

**A importância do vocabulário
controlado na representação de
assuntos da produção científica
nos sistemas automatizados
Uma abordagem geral**

**Monografia apresentada ao Curso de
Especialização em organização da
Informação em Contextos Digitais.
Orientadora: Professora Madalena
Martins Lopes Naves.**

**Belo Horizonte
Escola de Ciência da Informação – UFMG
2006**

Agradecimentos

À Deus por estar sempre presente na minha vida.

À professora Madalena Naves Lopes pela dedicação oferecida desde o início a este trabalho como orientadora e amiga.

À professora Gercina Angela Borém de Oliveira Lima pela oportunidade da minha formação em especialista.

Aos colegas da biblioteca pela amizade, companheirismo, tornando esta trajetória tão agradável.

Aos colegas do Curso de Especialização OICD, pela amizade e estímulo.

Ao meu marido Wilder, pelo apoio, incentivo e amor.

Aos meus filhos queridos, Débora e Leonardo, pela compreensão e carinho.

Ao meu pai, a toda minha família pelo incentivo, torcida.

A todos o meu muito obrigada.

SUMÁRIO

1 INTRODUÇÃO	7
2 PRODUÇÃO CIENTÍFICA ACESSÍVEL EM MEIOS ELETRÔNICOS.....	12
3 SISTEMAS DE INFORMAÇÃO AUTOMATIZADOS.....	16
4 INDEXAÇÃO	20
5 LINGUAGENS DE INDEXAÇÃO (LI).....	23
5.1 Conceito e funções das LI.....	24
5.1.1 Relações entre os termos nas LI.....	29
5.1.1.1 Relações semânticas	29
5.1.1.2 Relações sintáticas	32
5.1.2 Uso do vocabulário controlado nas LI.....	33
5.1.3 Tipos de LI.....	36
5.1.3.1 Linguagens pré-coordenadas	36
5.1.3.2 Linguagens pós-coordenadas.....	38
5.1.3.3 Linguagens verbais.....	39
5.1.3.3.1 Tesouros	39
5.1.3.3.2 Listas de cabeçalhos de assunto	43
5.1.3.3.4 Linguagens simbólicas (CDD e CDU)	45
6 RECUPERAÇÃO DA INFORMAÇÃO NOS SISTEMAS AUTOMATIZADOS, ATRAVÉS DO VOCABULÁRIO CONTROLADO	48
6.1 Estratégia de busca.....	52
6.2 Interação com o usuário.....	56
7 O PAPEL DO PROFISSIONAL DA INFORMAÇÃO NO CONTEXTO DESTES TRABALHO.....	60
8 CONSIDERAÇÕES FINAIS	64
REFERÊNCIAS	69

RESUMO

O vocabulário controlado se faz presente na indexação de assuntos, e no planejamento da estratégia de busca em bases de dados de informação bibliográfica e bibliotecas digitais. A linguagem de indexação é um vocabulário controlado e instrumento no processo de indexação, construído a partir de um conjunto de regras, que servem para representar abreviadamente um conteúdo de um documento, de modo a organizar e facilitar o acesso e a transferência da informação na comunidade científica. No processo documentário, representar conteúdos se relaciona a propósitos e concepções definidas institucionalmente que perseguem objetivos específicos. Uma linguagem pressupõe a noção de arranjo, e conseqüentemente de organização. O controle é o fator que diferencia a linguagem natural da linguagem de indexação. A linguagem de indexação, pode ser pré-coordenada, pós-coordenada, simbólica (classificação, CDU e CDD) e verbal (listas de cabeçalhos de assunto e tesouros), de modo a representar o conteúdo dos documentos nas bases de dados. A linguagem de indexação atua no sistema, orientando o indexador na escolha do melhor termo para representação do assunto e orientando o pesquisador na busca de documentos relevantes. O vocabulário controlado configura-se como instrumento facilitador da comunicação em contextos documentários específicos e conseqüentemente a disseminação da produção científica em sistemas automatizados. A recuperação da informação é facilitada pelo uso das linguagem de indexação, e depende de uma estratégia de busca adequada e eficiente para ser bem sucedida. Cabe ao profissional da informação viabilizar uma melhor interface entre usuário e os sistemas de informação.

Palavras - Chaves: Linguagens de indexação; Vocabulário controlado; Tesouros; Listas de Cabeçalhos de assunto; Sistemas de informação; Estratégia de busca; Profissional da informação.

ABSTRACT

Controlled vocabulary is present in subject indexing and planning a search strategy in database from bibliographic information and digital libraries. Indexing language is a controlled vocabulary and a tool in the indexing process built from a set of rules which serve to briefly represent the contents of a document so that it organizes and makes it easier to access and transfer information in the scientific community. In the documentary process, representing contents is related to institutionally fixed purposes and conceptions which pursue particular goals. A language presupposes the notion of arrangement and therefore organization. Control is a factor which makes natural language different from indexing language. Indexing language may be pre-coordinated, post-coordinated, symbolic (classification, CDU and CDD) and verbal (lists of subject headlines and thesaurus), in order to represent the contents of documents from databases. Index language plays a role in the system, guiding the indexer to choose the best term to represent the subject and guiding the researcher to search relevant documents. Controlled vocabulary is a facilitating tool of communication in particular documentary contexts and therefore spreading the scientific production in automated systems. Recovering information is made easy by the use of indexing languages and depends on an appropriate and effective search strategy in order to be successful. It is up to the professional to enable a better interface between the user and information systems.

Keywords: indexing languages; controlled vocabulary; thesaurus; lists of subject headings; information systems; search strategy; information professional.

1 INTRODUÇÃO

O estudo constitui-se de uma abordagem geral sobre a importância do controle de vocabulário no processo de indexação e recuperação da informação técnico-científica, em sistemas automatizados. Esses sistemas de informação visam conhecer as necessidades de seus usuários, de forma que possam ser caracterizadas, em função da atividade que se desempenham, retratando seus desejos e percepções, quanto às necessidades de informação científica.

Com a difusão das bases de dados, e da biblioteca digital, em virtude da explosão bibliográfica, constata-se a necessidade de controle da informação. Nesse contexto, certamente a indexação de assunto passou a ser considerada chave de acesso a documentos, tornando imprescindível o uso das linguagens, mais especificamente a controlada, como instrumento de recuperação da informação em sistemas automatizados.

Percebe-se que é recente o aprofundamento teórico conceitual sobre os princípios que fundamentam a concepção dos sistemas de classificação, a web e tantas outras redes digitais que abrangem o mundo como um todo, e que necessitam de uma troca de informação eficaz, rápida e objetiva. A demanda por instrumentos de organização e tratamento da informação aumenta significativamente, de forma a proporcionar o acesso sem barreira à informação científica.

Pode-se verificar que a necessidade de um controle de vocabulário se faz presente na utilização da informação local ou remota, armazenada em grande sistema de recuperação da informação. Assim, o tema desta monografia-se justifica pela necessidade de uso de linguagem de indexação que pode contribuir de forma substancial, para o sucesso dos serviços de acesso à informação científica, que são oferecidos pelos sistemas de informação.

Este estudo tem, como objetivo geral, - mostrar a importância da utilização do vocabulário controlado na indexação de assuntos em sistemas automatizados que visam agilizar e ampliar a recuperação da informação em publicações técnico-científicas.

E como objetivos específicos:

- Verificar a necessidade do controle de vocabulário, e a função de sua aplicação nos sistemas que tratam e organizam a informação para ser recuperada.
- Mostrar os tipos de linguagens como a natural e artificial, utilizadas como parte integrante da atividade de controle da terminologia no processo de indexação.
- Analisar os instrumentos utilizados para controle de vocabulário em sistemas de informação, que visam padronizar a indexação e a recuperação, tendo em vista os objetivos do sistema para o qual ela será criada.
- Mostrar a influência do controle de vocabulário na recuperação eficiente da produção científica.

Para contextualizar o tema do trabalho, foi analisada, no capítulo dois, a produção científica acessível em meios eletrônicos, que representam o universo para o qual se faz necessária a construção de linguagem de indexação, ou seja, instrumento controlador de vocabulário, na intenção de se conseguir manter certo domínio sobre a produção técnico-científica. Visa ainda a sua disseminação em bases de dados, buscando prover o acesso aos conteúdos dos documentos disponíveis nos sistemas de informação automatizados, tema do abordado no terceiro capítulo.

A organização da informação tornou-se uma atividade fundamental para aumentar a visibilidade da produção científica, com imediato acesso ao que foi publicado, e,

consequentemente, uma avaliação imediata sobre o que foi publicado. Essa visibilidade se torna possível através da indexação.

O quarto capítulo trata da indexação, como o processo responsável pela tradução de um documento em termos documentários, isto é, em descritores, palavras-chave, cabeçalhos de assunto, e que tem, por função, expressar o conteúdo do documento. O termo é representado no sistema de busca, através da utilização da linguagem de indexação.

Partindo-se dessa definição, no quinto capítulo são abordadas as linguagens, que podem ser natural ou artificial, dando-se mais ênfase à linguagem de indexação, evidenciando as relações semânticas e sintáticas da linguagem. A linguagem de indexação é um instrumento da indexação, elaborada para proceder à mediação entre os processos de organização e recuperação da informação, e resultam do trabalho coletivo desenvolvido entre os profissionais da informação e os especialistas das áreas ao qual se destinam.

Na literatura, os instrumentos de indexação são chamados de linguagem de indexação, linguagem artificial, linguagem controlada e linguagem documentária, que são termos sinônimos. Observa-se que, na literatura, não existe um consenso entre os autores quanto à terminologia, visto ser um procedimento muito comum na biblioteconomia, em que a interdisciplinaridade torna complexa a definição dos termos. Dentro do objetivo deste estudo, é dispensável procurar critérios pelos quais se venha a reconhecer qual o termo ideal entre tais designações. Como o trabalho é uma revisão da literatura, não será possível a adoção de uma determinada terminologia, pois serão abordados conceitos de vários autores.

A linguagem de indexação tem, como característica, o vocabulário controlado. Considera-se imprescindível definir o que é um vocabulário controlado na representação de assuntos, mostrando a sua função dentro de um sistema de

informação, no sentido de prover acesso ao conteúdo informacional de documentos.

O quinto capítulo também trata dos tipos de linguagem de indexação, que são as linguagens pré-coordenadas e pós-coordenadas, as linguagens simbólicas (CDU e CDD), as linguagens verbais, como tesouros, listas de cabeçalhos, que são utilizados para a entrada de dados e sua disponibilidade para a busca. A estrutura da linguagem de indexação está diretamente relacionada à forma de entrada de dados, quantos e quais campos ou pontos de acesso estão disponíveis, para consulta, isto é com base no instrumento utilizado, é possível determinar os pontos de acesso, tendo-se em vista os objetivos da instituição para o qual ela é criada e a área do conhecimento que se pretende cobrir na determinação de conceitos.

O sexto capítulo versa sobre a recuperação da informação, em sistemas automatizados, através das linguagens de indexação. As informações arquivadas em bases de dados devem ser organizadas, observando-se uma sistemática que permita sua posterior recuperação. A maior dificuldade que se pode observar quando um pesquisador, um estudante de uma determinada área do conhecimento está realizando um trabalho de levantamento bibliográfico, é a determinação do termo de busca. Normalmente, é realizada uma pesquisa com resultados bem exaustivos para que, a partir dessas respostas de buscas obtidas, poder-se chegar à descoberta do termo que seria a chave de busca para suas pesquisas. E outra forma de ajuda são os tesouros ou índices existentes nas bases que ajudam muito na definição do perfil da pesquisa, nesse momento o sistema passa a interagir de certa forma com usuário, ajudando-o a formular sua estratégia de busca.

O sétimo capítulo aborda sobre o papel do profissional da informação no contexto deste trabalho. Acredita-se que o profissional da informação deva conhecer a

dupla função da linguagem documentária na mediação da comunicação do conteúdo do documento pela indexação e da expressão de busca pelo usuário. Com a evolução tecnológica dos serviços de informação, cabe ao profissional da informação acompanhar a necessidade do usuário, com o compromisso de capacitá-lo e orientá-lo no uso da tecnologia disponível.

