNQUIOS—CARCINOMA, usado para câncer brônquico

câncer brônquico, veja **BRÔNQUIOS—CARCINOMA**

Chagas, doença, veja **TRIPANOSOMÍASE—SUL-AMERICANA**

TRIPANOSOMÍASE—SUL - AMERICANA, usado para Chagas, doença.

Varizes, veja VEIAS—VARICOSAS.

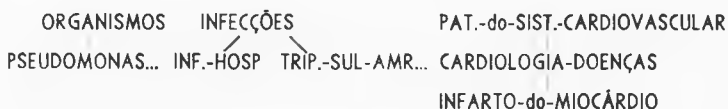
VEIAS—VARICOSAS, usado para varizes

VITAMINA—C, usado para ácido ascórbico.

A revocação também aumentará se usarmos tanto relacionamentos hierárquicos, gênero/espécie, quanto termos correlatos. Exemplifiquemos com os seguintes descritores:

CARDIOLOGIA—DOENÇAS, INFARTO—DO—MIOCÁRDIO, INFECÇÕES—HOSPITALARES, PATOLOGIA — DO — SISTEMA — CARDIOVASCULAR, PSEUDOMONAS, TRIPANOSOMÍASE—SUL - AMERICANA e ORGANISMOS.

Tais descritores podem ser agrupados, constituindo as hierarquias:



Além do relacionamento dentro de suas hierarquias, existe uma ligação de correlacionamento entre descritores pertencentes a hierarquias diferentes (mostrado no exemplo acima pela linha pontilhada). Modificando a estratégia da busca, produziremos aumento na revocação. Segundo GILCHRIST & ATCHINSON³ os mecanismos que influem na revocação são:

- (a) controle do vocabulário de entrada, podendo levar o usuário dos conceitos específicos aos descritores no sistema, p. ex. nefrite, veja, NEFROLOGIA-DOENÇAS;

- (b) controle da forma da palavra (número, forma gramatical). Evita que documentos importantes tenham seus conceitos dispersos sob diferentes formas, mas que dizem respeito ao mesmo termo;
- (c) controle de sinônimos, impedindo que os mesmos fiquem espalhados sob palavras-chaves diferentes, e transformação de sinônimos em termos-guias que remeterão a uma só palavra-chave, ex. Chagas, doença, veja TRIPANOSOMÍASE—SUL-AMERICANA;
- (d) relacionamentos estruturais, compreendendo relacionamentos hierárquicos e não hierárquicos. A presença de tais relacionamentos aumenta a revocação pela ampliação da busca; quanto à relevância, seus mecanismos são:
 - I. especificidade da linguagem: quanto mais detalhadas forem as palavras-chaves, tanto maior será a habilidade do sistema em separar informação não relevante;
 - II. coordenação: é o mecanismo mais poderoso. Aumentando o número de palavras-chaves combinadas na indexação ou na busca, conseguem-se conceitos mais específicos e conseqüentemente serão eliminados os documentos não desejados;
 - III. nível de pré-coordenação de termos: é outra maneira de coordenação. A pré-coordenação é formada no “thesaurus” — antes de ser uma característica operacional do sistema. Os conceitos formados de termos compostos asseguram a identificação minuciosa do documento, separando, assim, documentos não importantes;
 - IV. elos, funções e indicadores relacionais: mecanismos que vencem falsas coordenações e relacionamentos incorretos, codificando grupos de ter-

mos associados ou indicando as funções dos termos;

- V. pesos: mecanismos de diferenciação entre conceitos maiores e menores.

Os mecanismos de revocação operam como parte integrante do vocabulário, enquanto que os de relevância atuam separadamente, e, por este motivo, são chamados de mecanismos auxiliares. Existe um relacionamento inverso entre revocação e relevância determinando que um aumento na revocação é acompanhado de uma diminuição de relevância. A recíproca é verdadeira. Assim, no planejamento de um "thesaurus", deve-se ter em mente que a introdução de alta especificidade como termos compostos, acarretará uma diminuição no número de documentos provavelmente úteis, ao passo que termos simples, relacionamentos estruturais etc, aumentarão a revocação.

4. CARACTERÍSTICAS DO "THESAURUS" CONSTRUÍDO COMO LINGUAGEM DE INDEXAÇÃO E RECUPERAÇÃO DA BIBLIOGRAFIA SUL-RIO-GRANDENSE DE CIÊNCIAS BIOMÉDICAS.

