

PERSPECTIVAS PARA O ESTUDO DA ÁREA DE REPRESENTAÇÃO DA INFORMAÇÃO

Maria Luiza Almeida Campos

Resumo

Síntese do desenvolvimento das pesquisas que estão sendo realizadas na área de representação/recuperação da informação, visando à elaboração de sistemas de conceitos. Apresenta questões que envolvem o conceito/termo e as relações entre conceitos. Maior ênfase é dada às abordagens que evidenciam os aspectos comuns entre a teoria da classificação facetada, a teoria do conceito e a teoria da terminologia, pois as três teorias estabelecem princípios que permitem a estruturação dos conceitos em um sistema de conceitos.

Palavras-chave

Representação da informação, Recuperação da informação, Sistemas de conceitos, Estruturação de conceitos

A identificação de princípios que possam nortear a área de representação e recuperação da informação tem sido objeto de análise de diversas áreas. Este trabalho tem por finalidade apresentar os esforços realizados no sentido de contribuir para a fundamentação teórica desta área.

Instrumentos como a tabela de classificação, o tesouro e as terminologias, sistematizam os conceitos de uma área de conhecimento na perspectiva de representar e possibilitar a recuperação das informações e permitir, também, comunicações mais precisas no campo da ciência e da técnica.

Uma pesquisa na área de representação e recuperação de informação (CAMPOS, 1994), mostrou a existência de princípios comuns entre aspectos teóricos de instrumentos como esses, que têm por base a Teoria da Classificação Facetada, a Teoria do Conceito e a Teoria Geral da Terminologia, para a formação de estruturas sistemáticas, pois os conceitos de uma área de conhecimento se encontram relacionados entre si. Estes princípios são constituídos por elementos que estão na base da formação de tal estrutura, que são os conceitos, as relações entre os conceitos e a própria apresentação do sistema de conceitos.

O que efetivamente a literatura vem mostrando é que a classificação está na base das três teorias. Na área da Documentação Shiyali Ramamrita Ranganathan elabora a Teoria da Classificação Facetada (Ranganathan, 1963; Ranganathan, 1967), na qual apresenta princípios para a organização de conceitos hierarquicamente estruturados. Os estudiosos dessa área começam a perceber que os tesouros, como os esquemas de classificação, também possuem termos que representam conceitos que estão ligados entre si formando um sistema de conceitos. Campos (1986, p. 85) afirma, inclusive, que “qualquer autêntico tesouro contém em si os elementos básicos de uma classificação e esses elementos poderão assumir a forma de uma tabela de classificação”.

Porém a classificação hoje em dia não pode mais ser vista em seu sentido restrito de estruturas hierárquicas. Segundo a FID/CR – Comitê Técnico de Pesquisa de Classificação (1973), “classificação” é “qualquer método de reconhecimento de relações genéricas ou outras, entre itens de informação, não importando o grau de hierarquia usada, nem se aqueles métodos são aplicados em conexão com sistemas de informação tradicionais ou computadorizados”.

Uma das áreas que tem relação estreita com a classificação é a terminologia. Wüster (1981, p. 106), autor da Teoria Geral da Terminologia, observa a “semelhança das tarefas realizadas na elaboração de um tesouro e na normalização terminológica em geral” e reconhece que deveria existir maior intercâmbio entre as áreas. Em outro trabalho (Wüster, 1971), aborda as diferenças entre os sistemas de conceitos e as tabelas de classificação, enfatizando inclusive o que possuem de semelhante. Ele analisa as tabelas de classificação à luz da Teoria Geral da Terminologia. Em ambos os trabalhos, recomenda maior aproximação entre documentalistas e terminólogos.

Dahlberg (1993, p. 225), estudiosa da área da filosofia, estabelece relações não mais entre uma teoria e um instrumento, mas entre a Teoria da Classificação Facetada e a Teoria do Conceito, por ela desenvolvida. A Teoria do Conceito apresenta princípios que podem auxiliar na determinação do conceito e de suas relações, seja para elaborar tabelas de classificação, seja para elaborar tesouros (Dahlberg, 1978).