Inicia-se este estudo com a abordagem sobre a organização de um universo de documentos que são frutos da produção científica, tendo em vista sua posterior recuperação em sistemas automatizados, tema abordado a seguir. Optou-se por este tipo de literatura por reunir documentos que resultam, em sua maioria, de pesquisas, que contribuem para o corpo teórico das áreas do conhecimento. Além disso, os textos produzidos com base científica apresentam uma estrutura padronizada, o que facilita o trabalho de indexar assuntos dos mesmos.

2 PRODUÇÃO CIENTÍFICA ACESSÍVEL EM MEIOS ELETRÔNICOS

Em todos os níveis de atuação da sociedade está acontecendo a geração de informação em velocidade incalculável. Sabe-se que a informação é hoje fator importantíssimo para a sociedade e para a economia, e a busca por informação é freqüentemente estimulada pela necessidade de se tomarem decisões na execução diária do trabalho do cientista, ou na construção do conhecimento. Os cientistas necessitam divulgar suas pesquisas, para que os dados manipulados se transformem em informação, que, por sua vez, vão gerar conhecimentos.

O volume de produção de informação é gigantesco, processando-se em um ritmo assustador, retratando o fenômeno da explosão bibliográfica, gerando documentos convencionais (periódicos, publicações seriadas e livros) e não convencionais (teses, relatórios, patentes, preprints, prepapers), e as diferentes formas de multiplicação desses documentos, em todas as áreas do universo científico e tecnológico. Necessita-se, cada vez mais, do trabalho de armazenamento, identificação, localização e disponibilização desse registro, de forma a possibilitar a circulação e a divulgação mais eficiente.

O periódico científico é considerado, atualmente, o principal canal da comunicação científica formal, e, para Oliveira (2005, p.34), "o crescimento da produção científica tem explicado a importância da indexação dos periódicos em bases de dados para que a informação se torne visível à comunidade científica de forma rápida e sistemática". Conseqüentemente, hoje são vários os periódicos de resumos, que se constituem em versões condensadas de artigos publicados em revistas científicas. O surgimento dos computadores, com capacidade de processamento, veio facilitar a vida do profissional da informação e, conseqüentemente, a dos seus usuários.

A produção científica é um dos componentes básicos do avanço da ciência, e merece um entendimento acerca da sua dinâmica. Ao se analisar a importância do vocabulário controlado na representação de assuntos dos documentos em bases de dados, está sendo dada uma maior visibilidade à produção científica. Essa visibilidade e o seu acesso em meio eletrônico são possíveis, devido ao processo de indexação. É importante destacar que as bases de dados tornaram-se instrumentos essenciais na disseminação da informação científica.

"A problemática da explosão da informação, contudo, não envolve apenas seu aspecto físico, se considerarmos que o desenvolvimento da ciência, de um modo geral, é acompanhado do crescimento complexo de assuntos multifacetados, combinado com a criação de novos conceitos e eliminação de outros. Em consequência, o comportamento sempre mutável do aspecto temático da geração da informação exige, dos sistemas envolvidos com sua organização bibliográfica, flexibilidade suficiente para comportar uma contínua e, de certa forma, acelerada adaptação de seus canais de recuperação". (FUJITA, 1988, p.22)

Tornou-se, cada vez mais necessária a atividade de organização bibliográfica, elaboração de índices de assuntos, principais canais de recuperação da informação, condicionados à utilização de um sistema de indexação capaz de proporcionar, de forma flexível e consistente, a descrição do assunto de um documento. Para Naves (2001, p.189), "o aumento da produção científica aliado à crescente interdisciplinaridade entre as áreas do conhecimento que vem ocorrendo nas últimas décadas, tem tornado o trabalho de organização e tratamento da informação cada vez mais árduo e complexo".

Em todas as áreas, nota-se uma preocupação, com a produção constante de conhecimento e sua organização, "em um ciclo produtivo contínuo, assim como a necessidade constante da adequação dessa organização com as tecnologias de ponta, onde o organizar, recuperar e divulgar passa a ser a função principal, onde se encontra o uso de linguagens documentárias" (TRISTÃO, 2004, p.163).

A tarefa de organização e de termos nas áreas da humanidade é complexa, sendo importante conhecer os fatores envolvidos no processo, bem como suas conseqüências para o tratamento da informação.

Para Lara (2004, p.91) "a organização da informação via linguagem documentária tem sua importância aumentada contemporaneamente, quando se deseja contar com mecanismos que desempenhem o papel de filtros para a recuperação, dado o grande volume de informações recuperadas na internet. A linguagem documentária, entretanto, para exercer esse papel, não pode ser formulada de modo aleatório".

A organização da produção científica e sua disponibilização acarretam mudanças profundas e reais, provocadas por idéias, informação, conhecimento, e desejo de inclusão e ascensão social. Ponto favorável marcado pela intensificação do uso das tecnologias de informação, isto impulsiona o desenvolvimento e o crescimento das pessoas. Mas a qualidade da informação disponível é fator determinante do sucesso ou insucesso dos cientistas.

A comunicação eletrônica veio trazer grandes vantagens para os cientistas, uma vez que os pesquisadores dos países em desenvolvimento têm grandes dificuldades em acompanhar a frente de pesquisas. Outra vantagem é que esse cenário coloca os cientistas dos países em desenvolvimento em condições de igualdade com os cientistas dos países desenvolvidos.

Percebe-se a necessidade de utilização de um controle de termos, para a organização do conhecimento, em todas as áreas, principalmente com os avanços constantes da tecnologia, da informação e comunicação. Conforme explica Tristão (2004, p.162), "a atividade de organização da informação é necessária, para a evolução dos indivíduos, organizações e da sociedade em geral". É necessário o estudo sobre as ferramentas de indexação de forma a adequá-las aos objetivos para os quais são criadas devendo-se observar os padrões internacionais, regionais e nacionais.

O próximo capítulo trata dos sistemas automatizados, que são responsáveis pela comunicação entre os sistemas de informação e, ao mesmo tempo, possibilita a comunicação na comunidade científica, na disseminação da informação. Realiza a transferência de dados bibliográficos dentro de algum padrão terminológico, proporcionando o progresso da ciência com a geração de novos conhecimentos ou utilização de conhecimentos já produzidos.

3 SISTEMAS DE INFORMAÇÃO AUTOMATIZADOS

Na otimização dos fluxos e da transferência de informação científica, a Ciência da Informação desenvolveu, ao longo de décadas, diversas formas de catalogar, classificar e indexar as áreas do conhecimento humano. Estudiosos dessa área trouxeram esse conhecimento para a web. Criaram-se aplicativos de *software* para biblioteca digital, objetivando organizar os registros desse conhecimento, facilitando o acesso, para que os usuários consigam rapidamente explorá-los na realização de suas tarefas em sistemas de informação automatizados.

A internet é um mecanismo de comunicação de alcance mundial, que facilita enormemente o acesso a sistemas de informação e de recuperação da informação.

"A internet vem proporcionar facilidades que extrapolam o conceito tradicional de informação bibliográfica baseada em documentos, como artigos de periódicos, trabalhos em congressos teses etc. Novos recursos informacionais estão à disposição da comunidade de pesquisa além desses tradicionais, agora em versão eletrônica, como documentos multimídias, listas de discussão, fóruns eletrônicos, conferências em linha, imagens (de satélites, de microscópios, em tempo real), bancos de *preprints* eletrônicos, os *e-prints* etc. Estes recursos tanto servem de subsídio à pesquisa quanto de canais de comunicação e publicação dos resultados e de garantia de primado e originalidade intelectuais dos mesmos". (MARCONDES, 2001, p.24)

Cada vez mais se verifica a consolidação da internet e da web como meios de disponibilização e de acesso à informação, caracterizando a necessidade de adoção de mecanismos que maximizem a integração entre os serviços de informação. As tecnologias digitais e virtuais rompem barreiras geográficas, reduzindo o tempo de acesso à informação, não importando aonde tal informação se encontre desde que disponível por um computador ligado à rede, o acesso é em tempo real. A informação é considerada o insumo fundamental para o desenvolvimento da ciência e da tecnologia.

O uso integrado das tecnologias de comunicação, de computação e de conteúdos em formato digital, cujo paradigma é a internet, tem contribuído atualmente para criar um novo ambiente de acesso, disseminação, cooperação e promoção do conhecimento em uma escala global.

"Estas transformações têm exercido profunda influencia sobre a concepção e funcionamento dos sistemas de informação automatizados, especialmente aqueles voltados para as atividades de pesquisa. O rompimento de barreiras tecnológicas importantes, experimentadas na última década, permitiram o surgimento de um novo patamar para esses sistemas : antes orientados basicamente para recuperação de referências bibliográficas em bases de dados isoladas e textos em papel, voltam-se hoje para a recuperação distribuída de objetos digitais - textos completos, imagens em movimento, som etc.-, estabelecendo como palavras de ordem a publicação na internet e a interoperabilidade entre fontes de informação heterogêneas e globalmente distribuídas". (MARCONDES, 2001, p.25)

A importância da interoperabilidade entre bibliotecas digitais e outros recursos informacionais digitais citados pelo autor, seria a viabilização do acesso a todas essas fontes de forma integrada e transparente, podendo-se perceber que este é um caminho a ser trilhado pelos sistemas de informação e bibliotecas. A biblioteca digital oferece serviço mais amplo de acesso à informação no que se refere aos dados digitais

Tem sido uma preocupação para os estudiosos da área de informação na web, segundo Tristão (2004, p.169), "a questão da adoção de padrões internacionais no desenvolvimento, construção e disponibilidade da informação na web". A definição desses padrões permite as atividades de intercâmbio de dados bibliográficos entre os diversos sistemas de informação ou bibliotecas digitais. Os padrões facilitam a automação e aumentam o custo benefício dos processos automatizados, em suma, os padrões devem ser adotados tanto nacional como internacionalmente.

As novas tecnologias proporcionam maior racionalidade nas rotinas diárias de trabalho das instituições que tratam a informação, facilitando o compartilhamento de recursos informacionais entre as mesmas.

Esse processo instituiu em um envolvimento responsável e consciente entre os sistemas de informação, quanto ao acesso à informação. A aplicação dos computadores, no processamento das informações bibliográficas trouxe, como vantagens, o armazenamento de grandes quantidades de informação, a capacidade de ordenar os dados com rapidez e efetuar buscas por palavras-chave, tornando-se necessário o uso de vocabulário controlado para o armazenamento, para uma posterior recuperação por parte dos pesquisadores nos sistemas de informação automatizados.

Os sistemas de informação automatizados possibilitam a coleta, o armazenamento, o processamento; a recuperação; e a disseminação de informação via redes tecnológicas, tornando-se parte integrante e irreversível para as bibliotecas, bases de dados, e outras instituições que lidam com a informação científica.

Sobre bases de dados, Coimbra (1999, p.884), aborda visão diferente dos autores e leitores, e dos editores, para os primeiros,

"autores e leitores, as bases de indexação são os meios mais eficientes de disseminação de resultados de pesquisas e de realização de levantamentos bibliográficos, pesquisadores de diferentes países interessados em um determinado tema, por mais que não tenha contato, têm a oportunidade de se comunicar, solicitar separata e trocar informações". Já para os editores, "a indexação de um periódico nas principais bases de dados internacionais representa muito mais do que um veículo de disseminação de informação científica. A inclusão de revistas em determinadas bases consideradas como de maior prestígio, como parâmetro indicativo de qualidade de um periódico, gerando acirrada competição entre editores, autores e instituições financiadoras de pesquisa".

Nas duas visões, percebe-se que as bases de dados aumentam a visibilidade da produção científica em um meio organizado, permitindo o imediato acesso e uma avaliação imediata sobre o que foi publicado. A comunicação científica é cada vez mais fortemente dependente das tecnologias de informação.

