

## A prática de indexação: análise da evolução de tendências teóricas e metodológicas<sup>1</sup>

### *The indexing practice: a development analysis of theoretical and methodological trends*

Maria dos Remédios da SILVA<sup>2</sup>

Mariângela Spotti Lopes FUJITA<sup>3</sup>

#### R E S U M O

Almejando contribuir com uma reflexão sobre a prática do indexador e tendo em vista a evolução do processo de indexação, os autores procuram identificar, pela revisão da literatura, o desenvolvimento teórico e metodológico de tal processo, extraíndo as principais tendências e influências ali presentes. A revisão da literatura foi sistematizada por duas análises: a da literatura fundamental e a dos relatos de experiência. A primeira análise divide-se em três categorias: definição, história e aspectos teóricos e metodológicos da indexação, as quais permitiram evidenciar a constante preocupação, por parte dos estudiosos, com relação à abordagem do assunto do documento, tendo em vista sua recuperação. Com isso, conclui-se que a maior ênfase nesse processo está relacionada à indexação acadêmica – resultante da concepção de análise de assunto assumida pelo indexador. Este segundo tipo de análise tem suas influências mais diretamente relacionadas às áreas de interface da Análise Documentária – a Lingüística, a Lógica e a Psicologia Cognitiva, mais recentemente abordada pelos estudos teóricos, ainda não aprofundados.

<sup>1</sup> Parte integrante do Projeto integrado "Leitura em análise documentária: uma contribuição à formação do indexador", sob coordenação de M.S.L. FUJITA. Apoio do CNPq.

<sup>2</sup> Mestranda, Curso de Pós-Graduação em Ciência da Informação, Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, SP, Brasil. E-mail: <mariaarsilva@yahoo.com.br>.

<sup>3</sup> Livre-Doente, Departamento de Ciência da Informação, Faculdade de Filosofia e Ciências, Universidade Estadual Paulista. Av. Hygino Muzzi Filho, 737, Campus Universitário, 17525-900, Marília, SP, Brasil. Correspondência para/Correspondence to: M.S.L. Fujita. E-mail: <goldstar@flash.tv.br>.

Recebido em 8/7/2003 e aceito para publicação em 28/6/2004.

Quando estes aspectos lógicos, lingüísticos e cognitivos estiverem mais esclarecidos, permitirão ao indexador a realização de uma análise conceitual do documento mais efetiva, pois são estes aspectos que vão auxiliar na determinação da tematicidade do documento.

**Palavras-chave:** indexação, história da indexação, metodologia de indexação, análise documentária, leitura documentária.

## ABSTRACT

*To contribute with a reflection on the indexer's practice, having in mind the evolution of the indexing process, the authors of this paper sought to assess the theoretical and methodological development of such process, through a survey of the literature, extracting from it the main trends and influences. The scrutiny of the relevant literature was systematized in two types of analyses: one, the analysis of fundamental literature in the area, and the other, that of experience stories. The first analysis was divided in three categories: theoretical and methodological definition, history, and aspects of indexing. These have allowed the authors to substantiate a constant concern on the part of the scholars, regarding the approach to the document as a subject and having in mind its recovery. The authors conclude that the emphasis in this process rests in the academic indexing-resultant of the outset of subject analysis carried out by the indexer. This second type of analysis is more directly influenced by the interfacing areas of Documentation Analysis - Linguistics, Logic and more recently, of sciences such as the Cognitive Psychology, which was approached in theoretical studies that did not prove deep enough yet. Once the logical, linguistic and cognitive aspects affecting document contents be better clarified, they will assist the indexing professionals in determining the topics actually involved in a document, and will allow them to accomplish a more effective conceptual analysis of such document.*

**Key words:** *indexation, history of the indexation, methodology of indexation, documentary analysis, documentary reading.*

## INTRODUÇÃO

Temos constatado pela história da Indexação que o ato de construir índices define-se como uma prática bastante antiga no processo de tratamento de documentos. A atividade de indexação, como processo, é realizada mais intensamente desde o aumento das publicações periódicas e da literatura técnico-científica, surgindo a necessidade de criação de mecanismos de controle bibliográfico em centros de

documentação especializados. Dessa forma, o aparecimento de mecanismos de controle bibliográfico aconteceu fora do âmbito das bibliotecas tradicionais, representando uma evolução no processo de tratamento da informação e dando origem teórico-prática, naquela ocasião, a uma nova área, a Documentação.

Dentro da perspectiva evolutiva do tratamento da informação, está vinculada a Indexação como operação do tratamento temático que comporta a análise, síntese e representação. A

leitura documentária representa principal fase dessa operação, por ser inicial e desencadear todas as posteriores. Dessa forma, exerce uma influência contundente no que concerne à qualidade das outras fases, resultando nos produtos desse processo.

Nas pesquisas sobre leitura documentária, desenvolvidas por Fujita (1999; 2003), para observação da leitura do indexador, verifica-se uma constante preocupação com os procedimentos de análise do documento. Os resultados revelaram que o leitor indexador apresenta dificuldades quanto à identificação e seleção de conceitos representativos do assunto do documento. Tal constatação proporcionou a elaboração de diretrizes que contém proposta de um modelo de leitura documentária para textos científicos, combinando as estratégias de exploração da estrutura textual e abordagem sistemática de identificação de conceitos, visto que a compreensão de como se dá essa leitura poderá auxiliar o leitor-indexador a obter uma análise conceitual efetiva, que represente adequadamente o assunto tratado no documento.

As pesquisas realizadas no âmbito dessas pesquisas, também tiveram como objetivo investigar os procedimentos de leitura e estratégias de leituras com leitores indexadores especialistas e não-especialistas. O diagnóstico apontou resultados similares para ambos os casos. No entanto, os indexadores especialistas foram mais rápidos na realização da tarefa de indexar por dominarem a área de assunto dos documentos, ao passo que os não especialistas fizeram mais usos de estratégias, em especial a associação com a linguagem do sistema de indexação.

Outro aspecto a ser lembrado é que a análise de assunto feita com base unicamente na linguagem do sistema poderá acarretar problemas para a verdadeira representação do conteúdo do documento; um deles é a incompatibilidade entre a terminologia empregada no documento e a linguagem do sistema. Lancaster

(1993) esclarece que a terminologia usada pelo autor pode não corresponder exatamente aos termos do vocabulário controlado, o que pode ocasionar uma atribuição errônea do termo. Complementa, ainda, que devem ser indexadas as idéias do autor do texto e não as palavras.

Consideramos que esses aspectos estão, provavelmente, relacionados à falta de consolidação teórica da área de Indexação, uma vez que está muito articulada com o desenvolvimento da prática. Por isso, propõe-se realizar a identificação e análise das tendências na literatura publicada em Ciência da Informação sobre o desenvolvimento teórico e metodológico de indexação com ênfase em leitura documentária, tendo como objetivo contribuir com uma reflexão sobre a prática do indexador.

A metodologia adotada constituiu-se de revisão da literatura sistematizada por duas análises: a de literatura fundamental e outra de relatos de experiência. Denominamos revisão de literatura fundamental aquela que enfocou os textos sobre a temática em estudo e, identificou variáveis do processo de indexação para posterior comparação com a literatura de relatos de experiência sobre prática de indexação, a fim de, relacioná-los com os conteúdos teóricos-metodológicos.

Para a revisão de literatura da prática de indexação foram selecionados 10 textos (artigos) contendo relatos de experiência de sistemas de informação brasileiro. Os textos selecionados foram separados por décadas e o conteúdo foi analisado para identificar as variáveis influentes na leitura documentária anteriormente ressaltadas pela revisão de literatura fundamental. Cada artigo foi descrito fisicamente por referência bibliográfica seguida de uma descrição baseada no resumo de cada texto, com acréscimos de informações retiradas de seu conteúdo e, quando não havia resumo a descrição foi realizada com base no texto. Em seguida à descrição, procedeu-se à análise para identificação das variáveis apontadas na literatura fundamental e

discussão decorrente da categorização das variáveis, cuja avaliação com a literatura de relatos de experiência, encontra-se no item resultados.

### **A Indexação em revisão de literatura fundamental**

A revisão de literatura fundamental destacou subsídios para a discussão de temas presentes na evolução da indexação e os agrupou em três categorias: “Conceituação”, “História” e “Aspectos teóricos e metodológicos de indexação”. Após essa categorização incluíram-se em “Aspectos teóricos e metodológicos da indexação” outras sub-categorias por tratar-se de um tema amplo que justifica uma abordagem mais sistematizada para compreensão dos diferentes aspectos encontrados nas evoluções da indexação e que, certamente, demonstram o seu grau de importância na formação e atuação do profissional bem como para atendimento da demanda da comunidade usuária durante a recuperação da informação.

### **Conceituação de indexação**

O conceito de indexação surgiu a partir da elaboração de índices e atualmente está mais vinculada ao conceito de análise de assunto. Com a evolução da prática, em decorrência da necessidade de recuperação cada vez mais rápida, precisa e especializada por parte de Instituições informacionais, a construção de índices passou a contar com um aparato metodológico e instrumental mais diversificado e muito mais voltado para o contexto de cada documento. Ressalta-se que, a partir da evidência da Documentação como área científica na década de 60 e do surgimento dos serviços de informação em áreas especializadas, a indexação e a elaboração de resumos utilizados na elaboração dos serviços bibliográficos para recuperação de artigos de periódicos científicos,

ganham espaço e notoriedade mantidos até hoje. Bradford (1961), em seu clássico livro, “Documentação”, até hoje considerado como referência para a Documentação, destaca indexação para a análise de documentos. O termo, então, passou a ter um significado mais abrangente.

No âmbito da Análise Documentária, segundo a linha teórica de Gardin, a Indexação é vista como uma operação de representação documentária com a finalidade pragmática de Recuperação da Informação. Contudo, sob a perspectiva de outros teóricos, principalmente ingleses e norte-americanos, a Indexação é a própria Análise documentária, composta das mesmas etapas operacionais com o objetivo de representação do conteúdo informacional de documentos para a elaboração de índices.

A partir da evolução que determinou a importância do contexto do documento para a recuperação da informação, a área de indexação passa a incorporar os estudos dirigidos à compreensão do conteúdo dos textos a serem analisados. Esses estudos, porém, estão claramente inseridos em correntes teóricas e é fácil confundir, na literatura, a função da indexação perante a necessidade de análise de conteúdo. Na literatura observa-se a existência de duas correntes teóricas: a francesa e a inglesa.

A expressão “Análise Documentária” foi formalmente conceituada por Gardin (1981, p.29) como “um conjunto de procedimentos efetuados com a finalidade de expressar o conteúdo de documentos científicos, sob formas destinadas a facilitar a recuperação da informação”.

A corrente francesa adota a expressão Análise Documentária, introduzida por Gardin (1981) e este, tem seus seguidores como: Chaumier, Kobashi, Smit, Tálamo, Ginez de Lara, Cintra, Cunha, Guimarães, Fujita, Gil Leiva, Ruiz Perez, Pinto Molina, entre outros. Segundo essa concepção, a Análise Documentária é um macro universo no qual a indexação está inserida. A indexação é, então, o resultado da fase de

representação, fase final da análise documentária, em que se utilizam as linguagens documentárias para a geração de produtos documentários (índices e notações classificatórias, etc.) (GUIMARÃES, 2000).

Para Chaumier (1980) a indexação é a parte mais importante da análise documentária. Conseqüentemente, é ela quem condiciona o valor de um sistema documentário. O autor nos adverte que uma indexação inadequada ou uma indexação insuficiente representa 90% das causas essenciais para aparição de 'ruídos' ou de 'silêncios'.