Os textos foram examinados para a seleção do termo ou grupo de termos que coletivamente representavam o assunto dos documentos. Decidimos quais das descrições eram válidas registrar, por serem importantes para os objetivos deste sistema de recuperação. As descrições relevantes foram consideradas como termos-candidatos e estudadas, utilizando-se, como fontes bibliográficas, o "Medical Subject Headings" (MESH) de 1972 e o "Dicionário Médico" de Hugo Fortes & Genésio Pacheco de 1968. Tal estudo produziu uma classificação de termos, nos seguintes agrupamentos:

1. conhecimentos correlatos;
2. conhecimentos básicos;
3. conhecimentos núcleos.

Esta análise classificadora foi a estrutura básica da linguagem construída.

Na realização destas etapas usamos o seguinte padrão:

FORMULARIO PARA THESAURUS	CLASSIFICAÇÃO
TERMO:	NOTA DE ESCOPO:
USADO PARA:	SUBDIVISÕES
TERMOS RELACIONADOS:	
TERMOS GENÉRICOS:	FONTES:
TERMOS ESPECIFICOS:	

FIG. 1 — Formulário para thesaurus

Objetivando maior relevância, admitimos que conceitos únicos fossem representados por mais de um termo, p.ex.: CIRURGIA - PEDIÁTRICA. Para conseguirmos melhor revocação do sistema, controlamos a forma dos termos e a sinonímia. Também evidenciamos as relações hierárquicas gênero-espécie e correlacionamentos, isto é, relações entre termos de

diferentes hierarquias. Considerando como usuários em potencial os profissionais biomédicos em geral, procuramos estabelecer a linguagem usada por tal grupo. A função do "thesaurus", no que concerne a bibliografia impressa, é de guia para a linguagem de indexação, mas, tendo em vista levantamentos bibliográficos a serem realizados com o auxílio do computador, o "thesaurus" sugerirá palavras-chaves e, pela coordenação das mesmas, obter-se-ão referências armazenadas.

Consideramos os termos-candidatos quanto à forma, dando preferência ao uso de substantivos, em lugar de formas adjetivas ou verbais, mas, no caso de constarmos preferência pela forma adjetiva, a adotamos. Assim: INFECCÕES-HOSPITALARES em lugar de infecções-hospitais, mas ADENÓIDES-DOENÇAS em lugar de patologia-adenóideana. Quanto ao número, escolhemos o plural para conceitos que guardam um sentido quantitativo, respondendo a pergunta "quantos?" Usamos o singular para respostas unitárias, por ex.: ALCOOLISMO, GERIATRIA, mas ALIMENTOS, GESTANTES.

Baseados nas fontes de consultas já mencionadas, selecionamos a sinonímia e o termo preferido entre os sinônimos. Outras relações entre os termos, além daquelas derivadas da ordem classificatória, correlações, portanto, logo que identificadas foram anotadas no sistema.

Quanto à classificação, tornamos mais específicos os conhecimentos núcleos, procurando subdividi-los de acordo com as seguintes características: 1. LOCALIZAÇÃO; 2. PROCESSO PATOLÓGICO; 3. AGENTES DE PROCESSOS e de OPERAÇÕES; 4. OPERAÇÕES. Tais divisões procederam-se dando origem a grupos onde os termos (focos) se aproximavam pela seme-

lhança guardada. Embora pouco específica, a classificação obtida foi satisfatória para a massa informacional a que se destinava estruturar. Introduzimos no sistema um número significativo de palavras compostas, designando conceitos únicos. Com isto, facilitamos a relevância e não prejudicamos a combinação de conceitos para a obtenção de documentos que os contenham. Uma vez que utilizamos termos compostos, decidimos sobre a ordem de importância dos mesmos, com o intuito de colocarmos na frente o termo mais provável a ser procurado. Levando em consideração a ordem direta dos termos, aconselhados por GILCHRIST & ATCHINSON,³ procuramos conciliar com a ordem de citação. Optamos pela ordem de citação utilizada na "Classificação Facetada em Ciências Médicas".* Caracteriza-se por não ser a mesma para toda a ciência médica. No caso de operações em geral teremos OPERAÇÃO: LOCALIZAÇÃO ou CIRURGIA: ÓRGÃO, ex.: ANASTOMOSE-SAFENO-CAVERNOSA, CIRURGIA-do-FÍGADO, CIRURGIA-OBSTÉTRICA.

