



XI Encontro Nacional de Pesquisa em Ciência da Informação
Inovação e inclusão social: questões contemporâneas da informação
Rio de Janeiro, 25 a 28 de outubro de 2010

GT 2 - Organização e Representação do Conhecimento
Modalidade de apresentação: pôster

**PROPOSTA DE MODELO DE LEITURA PARA EXTRAÇÃO DE CONCEITOS:
CRIAÇÃO DE TAXONOMIA FACETADA PARA NAVEGAÇÃO EM BIBLIOTECAS
DIGITAIS DE TESES E DISSERTAÇÕES**

Benildes Coura Moreira dos Santos Maculan

Universidade Federal de Minas Gerais

Gercina Angela Borém de Oliveira Lima

Universidade Federal de Minas Gerais

Resumo: Apresenta um Modelo de Leitura Técnica para a sistematização da análise de assunto em documentos do tipo teses e dissertações, para a identificação dos conceitos representativos de seu conteúdo. É produto de uma pesquisa em andamento que tem como objetivo criar uma taxonomia facetada para navegação, como interface de busca, visando facilitar a representação e posterior recuperação de informações em bibliotecas digitais de teses e dissertações. A fundamentação teórico-metodológica para o desenvolvimento do modelo utilizou os estudos de Fujita e Rubi (2006), que criaram um modelo de leitura para análise de artigos da área biológica. As autoras basearam o modelo proposto na Norma 12.676 da Associação Brasileira de Normas Técnicas (1992), na metodologia proposta por Kobashi (1994) para indexação de documentos, e no sistema PRECIS (FUJITA, 2003) para identificação de conceitos. O uso do modelo elaborado visa minimizar os problemas de ocorrência de inconsistência na indexação dos documentos apontados pelo resultado parcial da pesquisa. Com a criação da taxonomia facetada pretende-se dar maior visibilidade ao conhecimento acumulado nas teses e dissertações, principalmente facilitando sua recuperação.

Palavras-chave: Modelo de Leitura Técnica; Taxonomia facetada; Interface de busca; Recuperação de informação.



1 Introdução

Este trabalho é parte de uma pesquisa em andamento, que propõe criar uma taxonomia facetada para navegação, como interface de busca, em uma biblioteca digital (BD) de teses e dissertações, para facilitar a representação e a organização das informações, visando sua posterior recuperação. Os primeiros resultados apontam inconsistência na indexação de documentos, sinalizando a necessidade de um Modelo de Leitura Técnica, visando minimizar tais problemas.

Destaca-se o aumento na implantação de BDs de teses e dissertações (BDTD), por meio de diferentes iniciativas para compartilhar e divulgar informações acadêmico-institucionais. Esse fato pode ser observado principalmente após a Portaria 13/2006, instituída pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), estabelecendo a obrigatoriedade da divulgação digital da produção dos programas de pós-graduação *stricto sensu* no país.

Nesse contexto, as questões que envolvem a recuperação da informação são objeto de investigação desde a origem da Ciência da Informação (CI), que investiga a informação nos mais diferentes ambientes de fluxo, objetivando facilitar seu acesso e uso (BORKO, 1968). Para a CI, o documento¹ é o produto tangível nesse processo, que, segundo Rabello (2009, p.12), “apresenta a materialidade e a susceptibilidade de [...] registro informacional como características principais e apresenta [...] uma estrutura e uma disposição que facilita e/ou propicia a identificação e a apreensão das informações nele registradas”.

Documentos do tipo teses e dissertações têm caráter didático (MARCONI; LAKATOS, 1988), apresentando metodologia e resultados oriundos da análise de dados de pesquisa científica. Usam linguagem formal e específica para a área ou domínio², usando uma terminologia própria, porém, o documento textual das teses e dissertações é produzido em linguagem natural, permitindo o uso de diferentes termos para um mesmo conceito. Esse fato dificulta a comunicação e compartilhamento do conhecimento registrado nesses documentos, o que motivou esta pesquisa.

¹ Qualquer unidade, impressa ou não, passível de catalogação ou indexação (NBR-12676/92). Neste artigo, considera-se documento em sentido amplo, qualquer que seja a sua natureza ou suporte.