Para a organização e a recuperação da informação produzida, torna-se necessário o processo de representação através da indexação dos assuntos dos documentos, para sua posterior recuperação, de forma rápida e precisa.

A indexação, tema abordado no próximo capítulo, tem por objetivo estabelecer mecanismos para representar, de maneira mais fiel possível, os elementos que pertencem ao conteúdo de um documento, a fim de que o mesmo possa atender às necessidades dos usuários de um sistema de recuperação da informação.

Para Cesario (1978, p.288), a indexação

realiza-se de duas maneiras:

1. indexação manual;

2. indexação automática;

3. indexação híbrida;

4. indexação por

5. indexação por

6. indexação por

7. indexação por

8. indexação por

9. indexação por

4 INDEXAÇÃO

A indexação é definida no documento UNISIST (1981, p.84), "como a ação de descrever e identificar um documento de acordo com seu assunto". A transcrição dos conceitos em uma linguagem documentária, após tê-los extraído do documento através de uma análise e conseqüentemente a sua disponibilidade em meio eletrônico, caracteriza o sucesso das bases de dados no controle e na disseminação da produção científica.

"É importante indicar que a indexação da bibliografia científica não se constitui em 'invenção' recente, fruto da era da informática. Tampouco foi criada com o objetivo de avaliar a qualidade de revistas e, muito menos, de artigos. A indexação bibliográfica teve início no final do século XIX e no início do XX, resultante da iniciativa de grupos de pesquisadores que, impossibilitados de acompanhar de seus *bureaus* a crescente quantidade de materiais que se publicava ao redor do mundo sobre um determinado assunto, viram-se impelidos a desenvolver sistemas que racionalizassem e facilitassem o acesso às informações" (COIMBRA Jr., 1999, p.884).

Para Cesarino (1978, p.268), a indexação de documentos, em qualquer sistema de recuperação, se realiza de duas maneiras:

a) bibliograficamente ou objetivamente – este tipo de análise pretende a descrição do documento através de suas características físicas, com o objetivo de dar resposta à questão: <<Qual a aparência física deste documento>>?

b) intelectualmente ou subjetivamente - este tipo de análise pretende a descrição do documento em termos de suas características de conteúdo, com o objetivo de dar resposta à questão: <<Sobre o que é este documento>>?

A descrição física de documentos apresenta características fáceis de serem identificadas e existem normas que padronizam a entrada de dados. O problema maior se efetiva na descrição dos conteúdos, por ser menos reduzível a regras e por não se tratar exatamente como um identificador de uma determinada obra e sim de um conjunto de obras que tratam de um determinado assunto. Este

processamento de análise e determinação do conteúdo informativo de obras é comumente chamado indexação.

A indexação inclui dois estágios, sendo a primeira etapa a descrição física e a segunda etapa do processo de indexação, a tradução de assuntos extraídos do documento, através da análise de assunto, em linguagem de indexação utilizada no sistema. O indexador traduz os conceitos extraídos de um documento, expressos em uma linguagem natural, para uma linguagem controlada. Neste sentido, o indexador está utilizando o vocabulário controlado, que é uma lista de termos autorizados, usados na linguagem de indexação, tema de estudo desta monografia.

A palavra termo, para Cesarino (1978, p.270), “é usada de forma generalizada para representar qualquer forma de classe, subclasse, cabeçalhos de assunto, unitermos, palavras ou frases compostas”. Na indexação, define-se o assunto contido em um documento e a determinação de um ou mais termos que expressam o assunto desse documento. O conjunto de termos indexadores é chamado de linguagem de recuperação de informação ou linguagem de indexação.

Para Lancaster (1993, p.14), “o indexador somente pode atribuir a um documento termos que contêm na lista adotada pela instituição para a qual trabalha”. É necessário conhecer a terminologia da área, entender como os usuários procuram a informação, para melhor representar os elementos dos documentos como um todo. Os termos de uma área do conhecimento formam entre si um sistema de conceitos que são parte integrante do processo de controle da terminologia.

O tratamento da informação nos sistemas de informação e recuperação da informação é definido, tanto do ponto de vista físico quanto do ponto de vista temático. Esse processo resulta na produção de representações documentais,

como: termos de indexação, índices, referências bibliográficas etc. Para que isso possa ser feito, outras atividades são necessárias. É o caso da criação e manutenção de linguagens de indexação, listas de cabeçalhos de assuntos, sistemas de classificação, tesouros, e códigos de catalogação.

"As bases de indexação são os meios mais eficientes de disseminação de resultados de pesquisas e de realização de levantamentos bibliográficos. Portanto, pesquisadores de diferentes países interessados em um determinado tema, por mais que não tenham contato, têm a oportunidade de se comunicar, solicitar separatas e trocar informações". (COIMBRA JR. 1999, p.884)

O controle de vocabulário em sistemas de organização da informação é realizado mediante o uso de linguagens. Existem dois tipos de linguagem: a natural e a artificial, sendo que a linguagem artificial pode ser verbal e simbólica. As verbais são representadas por tesouros, listas de cabeçalho de assuntos, etc. Nas simbólicas, os conceitos são representados por símbolos de classificação, são as classificações, CDD (Classificação Decimal de Dewey), CDU (Classificação Decimal Universal), etc. Ambas podem co-existir em um sistema de recuperação de informação, uma complementando as deficiências da outra. Esses temas são tratados no quinto capítulo, em que são abordadas todas as características das linguagens de indexação.

5 LINGUAGENS DE INDEXAÇÃO (LI)

A linguagem é utilizada para representar o conteúdo, é um meio organizado em torno de uma área temática, que possibilita as operações de representação e de acesso à informação. No processo documentário, representar conteúdos se relaciona aos propósitos e concepções definidas institucionalmente, que perseguem objetivos específicos.

Para Lara (2002, p.132), “representar conteúdo para constituir informação significa recortar, segmentar de forma a procurar viabilizar um fluxo informativo efetivo e um mesmo universo documental pode ser organizado e representado de diversas maneiras”. Uma linguagem pressupõe a noção de arranjo, e, conseqüentemente, de organização.

A linguagem de indexação deve ser rigorosamente formulada, para se constituir em sistemas de organização e de comunicação da informação, qualquer que seja seu universo (bases bibliográficas, sites na web, conteúdos de manuais técnicos). No sub-item a seguir, será abordado o conceito e as funções da linguagem natural e da linguagem de indexação, que são considerados instrumentos necessários da indexação, na representação de conteúdo dos documentos bibliográficos resultantes da produção técnico-científica.

5.1 Conceito e funções das LI

A linguagem de indexação é caracterizada como a que utiliza campos de descritor, termo de indexação e identificadores. A linguagem natural abrange os termos do título e do resumo dos documentos. A decisão sobre qual linguagem a ser empregada, se a natural ou se a controlada, envolve escolhas que irão determinar a obtenção eficaz daquelas informações específicas solicitadas pelo usuário, ora restringindo, ora ampliando os resultados, para a obtenção de informações mais relevantes, conforme seu pedido de busca da informação.

A linguagem natural pode ser definida como a linguagem do discurso técnico científico, e, no contexto da recuperação da informação. A expressão linguagem natural, para Lancaster (1993, p.200), é considerada, sinônimo de discurso comum, linguagem utilizada habitualmente na escrita e na fala, o contrário de vocabulário controlado. "A linguagem natural é uma expressão que normalmente se refere às palavras que ocorrem em textos impressos, considerando-se como seu sinônimo a expressão 'texto livre'".

Nessa mesma linha de pensamento, pode-se afirmar que, nas bases de dados, a linguagem natural é representada nos termos extraídos diretamente dos documentos (título e resumo). "Trabalha com o uso de termos e a representação do assunto está ligada às palavras utilizadas pelo autor, esta linguagem também é conhecida por linguagem livre" (SOUTO, 2003, p.76).

A linguagem natural é aquela linguagem retirada do próprio documento, sem nenhum controle de vocabulário, ou seja, quando se usa uma linguagem retirada do documento sem nenhuma modificação e se emprega termos do texto na indexação, está sendo utilizada uma linguagem natural. A linguagem natural é a linguagem humana, que é composta por imensurável número de conceitos, que são expressos por meios de termos.

A linguagem natural dificulta a concordância entre os termos indexados, deixando espaço para que o mesmo assunto seja representado de formas diferentes e, por isso, possibilitando que informações sejam perdidas no momento da busca.

A esse respeito, Pinto (1985, p.172) afirma que, na recuperação de informações:

"esse procedimento pode causar vários problemas, como no caso de diferentes autores usando palavras diversas para expressarem o mesmo conceito (dispersão terminológica) ou o caso do uso de diferentes estruturas para expressarem a mesma idéia (dispersão sintática) por essa razão, a maioria dos sistemas tem adotado algum tipo de controle sobre os termos usados".

"Os campos de título e resumo registram os termos da linguagem natural, enquanto os campos de descritores, termos de indexação, ou identificadores registram os termos da LI". A LI pode ser definida, "como um conjunto limitado de termos autorizados para uso na indexação e busca de documentos". (LOPES, 2002, p.42)

Uma linguagem de indexação, segundo Pinto (1985, p.171), "é qualquer conjunto de termos e ou símbolos usados para denominar assuntos em sistemas de recuperação da informação". Representa a comunicação escrita através de um sistema de símbolos utilizados na indexação dos documentos, equivalente à análise documentária em que se exprime o conteúdo dos documentos, mediante descritores.

As linguagens de indexação são instrumentos auxiliares da análise documentária e são determinantes para a recuperação da informação. As frases de indexação, depois de traduzidas para uma linguagem de indexação, passam a ser chamadas de descritores de assuntos, cabeçalhos de assunto, palavras-chave, termo de indexação ou enunciados. São termos diferentes que exercem a mesma função dentro de sistema de informação.

“O uso de uma linguagem documentária para a escolha do descritor ou cabeçalho de assunto reduz a diversidade e a ambigüidade do vocabulário e estabelece uma uniformidade de representação dos termos selecionados pelo indexador para descrever o assunto dos documentos, já que vários autores podem utilizar diferentes palavras para expressar uma mesma idéia, assim como os usuários podem apresentar diversidade de vocabulário quando da expressão de uma estratégia de busca” (FUJITA, 2004, p. 17).

Quando se adota a linguagem controlada, ou vocabulário controlado, está se empregando uma linguagem artificial, através de um vocabulário controlado. As finalidades do mesmo são atuar sobre o vocabulário (termo propriamente dito) e sobre a sintaxe (estrutura frásica). Como qualquer língua, a linguagem de indexação é composta de vocabulário e sintaxe. O vocabulário é a relação entre os termos e ou símbolos indexadores. E a sintaxe é a coordenação de termos indexadores para representação de assuntos.

O componente básico da linguagem de indexação é o vocabulário e a sintaxe: “vocabulário (conjunto dos descritores), sintaxe (função/ ordem do vocabulário em relação ao outro) e a semântica (sentido do vocabulário)” (SOUTO, 2003, 77). O conhecimento desses componentes básicos ajuda a aumentar a objetividade no processo de indexação, reduzindo a subjetividade a um mínimo aceitável. São formadas de palavras preferenciais, cumprindo função intermediadora, entre palavras de vocabulários de determinados domínios do conhecimento, e, palavras utilizadas pelos usuários.

A função da linguagem não é somente orientar o indexador sobre os termos indexadores a serem usados, mas também instruí-lo sobre quais orientações deve ser estabelecidas no catálogo de índice. Para Pinto (1985, p. 171) “essas orientações vão desde o simples controle de termos sinônimos, até a construção de complexas redes de referências cruzadas, dando ao usuário outras opções de

busca em relação ao assunto procurado". A linguagem assume o papel de comunicação de informações.

A linguagem de indexação está diretamente comprometida com o desempenho do sistema de informação para o qual a originou.