Os processos patológicos específicos de um órgão tem a seguinte ordem:

LOCALIZAÇÃO: PROCESSO PATOLÓGICO ou ÓRGÃO-DOENÇA ex: BEXIGA-CARCINOMA, CÉREBRO-HEMORRAGIA. Processo patológico com localização não específica será: PROCESSO PATOLÓGICO: LOCALIZAÇÃO, ou, DOENÇA: ÓRGÃO. Por nos termos informado que a infecção urinária é característica do sistema urinário, optamos pela entrada SISTEMA-URINÁRIO-INFECÇÃO.

Similarmente preferimos NEUROLOGIA-INFECÇÃO. Sempre que um conceito podia ser representado

* Resultado do projeto Wessex, feita para a Biblioteca do Hospital Regional de Wessex, Southampton, Inglaterra, de autoria de Ruth Daniel.

por um só termo em lugar de dois, preferimos a primeira forma; assim, APENDICITE em vez de apêndice-doenças. Utilizamos de preferência o termo indicador da especialidade, por ser mais abrangente e o subdivimos de acordo com o conceito que representava.

Exemplificando: OTOLOGIA-DOENÇAS, PEDIATRIA-DOENÇAS.

Na parte referente a ODONTOLOGIA seguimos integralmente a ordem de citação adotada pela Classificação Facetada em Odontologia,* isto é, OPERAÇÃO: AGENTE: ENTIDADE: PROCESSO, por ex: CAPEAMENTO-PULPAR, DENTE-REABSORÇÃO, ESMALTE-DENTÁRIO — HIPOPLASIA. Das formas não preferidas, consideradas termos-guias, foram feitas remissivas para as palavras-chaves ou para os descritores. (veja fig. 2).

crianças, cirurgia, veja CIRURGIA-PEDIÁTRICA (Q1)
criança, doenças, veja PEDIATRIA-DOENÇAS (N13)
criança, nutrição, veja NUTRIÇÃO-INFANTIL (K)
criança, psiquiatria, veja PSIQUIATRIA-INFANTIL (L)
criança, reeducação, veja EDUCAÇÃO-ESPECIAL (B2)
criança, tratamento psiquiátrico, veja PSIQUIATRIA-INFANTIL (2).

FIG. 2 — Amostragem do "thesaurus"

Usamos o hífen como ligação entre termos que representam conceitos unitários. Ex.: EDUCAÇÃO-ODONTOLÓGICA.

As palavras-chaves que são descritores aparecem grifadas e em caixa-alta, porém aquelas que apenas

* Dissertação para obtenção do título de mestrado de autoria de MALVINA VIANNA ROSA.

são palavras-chaves somente aparecem em caixa-alta. Quanto aos termos-guias, estes são representados por letras minúsculas.

Cada palavra-chave, cada descritor é acompanhado de uma notação mista, constituída por letra maiúscula seguida de um número, que a coloca dentro da categoria classificada a que pertence.

Alfabetamos palavra por palavra, embora o hífen tenha um significado de união e não de separação. Com o objetivo de não prejudicarmos a sintaxe, usamos preposições, contudo não as levamos em consideração na ordem alfabética (veja fig. 3).

ALIMENTOS (J)

Tr: DIETA (Q4)

NUTRIÇÃO (K)

Tg: TECNOLOGIA-COMÉRCIO-INDÚSTRIA (J)

CIMENTO-de-SILICATO (P3)

Tr: MATERIAIS-ODONTOLÓGICOS (P3)

DIETA (Q4)

Tr: ALIMENTOS (J)

NUTRIÇÃO (K)

Tg: TÉCNICAS-TERAPÊUTICAS (Q4)

NUTRIÇÃO (K)

Tr: ALIMENTOS (J)

DIETA (Q4)

Tg: FISIOLOGIA (K)

Te: NUTRIÇÃO-ENSINO (B2, K)

NUTRIÇÃO-INFANTIL (K)

FIG. 3 — Amostragem do "thesaurus"
(ordem alfabética)