As primeiras constatações de semelhança entre tesouros e sistemas de classificação se dão no âmbito da documentação, como é de se esperar. Gupta e Tripathi (1975, C40) analisam o tesouro de exploração e produção de petróleo da Universidade de Tulsa, o qual tem uma estrutura facetada, e comprovam a utilidade desta estrutura na preparação de um esquema especial de classificação para geofísica. Este método de trabalho é possível pela adoção de princípios compatíveis. Gopinath & Prasad (1975, A37) apontam para as diferenças essenciais entre esses dois instrumentos. Observam eles que o tesouro tem dois planos de trabalho, – o plano das idéias e o plano verbal –, enquanto um esquema de classificação abrange os três planos de trabalho, ou seja, o plano das idéias, o plano verbal e o plano notacional. O tesouro e o esquema de classificação apresentam diferentes níveis e profundidade de organização e podem coexistir em um sistema de recuperação de informação, complementando um as deficiências do outro.

Outros autores, no entanto, realizaram, efetivamente, um trabalho de integração dos dois instrumentos. Em 1968 Davis, (*apud* Gopinath, 1987, p. 211) publica um artigo sobre a integração de vocabulário com um esquema de classificação. Em 1969, a English Electric Company publica seu *Thesaurifacet: a thesaurus and faceted classification for Engineering and related subjects*. Este trabalho foi desenvolvido por Jean Aitchison, membro do Classification Research Group (CRG). Até o final dos anos 50 e princípio dos anos 60, os tesouros eram estruturados puramente em ordem alfabética.

“As limitações do arranjo alfabético levaram ao emprego de meios auxiliares da classificação que iam desde os dispositivos mais amplos até os mais detalhados e dos dispositivos auxiliares aos integrados”. (Aitchison, 1972, p. 72)

Aquela classificacionista utilizou ainda outros dispositivos classificatórios para os relacionamentos puramente hierárquicos.

“O método tradicional de tesouros para indicar as hierarquias de termos mais gerais até os mais restritos se mostrou insatisfatório porque nem todos os níveis da hierarquia podem ser dispostos alfabeticamente de uma única vez, ou, se os termos forem dispostos alfabeticamente, não é possível distinguir os diferentes níveis hierárquicos entre eles”... “Finalmente a análise de faceta pode ser usada como um dispositivo da classificação na construção de tesouros” ... “Do uso de facetas amplas na construção de tesouros para o uso de uma classificação totalmente facetada junto com um tesouro foi um passo. O Thesaurifacet ... foi provavelmente um desenvolvimento inevitável” (Aitchison, 1972, p. 74).

Este novo tipo de instrumento tem múltiplo uso: serve para catálogos convencionais e organização nas estantes, bem como para indexação coordenada e uso em sistemas

computadorizados (Aitchison, 1970, p. 187). Gopinath (1987, p.221) ressalta a complementaridade das duas abordagens, pois existe uma “relação simbiótica entre um esquema de classificação e um tesouro”.

O Thesaurofacet é um marco no desenvolvimento das linguagens documentárias pela integração da tabela de classificação com o tesouro. Na literatura um novo nome tem sido usado para instrumentos que integram tesouro e classificação, o ‘classaurus’(Bhattacharyya, 1982, p. 139; Fugmann, 1990, p. 133).

Se foi possível adotar princípios classificatórios para solucionar problemas de estruturação de conceitos tanto nas classificações bibliográficas quanto nos tesouros, o mesmo não acontece com respeito aos descritores ou termos, sua forma, sua definição. As respostas para estas questões têm sido soluções estritamente lingüísticas. As palavras compostas sempre apresentaram problemas nos sistemas pós-coordenados, e os tesouros, em um certo momento, foram vistos como fonte para a solução desses problemas. A lingüística foi utilizada como base para o estabelecimento de palavras compostas em sistemas de recuperação (Jones, 1981, p. 54). Porém a utilização dos pressupostos teóricos da lingüística na formação das palavras compostas no interior dos tesouros não tem conseguido resolver o problema deste tipo de descritor, pois as soluções ficam apenas no plano da língua e não da representação de conceitos. Por outro lado, as diretrizes e normas para tesouros oferecem propostas de solução para “termos compostos” por meio do que se chama de “fatoração”(IBICT, 1985; BSI 5723: 1979; AFNOR Z47-100, 1981; Unesco, 1973; ISO 2788, 1986). Em um artigo sobre o tema, Jones (1981) relata várias propostas para tratamento de palavras compostas, desde Coates até Austin, passando por Farradane, Lee & Jespersen, os quais abordam a questão ora do ponto de vista lingüístico, ora do ponto de vista conceitual. Seetharama (1975) reforça a falta de critérios consistentes para o estabelecimento dos termos nos tesouros. No entanto, propõe, também, a fatoração como solução. Na verdade, o termo necessita de um tratamento terminológico e não lingüístico (DROZD, 1981).