A expressão Análise Documentária, do ponto de vista dos teóricos espanhóis, comporta dois níveis de divisão: o da forma – análise descritiva ou bibliográfica – refere-se ao tratamento físico da informação ligado com o suporte; e o do conteúdo, que se refere ao tratamento temático da informação e destina-se à representação condensada do assunto intrínseco ou extrínseco tratado em um determinado documento.

Há, portanto, uma diferença da concepção francesa de Análise Documentária entre espanhóis e franceses. Para os teóricos franceses, a Análise Documentária refere-se somente ao tratamento do conteúdo do documento, não adotando a divisão em forma e conteúdo, ou descrição física e temática do documento.

A corrente inglesa, representada por autores como Foskett, Lancaster, Campos, Van Slype, Farrow, entre outros, faz o uso da expressão indexação, entendendo-a como um processo.

Segundo Borko e Bernier (1978, p.8) a indexação é definida como “o processo de analisar o conteúdo informacional dos registros do conhecimento e sua expressão na linguagem do sistema de indexação”. Para Van Slype (1991) é “a operação que consiste em enumerar os conceitos sobre os quais trata um documento e representá-los por meio de uma linguagem

combinatória: lista de descritores livres, lista de autoridades e o thesaurus de descritores”.

Na concepção da corrente inglesa, análise documentária e indexação compreendem processos idênticos, incluindo-se a análise de assuntos como etapa inicial da indexação.

A existência de diferentes correntes teóricas explica o uso de termos como análise de assuntos, análise de conteúdos documentários e análise documentária. Observa-se que, em estudos da área, os termos “Indexação”, “Indexador” e “Análise de assunto” aparecem com mais freqüência do que “Análise documentária” e “Documentalista”. A mais importante bibliografia da área, a *Library and Information Science Abstracts* (LISA), traz em seu índice de assunto os termos Indexing e Indexer, mas não *Documentary analysis ou Documentalist*, o que justifica uma boa quantidade de publicações utilizando aquela nomenclatura.

A despeito dessas divergências, advindas de correntes teóricas, é preciso considerar que, dentro de uma perspectiva histórica, a Indexação tinha uma finalidade específica de construção de índices e o termo “Indexação” se ajustava perfeitamente à atividade, porém, com o uso de tecnologias de recuperação da informação a necessidade de elaboração de índices foi sendo substituída pela necessidade de representação do conteúdo documentário por termos de indexação em decorrência da análise de assunto. O conceito de indexação, elaborado por Esteban Navarro (1999, p.70), expõe essa necessidade de maneira bem detalhada:

A indexação consiste em um processo destinado a identificar e descrever ou caracterizar o conteúdo informativo de um documento mediante a seleção das matérias sobre as quais versa (indexação sintética) ou dos conceitos presentes (indexação analítica) para sua expressão da língua natural e sua reunião em índice, com objetivo de permitir posterior recuperação dos

documentos pertencentes a uma coleção documental ou conjunto de referências documentais como resposta a uma demanda acerca do tipo de informação que este contém.

Ainda para Esteban Navarro (1999) a indexação tem como objetivo a:

[...] representação do conteúdo dos documentos que formam parte de um conjunto para garantir sua eficaz recuperação durante o processo de busca nesse grupo.

Para o autor, o processo de indexação se constrói a partir do exame tanto da atividade que é realizada durante o exercício dessa técnica, como também em um sistema de informação documentária.

Dessa forma, compreendemos a análise documentária como área teórica e metodológica com o objetivo de tratamento temático de documento que abrange as atividades de Indexação, Classificação e elaboração de resumos, considerando as diferentes finalidades de recuperação da informação.

A indexação em análise documentária, sob o ponto de vista dos sistemas de informação, é reconhecida como a parte mais importante porque condiciona os resultados de uma estratégia de busca. O bom ou mau desempenho da indexação reflete-se na recuperação da informação feita pelos índices.

Por isso, segundo o UNISIST (1981), há de se considerar a indexação sob dois pontos de vista distintos: enquanto processo que consiste em descrever e identificar um documento com ajuda de representações dos conceitos nele contidos e quanto à sua finalidade, permitindo busca e acesso à informação armazenada.

Maiores considerações a respeito do desempenho da indexação e sua importância para a recuperação demandam a verificação de como esta é operacionalizada, haja visto os procedimentos de análise existentes.

## História da Indexação

A indexação surgiu com a atividade de elaboração de índices. Gomes e Gusmão (1983, p.12) afirmam que o índice, como um instrumento de armazenagem e recuperação da informação, tem sua origem a partir do momento em que o homem passou a se preocupar em tornar acessível a informação registrada em um documento e para isso resolve ordená-la de alguma forma.

A forma mais antiga de armazenagem de informação de que se tem conhecimento foi encontrada nas tábuas de argila produzidas pela extinta Mesopotâmia no século II a.C. Nelas foi grafada uma espécie de resumo dos livros antigos considerada como forma de representação condensada do conteúdo informacional que dava acesso ao assunto dos livros (WITTY, 1973).

No histórico da indexação, Collinson (1971) indica que o primeiro tipo de indexação existente era baseado na memória. Textos célebres, como as grandes epopéias, por exemplo, eram transmitidos oralmente. Depois disso, os primeiros índices de que se têm notícia eram arranjados pela primeira sentença de cada parágrafo.

Na Biblioteca de Alexandria, organizada pela classificação de Calímaco, seu catálogo era arranjado em ordem alfabética de autores e subordinados a assuntos mais gerais. Várias obras, principalmente as histórias e peças dos grandes dramaturgos da época, eram condensadas.

No século II, Cláudio Galeno compilou *De Libris Propiis Líber*, determinando o aparecimento dos primeiros guias para obras isoladas: cabeçalhos de capítulos, tábuas de matéria ou sumários, cabeçalhos nas margens dos parágrafos, cabeçalhos descritos no alto das páginas.

No século V, a obra anônima *Apothegmata*, apresenta-se como o trabalho que mais se aproximou do índice alfabético de assunto porque consistia de uma listagem de provérbios gregos

sobre tópicos teológicos. É nessa época que as obras aparecem já arranjadas em capítulos e seções numeradas permitindo, portanto, a localização de suas partes.

No século XIV era comum a elaboração de inventários ou catálogos dos livros existentes nos mosteiros e a elaboração de listas completas era um objetivo definido.

A partir disso passaram a surgir os guias para cada livro e a esse respeito Collinson (1971, p.8) pontua, mas não descreve, alguns métodos que se fixaram e ficaram permanentes, como:

- Os cabeçalhos de capítulos de livros são antigos e os livros sempre apresentavam, desde então, sumário ou tábuas de matéria;

- Para a inserção de cabeçalhos em parágrafos nas margens, ou como entradas nos próprios parágrafos cita as obras *Short History of English Literature de Saintsbury* e *History of Greece de Bury*;

- A impressão de um cabeçalho descritivo no alto de cada página muda de página para página em alguns livros, porém, o mais freqüente é repetir apenas o título do livro ou do capítulo.

A noção de índice nessa época significou uma lista de conteúdo, lista de resumos ou várias notas e muito raramente essas listas representavam o que se conhece de índice atualmente.

Os copistas, na tentativa de esclarecer ou indicar os pontos principais do assunto tratado em trechos ou parágrafos mais longos, escreviam às margens dos livros algumas palavras ou sentenças que indicassem o conteúdo. Isso acontecia de acordo com o grau de entendimento de cada copista, sendo mantido o critério de relevância dos pontos principais tratados nos livros até que um copista fosse substituído por outro. Temos, aqui, a primeira afirmação de que a indexação realizada em épocas diversas e por pessoas diferentes diferia, também, quanto à qualidade.

Em seqüência, a necessidade de elaboração de índices, segundo Collinson (1971), apresentou-se logo que surgiu a Bíblia inglesa e a indexação surgiu, então, em grande escala em 1737 com a compilação da primeira concordância completa da Bíblia por Alexandre Cruden. Esse foi o primeiro passo para que os próximos índices adquirissem grande valor, pois relacionavam citações com sua localização no texto. O século XVII foi, portanto, o início da grande época do índice facilitada pela Reforma Protestante que possibilitou a tradução da Bíblia e, portanto, franqueada ao público em geral.

Cruden indexou a Bíblia e Johnson no século XVIII indexou a língua inglesa. Estabeleceu as passagens a serem indexadas e o termo pelo qual deveria ter entrada. Foi um trabalho feito, a partir das coordenadas de Johnson por seis escribas. É atribuído a Cruden e a Johnson o estabelecimento de verdadeiros padrões de clareza e consistência para a indexação.

Com a difusão dos procedimentos de indexação, surgiu na Alemanha a contribuição da idéia de palavra-chave na representação de um item com o sistema de escolha da "schlagwort" (palavra-chave), que representou uma melhoria para a busca de informações.

Para Kobashi (1994) a documentação como é praticada hoje, nasceu no século XVII com a edição de *Le Journal des Sçavans* publicado em Paris no ano de 1665. Tratava-se de um periódico semanal que trazia os resumos dos trabalhos científicos, filosóficos e artísticos. Esse periódico deu origem a uma série de outros posteriores de mesma natureza que surgiram na Europa. Nos séculos seguintes, XVIII e XIX, aconteceu o crescimento com mais intensidade de periódicos referenciais que atualmente encontram-se no formato eletrônico denominados base de dados.

Até o surgimento da imprensa, os índices eram a única forma de acesso aos livros encontrados nas bibliotecas dos mosteiros, a partir do

registro dos títulos dos livros. A partir de então, houve um significativo aumento da literatura que impulsionou o aparecimento de várias listas com diferentes finalidades. Konrad Gesner elaborou um repertório geral e europeu – o *Bibliotheca Universalis* – no qual relacionava cerca de 12 mil títulos de todos os livros latinos, gregos e hebraicos de seu conhecimento. Mais tarde foi publicado o índice alfabético de assunto do referido repertório, cujo nome era *Pandectarum sive partitionum uníversalium, libri XXI*.

A indexação, hoje conhecida como ação de descrever e identificar um documento de acordo com o seu assunto (UNISIST, 1981, p.84) nos permite afirmar que os índices antigos não tinham uma indexação, ou “indexação crítica”. A literatura da área registra que isso se justifica pelo fato de que as pessoas da época desconheciam o processo analítico que atualmente consiste a indexação e por esse motivo, os índices apresentavam uma entrada limitada e simples como nomes próprios ou entradas por acontecimentos diretos.

A literatura da área registra que a história da indexação tem seu início com a história da bibliografia e que a indexação teve maior atenção a partir do surgimento das publicações periódicas. Esse fato fez com que surgisse a necessidade de elaboração de uma técnica para organização por assunto do conteúdo desse tipo de publicação.

Os trabalhos que contribuíram mais significativamente para o aprimoramento desse processo, são:

- década de 40 do século XIX, apareceu o Periódico *Punch e Illustrated London News* que apresentava ainda índices pobres;

- século XIX, Willian Fredrick Poole publica *An Alphabetical Index to Subjects Trated in the Review and other Periodicals* na Universidade de Yale;

- 1882, Willian Fredrick Poole criou o *Poole's Index* onde introduziu os índices cumula-

tivos de periódicos, usando para artigos de jornais entradas de assunto representados pelas palavras-chave retiradas dos títulos dos artigos indexados;

- 1876, Melvil Dewey publica o seu índice relativo da Classificação Decimal, tido como a mais expressiva contribuição à Indexação.

É nos Estados Unidos que surgem os melhores índices de periódicos e onde foi possível observar os mais importantes progressos, enquanto que, na Europa, W.F. Poole foi responsável pelos grandes índices dos periódicos do século XIX.

Consideramos que o século XIX foi o período em que a indexação começou a apresentar um aprimoramento de sua execução e ao mesmo tempo ser apreciado pelo público, que sentia necessidade de encontrar uma fórmula para o controle da massa documental que crescia em demasia. Temos com isso instituições particulares indexando livros, índices antigos sendo refeitos, elaboração de índices retrospectivos, índices cumulativos, índices cooperativos, entre outros.