² “Domínio” pode ser uma área de conhecimento (especialidade), um conjunto literário ou ainda a tarefa de um grupo de pessoas trabalhando juntas num projeto ou numa organização (HJORLAND; ALBRECHTSEN, 1995).



2 Modelo de Leitura Técnica (MLT)

O documento textual das teses e dissertações tem como diferença básica a profundidade de conteúdo. Contudo, pode-se considerá-las padronizadas, pois contêm uma estruturação bastante similar, numa ordem lógica: 1) seção inicial, com a introdução, as justificativas e os objetivos; 2) desenvolvimento, geralmente com a revisão da literatura, a metodologia, uma discussão (argumentação teórica, conceitual ou analítica), descrição dos dados, apresentação dos resultados; e 3) última seção, apresentam-se as conclusões.

No processo de indexação de documentos, a análise conceitual, segundo Fujita e Rubi (2006, *on-line*), segue os seguintes procedimentos: 1) leitura do documento para exploração da sua estrutura; 2) identificação dos conceitos; e 3) seleção dos conceitos.

A partir das características dos documentos do tipo tese e dissertação, foi criado o MLT, para sistematizar a análise conceitual deles. O MLT foi elaborado a partir do modelo apresentado por Fujita e Rubi (2006). Essas autoras desenvolveram um modelo para a análise de artigos científicos da área biológica, consistindo na “combinação das sistemáticas de identificação de conceitos análise conceitual (primeira coluna) [metodologia proposta por Kobashi (1994)] e abordagem sistemática da Norma 12.676 (segunda coluna) com a localização dos conceitos em parte da estrutura textual (terceira coluna) [baseado no sistema PRECIS (FUJITA, 2003)]” (FUJITA; RUBI, 2006, *on-line*). A seguir, o modelo desenvolvido na presente pesquisa:



QUADRO 1 – Modelo de leitura documentária para textos científicos (teses e dissertações)

CONCEITO (ANÁLISE CONCEITUAL)	QUESTIONAMENTO (NORMA 12.676)	PARTE DA ESTRUTURA TEXTUAL
TEMA	Qual o assunto de que trata o documento?	INTRODUÇÃO
OBJETO	Qual o objeto de estudo em questão?	INTRODUÇÃO (PROBLEMA/ JUSTIFICATIVA/ OBJETIVOS)
AÇÃO	O que pretende a pesquisa, de forma geral e específica, que seja relevante determinar?	INTRODUÇÃO (PROBLEMA/ JUSTIFICATIVA/ OBJETIVOS)
DIMENSÃO	O tema, objeto e/ou ação são considerados no contexto de um lugar específico ou ambiente?	INTRODUÇÃO (PROBLEMA/ JUSTIFICATIVA/ OBJETIVOS)
COLETA DE DADOS	Instrumentos específicos para realizar a ação?	METODOLOGIA
MÉTODOS	Modos específicos para realizar a ação (por exemplo, instrumentos, técnicas ou métodos)?	METODOLOGIA
FUNDAMENTO TEÓRICO	Como se define o assunto em termos de teorias, hipóteses, etc.?	REVISÃO DE LITERATURA; FUNDAMENTAÇÃO TEÓRICA
CAUSA E EFEITO	São identificadas algumas variáveis dependentes ou independentes?	RESULTADOS; DISCUSSÃO DE RESULTADOS; CONCLUSÕES
PRINCIPAIS RESULTADOS ALCANÇADOS	Quais pontos a pesquisa alcançou?	RESULTADOS; DISCUSSÃO DE RESULTADOS
PONTO DE VISTA DO AUTOR; PERSPECTIVA	O assunto foi considerado de um ponto de vista pessoal? Quais as perspectivas futuras de estudo?	CONCLUSÕES

Fonte: Adaptado de Fujita e Rubi (2006, *on-line*).