A "Bibliografia Sul-Rio-Grandense de Ciências Biomédicas", em seu primeiro volume, traz o seguinte sumário:

1. Introdução
2. Relação de abreviaturas usadas na bibliografia
3. Relação de títulos das publicações indexadas
4. Vocabulário controlado dos termos indexadores dos assuntos da bibliografia (Thesaurus)
 - 4.1 Listagem alfabética do vocabulário controlado
 - 4.2 Listagem classificada dos termos indexadores
5. Bibliografia em ordem alfabética de termos indexadores
6. Índice alfabético de autores com endereços

Toda a informação coletada foi transferida para cartões perfurados IBM e deu entrada no computador B 6.700, pertencente ao (CPD) Centro de Processamento de Dados da UFRGS. Este processamento nos facilitou a obtenção de diferentes listagens, datilografadas e remetidas à Gráfica da UFRGS. As informações foram transferidas para fita magnética e assim armazenadas para futuras pesquisas a partir desse volume nº 1. Serão possíveis, além de abordagens por assunto, outras como: autor, título do documento, título da publicação, — volume-e-paginação. Congresso em que foi apresentado o trabalho, forma: artigo de periódico, dissertação, tese, livro ou atlas, idioma, notas especiais.

Usamos no "thesaurus" as seguintes abreviaturas: up., usado para, tr., termo correlato, tg., termo genérico e te., termo específico.

A "Bibliografia Sul-Riograndense de Ciências Biomédicas" arrola 46 títulos de publicações periódicas. Indexa 316 referências bibliográficas. A maior parte desse material refere-se a artigos de periódicos, incluindo também dissertações, teses, trabalhos apresentados em Congressos e livros. (veja fig. 4).

ALIMENTOS (J)

DUARTE, L.J.V. Alimentação no período perinatal: retificação. *R. Med HED*, 2 (1):59, 73.

DUARTE, L.J.V. Roda dos alimentos. *R. Ens*, 20(146) : 59/(147): 46/(148):59-60, 73.

DIETA (Q4)

WEISS, L.R. Manual de dietas, primeira parte. *R. Med HED*, 2(2):97-120, 73 (caderno de Informações Médicas, 1)

WEISS, L.R. Manual de dietas: segunda parte. *R. Med HED*, 2(3):161-75, 73 (caderno de Informações Médicas, 2)

NUTRIÇÃO-ENSINO (B2, K)

DUARTE, L.J.V. O ensino da nutrição na escola fundamental. *R. AMERIGS*, 17(1):23-32, 73

NUTRIÇÃO-INFANTIL (K)

MARTINO, A. de. Alimentação do recém-nascido prematuro e de peso baixo. *Pesq. Med*, 9(1):204-5, 73.

FIG. 4 — Amostragem da Bibliografia

5. CONCLUSÃO

O "thesaurus" é um meio valioso de organização da informação, tendo em vista indexação e recuperação eficientes. Do ponto de vista do elaborador, o "thesaurus" permite um domínio maior do assunto, colocando o bibliotecário em sua posição mais verdadeira, isto é, auxiliar indispensável do usuário.

Describes standard language type thesaurus chosen as means of indexing. Presents problems faced by thesaurus planners. Emphasizes the basic classification structure, as well as the chosen citation order. Explains the followed steps in the construction of indexing system.

BIBLIOGRAFIA

1. FOSKETT, A. C. *A abordagem temática da informação*. São Paulo, Polígono, 1973.
 2. GILCHRIST, A. *The thesaurus in retrieval*. London, ASLIB, 1971.
 3. GILCHRIST, A. & ATCHINSON, J. *Manual of thesaurus construction*. London, ASLIB, 1971.
 4. HOLM, B. E. & RASMUSSEN, L. E. Development of a technical thesaurus. *American Documentation*, 12 (3):184-90, July, 1961.
 5. SHARP, J. R. Indexing languages; the thesaurus. In: *Information retrieval*. London, Deutsch, 1970, p. 40-50.
 6. VICKERY, B. C. *Techniques of information retrieval*. London, Butterworths, 1970.
 7. VICKERY, B. C. Thesaurus: a new word in documentation. *Journal of Documentation*, 16(4):181-9, Dec. 1960.
- R. Esc. Bibliotecon. UFMG, Belo Horizonte, 5(1) : 27-41, mar. 1976