Em 1901, com o lançamento de *Reader's Guide to Periodical Literature* por H.W. Wilson, os métodos de indexação ganharam um grau de aprimoramento ainda mais significativo. Nesse periódico, cada artigo foi indexado pelo seu autor e por assunto em específico. Trazia inúmeras remissivas que ligavam um assunto a outros correlatos, mantendo, desde o início, um padrão muito alto de uniformidade e exatidão (COLLINSON, 1971, p.11).

A partir de então, surgiram muitos bons índices, tanto gerais como específicos. O aprimoramento da técnica de indexar acompanhava o aumento da publicação de livros. Como exemplo da melhoria da técnica de indexação, encontramos o índice da *Encycloepadia Britânica* considerada maior representante do aperfeiçoamento da indexação no século XIX dada a clareza e apresentação gráfica dos seus índices.



Diante do exposto, podemos concluir que o século XIX foi a grande fase da indexação, dado o aumento significativo da massa documental. A indexação evoluiu dos índices das obras isoladas para os índices de vários volumes e para os índices cooperativos e em nível internacional.

Podemos dividir a história da indexação associando-lhe trabalhos que originaram sua prática e que na maioria coincidem com a história da documentação/bibliografia, porém, os títulos de trabalhos, datas, locais, instituições e autores presentes da literatura de história da documentação diferem daqueles presentes na história da indexação.

No caso da documentação, encontramos como autores mais destacados Paul Otlet e Henri La Fontaine criadores do *Office International de Bibliographie*, em Bruxelas no ano 1892. A história da documentação tem aproximação com o que se convencionou chamar de “explosão bibliográfica” ou “explosão documental” com ocorrência no século XX, portanto, posterior ao aparecimento dos índices e da indexação.

Segundo Chaumier (1971) foi em 1931 que a palavra “documentação” começou a ser usada. Os organismos criados pelas atividades de Documentação foram o *Institute International de Documentation* (FID) e a *Union Française des Organismes de Documentation* (UFOD). Os principais instrumentos de organização documental criados, foram os sistemas de classificação bibliográfica com destaque para a Classificação Decimal Universal (CDU), os estudos para criação de sistemas classificatórios realizados em 1929 e 1933 por H.G. Bliss e R.S. Ranganathan, a criação da Classificação Decimal de Dewey (CDD), além dos repertórios documentais que incluíam as bibliografias, códigos de abreviaturas dos títulos de periódicos e catálogos bibliográficos. É importante ressaltar que La Fontaine e Paul Otlet introduziram o princípio de pré-coordenação, ao criarem na CDU o uso de dois pontos (:) para relacionar duas classes de assunto.

Apesar de atualmente presenciarmos uma evolução ainda maior da indexação, é importante considerarmos a importância do índice enquanto ferramenta de busca. Robredo (1994, p.202) classifica índice em dois sentidos: no sentido tradicional e no amplo. No primeiro sentido afirma ser uma listagem alfabética ou sistemática de tópicos que indicam a existência e localização de cada um deles num documento ou em uma coleção de documentos. No segundo sentido,

[...] um conjunto ordenado de códigos representativos de assuntos, tópicos ou conceitos (por exemplo, códigos de classificação, grafismos diversos, incluindo palavras ou frases), os quais podem servir como critérios de busca relacionando com alguma chave de acesso que permita localizar os documentos – ou suas partes ou representações – relativos a cada assunto.

O autor complementa dizendo que o índice é o mais significativo instrumento para recuperação da informação Sendo definido como uma ‘chave’ condensada que dá acesso à informação contida nos documentos, ou como uma ponte entre o conteúdo de um acervo de informação e os usuários (ROBREDO, 1994, p.244).

### Aspectos teóricos e metodológicos da indexação

Após a perspectiva histórica, realizou-se uma exploração dos princípios teóricos e metodológicos da indexação a partir da revisão da literatura fundamental, obtendo-se subsídios para discussão de temas presentes na evolução da indexação. A partir desses temas a revisão de literatura fundamental dividiu-se em: **indexação alfabética de assunto**; a **indexação coordenada**, sobretudo a pré-coordenação dos cabeçalhos de assunto; a **indexação automática** e a relação entre o indexador humano e o

automático; a **leitura em indexação** e os aspectos ligados a ela como: a) **tematicidade**; **os aspectos b) lógicos**, c) **lingüísticos** e os d) **cognitivos** que influem na identificação dos conceitos que irão representar o conteúdo do documento; e os **procedimentos de indexação**, com base nos princípios do UNISIST, normas do **International Standard Organization (ISO)** e a Associação Brasileira de Normas Técnicas (ABNT) e na abordagem sistemática de identificação de conceitos do sistema de indexação PRECIS. As explicações que se sucedem dão ênfase aos itens de maior importância.

### **Indexação alfabética de assunto**

A indexação alfabética de assunto está vinculada à determinação de cabeçalhos de assuntos e por isso é, em alguns casos, também denominada de catalogação de assuntos. Apesar das divergências sobre semelhanças e diferenças entre os termos, a indexação alfabética de assuntos e a catalogação de assuntos são equivalentes porque são resultados de um mesmo processo: a análise de assunto. As diferenças que induzem à uma distinção referem-se à utilização de linguagens documentárias distintas quanto à origem e estrutura: listas de cabeçalhos de assunto e tesouros a serem empregadas para a elaboração de diferentes instrumentos de recuperação (índices de assunto e catálogos de assunto).

Por outro lado, Lancaster (1993, p.16) argumenta que

O processo que consiste em decidir do que trata um documento e de atribuir-lhe um rótulo que represente esta decisão é conceitualmente o mesmo, que o rótulo atribuído seja extraído de um esquema de classificação, de um tesouro ou de uma lista de cabeçalhos de assuntos, que o item seja uma parte bibliográfica completa

ou parte dela, quer o rótulo seja subsequente arquivado em ordem alfabética ou em alguma outra seqüência ou, com efeito, não arquivado de modo algum, quer o objeto do exercício seja organizar documentos em estantes ou registros em catálogos, índices impressos ou base de dados legíveis por computador.

Para completar, o autor acima citado afirma que a indexação de assunto é idêntica à catalogação de assunto. A catalogação de assunto é considerada um método de expressar o conteúdo informacional do documento, usando um número limitado de descritores. O uso do termo “catalogação de assunto” é influência dos Estados Unidos da América reforçado pelo uso das listas de cabeçalhos de assunto por diversas instituições. Assim, caracterizou-se o termo catalogação de assunto, basicamente, como atribuição de cabeçalhos de assunto para representar o conteúdo total dos documentos em catálogos de bibliotecas. A origem do termo catalogação de assuntos está ligada a construção dos catálogos de bibliotecas, principalmente, do catálogo de assuntos que é organizado mediante determinação de cabeçalhos de assuntos que funcionam como enunciados de assuntos formados a partir da composição ordenada de palavras.

Milstead (1983) afirma que a catalogação de assunto e a indexação são conceitualmente a mesma atividade e são tratadas como se fossem separadas, atividades distintas. A autora assume uma posição ao discutir o assunto, ou seja, de que catalogação de assunto e indexação são a mesma coisa. Afirma que considera aquela como uma forma dessa, e se quisermos pensar o contrário será a mesma coisa. Contudo, do ponto de vista da prática atual, a autora considera que a catalogação e a indexação são de fato diferentes e esses dois ramos da análise da informação podem aprender um com o outro.

Entendemos que a catalogação de assuntos deve ser considerada como um ramo dentro do âmbito maior que é a indexação alfabética de assuntos, pois existem cem anos de estudos realizados em torno da indexação alfabética de assunto desde a publicação, em 1876, da obra básica de Charles Ammi Cutter *Rules for a dictionary catalog* até a idealização do sistema de indexação PRECIS por Derek Austin em 1974.

Os cabeçalhos de assunto são a primeira tentativa de sistematização para a representação alfabética de assuntos. O aparecimento dos cabeçalhos de assuntos foi determinado pelos seguintes fatores: os títulos das obras não representavam de forma adequada o assunto nelas tratado; problemas relativos às subdivisões de assunto; existiam obras com mais de um assunto; a interdisciplinaridade dos assuntos em uma mesma obra; obras que relacionavam os assuntos a lugares e épocas diferentes.

Os princípios básicos que nortearam Cutter a estabelecer um cabeçalho de assunto estão pautados em princípios fundamentais que, segundo (CESARINO; PINTO, 1978, p.274), podem ser resumidos em: a) princípio específico: os assuntos são representados pela entrada do termo mais específico, desconsiderando a classe de assunto a qual está subordinado; b) princípio de uso: prevê a necessidade do usuário; c) princípio sindético: desenvolvimento nas listas de cabeçalhos de assunto de estruturas sindéticas, entendidas como uma rede de remissivas cruzadas para superar o problema de entrada alfabética dos cabeçalhos de assunto.

Foskett (1973) aponta alguns problemas referentes à metodologia de indexação proposta por Cutter, como:

- a entrada dupla para superar necessidade de especificidade revelava uma prática que prejudicava a facilidade do uso e o desenvolvimento de índices alfabéticos;
- o catálogo alfabético de assunto não se presta ao tipo de estratégias de pesquisa que envolve a procura sistemática de cabeçalhos.

Cutter, também, limitava o uso de remissivas descendentes, ou seja, a relação de assuntos genéricos para assuntos específicos e vice-versa. Contudo, a obra de Cutter representa o primeiro estudo dos problemas relativos à indexação alfabética de assunto, quanto às implicações de entradas específicas de assunto e as dificuldades da aplicação dos termos simples, compostos e geográficos.

Por uma síntese evolutiva, abaixo demonstrada, é possível observar que, embora a preocupação principal dos estudos seja o produto final, ou seja, a geração do índice, a análise que envolve a transformação do conteúdo em índice está expressa tanto pela proposição de categorias quanto pelos sistemas para indexação, como se observa na evolução dos principais estudos teóricos:

- Kaiser (1911) – com a publicação do trabalho *Systematic Indexing* propõe a análise de assuntos compostos pela combinação de três categorias: um “concreto”, um “processo” e “lugar”;

- Ranganathan (1965) (1933 – primeira edição de *The Colon Classification*) – desenvolveu um esquema de classificação baseado na análise de facetas e o uso de cinco categorias fundamentais: Personalidade, Matéria, Energia, Espaço e Tempo;

- Coates (1960) – em seu livro “Subject Catalogues”, apresenta a formulação de cabeçalhos de assunto específicos por categorias: coisa – parte – matéria – ação;

- Metcalfe (1959) – admite que a entrada deve ser direta e discute o propósito da catalogação de assunto como sendo o de indicar somente a classe de assunto em que está inserido;

- Lynch (1973) – criou e desenvolveu os índices articulados de assunto num estudo de índices para o Chemical Abstracts;

- Farradane (1977) – idealizou um sistema de indexação que adota nove operadores relacionais, para indicar as relações entre termos

em etapas de discriminação no tempo e no espaço;

- *Postulated-based Permuted Subject Indexing Language* (POPSI), idealizado por Neelameghan e Gopinath (1975), é um sistema inteiramente baseado em princípios classificatórios e que utiliza cabeçalhos de classificação como termos de entrada na produção dos índices cuja padronização é derivada das categorias da classificação de dois pontos de Ranganathan;

- Craven (1978) – idealizou inicialmente o sistema NEPHIS *Nested Phrase Indexing System* e depois, em consequência de uma evolução experimental, o sistema LIPHIS *Linked Phrase Indexing System* (LIPHIS). Ambos são sistemas de indexação automática;

- Austin (1974) – idealizou para a *British National Bibliography* (BNB) o PRECIS, cujo funcionamento se fundamenta em estruturas semântica e sintática e em esquema de operadores de função.