"A linguagem documentária deve estar de acordo com as políticas de indexação definidas pelo sistema, intermediando o acesso à recuperação da informação e este, por outro lado, deve considerar a instituição aonde se desenvolve; as expectativas e necessidades do usuário; as características do assunto tratado; os recursos humanos, físicos e financeiros; os produtos e serviços visados e a relação custo/desempenho" (BOCCATO, 2006, p.268).

Todavia, a linguagem de indexação processa a comunicação especializada, entre os pesquisadores e a área de domínio. A representação adequada desses termos técnicos da área resultará em uma indexação de qualidade, bem como de buscas mais consistentes por parte do usuário pesquisador. Seria, então, uma atuação eficaz da linguagem nos sistemas de informação.

As linguagens de indexação e recuperação da informação exercem uma função muito importante na construção dos sistemas automatizados nacionais e internacionais de informação técnica científica. A qualidade e o nível dos atuais serviços de informação dependem, diretamente, da qualidade das linguagens de indexação e recuperação da informação e do *software*.

"Há 20 anos um número muito grande de linguagens de indexação vem sendo utilizado em sistemas de recuperação de informação. São variações das linguagens naturais x controladas; linguagens alfabéticas x classificadas; linguagens pré x pós - coordenadas, etc. Não se pode afirmar, simplesmente, quais as melhores ou as piores. Não há linguagem perfeita. Existe sim, a linguagem que melhor se adapta a um sistema de recuperação de informação, com determinadas características" (CESARINO, 1985, p.162).

A autora mostra que a melhor linguagem é aquela que atende às necessidades de informação do sistema, no cumprimento de seus objetivos e metas, na prestação dos serviços de indexação e recuperação da informação.

Nessa mesma linha de pensamento tem-se Lopes (2002, p.51), ao relatar suas conclusões depois que:

"Inúmeros estudos sobre o uso da linguagem controlada e da linguagem natural na recuperação da informação tem se concentrado na utilização conjunta das duas linguagens na estratégia de busca, comprovando que o uso simultâneo dessas linguagens proporciona melhor desempenho nos resultados. Observou-se, também, a predominância de investigações que descrevem as vantagens e desvantagens do uso da linguagem natural e da linguagem controlada nas estratégias de busca em bases de dados. Essa observação reflete a ampla difusão do tema, sem, contudo, apontar especialmente a melhor linguagem, pois vários outros fatores intervenientes afetam os resultados finais das buscas em bases de dados".

Isso retrata o problema enfrentado pelo sistema de informação na representação dos documentos, podendo-se perceber que a utilização de algum controle de vocabulário se faz necessário também na indexação em bases de dados. Todavia, as linguagens de indexação, além de suas estruturas formais e suas relações hierárquicas, devem estar compatíveis aos objetivos do sistema de informação que a utiliza, como políticas de indexação bem definidas, devendo estas também refletir as reais necessidades do usuário.

A linguagem, seja natural ou artificial, exerce uma finalidade essencial na descrição da informação, para a armazenagem, a recuperação, o processamento e a comunicação em sistemas de informação.

Neste contexto, torna-se pertinente estudar as relações que ocorrem entre os termos dentro das linguagens de indexação, que são relações semânticas (sentido do vocabulário) e a sintática (função/ordem do vocabulário em relação ao outro), abordadas no próximo sub-item.

"O controle sobre o vocabulário evidenciará as relações semânticas, ou de significado, que existem entre os termos indexadores. O controle sobre a sintaxe mostrará as relações estruturais ou frásicas entre os conceitos, evidenciando a possibilidade de combinação de elementos para expressar assuntos compostos". (PINTO, 1985, p.172)

As relações semânticas podem-se subdividir em três subtipos:

- a) equivalência;
- b) hierarquias ;
- c) afinitivas ou associativas.

a) As relações de equivalência (termos similares) podem ser representados por mais de um termo. Ainda de acordo com Pinto (1985, p. 173):

"quando a linguagem mostra as relações de equivalência, fornece termos substitutivos de indexação e busca, que podem ser de várias espécies:

- sinônimos e antônimos (adubos x fertilizantes)
(diminutivos x aumentativos)
- quase sinônimos (datiloscopia x impressão digital)
- grafias diferentes (contato x contacto)
- abreviaturas e acrônimos (ONU x Organização das Nações Unidas)
- traduções (recall x revocação)
- ordem de citação (Automação de bibliotecas x Bibliotecas, automação de)".

b) As relações hierárquicas e afinitivas permitem a busca por termos alternativos. Pinto (1985, p.173) explica que:

"as relações hierárquicas exprimem o posicionamento de um conceito em relação a termos mais amplos ou mais específicos. Já as relações afinitivas revelariam termos indexadores que desencadeariam uma ampliação da estratégia de busca. Mostrariam termos afins, aonde <<um lembra o outro>> sem, no entanto, expressarem uma noção de hierarquia".

Essas relações hierárquicas e afinitivas são mais difíceis de serem estabelecidas e mostradas na construção de uma linguagem.

Também para Pinto (1985, p. 174), as relações hierárquicas podem ser estabelecidas em dois tipos:

“- gênero x espécie - quase genérico (gato x animal de estimação) e verdadeiramente genérica (gato x mamífero)
todo x parte (aparelho respiratório x pulmão)”

c) As relações afinitivas, para Pinto (1985, p. 174), possibilitariam a busca de termos alternativos, em relação a:

"coordenação (portas x janelas) (terra x marte)
genética (pai x filho)
concorrente (educação x ensino)
causa e efeito (ensino x aprendizagem)
instrumental (medicina preventiva x vacinação)
material (esquadilhas x alumínio)
similaridade de processo (classificação x indexação)."

É interessante ressaltar que, nas relações hierárquicas e afinitivas, deve-se mostrar somente aquelas que são importantes na recuperação da informação. Essas relações possibilitam ampliar a pesquisa para termos mais amplos, mais específicos e associados, através de referências "VER TAMBÉM".

As relações semânticas podem ser mais complexas ou mais simples, dependendo da estrutura da linguagem de indexação. A linguagem de indexação pode ser alfabética ou classificada. Quanto ao arranjo alfabético, pode-se observar que existe uma dispersão de assuntos por causa da ordem alfabética, dessa forma, as relações de equivalência seriam mostradas através da expressão "VER" e as relações hierárquicas e afinitivas seriam mostradas através do "VER TAMBÉM". Nesse tipo de sistema, existe uma rede de referencia cruzada bastante complexa. Já no sistema de arranjo sistemático, a estrutura de referências é bastante diminuída.

A seguir, observam-se explicações sobre as relações sintáticas que são evidenciadas no controle de vocabulário.

5.1.1.2 Relações sintáticas

A sintática, na definição de Ferreira (1986), "é o que está de acordo com as regras da sintaxe". Pode-se conceituar sintaxe, então, "como parte da gramática que estuda a disposição das palavras na frase e a das frases no discurso, bem como a relação lógica das frases entre si e a correta construção gramatical".

A sintaxe é uma série de regras, para combinar os termos, reduzindo a uma espécie de formato único às diferenças nos processos de expressão sintáticas.

O estudo da sintaxe mostra a possibilidade da coordenação de termos indexadores para a representação de assuntos compostos. Pode-se notar que existe uma tendência à especialização da ciência e tecnologia e, conseqüentemente, a literatura espelha essa especialização, e os documentos tratam mais de assuntos compostos do que assuntos simples.

No processo de análise do documento para determinar o seu conteúdo, pode-se ser observado, segundo Pinto (1985, p.176),

"que muitas vezes esse conteúdo refere-se a conceitos de natureza e complexidade diversas. Por exemplo, documentos sobre: a) Conceito: trabalhador metalúrgico; Natureza: personalidade; b) Conceito: usina siderúrgica; Natureza: entidade; c) Conceito: alumínio; Natureza: material; etc. Pode ocorrer ainda que o conteúdo não se refira a um conceito único, mas à combinação de conceitos simples ou primários que corresponderiam à representação de um assunto composto."

As relações sintáticas entre os termos de uma linguagem de indexação é que possibilitam a recuperação da interseção entre duas ou mais classes de conceitos distintos. Para exemplificar melhor, Pinto (1985, p.177) mostra que:

"Pode existir um interesse por <<fundição do alumínio>> e não por <<fundição>> e <<alumínio>>. Essas relações constituem-se da associação de termos para representar assuntos compostos que, geralmente, aparecem quando se analisam os documentos: são dependentes dos documentos e por isso são temporários, ao contrário das relações semânticas que são permanentes".

A linguagem artificial abriu amplas perspectivas na técnica de armazenar e recuperar as informações nos sistemas automatizados, necessitando, para isso, de um vocabulário controlado que é o tema abordado no próximo sub-item. A linguagem de indexação é, então, um instrumento de controle de vocabulário utilizado nos sistema de recuperação da informação.

5.1.2 Uso do vocabulário controlado nas LI

O vocabulário controlado é uma forma de organização e comunicação da informação em vários domínios do conhecimento. O processo de organização persegue invariavelmente modelos orientadores que permitem comunicar e compartilhar idéias. O vocabulário controlado pode ser definido como "um conjunto de termos organizados de forma hierarquizada e/ou alfabética, com o objetivo de possibilitar a recuperação de informações temáticas, reduzindo substancialmente a diversidade de terminologia" (LOPES, 2002, p. 47).

Quando se procura adotar um vocabulário controlado num sistema de informação, o que se quer, de fato, é poder, no final, satisfazer às necessidades expressas de busca dos usuários do sistema. Para definir o que é vocabulário controlado, é relevante conceituar o que é vocabulário. Existem várias definições, mas a que

mais se aproxima da visão deste estudo é o vocabulário definido como "conjunto das palavras especializadas em qualquer campo de conhecimento ou atividade, nomenclatura, terminologia" (FERREIRA, 1986).

O vocabulário controlado é um banco de dados que define os termos mais utilizados num determinado assunto ou porção de conteúdo e suas relações. Sua função é manter a consistência entre o conteúdo e servir como referência para as diferentes entidades que alimentam o banco de dados.

O vocabulário controlado se faz presente na hora da tradução dos assuntos dos documentos para uma linguagem padronizada, visando uma melhor escolha dos termos indexadores, que mais representam os assuntos dos documentos, tendo em vista alguma padronização. Esse procedimento é refletido no momento da recuperação da informação, mostrando as relações existentes entre os termos indexadores, ajudando os usuários a montarem estratégias de busca mais eficientes.

A necessidade de controle de vocabulário, para Cesarino (1978, p.272), pode ser evidenciada por vários fatores:

a) fatores humanos – diferenças de cultura, de experiência dos autores e domínio de terminologia entre indexadores, autores e usuários;

b) fatores referentes à própria linguagem natural aonde:

-vários termos podem designar o mesmo conceito (sinônimo);

-um mesmo termo pode designar conceitos diferentes (homógrafos);

- os sentidos dos termos se alteram pela posição que ocupam na frase (sintaxe);

c) fatores hierárquicos – um conceito implica em outros mais amplos e/ou mais restritos.

Normalmente esse controle tem sido exercido através de: seleção de termos preferidos; distinção dos homógrafos; estabelecimento formal das relações entre os termos".

Uma outra questão para ser pensada é em relação à organização da produção bibliográfica, que não pode existir sem o controle do vocabulário. É um desafio que, até o momento, ainda não foi solucionado, mesmo com a utilização de moderna tecnologia de computadores, na tentativa de organização desse conhecimento técnico-científico.

O controle de vocabulário implica um processo classificatório, que pode ser usado na recuperação para ampliar o número de documentos relevantes recuperados. Esse processo é dividido em duas etapas distintas de acordo com Lopes (2002, p.42):

“A primeira refere-se à classificação de variantes gramaticais do mesmo termo e ou conceito, significado singular e plural, variantes gramaticais e diferentes flexões dos tempos verbais; em uma segunda etapa, os termos e ou conceitos são agrupados por descreverem o mesmo conceito ou um similar, isto é, sinônimos ou palavras que são equivalentes em seus significados”.