Com base na Teoria do Conceito, Dahlberg (1983) desenvolve um estudo sobre definições terminológicas que vai se mostrar útil nos tesouros do futuro. Segundo Rahmstorf (1993, p. 39), as definições terminológicas abrem um campo de estudo de complexidade crescente porque elas podem ser úteis não apenas para cientistas da informação, indexadores, especialistas em recuperação e outros especialistas da organização do conhecimento, mas também para tradutores, cientistas, engenheiros, especialistas em elaboração de normas, epistemólogos, psicólogos, engenheiros do conhecimento, lingüistas e terminólogos. Ele apresenta, também, de forma esquemática, ao lado das classes de usuários, a finalidade das definições para cada classe, os aspectos principais (se se trata da estrutura conceitual, ou da sintaxe etc.) e o papel da definição para cada classe de usuário.

O primeiro trabalho de comparação sistemática entre terminologia e tesouro cabe ao terminólogo Leska (1981, p. 583). Ele observa que os tesouros de recuperação precisam “expandir suas referências lexicais tornando os conceitos mais precisos, bem como definindo-os e qualificando-os de acordo com seu relacionamento com outros descritores”. São os seguintes os aspectos comuns:

- “1) Os sistemas de conceitos são criados para sistematizar os conceitos de uma área (...)
- 2) Os conceitos do sistema de conceitos são definidos por meio de suas características (...)
3. Os sistemas de conceitos, como os tesouros, visam a abranger todos os conceitos e/ou termos de um campo de assunto (...)
- 4) A estrutura básica do sistema de conceitos de uma área é o esquema estrutural no qual todos os conceitos relevantes devem encontrar seu lugar apropriado (...)
- 5) Cada sistema de conceitos, especialmente com relação a desenvolvimentos futuros, visa ainda à complementação no quadro de seu âmbito temático (...)

- 6) A atividade de desenvolvimento e aperfeiçoamento do sistema de conceitos não fica fora da influência das regras gramaticais que governam os nomes que representam esses conceitos (...).

As terminologias devem se apresentar de forma sistemática e não alfabética. Este aspecto tem levado à necessidade de empregar notação, aproximando a terminologia da classificação. O conteúdo de um conceito é estabelecido a partir da área de conhecimento e do propósito da terminologia. Por sua natureza sistemática, o

“código de assunto é um dos elementos mais importantes na entrada dos bancos de dados terminológicos... Uma lista alfabética não ajuda... Somente um esquema de classificação pode mostrar em que detalhe um campo de assunto foi estruturado e o código ajuda a verificar a amplitude correta do conteúdo de um conceito, especialmente quando usado como termo de indexação, e facilita o intercâmbio entre vários campos de termos.” (Nedobity, 1986/1987, p. 12).

Os bancos de dados terminológicos são, na verdade, sistemas de classificação, à medida que agrupam conceitos ligados hierarquicamente. E é a área tecnológica, responsável pelo desenvolvimento destes bancos, que está suscitando a discussão em torno do caráter sistemático da terminologia e, com isso, mostrando a atualidade da classificação. Um recente

“levantamento da atividade terminológica... mostra a necessidade da abordagem sistemática para melhorar a confiabilidade de produtos, processos e serviços, bem como facilitar a cooperação técnica e a eliminação das barreiras comerciais. Como é comum na atividade terminológica, a estruturação conceitual e a classificação são os principais componentes da atividade.” (Strehlow e Wright, 1993, p. 3).