A noção de índice sempre esteve muito ligada ao processo de indexação. Os índices outrora existentes em sistemas de recuperação da informação, tais como os antigos catálogos de fichas de biblioteca, foram considerados dentro de uma perspectiva classificatória, porque os chamados cabeçalhos de assunto eram compostos sob influência da terminologia classificatória e não do texto e seu conteúdo.

O grande elemento transformador dentro da indexação alfabética a marcar os estudos teóricos foi a análise em facetadas proposta por Kaiser, Ranganathan e seus seguidores, demarcando a possibilidade de maior especificidade e uniformidade com o uso dos conceitos essenciais: espaço, tempo, processo, concreto, coisa, ação etc.

Após Ranganathan, o *Classification Research Group* desenvolveu a aplicação dos

princípios da análise em facetadas assumindo a influência da classificação facetada e passou a utilizar e desenvolver uma metodologia facetada (PIEADA, 1983). Vickery (1975, p.181), citado por Esteban Navarro (1999, p.74), por exemplo, ampliou a quantidade de facetadas propostas por Ranganathan: Personalidade, Matéria, Energia, Espaço e Tempo (PMEST) para Tipo, Estrutura, Constituintes, Propriedades, Processos, Operações, Técnicas, Generalidades.

Segundo Esteban Navarro (1999, p.73) "...a faceta permite descobrir as relações que mantêm entre si os conceitos mediante a formulação de uma série de perguntas peculiares para o domínio disciplinar em que se situa o assunto do documento..." Nesse sentido, as facetadas relacionadas ao assunto "materiais dentários", por exemplo, seriam reveladas a partir dos seguintes conceitos:

- Tipo de materiais dentários: Materiais dentários metálicos e Materiais Dentários não-metálicos.

- Constituintes: Ouro, alumínio, porcelana, prata.

- Propriedades: resistência à fratura, fotoelasticidade, rigidez.

- Processos: Amalgamação, polimerização.

- Operações: Vibração.

- Técnicas de laboratório: Fase Gama etc.

Na visão de Esteban Navarro (1999, p.79) a identificação de conceitos na indexação deve utilizar questões construídas a partir da "análise das facetadas que caracterizam um conjunto de assuntos relacionados entre si". Se retomarmos a recomendação dos "Princípios de indexação" do UNISIST e da Norma 12.676 (ASSOCIAÇÃO BRASILEIRA..., 1992) para identificação de conceitos<sup>4</sup>, veremos que a base é a mesma, ou seja, a influência da análise em facetadas permeia

<sup>4</sup> "A escolha dos conceitos pode obedecer a um esquema de categorias reconhecidas como importantes no campo coberto pelo documento, ex.: o fenômeno, o processo, as propriedades, as operações, o material, o equipamento, etc." (UNISIST, 1981, p.87).

o processo de análise em indexação. Os autores citados mencionam a identificação do tema referindo-se a conceitos, categorias e facetas, que podemos considerar como a mesma coisa, porque o tema é constituído da presença de conceitos.

As publicações sobre indexação alfabética de assunto privilegiam a discussão em torno da ordem de citação, de entrada do assunto nos catálogos e não do processo de análise do assunto.

### **Indexação coordenada**

Esse tipo de indexação caracteriza-se pela composição de assunto usada para representar o conteúdo informacional de um documento. A pré-coordenação em cabeçalhos de assunto remonta aos princípios de Cutter em 1876 que consiste em recolher de um documento um ou mais aspectos dominantes tendo em vista certas subdivisões do assunto.

Para Robredo (1986, p.80) a indexação coordenada baseia-se

[...] na suposição de que o conteúdo substancial de um documento e de uma pergunta podem ser representados com suficiente precisão e de maneira suficientemente completa mediante um certo número de descritores ou palavras-chave, explícita ou implicitamente contidos no documento ou na pergunta.

A indexação pré-coordenada apresenta-se nos chamados sistemas de recuperação da informação tradicionais, que usam catálogos de assuntos alfabéticos ou classificados pelos quais o assunto de um documento é representado de modo mais geral. A diferença básica entre pré e pós-coordenada é que, na primeira, os termos são combinados no momento da elaboração do índice e na segunda os termos são combinados no momento da busca para recuperação da

informação. O assunto é reunido numa dada entrada que deve obedecer às formas específicas de entradas nos catálogos ou etiquetas numéricas de programas de automação de bibliotecas.

### **Indexação automática**

A indexação automática é segundo Robredo (1986, p.96) qualquer procedimento que permita identificar e seleccionar os termos que representem o conteúdo dos documentos, sem a intervenção direta do documentalista. Faz-se, entretanto, necessária uma distinção entre indexação automática e indexação automatizada.

Comparando o processo de indexação automática com a indexação manual, Vieira (1988, p.48) afirma que aquela refere-se à operação que identifica palavras ou expressões significativas dos documentos para descrever seu conteúdo de forma condensada por meio de programas de computador.

Guimarães (2000, p.1) apresenta o processo de indexação que envolve o uso do computador em três concepções - a primeira está relacionada pelo uso de programas informáticos que dão suporte ao armazenamento dos termos de indexação obtidos pela análise conceitual; a segunda pelo uso dos sistemas que analisam documentos de forma automática com validação dos termos por um profissional (indexação semi-automática) e, a terceira, refere-se a indexação automática propriamente dita conforme definição de Robredo (1986) e Vieira (1988) e classificada por Guimarães (2000) como aquela realizada pelos programas de computador sem nenhum tipo de validação por profissionais.

A indexação automatizada seria, portanto, aquela resultante do trabalho intelectual de um profissional para checagem do valor dos termos atribuídos a um documento por um programa de computador.

A indexação automática abrange uma diversidade de experimentações e modelos, mas nesta revisão de literatura será centrada em *Key-Word-in Context* (KWIC) e *Keyword-out-of-Context* (KWOC) por questões históricas.

Foi na década de 1960 que o índice KWIC, indexação pela palavra-chave no contexto, apareceu propagando um novo método de indexação: a indexação pela palavra. Representa a primeira aplicação de indexação automática de documentos técnicos, tendo por base as palavras significativas dos títulos. A história da indexação evidencia que a sua criação foi atribuída a William Frederick Poole que em 1882 com a publicação de "Poole's Index", criou um índice dando entrada do assunto pela palavra-chave do título dos artigos desse periódico. É atribuída a Poole a criação do índice KWIC (BORKO; BERNIER, 1978, p.8).

Foskett (1973) afirma que esse método de indexação dispensa o uso de nenhum esforço intelectual. Foi utilizado por Cretadoro na compilação do catálogo da Biblioteca Pública de Manchester. Esse índice é mais conhecido como índice permutado por computador e, nesse caso, é atribuído a Hans Peter Lunh, a sua criação em 1953, coincidindo com os primeiros e promissores trabalhos de Chomsky que abriam um novo caminho investigador sobre a estrutura da frase (PINTO MOLINA, 1993, p.225).

Para a autora esse foi um dos primeiros métodos de indexação automática baseado em métodos estatísticos de recuperação da informação superficiais e pouco rigorosos, mas que deu origem aos primeiros trabalhos de caráter morfológicos e sintáticos.

O índice KWIC é caracterizado pelo uso da linguagem natural, conseqüentemente não há controle de termos significativos e os sinônimos não são identificados.

Em suma, sua metodologia consiste no deslocamento dos títulos de vários documentos tanto à direita como à esquerda de forma a ordenar alfabeticamente as palavras tomadas como significativas. Nesse índice, cada título aparece na listagem alfabética igualmente ao número de vezes das palavras significativas.

O índice KWOC – palavra fora do contexto, é para Lancaster (1993, p.48) aquele em que a palavra-chave usada como ponto de entrada não se repete no título mas é substituído por asterisco (\*) ou outro símbolo. Nesse tipo de índice, as palavras significativas do título são extraídas e colocadas na ordem alfabética e não há uma permutação do título como no KWIC; o título aparece na mesma ordem da sua apresentação e a "palavra significativa" em linguagem natural fica na ordem alfabética como cabeçalho. Suas limitações são comparadas às do índice KWIC.

Nesse tópico, além da abordagem sobre a metodologia empregada para realização dos índices automáticos KWIC e KWOC, são discutidas as vantagens e desvantagens acerca da indexação automática.

De acordo com Ward (1996, p.217) autor do artigo intitulado *The future of the human indexer*<sup>5</sup> com objetivo de esclarecer como o indexador automático poderia suplantou ou completar o trabalho do homem, considerando-se os princípios de indexação e as habilidades intelectuais que ela envolve, afirma que uma boa indexação requer:

- 1) conhecimento prévio da literatura;
- 2) capacidade para avaliar o que deve ser indexado, e em que profundidade;
- 3) habilidades de leitura - a) incluindo dados não verbais, b) análise e avaliação de dados;
- 4) criação de um sumário, incluindo a formação de relações intertextuais;

<sup>5</sup> Indexador técnico da *PAWERLINK textbase* – índice interno da Biblioteca de Recardo *Consulting Engineers*. Nesse artigo o autor procura responder a questão a qual seu texto faz e suas considerações vêm da sua participação naquele índice automático.

5) habilidades para catalogar e classificar necessárias para criação de texto-base.

Além dos itens responsáveis por uma boa indexação, Ward (1996) indica as vantagens e desvantagens do indexador automático:

- Desvantagens de um indexador automático: 1) funciona somente em documentos separadamente; 2) não consegue fazer relações entre os textos ou entre um texto e uma visão de mundo; 3) fica amarrado ao vocabulário e à gramática usada no documento indexador; 4) não consegue lidar com dados gráficos; 5) não consegue lidar com línguas estrangeiras; 6) não consegue avaliar textos; 7) não consegue criar relações intertextuais; 8) só consegue indexar o que está explícito; não consegue indexar o que está implícito; 9) não é capaz de imitar o questionamento, a resposta humana a um texto, o que acrescenta valor à indexação; 10) requer constante aprimoramento para manter-se em dia com os novos desenvolvimentos; 11) não consegue catalogar ou classificar.

- Vantagens de um índice automático: 1) leitura instantânea de todo texto; 2) diz-se que é mais coerente do que um indexador humano; 3) não é tendencioso.

Os problemas dos índices automáticos são que não representam os assuntos dos documentos da mesma forma que a indexação humana o faz. Isso se justifica pelo fato de que ainda se desconhece o processo mental de análise de assunto envolvido durante o processo de indexação. Consideramos que, enquanto não se conhecer tais processos realizados pelo indexador humano, não será possível atribuir ao computador indexação semelhante. O computador representa maior agilidade em tarefas repetitivas mais simples, em que a análise conceitual não se faz necessária.

### **Leitura documentária**

A leitura documentária faz-se presente no momento em que o indexador realiza a análise

do assunto tratado em um documento a fim de representá-lo em termos de indexação. É nesse momento que se inicia a identificação de conceitos - principal etapa da análise de assunto - por meio da qual o indexador compreende os conceitos tratados em um documento, bem como verifica sua importância para seu sistema de informação. É aqui que os aspectos lógicos, lingüísticos e cognitivos, envolvidos na indexação, representam fatores de interferência, cabendo ao indexador a habilidade necessária para poder realizar a análise conceitual efetiva do documento.

A identificação de conceitos também está atrelada ao seu contexto, pois é necessário que o indexador verifique, por meio da leitura, qual a importância dos conceitos selecionados para o sistema de informação. Nesse sentido, a leitura do indexador está condicionada a determinados objetivos e ao contexto do sistema de informação. Torna-se, pois, difícil, dissociar esses aspectos da atividade de leitura.