Todavia, o vocabulário controlado torna-se o ponto de convergência entre as linguagens utilizadas por indexadores, pesquisadores e outros, principal ponto para comunicação de informações dentro de um sistema, e entre sistemas. “A linguagem artificial trabalha com o uso de conceitos e a representação do assunto está ligada aos descritores permitidos pelo instrumento de indexação utilizado pelo sistema” (SOUTO, 2003, p.77).

A elaboração de um instrumento com o objetivo de padronizar, exigiu a utilização de técnicas combinadas de documentação e de terminologia, afim de se obterem representações semelhantes de um mesmo documento, submetido a diferentes bibliotecários analistas, e também para tornar este instrumento de trabalho eficaz, ao evitar ambigüidades na comunicação documentária.

O uso da terminologia, para referendar o processo de construção de linguagens documentárias e possibilitar melhor representação e recuperação da informação

não garante por si só o sucesso da comunicação documentária. A atividade de informação requer, além dos parâmetros de um sistema conceitual, a observação das necessidades de uso e recepção da informação.

Os sistemas que visam à organização do conhecimento incluem uma variedade de esquemas que organizam, gerenciam e recuperam a informação, e estão presentes em todas as áreas do conhecimento humano, dos mais simples aos mais complexos. Dentre eles, pode-se citar: classificação, tesouros, listas de cabeçalhos de assunto, etc., que utilizam a pré e pós-coordenação, na combinação dos termos para representar o conteúdo dos documentos, sub-item abordado a seguir.

5.1.3 Tipos de LI

Os tipos de linguagens de indexação, abordados neste estudo, são as linguagens pré-coordenadas e as linguagens pós-coordenadas, as verbais e as simbólicas.

A complexidade dos assuntos está intrinsecamente relacionada com os aspectos de pré e pós-coordenação de conceitos na recuperação de informações.

5.1.3.1 Linguagens pré-coordenadas

Na pré-coordenação, a representação dos assuntos compostos é feita pela combinação dos termos no momento da entrada desses em índices, e catálogos. Num sistema pré-coordenado, a representação do conteúdo do documento, para Pinto (1985, p.178), “revela a totalidade do seu assunto, através de um enunciado,

muitas vezes único, os vários tópicos relativos ao conteúdo conceitual do documento são reunidos numa entrada que resume todos os aspectos que compõem o assunto”.

Sistemas pré-coordenados são definidos de acordo com Pinto (1985, 178), como “sistemas que estabelecem a coordenação dos vários tópicos referentes a um assunto composto no momento de indexação”. A autora alerta que existe um problema sério com a pré-coordenação, referente ao fato da importância de um determinado termo em relação ao outro, ser julgado na hora da entrada do termo no sistema de recuperação da informação.

As pré-coordenadas são as linguagens de indexação que:

“Combinam ou coordenam os termos no momento da indexação, como por exemplo, os sistemas de classificação CDD, CDU, abrangem todas as áreas do conhecimento e, representam os conceitos através de notações compostas de números, letras ou a mistura de ambas. Atualmente, tem basicamente a função, de ordenar por assunto, os livros nas estantes. As listas de cabeçalhos de assunto, também linguagens pré - coordenadas, são listas gerais que arrolam termos de todas as áreas do conhecimento, alfabeticamente organizadas, com sinais como o traço, a vírgula e o parêntese para estruturar os cabeçalhos indiretos. Apresentam sinais de relação entre os cabeçalhos.” (AMARAL, 2002)

Para Lancaster (1993, p.43), os sistemas pré-coordenados apresentam as seguintes características:

- 1) “É difícil representar a multidimensionalidade das relações entre os termos;
- 2) Os termos somente podem ser listados numa determinada seqüência (A, B, C, D, E), o que implica que o primeiro termo é mais importante do que os outros;
- 3) Não é fácil (senão completamente impossível) combinar termos no momento em que se faz uma busca”.

Nos sistemas automatizados de recuperação da informação, normalmente a coordenação dos termos de um assunto composto, que seria no momento da

indexação, é transferido para o momento da busca. Neste caso, são chamados de pós-coordenados, tema tratado no sub-item a seguir.

5.1.3.2 Linguagens pós-coordenadas

Para Pinto (1985, p.183), “um sistema pós-coordenado consiste de entradas que normalmente são conceitos únicos (termos indexadores) e de saída que permite a comparação das entradas de assunto, com o objetivo de determinar coincidências que revelem documentos pertinentes ao assunto composto pesquisado”.

Nos sistemas pós-coordenados segundo Pinto (1985, p.182), “a coordenação dos elementos básicos de um assunto composto no momento da indexação, é transferido para o momento de busca”. De acordo com a autora nesse sistema não existe a preocupação com a importância dos elementos de um assunto composto na hora da entrada dos termos nos sistemas de recuperação da informação.

Um sistema pós-coordenado consiste em entradas que normalmente são conceitos únicos, termos indexadores. Essa entrada resulta numa saída que revele documentos pertinentes ao assunto composto pesquisado.

Nos sistemas pós-coordenados, também existem problemas em relação ao controle de vocabulário, ou seja, os problemas relativos ao controle de semântica e sintaxe.

“O problema semântico relativo às relações de equivalência, hierarquias e afinativas, deve ter uma solução de acordo com as mesmas diretrizes utilizadas num sistema de pré – coordenado: o uso de remissivas para

relações de equivalência e o uso de referências do tipo <<ver também>> para relações hierárquicas e afinitivas. O problema sintático, relativo aos assuntos compostos, apesar de simplificado, em relação à ordem de citação, permanece no que se refere à determinação das partes que irão compor um assunto composto”. (PINTO, 1985, p.184)

Em sistemas pré-coordenados têm-se criado e usado listas de cabeçalhos de assunto, e, em sistemas pós coordenados, a linguagem adotada tem sido os tesouros. O próximo sub-item trata especificamente dos instrumentos de indexação utilizados na descrição de assunto, que são as linguagens verbais: lista de cabeçalhos de assuntos e tesouros.

5.1.3.3 Linguagens verbais

Entre as linguagens verbais mais conhecidas destacam-se os tesouros e as listas de cabeçalhos de assuntos. Os tesouros e as listas de cabeçalhos de assuntos possuem diferenças quanto à origem e à estrutura, e são empregados para a elaboração de diferentes instrumentos de recuperação da informação. Os instrumentos, como os tesouros, são vocabulários controlados utilizados quando se pretende que um mesmo texto, submetido a analistas de diferentes lugares, apresente representações semelhantes, condição essa necessária à transmissão da informação.

5.1.3.3.1 Tesouros

Tesouros é um índice controlado utilizado tanto na indexação, quanto na recuperação de documentos. Segundo Tristão (2004, p.162),

"os tesouros são linguagens de indexação, ou seja, são sistemas artificiais de signos normalizados, que permitem representação mais fácil e efetiva do conteúdo documental, com o objetivo de recuperar manual ou automaticamente a informação que o usuário solicita".

Os tesouros, para Pinto (1985, p.185), constituem a linguagem adotada em sistemas pós - coordenados, automatizados e especializados, aonde não existe a preocupação com a importância dos elementos de um assunto composto, e com a ordem de citação desses elementos. A combinação dos descritores se faz por ocasião da pesquisa documentária em sistemas de informação.

Conforme Tristão (2004, p.167), tesouro pode ser definido como uma "seleção de termos, baseados em análise de conceitos, na qual se define o termo geral, de maior abrangência, e sua relação com termos mais específicos, que representam os conceitos menores". Para o autor, tesouros podem ser utilizados na organização dos mais diversos sistemas gerais e especializados de organização e representação do conhecimento. Propõe que os tesouros possam ser distribuídos da seguinte maneira:

- a)"os que dizem respeito a objetos, como minerais, plantas, animais etc. ... (fala-se de taxonomia);
- b)os que dizem respeito a uma disciplina, aqui se podem falar de classificações mono ou pluridisciplinares;
- c)os que dizem respeito a todas as disciplinas ou assuntos, que são as chamadas classificações universais".

Os tesouros mostram os conceitos de determinada área de conhecimento, fixando-os e dando-lhes autoridade.

"Os tesouros, são constituídos por listas de termos autorizados: descritores e não descritores, de um domínio do conhecimento, tendo em vista relações semânticas e lógicas, sendo a sua construção apoiada no *corpus* discursivo da área que abrange, isto é, a seleção de termos é feita dentre mais frequentes, buscando assim assegurar a semelhança entre o vocabulário a ser utilizado para a representação dos artigos e o universo do conhecimento expresso nos textos. Desta forma os tesouros cristalizam

os conceitos de determinada área de conhecimento, fixando-os, e dando-lhes autoridade". (LUCAS, 1999)

Os tesouros são mecanismos que formam uma estrutura de conceitos, ou seja, criam-se uma lista de termos relacionados entre si, em que o termo mais usado, é definido tanto de forma geral, como de forma mais específica e suas co-relações.

"O conceito não é apenas um elemento de significação do termo: o termo acaba sendo um elemento do próprio conceito - o terminum - que sintetiza o conceito como um todo e permite a comunicação, neste caso, verbal. Não importa se o termo é formado por mais palavras, se é constituído por um substantivo mais um adjetivo etc. O que importa é que ele denota um referente. Assim, tratar o termo como representante de um referente, com suas características, é dar a ele um tratamento terminológico". (TRISTÃO, 2004, P.168)

Neste processo, se estabelece uma equivalência entre o termo, e as características necessárias de um referente conceito, com o propósito de restringir o uso do termo em um contexto, em uma determinada área, resultando conseqüentemente, na definição desse conceito dentro de um sistema. Assim, a terminologia visa atender a uma exigência de padronização, para que haja comunicação sem ambigüidades, isto é, o termo descriptor deve expressar apenas um conceito e vice-versa.

A definição do termo fornece características do conceito, que permite seu agrupamento, formação das categorias e indica as relações. Um termo tem várias definições, escolhe-se a mais útil. A estrutura de um tesouro segue uma linha hierarquizada de relações entre os termos, e sua construção demanda um conhecimento técnico e especializado.

"A terminologia se ocupa, sempre, de uma área de conhecimento em que há uma seleção das características relevantes para aquela área e, também, para os propósitos do trabalho. Assim, conforme a área e conforme o ponto de vista abordado as características e conseqüentemente as relações entre os conceitos mudam e é por esse motivo que os tesouros são desenvolvidos de forma específica a cada área do conhecimento" (TRISTÃO, 2004, p.168).

Os tesouros são mecanismos utilizados de forma bastante intensa e, atualmente, estão entre as ferramentas mais estudadas, como recurso para a organização do conhecimento, na transferência da informação via sistemas automatizados, pelo fato de trabalhar o conceito, e apresentar-se apto a atender a demanda em todos os campos do conhecimento humano. Ocupa posição de destaque como elemento central na organização e divulgação da informação, possibilitando compartilhar os documentos de maneira rápida e fácil, deixando fluir o conhecimento com recuperação eficiente da informação.

Os tesouros, para Pinto, (1985, p.185) "exercem um controle menos rígido, mais sugestivo e flexível. Utilizam um grau de pré - coordenação mínimo, somente aquele necessário ao não comprometimento da significação dos descritores". Oferece muita ajuda aos intermediários que executam a busca, incluindo entre outros, o controle de sinônimos e quase sinônimos; a separação dos homógrafos.

A utilização de tesouro permite, ao sistema, ser mais objetivo na indexação de assuntos, diminuindo a subjetividade, visto que em sistemas automatizados, a junção dos elementos básicos de um assunto composto da indexação é transferida para o momento da busca.

"Os tesouros são baseados nas classificações facetadas, apresentam relações de equivalência, hierárquicas (gênero/espécie; coisa/tipo), relações parte/todo, poli hierárquicas, relações associativas e/ou não - hierárquicas. Possuem sinais para indicar a função dos descritores. Ex. TG (BT) = termo geral; TE (NT)= termo específico" (AMARAL, 2002).