O MLT desenvolvido será empregado para dar consistência ao processo de análise para a indexação do conteúdo das teses e dissertações, e permitirá reduzir os problemas de inconsistências já observados nessa atividade em diferentes pesquisas (VIEIRA, 1988; GIL LEIVA; FUJITA; RUBI, 2008; SOUZA, 2009). Posteriormente, para o registro dos conceitos identificados, será acrescentada uma quarta coluna ao modelo, nos quais serão alocadas as “respostas” a cada módulo. Os conceitos extraídos dos documentos, e considerados relevantes dentro do domínio, serão utilizados para a construção da taxonomia facetada (vocabulário controlado).



3 Vocabulário controlado (VC), taxonomia e taxonomia facetada

Um VC “é uma lista de termos elaborada para identificar o assunto ou os assuntos de um documento com especificidade bastante para permitir sua recuperação rápida e eficaz” (CURRÁS, 1995), pois permite controlar a terminologia de um domínio, instituindo regras para seu uso, compartilhamento e expansão. Segundo Smit e Kobashi (2003, p.14), esse controle permite organizar e recuperar informações de forma consistente. A norma ANSI/NISO Z39.19 (2005) indica que um VC é uma lista controlada de termos explicitamente listados, com definição clara e inequívoca dentro do propósito para o qual foi criado, com o objetivo de organizar a informação, apresentando a lista de assuntos, o anel de sinônimos³, o tesouro e a taxonomia, como tipos de VC.

De acordo com a norma ANSI/NISO Z39.19, uma “taxonomia é a coleção de termos de um vocabulário controlado, organizados em uma estrutura hierárquica [na qual] cada termo [...] está em uma ou mais relações tipo pai/filho (geral/específico) em relação a outro termo nessa taxonomia” (2005, p.9, tradução nossa). A forma tradicional de uma taxonomia é a hierárquica, porém, hoje a taxonomia pode ser representada na forma de listas, árvores, hierarquias, polihierarquias, matrizes, facetas e sistemas de mapas (LAMBE, 2007).

A taxonomia facetada é baseada na teoria da análise facetada e no método analítico-sintético de Ranganathan. A abordagem analítico-sintética visa à estruturação do conhecimento de forma que permita a inclusão de novos conceitos sem que modifique toda a estrutura do sistema classificatório (LAMBE, 2007). Dessa forma, temos um esquema de conceitos, pois “o resgate da taxonomia nos sistemas de informação considera a unidade sistemática (*taxon*) não mais família, gênero, espécie, mas conceitos. [...] segundo uma ordem lógica, apoiada igualmente em princípios classificatórios” (CAMPOS; GOMES, 2008, *on-line*).

A taxonomia facetada proposta será criada a partir desses princípios classificatórios. Os termos utilizados para representar o conteúdo informacional na estrutura da taxonomia facetada para navegação serão obtidos por meio da análise documental usando o MLT desenvolvido. A navegação pela estrutura facetada irá organizar as informações de modo a facilitar e orientar o usuário (cliente) na busca por informações (produtos e serviços) de sua

³ Anel de sinônimos (*synonym ring*) é um grupo de termos considerados sinônimos com o propósito de recuperação.



necessidade. Isso facilitará ainda o gerenciamento dos recursos informacionais pelos gestores da BD, como tem sido recorrentemente utilizado em sites corporativos (Americanas.com, Amazon.com, Shoptime, Submarino, entre outros).

3.1 A navegação pela taxonomia facetada

A navegação pela taxonomia facetada “permite combinar um ou mais valores [...] [sendo] possível obter todas as páginas escritas em português (faceta língua), escritas entre 2001 e 2002 (faceta data), do tipo texto (faceta tipo do conteúdo)” (PRAZERES *et al.*, 2006, p.2). De tal modo, o usuário estará usando uma interface de busca indireta, na qual “o espaço de informação é mapeado, resultando na construção de uma taxonomia que melhor represente o mesmo” (HEUSELER, 2010, p.44-45). “O usuário pode elaborar a questão buscada progressivamente, ao mesmo tempo em que pode observar o efeito que sua escolha por uma faceta determina quais outras facetas estarão disponíveis para ele” (TUNKELANG, 2009, p.23), facilitando o compartilhamento da informação.