A leitura para fins de indexação difere da leitura tratada de modo geral por possuir finalidades profissionais e pragmáticas. No entanto, os conhecimentos basilares necessários para uma boa compreensão de um texto são comuns a ambas. Segundo Beghtol (1986), para tal compreensão, é necessário: conhecimento do assunto primário e estrutura do conhecimento; o leitor indexador deve conhecer o assunto e possuir conhecimento extratextual tomando por base as instruções de uso do sistema lingüístico.

A leitura documentária caracteriza-se, de acordo com Ginez de Lara (1993), pela presença de operações seletivas voltadas para o processo de identificação e extração de informações. Nesse processo a autora afirma que a leitura documentária é realizada sob variáveis que lhe são específicas, além daquelas que são características de um leitor comum, não profissional. Dentre elas destaca: instruções textuais, condições de produção, circunstâncias

de enunciação, momento e lugar de enunciação etc.

Evidencia-se que a leitura documentária, realizada pelo leitor-indexador na fase de análise, corresponde à primeira fase de abordagem entre o leitor-indexador e o texto a ser analisado. Esta tem por finalidade, no primeiro momento, a identificação de conceitos que caracterize o assunto tratado no documento e no segundo momento, a seleção dos conceitos.

O indexador, na atividade de leitura para fins documentários, tem por objetivo identificar o tema ou assunto do documento. Kleiman (2000, p.42), observa que o estabelecimento de objetivos na leitura deixa claro três aspectos: “1) que se deve ler apenas para procurar as idéias principais, e por isso, não os detalhes; 2) serve para criar expectativa e, dessa forma, permitirá o reconhecimento de itens lexicais globalmente e 3) mantendo em mente os objetivos da leitura, não perderá de vista o texto em sua totalidade”.

A esse respeito, ressaltamos a relevância do objetivo de leitura para o leitor, ou seja, ter consciência do porquê estar lendo um texto. Essa consciência torna sua leitura metacognitiva<sup>6</sup>. A metacognição significa o conhecimento do conhecimento disponível para executar uma determinada tarefa e, é importante, porque podemos realizar um monitoramento de nossa atividade de leitura evitando erros e incompreensão. O leitor indexador, durante a atividade de leitura documentária, deve buscar a compreensão, pois somente assim terá condição de concretizar sua atividade de indexação.

A compreensão do texto pelo leitor indexador, segundo Farrow (1991, p.151), acontece na mesma medida em que leitores fluentes compreendem um texto. No entanto, acrescenta que a compreensão do leitor indexador tem implicações de quatro aspectos diferentes, a saber:

1) Indexador tem a pressão do tempo para indexar os documentos, o que faz com que leia o documento rapidamente ao invés de ler normalmente, ou extensivamente.

2) Muitos indexadores compreendem o texto apenas com o propósito de classificar, indexar ou resumir documento.

3) A compreensão de texto é seguida diretamente pela produção de um resumo.

4) Muitos indexadores trabalham dentro de um âmbito estreito de tipos de textos e campos de assunto e o elemento repetitivo conseqüentemente no trabalho deles conduz ao processamento automático.

Nesse sentido, a leitura do indexador está condicionada a determinados objetivos, voltados para uma determinada realidade e definida pela razão de ser da sua tarefa profissional.

A literatura sobre leitura revela uma tendência importante quanto à visão interacionista da leitura. Segundo essa visão, defendida por Giasson (1993) e Cavalcanti (1989), a interação em leitura concretiza-se no envolvimento de três variáveis intrínsecas ao processo de compreensão. São elas: o texto, o leitor e o contexto.

Na perspectiva da comunicação, Cintra (1987, p.30) considera que o texto como variável importante do processo de interação com o leitor é “produzido para determinados receptores e que a sua eficácia depende, em boa parte, da capacidade do autor em estabelecer com seus leitores potenciais uma relação cooperativa”.

Cintra (1987), também, concorda com Kato (1985) ao afirmar que a leitura é um processo interativo entre o leitor e o texto. Essa autora, aponta três fatores que atuam como suporte à legibilidade do texto: a qualidade, o

<sup>6</sup> Estratégias conscientes utilizadas na resolução de dificuldades durante a leitura (BROWN, 1980).



conhecimento prévio do leitor e o tipo de estratégias que o texto exige. No que diz respeito à qualidade do texto, Cintra (1987) pontua: a manutenção do tema, a adequação lexical e a estruturação do texto.

Com relação à leitura feita pelo indexador, de acordo com Cintra (1987, p.31) considera-se que a cooperação leitor/texto é rompida, tendo em vista que o autor não previu o indexador como leitor. Desse ponto de vista, o leitor-indexador nem sempre é o especialista da área de assunto do texto em análise, necessitando usar estratégias diversificadas e fazendo mais uso de estratégias metacognitivas.

A leitura do indexador, portanto, é guiada pelos seus objetivos, demanda do sistema de informação e de suas habilidades de leitor, definidas pelos seus conhecimentos prévios necessários à atividade de indexação. Esses conhecimentos prévios da indexação, entendemos que sejam um conjunto formado por conhecimentos profissionais e conhecimentos que estão relacionados à tematicidade do documento e aos aspectos: lingüísticos; lógicos e cognitivos.

### **Tematicidade**

Durante a leitura documentária, o indexador tem por objetivo identificar conceitos que compõem o tema do documento. Ao identificar os conceitos, também, poderá selecionar aqueles que considera mais representativo do conteúdo do documento. A seleção de conceitos, ressaltamos aqui, é determinada conforme o grau de interesse que o conceito apresenta perante as necessidades informacionais dos usuários de um sistema de informação, tendo em vista que o objetivo da indexação é tratar os documentos para que os mesmos possam ser recuperados e seus conteúdos disseminados.

A determinação do tema ocorre, na leitura documentária, mediante análise conceitual para

identificação dos conceitos presentes no conteúdo textual. Segundo Lancaster (1993, p.8), isso implica em decidir do que trata um documento. Essa questão sobre o que trata um documento é apontada na literatura de Análise Documentária como algo difícil de ser respondido ou demonstrado.

Para muitos estudiosos da área, quando se busca pesquisar sobre a problemática da identificação do tema, estamos pesquisando sobre *aboutness*. O termo originário da língua inglesa pode ser traduzido como “do que trata um texto”, em português. Foi introduzido em 1986 na Grã-Bretanha por Begthol, que estudou a noção de assunto mediante a lingüística textual e propôs que a tematicidade em classificação bibliográfica dos documentos fosse determinada pelas cinco macroregras do modelo de leitura de Van Dijk e Kintsch (1983), descritas logo a seguir.

No idioma português, há divergências entre os pesquisadores para se referir a *aboutness*; para alguns, o referido termo, pode ser “tematicidade”, por se considerar como um substantivo ligado ao termo temático, enquanto outros adotam “atinência”.

Na opinião de Begthol (1986), o documento tem uma tematicidade que lhe é relativamente permanente, porém um número variado de mensagens ou significados que podem ser medidos, conforme o uso exato do documento para o usuário. Destaca ainda, que o mesmo documento pode ter significados diferentes para o mesmo leitor, em épocas diferentes, mas, como o documento por si mesmo, é imutável, presume-se possuir uma tematicidade que lhe é fundamental.

Os pesquisadores Ellieker, Connel e Lancaster (1989) citados por Naves (2000, p.66) destacam que “a tematicidade poderá referir-se ao conteúdo do documento sendo caracterizada como tematicidade intrínseca e que questões externas de como o documento poderá ser usado,

por que ele foi adquirido, entre outras, são identificadas como tematicidade extrínseca”.

Elucidando a definição acima, acreditamos que a tematicidade sempre será o conteúdo relevante do documento, no entanto, algumas variáveis como os interesses informacionais dos usuários do sistema irão influenciar na determinação desse conteúdo, entre outras. Portanto, a escolha do tema de um documento sempre estará relacionada com os interesses de tais usuários, independente da quantidade de informações referentes ao tema selecionado.

Nesse sentido, Wilson (1985) citado por Todd (1992 p.102), afirma que podemos entender também que o grau de relação entre tematicidade e significado é variável porque depende do

[...] uso que a pessoa pode encontrar da tematicidade do documento numa certa época, e o mesmo documento pode vir a ter diferentes significados para o mesmo leitor em diferentes épocas, entretanto o documento possui uma tematicidade fundamental.

A determinação do tema do texto coincide com a identificação de sua estrutura temática. Segundo Tálamo (1994, p.24) o documento possui uma estrutura temática na qual o tema está representado pelos seguintes componentes: Quem? (ser), O quê? (tema), Como? (modo), Onde? (lugar) e Quando? (tempo). Conforme a autora, identificando essa estrutura temática encontra-se o objetivo principal do texto, isto é, as informações relevantes, separando-as assim das acessórias.

Entendemos aqui que os componentes fundamentais identificados no tema podem ser os propostos por Tálamo, pois Kobashi (1994) procurou identificar a estrutura temática proposta por Tálamo, destacando que a categoria “Quem?” não foi identificada em textos técnicos científicos, enquanto que a categoria “O quê?” é essencial por ser o “elemento nuclear da estrutura temática”. As categorias Quando?, Onde? e Como? são categorias acessórias da principal “O quê?”,

podendo, assim, aparecerem ou não, no texto, independente da ordem de procedência, entre elas.

Na literatura, encontramos também, um modelo de leitura de Van Dijk e Kintsch (1983) citado por Todd (1992, p.103) que orienta o leitor a identificar o tema por meio da “macroestrutura” de um documento, desencadeando as seguintes ações cognitivas:

- 1) **Regra de deleção fraca:** supressão de informação acidental, isto é, detalhes que não mudam o significado ou influenciam a interpretação das sentenças seguintes.
- 2) **Regra de deleção forte:** supressão de informação importante localmente: a informação que é suprimida especifica associações normais ou esperadas.
- 3) **Regra de deleção zero:** nenhuma redução ocorre. Todas as informações são consideradas relevantes, e são admitidas diretamente na macroestrutura.
- 4) **Regra de Generalização:** referência a diversos objetos ou propriedades da mesma classe superordenada, de forma global, pelo nome da classe superordenada.
- 5) **Regra de Construção:** a combinação ou integração de informação que denota propriedades, causas, componentes, conseqüências, etc. de um fato de nível superior.

Na opinião de Farrow (1996), apesar das macro-regras parecerem adequadas para a identificação do tema, é preciso ressaltar que a sua aplicação demandaria muito tempo, tornando-a inviável para a indexação, devido a grande quantidade de documentos a indexar. Essas macrorregas não foram testadas por indexadores na realização das atividades de indexação, logo elas exigem do indexador um conhecimento prévio “profundo” sobre o assunto a ser indexado. No entanto, para Begthol (1986)

a teoria de Van Dijk do processamento do discurso, apresentada pelas macrorregras, fornece uma definição sucinta e viável da tematicidade do documento e o modelo potencial de análise de assunto.

Podemos observar que o tema para os pesquisadores é a informação relevante abordada no texto, mas é preciso ressaltar que a seleção do tema sofre a influência da política do sistema de indexação. Portanto, observamos sobre a necessidade de existir uma equivalência da relevância do tema do documento, tanto para o indexador, como para o usuário, pois o objetivo maior da indexação é garantir a recuperação da informação. Aquele atingirá esse objetivo se elaborar informações documentárias (índices e resumos) consistentes, devendo para isso, de acordo com Kobashi (1994, p.103), integrar à essas (informações documentárias), as propriedades pertencentes aos princípios conversacionais de Grice (1975) que consistem em:

- **concisão** (princípio de quantidade): o resumo e o índice devem fornecer informação suficiente, não mais do que o necessário. Este aspecto é fundamental para a economia do sistema porque facilita a estocagem e a recuperação de informações;
- **pertinência** (princípio de qualidade e de relação): o resumo e o índice devem representar o conteúdo do documento tão fielmente quanto possível. Desse modo, não se pode integrar a eles informações que não estejam no texto original;
- **precisão e objetividade** (princípio de modo): o resumo e o índice não devem comportar ambigüidade, devendo ser formulados em termos precisos; no caso do resumo, deve-se acrescentar duas ordens de questões: 1) as informações devem ser apresentadas em uma ordem adequada; 2) o resumo não deve comportar julgamento.

