

PROCESSOS DE INDEXAÇÃO E ANÁLISE DE ASSUNTO: UMA ABORDAGEM BASEADA NA AVALIAÇÃO DOS FATORES INTERVENIENTES NESTES PROCESSOS¹

Elizabeth Andrade Duarte *

Resumo

O presente artigo promove reflexão à indexação e a análise de assunto em termos dos processos e características que os envolvem, considerando o usuário como parte do produto final. O texto se divide em três partes: a indexação – conceituação e caracterização, a análise de assunto e a representação da informação. Destaque para a indexação enquanto processo intelectual e subjetivo que tem como uma das suas etapas a análise de assunto. Aborda a recuperação da informação eficaz como consequência dos dois processos bem trabalhados.

PALAVRAS-CHAVE:

**INDEXAÇÃO
ANÁLISE DE ASSUNTO
RECUPERAÇÃO DA INFORMAÇÃO**

1 INTRODUÇÃO

A literatura na área da Ciência da Informação mostra que a indexação é uma forma de representação do conhecimento que faz uso da análise de assunto para identificar e selecionar os conceitos que representam a essência de um documento. Consiste em uma das etapas mais importantes do trabalho do indexador, isto porque, objetiva representar o conteúdo de um documento por meio de termos extraídos de um texto original. Para se atingir a recuperação da informação é necessário um trabalho criterioso de análise dos documentos para extrair os assuntos que deverão ser indexados.

Sendo assim, o objetivo deste artigo é sistematizar as principais contribuições da indexação e da análise de assunto em termos da ciência da informação, proporcionando uma visão geral da área apresentando ao leitor a relevância da mesma.

Neste contexto, pode-se ressaltar que a identificação de conceitos realizada durante a leitura documentária, envolve esforço adicional de compreensão de texto e as dificuldades apresentadas possibilitaram a motivação para investigar, por meio de revisão de literatura, a identificação de conceitos a partir da leitura documentária, da tematicidade e das concepções de análise de assunto. Pelas análises obtidas, considera-se que as concepções de análise orientadas para o conteúdo e para a demanda são

¹Trabalho apresentado à disciplina Estudo Especial, ministrada pelo Professor Doutor Eduardo Wense Dias. Curso de Mestrado do PPGCI da UFMG

* Mestre em Ciência da Informação pelo Programa de Pós-Graduação em Ciência da Informação da UFMG. Bibliotecária Coordenadora do Sistema Integrado de Bibliotecas da Faculdade Pitágoras de Belo Horizonte, MG. Pós-Graduada em Didática do Ensino Superior pela Pós-Graduação da Faculdade Pitágoras. calazbedt@yahoo.com.br

decisivas para o bom desempenho do indexador durante a leitura para a identificação de conceitos na medida em que se vincula à busca pela tematicidade do texto.

2 INDEXAÇÃO - CONCEITUAÇÃO, CARACTERIZAÇÃO

A indexação enquanto um processo de análise documentária, definido como ato de descrição ou identificação em um documento de termos de seu conteúdo temático, ou seja, consiste na representação dos documentos por meio de termos (descritores ou palavras-chave) extraídos tanto do texto original quanto dos termos escolhidos através de uma linguagem de informação ou indexação. É importante ressaltar que a indexação trata-se de um processo subjetivo, pois lida com análise, interpretação e definição do que será indexado, ou seja, com a tomada de decisão, envolvendo inclusive o contexto para o qual o documento está sendo indexado.

Nesse âmbito, é a atribuição de descritores que permite que se faça uma busca por assunto, ou seja, reflete diretamente na recuperação da informação. Desse modo, é considerável destacar que a recuperação da informação eficaz é resultado de uma indexação de qualidade com identificação de conceitos (termos) mais pertinentes com o conteúdo de um documento, produzindo uma correlação precisa entre o assunto pesquisado nos índices de busca.