4 Resultados Parciais

Foi feito o levantamento das teses e dissertações defendidas no PPGCI-UFMG, no período delimitado pela pesquisa, entre os anos 1998 e 2009, no qual foi identificado um total de 290 documentos. Desse total, 62 documentos se referem à linha de pesquisa Organização e Uso de Informação (OUI), que constituem o *corpus* desta pesquisa.

A pesquisa está ainda em fase de coleta de dados, mas aponta algumas constatações, tal como inconsistência na indexação, já discutida em estudos anteriores em BDTDs brasileiras (ELIEL, 2008). Observou-se que alguns documentos não possuíam palavras-chave, possuíam palavras-chave insuficientes para descrever o conteúdo dos documentos, ou apresentavam resumos incompletos. A partir desses problemas, percebeu-se a necessidade de análise conceitual nos textos completos dos documentos. Para sistematizar a análise, foi desenvolvido um Modelo de Leitura Técnica (MLT) a partir do modelo apresentado por Fujita e Rubi (2006), adaptado para a estrutura de documentos científicos do tipo teses e dissertações da área de ciências humanas aplicadas.



As definições estão sendo elaboradas a partir de glossários, dicionários técnicos da área e consulta aos especialistas do domínio delimitado pela presente pesquisa, obedecendo aos critérios: 1) ser preferencialmente positiva, clara, breve e não circular; 2) não usar linguagem figurativa; 3) ser ajustada à coisa definida; 4) conter a divisão da espécie (todo-parte), seja físico (quantitativo, essencial, potencial ou acidental), lógico (metafísico) ou moral; 5) eventualmente apresentar a diferença específica (diferença dentro do gênero/parte); 6) ser irreduzível; e 7) respeitar o mesmo princípio, a partir de elementos opostos.

5 Considerações Finais

As pesquisas em CI buscam compreender e solucionar problemas de natureza cada vez mais complexos, muitas vezes com cruzamento de técnicas e métodos, variando objetos de estudo e metodologias aplicadas, dependendo do olhar lançado ao objeto. Todos esses aspectos estão inscritos nas teses e dissertações aprovadas nos programas de pós-graduação, acumulando um conhecimento importante para o desenvolvimento da ciência e tecnologia.

Com a criação da taxonomia facetada pretende-se facilitar esse cruzamento, dando maior visibilidade a esse conhecimento acumulado, tornando-o mais perceptível ao usuário e facilitando a recuperação dos conteúdos dos documentos científicos.

O estudo e o desenvolvimento da “taxonomia facetada” têm importância na área da CI em relação aos temas da representação e organização da informação. Além disso, seu valor está principalmente ao que concerne o aperfeiçoamento da recuperação da informação em bibliotecas de teses e dissertações (BDTD), pois pode impactar positivamente o processo da comunicação científica. Nesse contexto, a taxonomia facetada poderá facilitar essa comunicação, à medida que o usuário poderá restringir progressivamente a busca durante a navegação.



REFERÊNCIAS:

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. *NBR 12676: Métodos para análise de documentos, determinação de seus assuntos e seleção de termos de indexação*. Rio de Janeiro, 1992.

BORKO, Harold. Information science: what is it? *American Documentation*, Washington, v. 19, n. 1, p. 3-5, Jan. 1968. Disponível em: <<http://www.cas.usf.edu/lis/lis6260/lectures/infosci.htm>>. Acesso em: 2 jul. 2007.

BRASIL. Ministério da Educação. CAPES. *Portaria 13*. 2006. Disponível em: <<http://www.abmes.org.br/legislacao/2006/portarias.asp>>. Acesso em 22 mar. 2010.
CAMPOS, Maria Luiza de A.; GOMES, Hagar Espanha. Taxonomia e classificação: o princípio de categorização. *DataGramaZero - Rev. Ci. Inf.*, v. 9, n. 4, ago. 2008. Disponível em: <http://www.datagramazero.org.br/ago08/Art_01.htm>. Acesso em: jun. 2009.