O indexador tem o objetivo de tornar o tema conhecido para os usuários interessados, portanto, é, também, função do “indexador aumentar a visão do que os outros podem ler em um texto” (HUTCHINS, 1977, p.19).

### Aspectos Lingüísticos

Os aspectos lingüísticos na indexação são evidenciados no momento da leitura, pois a “análise do conteúdo é um método de apreensão da informação transmitida por uma língua” (CHAUMIER, 1986, p.19), sobretudo no momento da tradução dos termos que representam os conceitos identificados no documento em linguagem natural para a linguagem de indexação.

Moreiro González (1994, p.206) classifica, da seguinte forma, as fases em que os fatores lingüísticos se fazem presente no fazer documentário:

- 1) **Leitura-análise**: que consiste em identificar e compreender o conteúdo dos documentos e sua distribuição; 2) *Síntese: mediante la cual se interpreta y simplifica el discurso a las líneas macroestructurales para poder manejarlo* e 3) *Representação: Re- expresión de la información en el uso social de la información manejada.*

Kobashi (1996, p.6) nos apresenta a interface entre a lingüística e a análise documentária, a partir das semelhanças dos processos documentários com os de tradução automática. Dessa forma, os estudos de interface entre lingüística e documentação começaram a surgir em fins dos anos 1960, com a implantação dos computadores nos trabalhos documentários, indicando que estavam ligados às análises das entidades que se manipulavam, como: as palavras, frases, resumos, descritores.

Navarro (1988, p.46) afirma que “somente a partir da necessidade de sistematizar as relações entre a linguagem natural e as linguagens documentárias, principalmente visando-se

procedimentos automatizados, a intersecção foi percebida e tratada em nível teórico”.

Durante a atividade de indexação e no momento de atribuir o termo que melhor represente uma palavra, o indexador deve considerar o contexto em que a palavra será usada, tomando como base, por exemplo, como os usuários interessados no documento recuperariam por “aquele” determinado termo.

Os aspectos lingüísticos estão presentes na leitura para indexação, por meio da lingüística textual que, segundo Fávero e Koch (1988, p.14) é a área responsável por “determinar o que faz com que um texto seja um texto e diferenciar as várias espécies de textos.”

O conceito de tipologia textual está, portanto, intrinsecamente associado à noção de que todo texto apresenta uma sintaxe que organiza as suas várias partes. É, a essa forma global de organização do texto que Van Dijk e Kintsch (1983) dão o nome de superestrutura.

Dessa forma, ao definir a estrutura textual das várias espécies de texto, a Lingüística textual possibilita à área de Análise Documentária conhecer e utilizar essas estruturas no momento da indexação, uma vez que o conhecimento da organização textual possibilita ao indexador identificar quais as partes que determinado texto apresenta, bem como, o conceito pertencente a cada parte. Outrossim, esse conhecimento permitirá que o indexador faça a indexação do documento realizando uma leitura mais rápida.

Os resultados obtidos no estudo sobre leitura documentária (FUJITA, 2003) comprovam que se os indexadores conhecerem e explorarem a estrutura textual, tendem a encontrar, mais facilmente, a informação relevante em determinado documento, para o que, salientamos a necessidade do conhecimento de estrutura textual, por parte do indexador, como um fator a ser considerado nas propostas de novas metodologias para a análise de conteúdo.

Finalmente, os estudos sobre estruturas textuais possibilitarão elaborar ou adaptar meto-

dologias para a identificação de conceitos visando elaboração de condensações (índices e resumos) de documentos a indexar.

### Aspectos lógicos

No momento da leitura do documento para indexação, o indexador realiza esforços mentais classificados dentro de uma das divisões da “Ciência Normativa Lógica”, que por sua vez pertence a uma ciência maior, a Filosofia.

A divisão da Lógica, conforme Santaella (1992) citada por Naves (2000, p.74), pode ser entendida como:

- 1- **Lógica Utens**: é direcionada para o raciocínio comum dos seres humanos visando formar opiniões, mas que, porém, não é suficiente por si somente.
- 2- **Lógica Docens**: exige do ser humano uma capacidade invertida, degeneração, de elaboração da teoria, ou seja, estudos dos processos de raciocínio e a investigação de métodos que dêem bases mais confiáveis ao pensar, e apremem o avanço do conhecimento para os resultados desejados.

Segundo Naves (2000, p.74), durante a identificação do tema, o indexador utiliza esforços mentais relacionados à Lógica Docens, porque nesse momento percebe-se a capacidade invertida de generalização e de construção teórica por parte do indexador, muito além da Lógica do bom senso do homem comum.

Para Cunha (1989, p.51), os procedimentos lógicos que ocorrem na identificação do tema durante a indexação do documento, pelo indexador são:

- Relações de inclusão/exclusão, todo/parte, gênero/espécie, embasadas em juízos e raciocínios, além da capacidade de programar a tradução do conteúdo do texto em etapas seqüenciais lógicas.

- No mesmo procedimento de identificação e tradução de informações significativas encontram-se, implícitas, as questões relativas à construção do discurso do autor/produtor, que quando científico envolve a identificação e avaliação (na medida em que o bibliotecário/analista da documentação decide o que é significativo ou não) dos métodos, hipóteses, leis, teorias e resultados, utilizados pelo autor/produtor com fim de chegar a determinada ‘informação nova’.

A Análise Documentária e a Lógica têm pouco a oferecer uma a outra, entretanto, é preciso mencionar que da Lógica a Análise Documentária pode apoiar-se na Lógica Formal porque, segundo Pinto Molina (1994, p.128) a Lógica Formal “se limita a um sistema de símbolos, relacionando o exercício do raciocínio lógico a um cálculo algébrico”.

### Aspectos cognitivos

O indexador, além dos processos lingüísticos e lógicos, utiliza, também, processos cognitivos que interagem na leitura.

Iniciando a abordagem dos processos cognitivos salientamos que são objeto de estudo da Psicologia Cognitiva, constituindo-se de:

[...] processos e estruturas mentais implicados na aquisição, no processamento e no uso do conhecimento ou da informação, entre os quais podem ser diferenciados os processos mentais básicos (memória e atenção), as representações mentais (imaginação, formulação de proposições e estabelecimento de categorias) e os processos mentais complexos (compreensão, raciocínio e solução de problemas) (PINTO MOLINA, 1994).

Os processos cognitivos envolvidos na leitura documentária, estabelecidos por Monday (1996), são: representação do conhecimento;

esquemas; unidades cognitivas; organização do conhecimento; compreensão e estruturas do texto; estruturas semânticas e esquemáticas do texto.

Consideramos, assim, que os processos cognitivos utilizados pelo leitor durante a leitura são: o seu conhecimento sobre a estrutura textual, visando identificar a informação que considera relevante; o conhecimento prévio sobre o assunto do texto; e a recuperação de esquemas de compreensão formados com sua experiência de vida que o permite inferir sobre o assunto abordado.

Naves (2000, p.85) identifica o processo de inferência como um dos muitos processos cognitivos, caracterizando-os em:

**Inferência lógica** – usada para estabelecer causas, motivações e condições que permitem fatos específicos. Inferência evolutiva – na qual analistas aplicam suas crenças às situações descritas. **Inferência integrativa** – executada no momento da compreensão e baseada nos conceitos e propriedades da organização hierárquica. **Inferência Construtiva** – baseada no conhecimento do indexador.

Shaw e Fonchereaux (1993), citados por Milstead (1994, p.578) esclarecem que existem dois processos cognitivos envolvidos na Análise Documentária, tanto na atividade de indexação como de classificação:

- 1) decidir sobre o que um texto ou uma informação fala (do que trata um texto ou uma informação e/ou que questão ele/ela responde);
- 2) traduzir essa decisão em termos usados no sistema de indexação.

Ressaltamos que o enfoque do processo cognitivo envolve os componentes básicos necessários: texto, contexto, dados do conhecimento do analista, objetivos documentários e um método de avaliação (PINTO MOLINA, 1994, p.130).

Em seu trabalho sobre a “abordagem interdisciplinar do conceito e prática da análise de conteúdo do texto documentário escrito (WTDCA)”, Pinto Molina (1994), esclarece enfaticamente que

[...] é inviável sem a contribuição de certas disciplinas tais como: a lingüística (textual), a lógica (formal) e a psicologia cognitiva. Há uma relação de superfície entre a lógica e a estrutura textual profunda. Quanto à psicologia cognitiva, e especialmente a psicologia do processamento da informação abre bons prospectos para o WTDCA. A lógica, especialmente a lógica formal permite a análise sintática essencial dos textos. A lingüística textual, fundada na gramática transformacional, contribui decisivamente para a transcrição difícil entre a forma e o conteúdo.

Consideramos, ao final de nossa abordagem dos aspectos lógicos, lingüísticos e cognitivos, que para adaptar ou propor metodologia que garanta uma eficiente análise de conteúdo dos documentos, é preciso estabelecer interface com a Lógica formal, a Lingüística textual e a Psicologia cognitiva que têm o texto escrito como objeto de estudo para leitura. Entretanto, o pesquisador da Análise Documentária deve ficar atento para estabelecer o devido “recorte” do conhecimento desenvolvido nas áreas interdisciplinares, destacando o que realmente apresenta viabilidade de aplicação para fins de tratamentos documentários, até porque é preciso verificar se não está se apropriando de todo o conhecimento desenvolvido por uma determinada área. (KOBASHI, 1994).

### **Procedimentos de indexação**

Analisaremos a seguir os procedimentos de indexação fundamentados na norma ABNT 12676 de 1992 que é uma tradução da Norma ISO 5696 de 1985 *International Organization For*

*Standardization* e, em seguida, a abordagem de identificação de conceitos propostos pelo sistema de indexação PRECIS a fim de verificar as influências que tiveram na determinação do assunto de um documento.

Do ponto de vista da norma ABNT 12676 de 1992, a Indexação corresponde ao ato de identificar e descrever o conteúdo de um documento com termos representativos dos seus assuntos. O processo de indexação, segundo a norma, possui três estágios: 1) exame do documento e estabelecimento do assunto de seu conteúdo; 2) identificação de conceitos presentes no assunto; 3) tradução desses conceitos em termos de uma dada linguagem de indexação (ASSOCIAÇÃO BRASILEIRA..., 1992).

Como já verificado, o objetivo da indexação é o de representar o conteúdo informacional do documento, tendo em vista sua recuperação, para tanto, realiza-se um exame do documento a fim de identificar conceitos pelos quais a tematicidade de um documento estará representada. Essa tematicidade é determinada pelo indexador através da leitura do documento, tendo em mente as necessidades informacionais da comunidade usuária do sistema de informação.

A norma ISO 5696 (International Organization For Standardization, 1985) é baseada nos princípios de indexação do UNISIST (1981, p.83) que considera a realização do processo de indexação em dois estágios: analítico – é realizada a compreensão do texto como um todo, identificação e seleção de conceitos válidos para indexação e tradução – consiste na representação de conceitos por termos de uma linguagem de indexação.