A terminologia, por causa de sua natureza sistemática, ao lado da classificação, tem sido vista em literatura mais recente como contribuindo para o desenvolvimento de outras áreas que, de alguma forma trabalham com representação da informação. Os princípios de sistematização de termos, comuns à terminologia e à classificação, são fundamentais aos bancos de conhecimento, uma vez que, neles, os conceitos são estruturados, classificados e sistematizados. (Dzincharadze, 1993, p. 127).

Nedobity (1985) identifica nos princípios teóricos da terminologia aspectos que devem ser observados no desenvolvimento de sistemas especialistas e pesquisa na área de inteligência artificial, áreas que lidam com conceitos, sistemas de conceitos, ligações de conceitos. etc. Considera também (Nedobity, 1986), os sistemas de classificação da ciência da informação e da terminologia, bem como os sistemas especialistas com base no conhecimento, como duas pontas de um *continuum* de recursos relevantes para síntese e interpretação do conhecimento.

Budin (1993, p. 1) vê a possibilidade de se estabelecer uma teoria da terminologia que resultaria da junção das teorias da ciência da informação aplicadas à construção e uso das linguagens de documentação (sistemas de classificação, tesouros etc.) com a teoria compreensiva da organização do conhecimento

A problemática relativa à representação da informação e do conhecimento está sendo abordada por estudiosos de diversas áreas, como mostra a literatura recente. Ela extrapola o domínio da documentação, o que não significa que ela abandone suas próprias teorias relacionadas com a representação, pois elas são parte integrante deste novo movimento, que tem em comum a organização do conhecimento.

A classificação, graças a Ranganathan, adquiriu um *status* de disciplina científica e os especialistas desta área não têm dúvida de entender o termo “classificação” no sentido de ordenação do conhecimento.

REFERÊNCIAS BIBLIOGRÁFICAS

- AFNOR NF Z47-100: 1981. *Règles d'établissement des thésaurus monolingues*. 20 p.
- AITCHISON, Jean (1970). The Thesurofacet: multipurpose retrieval language tool. *Journal of Documentation*, v. 26, n. 3, p. 187-203, Sept.
- AITCHISON, Jean (1972). Thesurofacet: a new concept in subject retrieval schemes. In: WELLISCH, H. ed. *Subject retrieval in the seventies*. Connecticut: Greenwood Publ. Co. p. 72-98.
- BATTACHARYYA, G. (1982) Classaurus: its fundamentals, design and use. *Studien zur Klassifikation*, n. 11, p. 139-148.
- BSI 5723:1987. *Guidelines for the establishment and development of monolingual thesauri*. 35 p.
- BUDIN, G. (1993) Knowledge organisation and modelling of terminology and knowledge engineering. In: TKE'93: Terminology and Knowledge Engineering, Cologne, Aug. 25-27, 1993. *Proceedings...* Frankfurt/M: Indeks Verlag. p. 1-7.
- CAMPOS, A.T. (1986) Linguagens documentárias. *Revista de Biblioteconomia de Brasília*, v. 14, n. 1, p. 85-88, jan./jun.
- CAMPOS, M. L. A. (1994) *Em busca de Princípios comuns na área de representação da informação: uma comparação entre o Método de Classificação Facetada, o Método de Tesouro-Baseado-em-Conceito e a Teoria Geral da Terminologia*. Rio de Janeiro: UFRJ/IBICT. 196 p. (Dissertação de Mestrado)
- DAHLBERG, I. (1978) A referent-oriented analytical concept theory of Interconcept. *International Classification*, v. 5, n. 3, p. 142-150.
- DAHLBERG, I. (1983) Terminological definitions: characteristics and demands. In: *Problèmes de la définition et de la synonymie en terminologie*. Québec: GIRSTERM. p. 13-51.
- DAHLBERG, I. (1993) Faceted classification and terminology. TKE'93. Terminology and Knowledge Engineering, Cologne, Aug. 25-27, 1993. *Proceedings...* Frankfurt/M.: Indeks Verlag. p. 225-234.
- DROZD, L. (1981) Some remarks on a linguistic theory. In: Theoretical and Methodological Problems of Terminology, Moscow, Nov. 27-30, 1979. *Proceedings...* München: Saur. p. 106-117.
- DZINCHARADZE, A. (1993) Le rôle de la terminologie lors de la création de bases de connaissances et de systèmes d'intelligence artificielle. In: TAMA'92. TermNet Symposium "Terminology in Advanced Microcomputer Applications". Avignon, 5-6 juin 1992. *Proceedings ...* Wien: TermNet, 1993. p. 125-141.
- FID (1973) *Committee on Classification Research*.