Como fatores que influenciam na qualidade da indexação Lancaster (2004, p. 81) destaca: fatores ligados ao indexador (conhecimento do assunto - conhecimento por parte do indexador sobre o conteúdo temático tratado e entendimento da terminologia abordada, das necessidades dos usuários, experiência, concentração, capacidade e compreensão de leitura), fatores ligados ao documento (conteúdo temático, complexidade, língua e linguagem, expressão, apresentação e sumarização), fatores ligados ao vocabulário (especificidade, ambigüidade, qualidade da estrutura, qualidade do vocabulário, disponibilidade de instrumentos auxiliares fins), fatores ligados ao processo (tipo de indexação, regras e instruções, produtividade exigida, exaustividade da indexação) e fatores ambientais (refrigeração, iluminação, ruído). Todos esses fatores quando bem aplicados refletem na recuperação da informação de qualidade por parte do usuário.

Nesse âmbito, Chaumier (1988) argumenta que a indexação enquanto operação que consiste na descrição e caracterização de um documento com o auxílio da representação dos conceitos nela contidos, caracteriza-se como importante parte da análise documentária, condicionando o valor de um sistema documentário. O mesmo autor enfatiza que a mesma pode ser compreendida dentro de três aspectos básicos que são: o seu conceito, as suas etapas e os instrumentos que utiliza.

Em termos de estruturação clara, a indexação possui como etapas: *o conhecimento do conteúdo do documento* (leitura rápida do documento através do título, subtítulo, introdução, conclusão, dentre outros), *a escolha dos conceitos a serem representados* - consistindo na análise conceitual do documento para condensar o documento (destaque para a seletividade e para exaustividade), *a tradução dos conceitos escolhidos e incorporação de elementos sintáticos eventuais* (levantamento dos descritores adequados considerando a importância dos conceitos em relação ao conjunto do documento a indexar). Nesta última etapa questionamentos importantes devem ser considerados: fontes dos termos de indexação, precisão, fidelidade e linguagem do usuário, uma vez que, todos interferem na representação dos conteúdos a serem indexados.

Como Chaumier (1988) considera a indexação como uma operação essencial para que se possa recuperar documentos do acervo documentário, respondendo de forma

adequada e eficaz a todo pedido ou questão solicitada por um usuário, sem que ocorram ruídos ou silêncios, o mesmo destaca a relevância dos então considerados instrumentos de indexação na composição deste processo, podendo então citar os sistemas de classificação e o thesaurus. Os tesouros são considerados importantes instrumentos que possibilitam o controle terminológico e uniformização das linguagens utilizadas por indexadores e usuários quando construídos, preferencialmente, conforme a realidade na qual se inserem.

De acordo com Silva e Milidiú (1991) a “Teoria de Funções de Crença” envolvendo o conteúdo semântico de documentos/consulta do usuário à base de documentos (representados por funções de crença distintas) e as propostas na busca de conveniente sistema de representação do conteúdo sistemático de documentos atuam com o uso de termos: “associados manualmente por leitor que lê o documento, de termos que aparecem no título do documento com o conceito de KWIC e KWOC, dos que aparecem nos nomes das seções do documento e dos que são estatisticamente significantes que aparecem no documento”. Para os autores esta função de crença possibilita a indexação automática e escalonamento de documentos e calcula o grau de concordância de cada documento com uma dada consulta ponderada do usuário, sempre se baseando na representação semântica de cada documento. Já Pinto (2001) considera a indexação como sendo:

Uma atividade que desconstrói o discurso montado pelo autor do conhecimento, à medida que faz recortes nestes discursos. A mesma coloca em cena três atores: o autor do documento, a indexação e o usuário. Para cada um a noção de pertinência informacional será percebida diferentemente segundo as suas experiências, os seus interesses no momento da produção ou da leitura do texto. Ao indexador e usuário é desejável que a taxa de cobertura dos termos seja ótima, pois isto influenciará a pertinência das respostas fornecidas no momento da recuperação da informação.

Essas considerações mostram que a indexação não é um processo isolado, sob a responsabilidade de somente um ator, mas de um conjunto de atores co-responsáveis. É como a autora salienta em suas análises: qualquer que seja o método de indexação utilizado, a indexação deverá permitir aos clientes o acesso ao documento que contém a informação de que necessitam. O seu resultado constituirá em um dispositivo chave entre o documento primário a ser lido e compreendido, e um documento secundário (índice) a ser construído, de tal modo que a representação dos elementos indicadores do conteúdo do primeiro sejam encontrados no segundo documento de maneira mais completa e fiel possível.