Segundo tais princípios, na operacionalização da indexação do assunto de um documento o primeiro estágio subdivide-se em três etapas: 1) Compreensão do conteúdo do documento como um todo, os objetivos do autor etc; 2) identificação dos conceitos que

representam este conteúdo este conteúdo, objetivos etc; 3) seleção dos conceitos válidos para recuperação. Na prática de indexação, segundo o UNISIST (1981, p.86), esses três estágios se superpõem para escolha dos conceitos e, com relação à fase de identificação dos conceitos, o “indexador deve adotar uma abordagem lógica, selecionando os conceitos que melhor expressarão o assunto do documento”.

Considerando que o momento de seleção é o mais delicado, o UNISIST (1981, p.90) recomenda que “o critério principal deve ser sempre seu valor potencial como um elemento de expressão do conteúdo do documento”. Nesta tarefa, o indexador deve ter em mente as prováveis questões que os usuários farão à unidade de informação. E nesse sentido as indicações do UNISIST (1981, p.90) apontam como critério: escolher os conceitos que são mais apropriados a uma determinada comunidade de usuários; modificar, se necessário, tanto os instrumentos de indexação quanto os procedimentos com base no *feedback* de questões.

Além disso, a seleção de conceitos deve ser feita, tendo em vista os objetivos para os quais as informações serão selecionadas destacando-se dois fatores que afetam mais diretamente essa escolha: exaustividade, especificidade.

Para a exaustividade na indexação, o indexador deverá procurar “todos os conceitos de um documento que possam ter um valor potencial para os usuários de um sistema de informação [...]” (UNISIST, 1981, p.88). Para a regra de especificidade, os conceitos devem ser os mais específicos possíveis e, os mais genéricos, podem ser selecionados com vista aos objetivos do sistema de informação.

A fim de assegurar a organização dos conceitos de uma forma que seja útil e acessível, “é necessário o conhecimento profundo dos instrumentos de indexação” (UNISIST, 1981,

p.90). Caberá, portanto, ao indexador ter familiaridade com os instrumentos para poder manipulá-los sabendo, de antemão, que estes podem impor algumas limitações na prática de indexação.

As recomendações da norma 12676 (ASSOCIAÇÃO BRASILEIRA..., 1992) baseadas nos princípios do UNISIST (1981), estabelecem que o indexador na análise de assunto, deverá realizar uma leitura do documento examinando as partes mais relevantes, visto que a leitura completa do item documental é impraticável. A análise do assunto na norma ABNT 12676 de 1992 é dividida em dois estágios: o primeiro é o exame do documento e o segundo é a identificação de conceitos.

No exame do documento a norma indica a verificação das seguintes partes do documento, tais como: título; resumo; lista de conteúdos; introdução (vendo as frases que iniciam capítulos e parágrafos); conclusão; ilustrações; diagramas; tabelas (e suas legendas); palavras em destaque.

Assim como o UNISIST, a norma ABNT, alerta que a indexação não deve ser guiada somente pelo exame do título ou resumo do documento. Isso se justifica, tendo em vista que os títulos, muitas vezes não são adequados e até ambíguos e, no caso dos resumos, alerta que podem ser inadequados. Em suma, esses dois itens não constituem únicas fontes para identificação do assunto. A seguir, a norma recomenda que a análise siga uma abordagem sistemática para a identificação dos conceitos considerados essenciais na descrição do assunto.

A abordagem para identificação dos conceitos deverá ser feita com base em um questionamento sistemático, proposto pela norma:

O documento possui em seu contexto um objeto sob efeito de uma atividade?  
O assunto contém um conceito ativo (por exemplo, uma ação, uma operação, um processo etc.)? O objeto é

influenciado pela atividade identificada? O documento possui um agente que praticou esta ação? Este agente refere-se a modos específicos para realizar a ação (por exemplo, instrumentos especiais, técnicas ou métodos)? Todos esses fatores são considerados no contexto de um lugar específico ou ambiente? São identificadas algumas variáveis dependentes ou independentes? O assunto foi considerado de um ponto de vista normalmente não associado com o campo de estudo (por exemplo, um estudo sociológico ou religioso)?

A etapa de seleção dos termos, segundo a norma 12676 (ASSOCIAÇÃO BRASILEIRA..., 1992) é feita em função da tradução dos conceitos identificados em descritores de uma linguagem de indexação. Para essa seleção consideram-se o sistema de informação e o usuário desse sistema.

O último item diz respeito à qualidade da indexação; nesse sentido, consta na norma, que a qualidade da indexação depende de factuais como: a competência do indexador e a qualidade dos instrumentos de indexação.

Um outro procedimento de indexação que merece destaque é o do PRECIS (FUJITA, 1989), um sistema de indexação que apresenta sua análise conceitual como metodologia de identificação de conceitos baseando-se em um questionamento dirigido unicamente ao texto, caracterizando, conforme Albrechtsen (1993) sua concepção de análise de assunto orientada unicamente para o conteúdo, a saber: O que aconteceu? (ação); A que ou a quem isto aconteceu? (objeto da ação – sistema chave); Que ou quem fez isto? (agente da ação); onde aconteceu? (local).

De outra forma, os procedimentos de indexação propostos pelo UNISIST (1981) e pela norma 12676 (ASSOCIAÇÃO BRASILEIRA..., 1992), estão direcionados, simultaneamente, para duas concepções de análise de assunto

identificadas por Albrechtsen (1993): a orientada para o conteúdo, representada pela indicação da abordagem sistemática de identificação dos conceitos e a orientada para a demanda, por orientar que a seleção dos conceitos seja feita com base no uso dos termos pelo usuário.

### **Revisão de literatura de relatos de experiência: resultados discutidos com os temas da revisão de literatura fundamental**

Com objetivo de leitura analítica da literatura de relatos de experiência, foram selecionados 10 textos (artigos) que tratavam da prática de indexação em sistemas de informação brasileiros. A sistematização desses artigos obedeceu à divisão por décadas a fim de verificar a existência da leitura documentária em prática de indexação.

Cada artigo foi descrito fisicamente por referência bibliográfica, seguida de uma descrição baseada no resumo de cada um, acrescida de informações retiradas de seu conteúdo. Na ausência do resumo, a descrição baseou-se na leitura do texto integral. Em seguida à descrição, segue uma análise em que se apresenta a existência ou não das variáveis apontadas na literatura fundamental e a discussão subsequente dá-se conforme categorização dessas variáveis.

A revisão de literatura fundamental em Indexação revelou que na prática de indexação encontram-se várias questões que interferem na sua realização, podendo-se destacar, em primeiro lugar, a questão da confusão terminológica entre catalogação de assunto e indexação alfabética de assunto.

O termo “catalogação de assunto” foi a expressão mais usada na prática de indexação nos relatos de experiência, em decorrência de muitas instituições, em décadas passadas e recentes, adotarem os cabeçalhos de assuntos



como linguagem de indexação nos seus sistemas de informação. Isso revela, apesar do processo de indexação ter evoluído, que ainda persiste a primeira concepção baseada em Cutter que formulou as primeiras tentativas de organização de assuntos representados pelos cabeçalhos. Talvez isso possa explicar-se pelo fato da Biblioteca do Congresso dos Estados Unidos ter dado continuidade aos cabeçalhos de assunto pela sua Lista de Cabeçalho de Assunto, atualizando-a e disponibilizando-a somente pelo preço de uma cópia.

Apesar da indexação existir há muito tempo, até hoje se desconhece o que lhe é fundamental, ou seja, a determinação do assunto ou tema do documento. Questão que volta à tona, sempre que se deseja a automação da indexação. Verificamos que, a partir da década de 1950, quando começaram os estudos de indexação automática, apareceram as preocupações com áreas de interface, como a Lingüística. Evidencia-se a partir de então, a importância do ser humano com relação ao juízo de valor da tematicidade do documento.

As análises dos textos que tratam da prática de indexação mostram a inexistência da descrição do processo de indexação, especialmente no que diz respeito ao momento de como se processa a análise de assunto, o que talvez se explica pelo fato do resultado da indexação estar sujeita à leitura que o indexador realiza para a determinação do assunto do documento. E isso, depende muito das ações mentais que cada indexador realiza no fazer documentário.

Os textos analisados na revisão de literatura fundamental evidenciaram a presença dos aspectos que interferem no processo de indexação, como os lingüísticos, lógicos, cognitivos e a questão da tematicidade. Os textos descritos e analisados na revisão da literatura de relatos de experiência mostraram uma tendência para a indexação automática, dando destaque aos aspectos lingüísticos e lógicos da indexação.

O aspecto lingüístico é um dos elementos que mais contribuem para a determinação do assunto e este somente é evidenciado mediante o processo de leitura, pelo qual o indexador identifica e seleciona conceitos relacionados ao assunto tratado no texto guiando-se pelo seu conhecimento de estrutura textual.

A contribuição desse aspecto foi demonstrada na leitura dos textos que versam sobre metodologia de indexação, ao indicarem duas concepções de análise de assunto: uma voltada para o conteúdo que é expresso pela língua em que está escrito o documento e outra para a demanda da informação que exige do indexador conhecimento do contexto no qual está inserida a instituição.

#### CONSIDERAÇÕES FINAIS

Ao realizarmos a pesquisa tínhamos como pressupostos que o fator mais importante na atividade de indexação era a concepção de análise de assunto assumida pelo indexador; que uma análise conceitual não poderia ser feita unicamente com base na linguagem do sistema de informação e, que esses fatores estão diretamente ligados à falta de fundamentação teórica na área de indexação, uma vez que está muito articulada com o desenvolvimento da prática.

Desenvolveu-se, então, no estudo, a análise e discussão das tendências na literatura com relação ao desenvolvimento teórico de indexação e sua influência na prática do indexador.

Pudemos verificar pela evolução da indexação, sobretudo na evolução dos cabeçalhos de assunto, que o fator de maior importância é a determinação do assunto, ainda que seja agravado pela dificuldade de uma linguagem que irá representá-lo.

Desde Cutter, a literatura publicada tem demonstrado a preocupação dos estudiosos com

essa questão, procurando não só o desenvolvimento de linguagens de indexação como também soluções na indexação automática. No entanto, o problema persiste, uma vez que a maior parte dos estudos não tem sido voltada para a questão da determinação do assunto.

Ora, se a indexação existe para que a recuperação da informação seja efetiva, então o que queremos recuperar? Pensamos que a resposta a essa pergunta seja a chave para as possíveis soluções futuras acerca da determinação do assunto do documento. E o que a prática tem mostrado é evidenciado pelos estudos que procuram saber como o indexador age na escolha de conceitos para representação do assunto do documento.

Na revisão de literatura de relatos de experiência, a análise dos textos de prática de indexação revelou que nenhum fez referência ao modo como a leitura foi realizada, denotando falta de reflexão sobre os procedimentos de indexação e a influência da leitura. Cogitamos que, talvez, não exista a devida ênfase nos programas de formação do indexador. Considerando que é por meio da ação de leitura que o indexador faz a análise conceitual, e por se tratar de uma atividade ligada à psicologia cognitiva, recomenda-se que os próximos estudos venham a focalizar esse aspecto da indexação, ainda inexplorado.

Verificamos pela revisão de literatura fundamental, que a indexação deve ser feita tendo em vista o conhecimento prévio do indexador, as necessidades informacionais do usuário, a política de indexação da unidade de informação e a estrutura textual dos documentos.

Um dos critérios da análise conceitual, além dos acima citados, é descobrir porque o documento foi adquirido pelo sistema de informação e examinar se os termos estabelecidos para a representação do documento são compatíveis com os termos que provavelmente

o usuário adotará ao formular a expressão de busca da informação. Sugerimos, portanto, examinar como o usuário formula uma questão de busca, além dos processos cognitivos, envolvidos durante a leitura pelo indexador.

Tendo em vista que o objetivo do estudo foi o de contribuir com uma investigação teórica em indexação para detectar quais as influências e tendências na prática de indexação, concluímos que as influências detectadas pela revisão de literatura fundamental estão parcialmente presentes na prática de indexação, sobretudo no que se refere à inexistência de detalhamento dos procedimentos de indexação.