O uso de tesouros em bases de dados pode ser analisado como um poderoso recurso facilitador para a busca de informação, diminuindo a inconsistência na recuperação da informação. Para Souto (2003, p.78) "o uso do tesouro como instrumento é sem dúvida uma estratégia eficaz para a busca de informações em

bases de dados.” Ele impede que documentos que se encontram indexados deixem de ser localizados, ou seja, posteriormente recuperados.

A seguir, será tratada uma outra ferramenta de controle de vocabulário, a lista de cabeçalho de assunto, também utilizada em sistemas automatizados.

5.1.3.3.2 Listas de cabeçalhos de assunto

As listas de cabeçalhos de assunto é um instrumento de controle utilizado em sistemas de recuperação da informação que adotam como linguagem o cabeçalho de assunto. A lista de cabeçalhos de assunto é um instrumento auxiliar da análise documentária, determinando a recuperação da informação em sistemas pré-coordenados. Segundo Amaral (2002, p.2), “são listas gerais que arrolam termos de todas as áreas do conhecimento, alfabeticamente organizadas, com sinais como o traço, a vírgula e o parêntese para estruturar os cabeçalhos indiretos e apresentam sinais de relação entre os cabeçalhos”.

Para contextualizar as listas de cabeçalhos de assunto, é possível relacionar algumas instituições internacionais e nacionais que trabalham com essa ferramenta, como por exemplo: LCSH (Library of Congress Subject Headings); Rede do Bibliodata Calco, FGV (Fundação Getúlio Vargas); BN (Biblioteca Nacional); Rede Pergamum, UFMG (Universidade Federal de Minas Gerais), etc.

As listas de cabeçalhos de assunto são classificadas e arranjadas alfabeticamente, e a combinação dos termos é fixada na hora da indexação, São dotadas de uma sintaxe controlada.

As listas de cabeçalhos de assunto, para Pinto (1985, p.179),

"Estabelecem a coordenação dos vários tópicos referentes a um assunto composto no momento de indexação, e com grande subjetividade, determinam uma prioridade na citação desses elementos. Mesmo que a meta seja atender os objetivos da instituição e as necessidades do usuário, essa tarefa é subjetiva. Obviamente, os elementos que não o primeiro, estariam ocultos, ou seja, o usuário não teria acesso a eles diretamente. Esse tipo de sistema requer o que se chama de "entrada múltipla" ou uma rede de referências ou remissivas bem elaboradas".

Assim, cada elemento referente a um assunto composto, passa a ser um ponto de acesso, ampliando as possibilidades de busca, mas, esse procedimento de "entrada múltipla", acarreta várias dificuldades como: "o sistema ficaria mais caro, o tempo de indexação seria maior, aumentar-se-ia o tamanho do índice, etc.". (PINTO, 1985, p.179)

Nesses sistemas, a coordenação dos vários tópicos referentes a um assunto composto é estabelecida no momento da indexação, aumentando muito a subjetividade na determinação de prioridades na ordem de citação, ou seja, ordem preferida de apresentação dos elementos de um assunto composto.

Para diminuir a subjetividade na citação de assuntos compostos, Pinto (1985, p.179) afirma que:

"Desde os tempos de Cutter, quando em 1876 publica sua obra "Rules for a dictionary catalog", existe a preocupação de se orientar o indexador sobre o julgamento da importância dos itens num enunciado de assunto composto."

Isto reflete na busca, em que o usuário só terá acesso ao primeiro termo num cabeçalho de assunto composto. Esse caso é problemático, a julgar a importância de cada um desses tópicos para criar a ordem de citação, que melhor atenda à necessidade do usuário.

Para Lancaster (1993, p. 15) “a tradicional lista de cabeçalhos de assuntos é similar ao tesouro pelo fato de sua base ser alfabética. Difere do tesouro por incorporar uma estrutura hierárquica imperfeita e por não distinguir claramente as relações hierárquicas das associativas”.

Tanto as listas de cabeçalhos de assuntos como os tesouros têm que se adequar às necessidades de informação da instituição para o qual foram criados, e permitir a adoção de novos termos que surgem na literatura científica, de forma a retratar esses conceitos, visando adequar seus termos às novas necessidades de busca dos usuários da informação. O mérito dessas linguagens verbais é poder absorver sempre os novos termos e possibilitar a busca pelos mesmos.

No próximo sub-item, são abordadas as linguagens simbólicas CDU e CDD, que são os sistemas de classificação bibliográfica mais utilizados internacionalmente.

5.1.3.4 Linguagens simbólicas (CDD e CDU)

Os sistemas de classificação bibliográfica, como o Dewey Decimal Classification (CDD) e a Classificação Decimal Universal (CDU), são considerados as mais antigas linguagens de indexação. Foram elaboradas com o objetivo de organizar os acervos de bibliotecas e facilitar o acesso dos usuários à informação contida nos acervos.

A CDU e a CDD segundo Amaral (2002, p.2), “abrange todas as áreas do conhecimento, e representam os conceitos através de notações compostas de números, letras ou a mistura de ambos. Atualmente, têm basicamente a função de ordenar por assunto, os livros nas estantes”.

A CDD surgiu primeiro e influenciou na construção da CDU, de acordo com Gigante (1995, p.1),

“a CDD é calcada na classificação das ciências de Francis Bacon, tem suas dez classes baseadas na divisão do conhecimento em três grandes grupos: memória, imaginação e razão. E depois da CDD, surgiu a CDU, construída também sob os princípios da divisão científica do conhecimento, que traz um avanço em relação à primeira : aumento da capacidade de síntese, ou seja, possibilidade de representar assuntos complexos e de classes diferentes por meio de mecanismos de combinação; incorporação do princípio de análise por facetas, princípio que permite uma análise multidimensional dos assuntos”.

As linguagens simbólicas (CDD e CDU) apresentaram transformações no decorrer do tempo, segundo a autora Lara (2002, p.133), “Com o crescimento do conhecimento, registrado e, conseqüentemente, das segmentações propostas para analisar a realidade, esses sistemas foram submetidos a inúmeras modificações para dar conta do novo”.

Conseqüentemente, surgiram publicações de edições na tentativa de acompanhar o desenvolvimento técnico-científico. De acordo com Lara (2002, p.1330),

“As sucessivas edições afastam-se, progressivamente, da hipótese inicial, mas sofrem os efeitos da ausência de alteração radical nas formas de segmentação do novo universo documental. Passam a conviver, nesses sistemas, segmentações de origem distinta, do que decorrem amálgamas de características provenientes de distintas segmentações; na mesma hierarquia. É por esse motivo que tais sistemas têm sido questionados quanto à eficácia para cumprir as funções de organização da informação”.

Esses sistemas promovem o acesso aos dados e os sistemas de informação, hoje, estão mais centrados nos usuários, de modo a torná-lo cada vez mais, independente dentro de um sistema de biblioteca ou centro de informação. Então, é necessário que se pense melhor na organização do acervo e seu sistema de acesso. Segundo Gigante (1995, p.4),

“Uma classificação bibliográfica como CDD ou a CDU pode até conseguir acomodar, em um nível de quase excelência, as obras de um acervo. Mas

a notação que produz, nos casos de alta especificidade (cuja ocorrência também é alta no momento histórico-científico em que vivemos), caracteriza uma interface muito pouco amigável para os usuários da biblioteca”.

As classificações são utilizadas como instrumentos de organização dos acervos bibliográficos e o objetivo primordial de todo esse processo é o acesso pelo usuário. No capítulo que se segue observa-se a questão da recuperação da informação em sistemas automatizados, em que a boa recuperação depende da qualidade dos instrumentos utilizados na indexação.

6 RECUPERAÇÃO DA INFORMAÇÃO NOS SISTEMAS AUTOMATIZADOS, ATRAVÉS DO VOCABULÁRIO CONTROLADO

Os sistemas de recuperação da informação, para Cesarino (1985, p.157), "podem ser definidos como um conjunto de operações consecutivas executadas para localizar, dentro da totalidade de informações disponíveis, aquelas realmente relevantes". São operações de extrema importância para os sistemas de informação que organizam o conhecimento com o objetivo de disponibilizar a busca por parte de seus usuários.

De acordo com Carneiro (1985, p.159), os sistemas de recuperação de informação podem ser estudados sob dois aspectos:

- a) como um conjunto de operações consecutivas, executadas para localizar, dentro da totalidade de informação disponível, aquelas que seriam necessárias ao usuário, ou seja, aumentar a transmissão de informação relevante e diminuir a transmissão de informação não relevante.
- b) como parte de um modelo de comunicação dentro de um contexto social-cultural-histórico".

Do ponto de vista da autora, os dois aspectos não se excluem, eles se complementam, o sistema de informação disponibiliza a informação dentro de um contexto social, sendo, ao mesmo tempo, um sistema de comunicação científica.

O importante é que os responsáveis pela construção e manutenção das bases de dados estejam conscientes que de seu trabalho inicial, depende todo o resto do processo.

"Um sistema de recuperação de informação precisa justificar e comprovar a qualidade dos seus serviços. Os bons resultados dos sistemas de recuperação de informação dependem muito da maneira como o usuário interage com o sistema, isto é, a interpretação que o usuário faz da capacidade e das limitações do sistema, a habilidade em expressar a própria necessidade de informação numa linguagem adequada, a

divulgação que o sistema de recuperação de informação faz da sua própria organização e serviços, e pela orientação e assistência que dá ao usuário.” (CESARINO, 1985, p.166).

A opinião expressa pela autora sobre a prestação de serviços de recuperação da informação pelos sistemas automatizados é uma realidade nos dias de hoje, e deve ser uma preocupação constante, para sistemas de buscas que querem prestar serviços de qualidade. A busca eficiente de informação em acervos digitais não dispensa o auxílio de um sistema de recuperação da informação.

Os sistemas de recuperação da informação, registradas em documentos, também denominados de bancos de dados, segundo Lopes (2002, p.60), “veio ampliar significativamente a qualidade das buscas bibliográficas, visto que essas bases proporcionam diversificados pontos de acesso à informação”.

A organização se dá em função do usuário, mostrando as relações existentes entre os termos indexadores, ajudando os usuários a montarem suas buscas mais eficientes, para satisfazer suas necessidades de informação e evitar a recuperação de itens inúteis. Os sistemas de recuperação da informação têm como objetivo, atender às necessidades de informação do usuário, diante de um acúmulo constante, e um volume sempre crescente da informação.

Para Cesarino (1978, p.273), todas as linguagens de indexação exercem a mesma função nos sistemas de recuperação da informação:

- a) representam o assunto de uma maneira consistente;
- b) permitem a coincidência entre linguagem do indexador e do pesquisador;
- c) possibilitam ao indexador modular o nível de sua pesquisa, do mais geral para o mais específico, e vice – versa, segundo a necessidade do usuário.”

As linguagens de indexação atuam nos sistemas de recuperação de informações em dois níveis: orientando o indexador sobre quais os melhores termos para a

representação do assunto de um documento, e orientando o pesquisador sobre a escolha dos termos indicadores que corresponderiam à representação do assunto por ele procurado. Além disso, orientaria a elaboração de uma estratégia de busca que o levasse com menor esforço e maior eficiência aos documentos procurados.

A recuperação mais pertinente à questão de busca é aquela cuja indexação proporcionou a identificação de conceitos mais pertinentes ao seu conteúdo, produzindo uma correspondência precisa com o assunto pesquisado em índices.

Um índice bem elaborado facilita muito o trabalho de busca do pesquisador. Representa o significado e o conteúdo dos documentos a serem recuperados. Conforme Pinto (1985, p.170) "Índice é o instrumento mais importante num sistema de recuperação de informação. Tem sido descrito como a chave para o acesso ao conteúdo dos documentos, ou como uma ponte ou elo entre o conteúdo de um arquivo de informações e seus usuários".