CURRÁS, Emilia. *Tesauros: linguagens terminológicas*. Brasília: IBICT, 1995.
ELIEL, Regiane A. Institucionalização da ciência da informação no Brasil: estudo da convergência entre a produção científica e os marcos regulatórios da área. *Transinformação*, Campinas, v. 20, n. 3, p. 207-224, set./dez. 2008. Disponível em: <http://biblioteca.ricesu.com.br/art_link.php?art_cod=6347>. Acesso em: 12 Abr 2010.

FUJITA Mariângela S. L.; RUBI, Milena P. Um modelo de leitura documentária para a indexação de artigos científicos: princípios de elaboração e uso para a formação de indexadores. *DataGramaZero – Rev. Ci. Inf.*, v. 7, n. 3, jun. 2006. Disponível em: <http://www.dgz.org.br/jun06/Art_04.htm>. Acesso em: 22 abr. 2010.

FUJITA, M. S. L. (2003). *A leitura documentária do indexador: aspectos cognitivos e linguísticos influentes na formação do leitor profissional*. 321f. 2003. Tese (Livre-Docência em Análise Documentária e Linguagens Documentárias Alfabéticas) – Faculdade de Filosofia e Ciências, UNESP.

GIL LEIVA, Isidoro; RUBI, Milena P.; FUJITA, Mariângela S. L. Consistência na indexação em bibliotecas universitárias. *Transinformação*, Campinas, v. 20, n. 3, p. 233-253, set./dez. 2008.

HEUSELER, Fábio M. Uma abordagem multifacetada para exploração integrada de dados estruturados e não-estruturados em ambientes OLAP. 2010. 131f. Dissertação (Mestrado em Informática) – Instituto de Matemática, Núcleo de Computação Eletrônica, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2010.

HJORLAND, B.; ALBRECHTSEN, H. Toward a new horizon in information science: domain analysis. *Journal of the American Society for Information Science-JASIS*, v.46, n.6, p. 400-425, 1995.

KOBASHI, N. Y. *A elaboração de informações documentárias: em busca de uma metodologia*. 1994. Tese (Doutorado em Ciências da Comunicação) - Escola de Comunicações e Artes, Universidade de São Paulo, São Paulo.



XI Encontro Nacional de Pesquisa em Ciência da Informação
Inovação e inclusão social: questões contemporâneas da informação
Rio de Janeiro, 25 a 28 de outubro de 2010

LAMBE, Patrick. *Organising Knowledge: taxonomies, knowledge and organizational effectiveness*. Oxford, England: Chandos, 2007.

MARCONI, M. A.; LAKATOS, E. M. *Técnicas de pesquisa*. 2. ed. São Paulo: Ed. Atlas, 1990.

NATIONAL INFORMATION STANDARDS ORGANIZATION. *ANSI/NISO Z39.19-2005: guidelines for the construction, format, and management of monolingual thesauri*. 2005.

PRAZERES, C. V. S. *et al.* Uma proposta de navegação na web utilizando facetas. In: ERBD'06: II Escola Regional de Banco de Dados. *Anais do ERBD 2006*. Passo Fundo, 2006, v. 1, p. 1-6.

RABELLO, Rodrigo. A face oculta do documento: tradição e inovação no limiar da Ciência da Informação. 2009. 331f. Tese (Doutorado em Ciência da Informação), Faculdade de Filosofia e Ciências da Universidade Federal Paulista, UNESP, São Paulo, 2009.

SMIT, Johanna W.; KOBASHI, Nair Y. Como elaborar vocabulário controlado para aplicação em arquivos. São Paulo: Arquivo do Estado, Imprensa Oficial, 2003. (Projeto Como Fazer, 10).

SOUZA, Vanessa I. de. *Indexação: teorias e práticas do corpo indexador da Universidade Federal do Rio Grande do Sul*. 2009. 36f. Monografia (Especialização). Universidade Federal do Rio Grande do Sul. Faculdade de Biblioteconomia e Comunicação. Curso de Especialização em Gestão de Bibliotecas Universitárias. Porto Alegre: UFRGS, 2009.

TUNKELANG, Daniel. *Faceted search*. Carolina do Norte: Morgan e Claypool, 2009.

VIEIRA, Simone B. *Indexação automática e manual: revisão de literatura*. Ciência da Informação, v. 17, n. 1, 1988.