Nesse sentido, pode-se afirmar que as tendências e influências dos estudos teóricos em prática de indexação provêm mais especificamente das áreas de interface da Análise Documentária, como a Lingüística, a Lógica e a Psicologia Cognitiva.

A Lingüística tradicional contribui para a indexação com seus aspectos semânticos e sintáticos e, agora, a preocupação com a necessidade do usuário, inserem-se, também, aspectos lingüísticos-pragmáticos (Lingüística Aplicada) como interação leitor-autor regida pela relevância-leitor, que em última análise, para a área de Indexação, é a relevância para o usuário-leitor final.

Já a Lingüística textual colabora com seus esquemas formais de diferentes tipos de texto; a Lógica com as inferências lógicas a partir de pistas do texto e a Psicologia Cognitiva com os mecanismos mentais e as estratégias, exclusivamente fundadas no conhecimento prévio do leitor.

Ao dominar esses aspectos, que auxiliarão na determinação da tematicidade do documento, o indexador conseguirá realizar uma análise conceitual com efetividade, realizada em termos da tematicidade do autor, do leitor e do usuário.

## REFERÊNCIAS

- ALBRECHTSEN, H. Subject analysis and indexing: from automated indexing to domain analysis. *The indexer*, v.18, n.4, p.219-224, 1993.
- 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. 4p.
- AUSTIN, D. *PRECIS: a manual of concept analysis and subject indexing*. London: Council of the British National Bibliography, 1974. 551p.
- BEGHTOL, C. Bibliographic classification theory and text linguistics: aboutness analysis, intertextuality and the cognitive act of classifying documents. *Journal of Documentation*, London, v.42, n.2, p.84-113, 1986.
- BORKO, H.; BERNIER, C. *Indexing concepts and methods*. New York: Academic Press, 1978. 261p.
- BRADFORD, S.C. *Documentação*. Rio de Janeiro: Fundo de Cultura, 1961.
- CAVALCANTI, M.C. I-n-t-e-r-a-ç-ã-o *leitor-texto*: aspectos de interpretação pragmática. Campinas: Universidade Estadual de Campinas, 1989. 271p.
- CESARINO, M.A.N.; PINTO, M.C.M.F. Cabeçalho de assunto como linguagem de indexação. *Escola de Biblioteconomia. Universidade Federal de Minas Gerais*, Belo Horizonte, v.7, n.2, p.268-88, 1978.
- CHAUMIER, J. *As técnicas de documentais*. Publicações Europa-América. 1971, 111p.
- CHAUMIER, J. *Travail et methodes du/de la documentaliste: connaissance du problème*. Paris: ESF/Libraries Techniques. 1980. Exposé 3, Chap.3: L'indexation, p.42-47.
- CHAUMIER, J. *Análisis y lenguajes documentales: el información documental*. Barcelona: Mitre, 1986. 170p. (Colección Ciencias de la Comunicación).
- CINTRA, A.M.M. Estratégias de leitura em documentação. In: Smit, J.W (Coord.). *Análise documentária: a análise da síntese*. 2.ed. Brasília: IBICT, 1987. p.29-37.
- COATES, E.J. *Subject catalogues: Headings and structures*. London: Library Association, 1960.
- COLLINSON, R.L. *Índices e indexação: guia para indexação de livros, e coleções de livros, periódicos, e coleções de livros, periódicos, partituras musicais, com uma seção de referência e sugestões para leitura adicional*. Trad. Antônio Agenor Briquet de Lemos. São Paulo: Polígono, [1971].
- CRAVEN, T.C. Linked phrase indexing. *Information Processing and Management*, New York, v.14, p.469, 1978.
- CUNHA, I.M.R.F. (Coord.). *Análise documentária: considerações teóricas e experimentações*. São Paulo: FEBAB, 1989.
- ESTEBAN NAVARRO, M.A.E. Elementos, actividades y criterios para la identificación, comprensión y selección de conceptos en la indización analítica. In: GARCIA MARCO, F.J.G.M. *Organización del conocimiento en sistemas de información y documentación*. Zaragoza: Capítulo Español de la ISKO. Universidad Carlos III de Madrid, 1999. v.3, p.69-93.
- FARRADANE, J.A. A comparison of some computer produced permuted alphabetical subject indexes. *International Classification*, Munich, v.4, n.2, p.94-101, 1977.
- FARROW, J.F. A cognitive process model of document indexing. *Journal of Documentation*, London, 47, n.2, p.149-166, 1991.
- FARROW, J.F. Propositional analysis and macrorules for indexing. *Library Review*, v.45, n.1, p.6-15, 1996.
- FÁVERO, L.L.; KOCH, I.G.V. *Linguística textual: introdução*. São Paulo: Cortez, 1988.
- FOSKETT, A.W. *Abordagem temática da informação*. Trad. Antônio Agenor Briquet de Lemos. São Paulo: Polígono, 1973.
- FUJITA, M.S.L. *Leitura documentária do indexador: aspectos cognitivos e lingüísticos influentes na formação do leitor profissional*. 2003. 321f. Tese (Livre-Docência nas disciplinas Análise Documentária e Linguagens Documentárias Alfabéticas) Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2003.

- FUJITA, M.S.L. *Leitura em análise documentária*. Marília: UNESP/CNPq, 1999. Relatório parcial de pesquisa.
- FUJITA, M.S.L. *PRECIS na língua portuguesa: teoria e prática*. Brasília: Universidade de Brasília, 1989.
- GARDIN, J.C. et al. *La logique du plausible: essais d'epistemologie pratique*. Paris: Maison de Sciences de L'Homme, 1981.
- GIASSON, J. *A compreensão na leitura*. Lisboa: Asa, 1993. 317p.
- GINEZ DE LARA, M.L. *A representação documentária: em jogo a significação*. São Paulo, 1993. 133f. Dissertação (Mestrado em Ciência da Comunicação) – Escola de Comunicação e Artes, Universidade de São Paulo. São Paulo, 1993.
- GOMES, H.E.; GUSMÃO, H.R. *Guia prático para a elaboração de índices*. Niterói: GBIDCSH da APB-RJ, 1983.
- GUIMARÃES, J.A.C. *Indexação em um contexto de novas tecnologias*. [S.l.: s.n.], 2000. 10p. Texto Didático.
- HUTCHINS, W.K. On the problem of aboutness in document analysis. *Journal of Information*, London, v.1, p.17-35. 1977.
- INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. *Documentation – methods for examining documents, determining their subjects, and selecting indexing terms*. Geneva: ISO. 5p. (ISO 5963-1985 (E))
- KAISER, J.O. *Systematic indexing*. London: Pitman, 1911.
- KATO, M.A. *O aprendizado da leitura*. São Paulo: Martins Fontes, 1985. 121p.
- KLEIMAN, A. *Texto e leitor: aspectos cognitivos da leitura*. 7.ed. Campinas: Pontes, 2000.
- KOBASHI, N.Y. *A elaboração de informações documentárias: em busca de uma metodologia*. 1994. 195f. Tese (Doutorado em Ciências da Comunicação) – Escola de Comunicação e Artes, Universidade de São Paulo, São Paulo. 1994.
- KOBASHI, N.Y. Análise documentária e representação da informação. *Informare*, São Paulo v.2, n.2, p.5-27, 1996.
- LANCASTER, F.W. *Indexação e resumos: teoria e prática*. Tradução de Antonio Agenor Briquet de Lemos. Brasília: Briquet de Lemos/Livros, 1993.
- LYNCH, M.F.; PETRIE, J.H.A program suite for the production of articulated subject indexes. *Computer Journal*, Oxford, v.16, p.46-51, 1973.
- METCALFE, J. *Subject classifying and indexing of libraries and literature*. New York: Scarecrow, 1959.
- MILSTEAD, J.L. Indexing for subject cataloguers. *Cataloging & Classification Quarterly*, New York, v.3, n.4, p.37-44, 1983.
- MILSTEAD, J.L. Needs for research in indexing. *Journal of the American Society for Information Science*, v.45, n.8, p.577-582, 1994.
- MONDAY, I. Les process cognitifs et la rédaction de résumés. *Documentation et Bibliothèques*, v. 42, n.2, p.55-62, 1996.
- MOREIRO GONZÁLEZ, J.A. Documentación y lingüística: conceptos de relación esenciales. *Ciencias de la Información*, v.25, n.4, p.202-210, 1994.
- NAVARRO, S. Interface entre lingüística e indexação: revisão de literatura. *Revista de Biblioteconomia e Documentação*, São Paulo, v.7, n.4/6, p.46-62, 1988.
- NAVES, M.M.L. *Fatores interferentes no processo de análise de assunto: estudo de caso de indexadores*. Belo Horizonte, 2000. 253f. Tese (Doutorado em Ciência da Informação) – Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2000.
- NEELAMEGHAM, A.; GOPINATH, M.A. Postulated-based permuted subject indexing (POPSI). *Library Science with a slant to documentation*, v.12, n.3, p.79-87, 1975.
- PIEIDADE, M.A.R. *Introdução à teoria da classificação*. 2.ed. rev. aum. Rio de Janeiro: Interciência, 1983. 221p.
- PINTO MOLINA, M. *Análisis documental: fundamentos y procedimientos*. 2.ed. rev. aum. Madrid: EUDEMA, 1993.
- PINTO MOLINA, M. Interdisciplinary approaches to the concept and practice of written text documentary content analysis (WTDCA). *Journal of Documentation*, v.50, n.2, p.111-133, 1994.

- RANGANATHAN, S.R. *The colon classification*. New Brunswick, N.J: University, 1965.
- ROBREDO, J. *Documentação de hoje e amanhã: uma abordagem informatizada de biblioteconomia e dos sistemas de informação*. Brasília, 1986.
- ROBREDO, J. *Documentação de hoje e amanhã: uma abordagem informatizada de biblioteconomia e dos sistemas de informação*. 2.ed. rev. ampl. Brasília: Edição de Autor, 1994.
- TÁLAMO, M.F.G.M. *Elaboração de resumos*. São Paulo: Escola de Comunicações e Artes, Universidade de São Paulo. *Cadernos de análise documentária*, São Paulo, n.1, p.23-31, 1994.
- TODD, R.T. Academic indexing: what's it all about? *The Indexer*, London, v.18, n.2, p.101-104, 1992.
- UNISIST. *Princípios de indexação*. *Escola de Biblioteconomia, Universidade Federal de Minas Gerais*. Belo Horizonte, v.10, n.1, p.83-94, 1981.
- Van DIJK, T.A.; KINTSCH, W. *Strategies of discourse comprehension*. New York: Academic Press, 1983. 337p.
- VAN SLYPE, G. *Los lenguajes de indización: concepción, construcción y utilización en los sistemas documentales*. Trad. Pedro Hípola e Félix de Moya. Madrid: Fundación Germán Sánchez Ruipérez; Pirámide, 1991. 200p. Tradução de: *Les languages d'indexation: conception, construction et utilisation dans les systèmes documentaires*.
- VIEIRA, S.B. *Indexação automática e manual: revisão de literatura*. *Ciência da Informação*, Brasília, v.17. n.1, p.47-57, 1988.
- WARD, M.L. The future of the human indexer. *Journal of the American Society for Information Science*, v.28, n.4, p.217-225, 1996.
- WILSON, P. Subject and the sense of position. In: CHAN, C. et al. *Theory of subject analysis: a manual*. Littleton, Colorado: Libraries Unlimited, 1985. p.306-23.
- WITTY, F.J. The beginnings of indexing and abstracting: some notes towards a history of indexing and abstracting in antiquity and the Middle Ages. *The Indexer*, London, v.8, n.4, p.193-198, 1973.