A eficiência dos sistemas de recuperação de informação depende muito da estrutura e composição da linguagem de indexação. Se o sistema for composto de classes muito abrangentes, gerais, a recuperação terá um grau muito superficial de especificidade, levando conseqüentemente a uma alta revocação e baixa precisão; em contra partida, se a linguagem retratar um número muito grande de classes, estas serão muito específicas e o sistema terá um maior grau de precisão e conseqüentemente um baixo grau de revocação.

"A revocação se relaciona com a capacidade do sistema em assegurar a recuperação de um número desejável de documentos relevantes e a precisão se relaciona à capacidade do sistema em impedir a recuperação de documentos não relevantes". (CARNEIRO, 1985, p.234)

É o usuário que determina a relevância num sistema de recuperação da informação. “A especificidade¹ e a exaustividade² na indexação estão diretamente ligadas à precisão e revocação na recuperação. A especificidade e a exaustividade estão relacionadas tanto à quantidade de cabeçalhos quanto ao uso de termos conceitos gerais e específicos” (SOUTO, 2003, p. 79). São critérios de avaliação de linguagens documentárias na recuperação da informação.

A linguagem de indexação, para Fujita (2004, p.17),

“faz-se necessária uma vez que a qualidade dos serviços de disseminação da informação depende da linguagem utilizada pelo sistema para a indexação e recuperação dos documentos. Isso se dá, quando a linguagem do sistema permite que se traduza a linguagem do autor sem que se perca a idéia principal, e quando permite que se traduza a linguagem do usuário de modo que satisfaça suas necessidades de informação”.

Os sistemas de recuperação de informação organizam e viabilizam o acesso aos dados de informação, desempenhando as atividades de: representar os assuntos contidos nos documentos, através da indexação, descrição temática; armazenamento da documentação com suas respectivas representações; recuperação das informações representadas e dos próprios documentos armazenados, de forma a atender as necessidades de informação dos usuários,. É necessário que haja uma interface do usuário com o sistema, em que o usuário irá descrever suas necessidades e questões, através da qual seja possível examinar os documentos recuperados e/ou suas representações.

A realização de um trabalho integrado entre o sistema, a linguagem e o próprio usuário, se realiza quando existe a busca de informação desejada pelo usuário e a

¹ É a extensão em que o sistema nos permite ser precisos ao especificarmos o assunto de um documento que estejamos processando (FOSKETT, 1973)

² É uma medida de extensão em que todos os assuntos discutidos em um certo documento são reconhecidos na operação de indexação e traduzidos na linguagem do sistema. (LANCASTER, 1993)

recuperação desta de acordo com sua pesquisa, ou seja, uma estratégia de busca bem elaborada tema do próximo sub-item.

6.1 Estratégia de busca

Estratégia de busca, para Kremer (1985, p.196), “é o procedimento através do qual se procuram documentos determinados ou dados sobre um assunto”. Pressupõe sempre a existência de um sistema de recuperação de informação que constitui um processo de busca da informação.

A estratégia de busca, para Lopes (2002, p.61), “caracteriza-se como uma técnica, ou conjunto de regras para tornar possível o encontro entre uma pergunta formulada e a informação armazenada em uma base de dados”. O uso das linguagens controlada e natural, nas estratégias de busca, proporciona uma perspectiva sobre a complexidade da busca de informação bibliográfica, referente à seleção de termos representativos.

Estratégia de busca é o procedimento através do qual se procuram documentos determinados ou dados sobre um assunto. Pressupõe sempre a existência de um sistema de recuperação da informação, também denominados de bancos de dados e, conseqüentemente, as suas bases de dados proporcionam diversificados pontos de acesso à informação.

"Esses sistemas possibilitam o planejamento de estratégias de busca com maior nível de complexidade envolvendo vários conceitos na mesma estratégia; permitem a utilização de busca de palavras apenas dos títulos e resumos dos documentos, isto é, termos da linguagem natural; buscam os termos específicos de linguagens controladas, nos campos de descritor; buscam por autores; por ano de publicação; por títulos de periódicos; por classificação; permitem, também, a busca de conceitos compostos ou simples e a possibilidade de truncagem de raízes de palavras e de substituição de caracteres no meio dos termos, dentre outros recursos de recuperação". (LOPES, p.60, 2002)

É preciso que haja uma perfeita identidade entre a linguagem adotada pelo sistema na entrada dos dados e a linguagem utilizada durante a busca, sendo necessário, definir claramente qual é a necessidade de informação do usuário, de forma a aproximar ao máximo a demanda expressa com a que deverá ser identificada. Esse tipo de busca é realizado normalmente em índices.

É óbvio que a recuperação por um determinado termo só se efetivará se os termos incluídos nas estratégias de busca corresponderem aos termos adotados na indexação de assuntos. A seleção de termos para a estratégia de busca é a chave de todo o processo. Deve-se estar atento sempre para a questão referente a que termo de busca deve ser selecionado para um determinado tema, de forma a representar efetivamente o problema, a necessidade de informação do usuário.

Para Ihadjadene (1998, p.199), “as estratégias de buscas se tornam simplificadas, porque a navegação nos vocabulários controlados, sob a forma de hipertexto, permite aos usuários designar os termos dos quais necessitem, ao invés de buscá-los mentalmente”. Mas o que é simples para um determinado usuário pode não ser para outro. Essa operação pode ser simples ou complexa, variando conforme o tipo de demanda do usuário, os recursos de busca oferecidos e o grau de sofisticação da indexação e armazenamento dos dados em determinado sistema de recuperação da informação.

No entanto, na opinião de Lopes (2002, p.63), a estratégia de busca é “um processo de alta complexidade envolvendo numerosos fatores e variáveis além de decisões e o entrelaçamento dos subprocessos inter-relacionados com a busca.” A autora afirma que a chave do processo de busca é a seleção de termos de busca, para isso é necessário alguma habilidade de pesquisa em bases de dados. Deve-se assegurar uma definição adequada do tópico, ou assunto, antes de partir

para a busca efetivamente. É necessário definir o problema, listar os termos mais apropriados do assunto pesquisado, para acentuar a qualidade da busca.

O alcance da qualidade na informação recuperada requer o planejamento de estratégias de busca específicas para cada tipo de base de dados. “Tanto as bases de dados referenciais, quanto as textuais, exigem um planejamento acurado da estratégia de busca, visando a uma recuperação de informação de acordo com as necessidades dos usuários”. (LOPES, 2002, p.60)

A escolha de aonde buscar, identificando qual a base de dados é a mais qualificada para atender a perguntas específicas do solicitante, depende de uma série de fatores de acordo com Lopes (2002, p.62), “fatores relacionados com a abrangência de assunto e período da base, dos tipos de documentos indexados, da linguagem de indexação, dos campos de busca disponibilizados em determinado banco de dados e outros”.

O comportamento de busca da informação é uma atividade individual, em que o indivíduo se empenha em identificar uma mensagem para satisfazer sua necessidade naquele momento. De acordo com Ferreira (1995, p.223),

“qualquer tentativa de descrever padrões de busca de informação deve-se admitir o indivíduo como o centro do fenômeno e considera a visão, necessidades, opiniões e danos desse indivíduo como elementos significantes e influentes que merecem investigação. As necessidades de informação, ainda, devem ser definidas em nível individual, destacando-se atenção para o tempo e espaço específicos experimentados pelo elemento em particular”.

Uma boa estratégia de busca assegura resultados mais relevantes de citações recuperadas. Com isso, o uso dos termos do vocabulário controlado é um recurso imprescindível para uma boa recuperação nos sistemas de informação automatizados. É preciso familiarizar-se bem com os recursos de cada base de dados, consultando sempre os índices ou tesauros do sistema ao elaborar uma

estratégia de busca. Conforme, Cuenca (1999, p.297), “na medida em que o usuário vai se aperfeiçoando no uso das bases, ele passa a ser mais exigente quanto à relevância dos resultados de busca, concluindo que o uso do vocabulário controlado, permite um produto mais refinado ao tema pesquisado”.

A navegação nos vocabulários controlados permite aos usuários designar os termos dos quais necessitam. Para Lopes (2002, p.41), “interfaces amigáveis foram desenvolvidas para permitir o uso maciço dos bancos e bases de dados”, no sentido de prover o usuário final com os instrumentos necessários para que o mesmo realize suas próprias buscas.

Lopes (2002, p.67), explica que a combinação de “um bom conhecimento das específicas bases de dados e com as linguagens de busca dos sistemas de recuperação da informação, permite a realização de buscas de qualidade nos sistemas”. A atividade de busca bibliográfica em um sistema de informação, necessita de uma linguagem de indexação que proporcione uma recuperação de informação condizente com as necessidades de informação desses usuários. E o usuário precisa saber utilizar os recursos de busca do sistema, para tornar-se capaz de construir uma estratégia de busca adequada a cada demanda ou necessidade de informação, de forma a obter o mais alto custo benefício, sem desperdício de tempo de conexão com o computador.

O acesso em rede aos grandes sistemas de recuperação da informação, conforme Lopes (2002, p.60), “veio ampliar significativamente a qualidade das buscas bibliográficas”, mas requer dos seus usuários determinados domínio da tecnologia eletrônica. O usuário precisa se interagir com os mecanismos que lhe são oferecidos, tema do próximo sub-item.

6.2 Interação com o usuário

A informação disponível no meio eletrônico está sob a forma de registros, contidos em bases de dados, bibliotecas digitais, que são acessados somente por meio das Interfaces de busca, disponibilizada por esses provedores de informação, o que pressupõe uma interação entre um usuário humano com o banco de dados. É interessante e confortável para o usuário poder submeter sua necessidade de informação e interagir com o sistema de informação, e obter o retorno de informações relevantes.

O sistema automatizado deve adaptar-se ao contexto e às necessidades e preferências dos usuários, tornando seu uso mais eficiente. “É necessário que sua interface seja flexível para realizar de diversas maneiras a mesma tarefa, correspondendo às características de cada tipo de usuário”. (Dias, 2003, p.37)

É um erro pensar que, com a disponibilização dos catálogos das principais bibliotecas e bancos de dados via web, as dificuldades de acessos mudaram os usuários. Esses muitas vezes não têm competência e paciência para fazer buscas em sistemas de recuperação.

No universo bibliográfico, normalmente, conforme Dias (, 2003, p.32), as interfaces são “projetadas para usuários especialistas, na área atendida pelo sistema, e podem adotar terminologia especializada e maior densidade de informações”. Dessa forma, os usuários leigos naquele assunto “necessitam de maiores explicações sobre as diferentes opções e termos empregados, sendo útil um glossário e telas de ajuda mais detalhadas”.

Além disso, pode-se observar que a maioria das pessoas, mesmo aquelas que possuem alguma familiaridade com computadores, não são capazes de

explorarem todas as facilidades disponíveis em um sistema de informação. Para Cuenca (1999, p.294),

“Além do desconhecimento, dificuldades como a existência de várias interfaces de busca para o acesso às bases de dados, tempo de busca, campos disponíveis para recuperação e a não-familiarização com o vocabulário especializado da área são colocados como motivo da não-utilização do acesso às bases automatizadas pelos usuários de buscas informatizadas”.

É interessante que haja uma interação do usuário com o sistema, e que lhe seja permitido, desenvolver e refinar sua estratégia de busca enquanto ela vai sendo realizada. Isto proporciona conseqüentemente, economia de tempo e dinamização da pesquisa.

Normalmente, nos bancos de dados, é realizada operação de busca em linha, aonde o usuário realiza uma pesquisa, e lhe é permitido rever suas estratégias, definidas anteriormente ao longo do processo de busca em qualquer etapa da mesma. Dessa forma, o usuário estaria utilizando os recursos de busca oferecidos pelos sistemas de recuperação de informação, atividade que, geralmente, necessita de um programa de capacitação.