Enquanto processo, a indexação possui três fases importantes para destaque que são: o exame do documento e estabelecimento do assunto de seu conteúdo (compreensão do conteúdo do documento); a identificação dos conceitos presentes no assunto; e a tradução desses conceitos nos termos de uma linguagem de indexação.

De acordo com Neves; Dias; Pinheiro² a indexação envolve a leitura de documentos, a identificação do conteúdo de um documento, a geração de respectivos textos dos assuntos/resumos. Já a análise de assunto é a etapa em que um documento é analisado para determinar de qual ou de quais assuntos trata.

Relacionando esse processo com questões do processo de leitura os autores argumentam que conhecimentos anteriores do indexador facilitam o processamento de um texto, bem como de sua compreensão. Assim, a leitura envolve processos como construção de imagens, monitoramento da compreensão, releitura e interpretação.

Os mesmos ressaltam ainda que o indexador necessita de uma leitura que exija a incorporação de uma série de atitudes, exigindo um profissional com adaptação

freqüente. Então é importante ressaltar que durante a leitura, os indexadores variam na sua expressão de julgamento sobre o texto e seu autor, constroem hipóteses (importância da experiência de uma leitura em lidar com a organização textual) e um procedimento marcado produz maior número de comentários do que um não marcado.

3 ANÁLISE DE ASSUNTO

Conforme argumenta Foskett (1973), a análise de assunto envolve paradigmas (um assunto pode incluir outro, fazer parte de outro ou ser da mesma categoria de outro) e sintagmas (são relações direcionais, indica relações entre os assuntos para se conseguir especificidade), sinônimos, homógrafos (uma mesma grafia é usada à vezes para diferentes palavras que podem ser pronunciadas de modo idêntico ou não), um palavras ou mais de uma, assuntos compostos, linguagens de indexação (vocabulário controlado, linguagem artificial, sintaxe, linguagem natural) sistemas pré-coordenados e pós-coordenados, sistemas abertos e sistemas fechados, enumeração e síntese, bem como a classificação crítica (imposição ao usuário do ponto de vista do indexador).

Para o mesmo autor uma coleção de documentos indexados não se modificará caso exista modificação no método de indexação, tampouco sofrerão mudanças às demandas feitas pelos usuários, isto porque, independente do sistema adotado, os critérios de avaliação deverão ser os mesmos. Conforme o autor argumenta, uma coleção indexada não se modificará caso exista modificação no método de indexação, como tampouco sofrerão mudanças as demandas feitas pelos usuários. Independente do sistema adotado, os critérios de avaliação deverão ser os mesmos.

Cesarino; Pinto (1980) consideram a análise de assunto como a operação base para processos de recuperação de informações. Segundo as autoras, no processo de análise de documentos ao se dar entrada em um sistema de recuperação de informações os documentos são analisados de duas maneiras: bibliográfica ou objetivamente e intelectual ou subjetivamente. A primeira tem foco para a descrição do documento em termos de suas características físicas. A segunda direciona para a descrição do documento em termos de seu conteúdo informativo (assunto abordado). As mesmas autoras apontam que no estabelecimento do assunto de um documento têm-se como etapas:

A compreensão do texto como um todo (relacionada à forma física) – atenção para título, subtítulos, introdução, capítulos, conclusões, ilustrações, tabelas, etc. A identificação de conceitos: seguir procedimento lógico na seleção dos conceitos que melhor expressarão o assunto do documento. A escolha dos conceitos poderá espelhar a configuração das categorias fundamentais reconhecidas como importantes para o assunto abordado pelo documento. E a seleção de conceitos: válidos para serem indexados. Dependerá exclusivamente do objetivo para o qual o documento está sendo indexado.

Para Dias (2004) a análise de assunto consiste em etapa do tratamento temático da informação na qual um documento a ser incorporado a uma coleção é analisado para se determinar de qual ou quais assuntos trata. De acordo com o autor, estudos sobre tal análise são feitos sob três principais perspectivas: a dos profissionais de informação, a dos usuários e a dos autores. A primeira é feita por classificadores e catalogadores, baseando-se no fato de que são estes que comandam o processo de análise de assunto e que têm conhecimento e vivência para a execução de tal processo. A segunda baseia-se no pressuposto de que os usuários são importante fonte para subsidiar ações que vivem o aperfeiçoamento do processo de análise de assunto. Já a terceira e última,

fundamenta-se na lógica de que ninguém melhor do que o próprio autor par identificar do que trata o documento que produziu.