- FUGMANN, R. (1990) An interactive classaurus on the PC. *International Classification*, v. 17, n. 3/4, p. 133-137.
- GOPINATH, M.A. (1987) Symbiosis between classification and thesaurus. *Library Science with a slant on Documentation*, v. 24, n. 4, p. 211-224, Dec.
- GOPINATH, M.A. & PRASAD, K. N. (1975) Thesaurus and classification scheme: a study of the compatibility of the principles for construction of thesaurus and classification scheme. In: Seminar on Thesaurus in Information Systems. Bangalore, Dec. 1-5, 1975. (Paper AF, p. A37-A50)
- GUPTA, D.K. & TRIPATHI, G.N. (1975) Exploration and production thesaurus: its helpfulness in indexing in a library and information system. In: Seminar on Thesaurus in Information Systems. Bangalore, Dec. 1-5, 1975. (Paper CF, p. C40-C42)
- IBICT (1984) *Diretrizes para elaboração de tesouros monolíngües*. Brasília: IBICT. 70 p.
- ISO 2788:1986. *Documentation guidelines for the establishment and development of monolingual thesauri*. 2. ed. 32p.
- JONES, Kevin P. (1981) Problems associated with the use of compound words in thesauri, with special reference to BS 5723:1979. *Journal of Documentation*, v. 37, n. 2, p. 53-68.
- LESKA, B. M. (1981) The connections between systems of concepts and thesauri. In: Theoretical and Methodological Problems of Terminology. Moscow, Nov. 27-30, 1979. *Proceedings...* München: Saur. p. 583-590.
- NEDOBITY, W. (1986) Classification, artificial intelligence and cognitive psychology. In: GAUL, W. & SCHADER, M. *Classification as a tool of research*. Amsterdam: Elsevier. p. 345-352.
- NEDOBITY, W. (1987) *Classification systems for terminological data bank*. Wien:INFOTERM. 12 p.
- RAHMSTORF, G. (1993) Role and representation of terminological definitions. In: TKE'93. Terminology and Knowledge Engineering. Cologne, Aug. 25-27, 1993. *Proceedings ...* Frankfurt/M. : Indeks Verlag. p. 39-49.
- RANGANATHAN, S. R. (1963) *Colon Classification*. Bombay: Asia Publ. House. 126 p.
- RANGANATHAN, S. R. (1967) *Prolegomena to library classification*. Bombay: Asia Publ. House. 640 p.
- SEETHARAMA, S. (1975) Term-Concept relationship in an information retrieval thesaurus. In: Seminar on Thesaurus in Information Systems. Bangalore, Dec. 1-5. (Paper AJ, p. A80-A87)
- STREHLOW, R.A. & WRIGHT, S.E. ed. (1993) Standardizing terminology for better communication. Philadelphia: ASTM. 390 p.
- UNESCO (1973) *Guidelines for the establishment and development of monolingual thesauri*. 37 p.
- WÜSTER, E. (1971) Les classifications de notions et de thèmes - Différences essentielles et applications. (Série INFOTERM 2-71 F)

WÜSTER, E. (1981) L'étude scientifique générale de la Terminologie, zone frontalière entre la Linguistique, la Logique, l'Ontologie, L'Informatique et les Sciences de Choses. In: FELBER, H. org. *Textes choisis de Terminologie*. Québec: GIRSTERM, p. 57-114.

Perspectives for studies on Information Representation and Retrieval

Abstract

Synthesis of the development of research in the field of information representation and retrieval aiming at building concept systems. Questions involving concept/term are presented as well as concept relations. Major emphasis is given to approaches showing evidences to common aspects among theories of Faceted Classification, of Concept and of Terminology, as they are all based on principles that permit structuring concepts in a concept system.

Keywords

Information representation; Information retrieval; Concept system; Structuring concepts.

Maria Luiza de Almeida Campos

Professora do Departamento de Documentação - UFF

Mestre em Ciência da Informação - UFRJ/IBICT