Segundo Kobashi (1996) a análise documentária pode ser compreendida sob a ótica da organização e da representação da informação organização. Essa representação documentária considera que um texto enquanto objeto, envolve as seguintes abordagens: a análise de conteúdo, a análise do discurso e a lingüística do texto, a análise logicista e a análise de texto na perspectiva da inteligência artificial. Em termos da análise de conteúdo, pode-se considerar um conjunto de procedimentos de análise e interpretação de textos de natureza diversa baseados na descrição e quantificação de palavras, termos ou frases desses textos. Tem-se aqui uma relação com a análise do discurso e da lingüística do texto. Estas se constituem na teoria que integra por meio da enunciação, uma abordagem interna do texto, indispensável para que se reconheçam os mecanismos e regras de engendramento do discurso.

A análise logicista faz uso da metodologia de análise de textos científicos produzidos pelas ciências humanas na qual a representação faz-se por meio de um sistema simbólico cujo produto se expressa sob forma de proposições, que articulam dados e raciocínios do texto. Finalmente a análise de texto na perspectiva da inteligência artificial é uma área que se preocupa com o comportamento inteligente e com sua simulação computacional. Procura descobrir formas de programar computadores para que os mesmos possam executar as funções que caracterizam a inteligência humana. O autor acredita que a análise documentária é capaz de aproximar de outras categorias de análise de textos, sempre propondo elaborar diversas categorias de informações sob a ótica de novos textos (resumos) ou de representações padronizadas (tesauros, sistemas classificatórios).

Em termos da indexação automática, Rodredo (1991) considera que cada vez mais esta ocupa espaço em obras de caráter geral (também livros didáticos) e não somente preocupação exclusiva de pesquisadores, na qual a abordagem de seleção por Exclusão permite desenvolver mecanismos simples e eficazes com grande versatilidade de aplicação. A associação dos avanços da inteligência artificial aos métodos de indexação automática tem aberto grandes perspectivas à aplicação de técnicas de análise e indexação automática de textos.

Considerando a indexação uma ação capaz de identificar e descrever um documento de acordo com seu assunto, Naves (2006) ressalta que os conceitos são extraídos do documento através de processo de análise e traduzidos para os termos de instrumentos de indexação (tesauros, listas de cabeçalhos de assunto, esquemas de classificação, etc), existindo um destaque para a análise de assunto, etapa esta considerada mais importante no trabalho do indexador (atividade intelectual e subjetiva).

De acordo com Naves (1996) na análise de assunto consiste no processo de extrair conceitos que traduzam a essência de um documento (análise de assunto ou análise temática). A mesma enfatiza que:

A compreensão diferentes concepções e abordagens sobre a análise de assunto é válida na medida em que serve como subsídio par estudos na área de tratamento da informação, mais especificamente em indexação. E a presença da subjetividade do analista, mais imprecisão terminológica torna a análise de assunto tema merecedor de atenção por especialistas, lingüistas e demais interessados.

Isto porque a análise de assunto, enquanto importante etapa da indexação lida com fatores cognitivos, lingüísticos, lógicos, já que é um processo desenvolvido por um

indivíduo que conforme Naves afirma segue um raciocínio lógico, numa determinada língua, tudo isso regado de uma grande dose de subjetividade, conforme já destacado anteriormente.

4 RECUPERANDO A INFORMAÇÃO - PROCESSO

Para Naves (1998) a classificação de assuntos pode ser considerada importante instrumento de organização de livros nas estantes e sua relação à técnica de busca chamada “browsing” que consiste no processo de exploração entre pilhas e pilhas de documentos os quais podem ser examinados pela utilização dos sentidos.

Tal técnica envolve a interação entre o estoque de informação do indivíduo e a assimilação da nova informação. O autor defende a idéia de que a classificação é uma ferramenta que favorece o browsing, isto porque o Browsing: técnica de busca que necessita das classificações, tanto na ordenação dos livros nas estantes, quanto no arranjo de registros bibliográficos em bases de dados automatizadas.

5 CONSIDERAÇÕES FINAIS

A indexação é um processo intelectual que envolve atividades cognitivas na compreensão do texto e a composição da representação do documento. A análise de assunto ao identificar e selecionar conceitos que representam a essência de um documento necessita atuar com o foco no usuário. Portanto, para se recuperar uma informação é preciso que dois processos anteriores sejam bem trabalhados: a organização da informação e a sua recuperação. De tal forma que uma entrada de dados ineficaz implicará uma saída de dados também ineficaz. Cabe aqui, o desenvolvimento de uma indexação de qualidade, permitindo aos clientes (usuários) o acesso ao documento que contém a informação que ele necessita.

INDEXING AND SUBJECT ANALYSIS: AN APPROACH BASED ON THE EVALUATION OF INTERVENING FACTORS IN THESE PROCESSES

ABSTRACT

The article promotes reflection to the indexing and the subject analysis in terms of the processes and characteristics that involve them, considering the user as part of the final product. The text has been divided in three parts: the indexing – conceptualization and characterization, the subject analysis and the representation of the information. Detach for the indexing while an intellectual and subjective process that has as one of their stages the subject analysis. It approaches the recovery of the effective information as a consequence of the two processes worked.

KEYWORDS:

**INDEXING
SUBJECT ANALYSIS
INFORMATION RETRIEVAL**

REFERÊNCIAS

- CESARINO, Maria Augusta da Nóbrega; PINTO, Maria Cristina Mello Ferreira. Análise de assunto. *Revista de Biblioteconomia de Brasília*, Brasília, v. 8, n.1, jan./jun. 1980. p. 32-43.
- CHAUMIER, Jacques. Indexação: conceito, etapas e instrumentos. *Revista Brasileira de Biblioteconomia e Documentação*, São Paulo, v. 21, n.1/2 p. 63-79, jan./jun. 1988.
- DIAS, Eduardo Wense. Análise de assunto: percepção do usuário quanto ao conteúdo de documentos. *Perspectivas em Ciência da Informação*, Belo Horizonte, v. 9, n.2, jul./dez. 2004. p.146-157.
- DIAS, Eduardo Wense, NAVES, Madelena Martins Lopes; MOURA, Maria Aparecida. O usuário-pesquisador e a análise de assunto. *Perspectivas em Ciência da Informação*, Belo Horizonte, v. 6, n. 2, p. 2005-221, jul./dez. 2001.
- FUJITA, Mariângela Spotti Lopes. A identificação de conceitos no processo de análise de assunto para a indexação. *Revista Digital de Biblioteconomia e Ciência da Informação*, Campinas, v. 1, n.1, 2003. Disponível em:< <http://server01.bc.unicamp.br/seer/ojs/viewarticle.php?id=9&layout=abstract>>. Acesso em: 10/12/06.
- KOBASHI, Nair Yumiko. *Informare – Cadernos do Programa de Pós-Graduação em Ciência da Informação*, Rio de Janeiro, v.2, n.2 p.5-27, jul./dez. 1996.
- LANCASTER, F. W. *Indexação e resumos: teoria e prática*. 2. ed. rev. atua. Brasília: Briquet de Lemos, 2004.
- NAVES, Madalena Martins Lopes. Análise de assunto: concepções. *Revista de Biblioteconomia de Brasília*, Brasília, v. 20, n.2, jul./dez. 1996. p. 215-226.
- NAVES, Madalena Martins Lopes. Aspectos conceituais do *browsing* na recuperação da informação. *Ciência da Informação*, Brasília, v.27, n.3 p.279-283, set./dez. 1998.
- NEVES, Dulce Amélia de Brito; DIAS, Eduardo Wense; PINHEIRO, Ângela Maria Vieira Pinheiro. Aspectos cognitivos na leitura do indexador. Artigo a ser publicado.
- PINTO, Virgínia Bentes. Indexação documentária: uma forma de representação do conhecimento registrado. *Perspectivas em Ciência da Informação*, Belo Horizonte, v.6, n.2 p.223-234, jul./dez. 2001.
- RODREDO, Jaime. Indexação automática de textos: uma abordagem otimizada e simples. *Ciência da Informação*, Brasília, v. 20, n.2, jul./dez. 1991. p. 130-136
- SILVA, Wagner Teixeira da; MILIDIÚ, Ruy Luiz. *Ciência da Informação*, Brasília, v.20, n.2 p.155-164, jul./dez. 1991.
- Biblionline, João Pessoa, v. 3, n. 2, 2007