“Nos sistemas de recuperação de informação, há usualmente interface através da qual o usuário traduz sua necessidade de informações em forma de questões ou palavras-chave, ou mesmo no momento em que ele examina os documentos na busca de informações pertinentes. Essas ações são caracterizadas como papel do usuário. Os dois modos de buscar informações são classificados em modelos de recuperação e os modelos de navegação. Nestes últimos, o usuário não propõe uma questão ou necessidade de informação ao sistema. Em vez disso, navega através dos documentos, buscando informações de interesse”. (SOUZA, 2006, p.165)

A interface pode ser realizada de duas formas, quando usuário propõe uma questão ou necessidade de informação ao sistema, e quando o usuário navega através dos documentos, buscando informações de interesse, procedimento

comum nas bases dados textuais. A busca em estruturas de arquivos ligados em rede é usualmente executada através de navegação do tipo hipertextual³.

Para Ihadjadene (1998, p.197), "os vocabulários controlados sob forma hipertextual, permitem aos usuários conhecer os descritores ou assuntos utilizados numa base de dados", e dessa forma, seus usuários são instruídos para que eles mesmos realizem suas próprias buscas, com interfaces amigáveis, programas baseados em menus, possibilitando ao usuário menos experiente uma orientação passo a passo, bem como navegar e consultar cada recurso individualmente e traçar o perfil de suas necessidades de informação.

Pode-se perceber que os sistemas de informação estão empenhados em facilitar ou melhorar, cada vez mais, a sua interface com os seus usuários. A interface apresenta constantes inovações tecnológicas, sendo mais flexível o bastante para realizar a mesma tarefa de diferentes maneiras, de acordo com o contexto e com as características de cada tipo de usuário.

"Deve-se fornecer ao usuário procedimentos e opções diferentes pra atingir o mesmo objetivo, da forma que mais lhe convier. Quanto mais variadas as maneiras de realizar uma tarefa, maiores são as chances que o usuário possui de escolher e dominar uma delas no curso de seu aprendizado"
DIAS (2003, p.37),

Existe, no momento, uma necessidade de definir o perfil dos usuários, determinando as suas características individuais e suas principais demandas e necessidades. A eficiência desse serviço está estritamente relacionada com a precisão na determinação desse perfil. É o próprio usuário que traça o seu perfil com ajuda da interface do sistema de informação, e deve refletir seus interesses

³ Na sua forma mais simples, um sistema hipertexto é constituído por um conjunto de textos ou de partes de textos interligados por elos. A navegação é o modo típico de consulta e de recuperação de um sistema hipertexto. (IHADJADENE, 1998, p.197)

correntes, resultando numa interface perfeita. Percebe-se que as bases de dados dos sistemas de informação estão adotando esse serviço como uma forma de melhorar o atendimento às necessidades de informação de seus usuários potenciais.

Outro ponto positivo que se pode observar é que as interfaces de uma base de dados para outra, muitas vezes não apresentam grandes alterações na sua estrutura de busca, uma vez que o usuário possui certo domínio de uma interface, ele não terá grandes dificuldades em pesquisar em outra base de dados. As interfaces devem de ser de fácil compreensão, para permitir a sua utilização em um tempo de conexão satisfatório, dessa minimizando o custo da pesquisa.

Constata-se que os sistemas de informação têm procurado se interagir de maneira mais eficiente possível com seus usuários, com intensivos programas de treinamentos. O profissional da informação torna-se cada vez mais necessário para o efetivo desempenho do fluxo da informação nesses sistemas, e quanto mais a tecnologia é desenvolvida, maior é o compromisso do profissional da informação na organização e na busca por informação, atuando como condutor dessas ferramentas no atendimento às necessidades dos usuários. Este tema é abordado no próximo capítulo.

7 O PAPEL DO PROFISSIONAL DA INFORMAÇÃO NO CONTEXTO DESTE TRABALHO

O papel que a informação exerce na atual sociedade pós-industrial, faz surgir um novo paradigma em relação ao profissional da informação para com os usuários de sistemas automatizados, destacando-se a necessidade de comunicação, do sistema para com o usuário, e a qualidade do serviço de informação centrado no usuário.

“É muito importante que a coordenação do processo de construção de bases de dados esteja sob a responsabilidade de um profissional da informação. Ele pode ser o responsável direto pelo sucesso do serviço. A competência desse profissional está relacionada com o domínio e o conhecimento das técnicas de indexação e de elaboração e uso de linguagens documentárias”. SOUTO (2003, 79)

O profissional da informação assume cada vez mais um papel importantíssimo nos sistemas de informação. Podendo-se considerar que o bibliotecário, devido à sua formação acadêmica, é o profissional mais preparado para prestar serviços de qualidade nesses sistemas. Quanto mais a tecnologia é disponibilizada para o usuário, maior é o compromisso dos sistemas de informação, como os bancos de dados e bibliotecas, em capacitar e orientar seus usuários, no sentido de prover o acesso aos conteúdos disponíveis.

Hoje, uma unidade de informação deve planejar e desenvolver os seus serviços com os usuários, tanto para o fornecimento útil à pesquisa quanto para a formação do usuário cidadão.

Para Boccato (2006, p.268)

"A integração entre o profissional da informação e o usuário no desenvolvimento participativo de suas atividades vai ao encontro de seus objetivos em atender às necessidades de informação solicitadas por seus usuários/pesquisadores com competência, qualidade, rapidez, visando à satisfação e à confiabilidade desses usuários nos seus serviços prestados".

O usuário desempenha um papel fundamental junto ao sistema de informação, sendo "ele quem utiliza a informação processada em seu nível descritivo e temático e, por conseguinte, quem também avaliará se o retorno obtido está compatível ou não com as suas necessidades iniciais de pesquisas". (BOCCATO, 2006, p.269)

O sistema de informação é totalmente dependente das necessidades de informação de cada usuário. Portanto, é necessário que o usuário seja capaz de identificar as etapas de busca em bases de dados, de modo que ele consiga focalizar o seu problema de pesquisa, para elaboração de uma busca com sucesso.

Os usuários diminuíram o contato com os bibliotecários devido ao desenvolvimento de sofisticadas ferramentas que oferecem diversas formas de realizar as buscas, então, com isso, mudou também o papel do bibliotecário, que passou a ser o educador, para acompanhar a necessidade do novo usuário e capacitar os usuários a se tomarem autônomos para fazer suas buscas nos sistemas de informação automatizados de forma eficiente e, sobretudo eficaz.

Programas educativos conforme Cuenca (1999, p.293), servem:

"Para que os usuários conheçam os sistemas automatizados de recuperação da informação, sejam capazes de elaborar uma estratégia de busca simples e saibam utilizar a nova biblioteca eletrônica e o recurso informacional de que dispõem, é necessário que as bibliotecas ofereçam treinamentos e cursos específicos, com modalidades de programas educativos".

Dessa forma, os usuários se tornariam aptos a usar os recursos da internet, banco de dados, bibliotecas digitais, e identificar estratégias eficientes para sanar necessidades de informação, bem como se tornarem e capacitados para o uso adequado dos recursos disponíveis.

Muitas vezes a razão da não utilização das bases de dados, de acordo com Cuenca (1999, 293), é devido ao desconhecimento, por parte do usuário, de sua existência, dificuldade com as várias interfaces de busca para o acesso as bases, e a não familiarização com o vocabulário especializado da área. Além disso, verifica-se grande dificuldade na escolha da ferramenta adequada a seu tema de pesquisa.

A atuação do profissional deve estar voltada para o estudo dos fluxos de informação, visando à promoção de seu uso, para explorar as interfaces de informação gerais ou especializadas, que são importantes para os seus usuários e oferecer treinamentos. "Através desse treinamento, o usuário será capaz de avaliar o sistema de informação, quanto a sua exatidão, atualidade, abrangência, formato disponível e saber manejar a pluralidade das necessidades de informação em ambientes virtuais". (BUFREM, 2004, p.176)

O ideal é que o usuário se torne cada vez mais independente na utilização dos sistemas de informação, provendo um efetivo e eficiente uso dos dados, que estão disponíveis nos sistemas automatizados de recuperação da informação, sejam capazes de elaborar uma estratégia de busca condizente com as suas necessidades e saibam utilizar os novos recursos informacionais que estão surgindo constantemente em todas as áreas do conhecimento.

O bibliotecário é o intermediário entre a base de dados e o usuário, dispõe de maior habilidade para utilizar as linguagens controladas na obtenção do melhor desempenho durante o processo de planejamento e operacionalização da

estratégia de busca. Deve ter um conhecimento continuamente aprofundado sobre as diferentes fontes de linguagens controladas e suas estruturas hierarquizadas ou não, para que os termos da estratégia de busca sejam os termos utilizados pelos indexadores no processo de entrada dos documentos na base.

“A responsabilidade desse profissional vai muito além do simples desenvolvimento e controle de mecanismos de recuperação da informação. Todas as suas atividades possuem um cunho de responsabilidade social, pois, nunca perderão de vista a totalidade do ser humano, que almeja uma informação rápida, confiável, precisa e com o menor custo e tempo”. (SOUTO, 2003, p.80)

O bibliotecário precisa avaliar as necessidades dos usuários e se auto-avaliar, práticas que devem estar sempre presentes na atuação de um bibliotecário envolvido com a superação e a otimização de seus serviços para o crescimento evolutivo na sua área. É através desta prática pensante e criadora que, inclusive, a disseminação do uso dos sistemas de informação automatizado se consolida.

8 CONSIDERAÇÕES FINAIS

Existe uma tendência ao uso de algum tipo de controle de vocabulário, nos sistemas de recuperação da informação, e a criação de instrumentos de indexação, adequados às necessidades dos sistemas, sendo esses de fundamental importância para a comunicação da informação. Quando se faz necessária a criação de um instrumento de indexação, deve-se permitir que essa criação seja realizada entre o usuário do sistema e o próprio sistema de informação visto que, esse instrumento, para ser criado, necessita de uma estrutura terminológica conceitual que será buscada em um sistema terminológico, sendo importantíssima a participação do usuário especialista, em que se requer um domínio da terminologia da área em estudo.

A escolha por um determinado vocabulário controlado afeta o desempenho do sistema de recuperação de informação tanto na indexação quanto na estratégia de busca. Deve-se optar pela linguagem que melhor irá representar terminologicamente os documentos, e as necessidades de informação da comunidade usuária, possibilitando a recuperação eficaz e eficiente.

A linguagem de indexação tem sido usada não só exclusivamente para universo bibliográfico, como também nos *sites* na web, que aumentaram significativamente a demanda por instrumentos de organização e tratamento da informação. A transferência da informação se dá no universo da linguagem de indexação ou natural.

A linguagem de indexação deve estar em constante avaliação de forma a atender às necessidades específicas de um determinado sistema, é isso que permite a sua atualização, é importante a adoção de mecanismo de adaptabilidade e flexibilidade que não os tornem obsoletos com o passar do tempo. O ideal é que a

linguagem de indexação utilizada seja um sistema aberto e que permita a inclusão de novos termos, para não colocar em risco a confiabilidade do sistema de informação. A linguagem que não contempla os novos conceitos torna-se desatualizada, conseqüentemente, não atinge os seus objetivos.

O desenvolvimento de linguagens controladas para indexação e recuperação da informação requer um trabalho em equipe, investimentos e tempo para gerar, testar e sedimentar um instrumento de trabalho que deve ser continuamente atualizado, por se tratar de um conjunto de termos que vão compor as estratégias de busca em bases de dados.

Contudo, o bom desempenho de uma linguagem de indexação na recuperação da informação é um fator essencial na obtenção de informações satisfatórias para o usuário, dessa forma, possibilitando a realização de investigações que contribuam para o desenvolvimento da ciência, almejando o bem estar da sociedade.

O avanço científico e técnico se fundamenta na divulgação, leitura, assimilação e utilização de pesquisas anteriores, sendo preciso que as informações sejam divulgadas e acessíveis aos cientistas. Essa visibilidade só é possível com o processo da indexação nas bases de dados.

Dentro do contexto digital, o tratamento temático da informação permanece como uma atividade crucial, pois cumpre função fundamental no trabalho de facilitar o acesso à informação. É interessante que novos processos e instrumentos venham a ser desenvolvidos no sentido de absorver as novas tecnologias. A linguagem documentária configura-se como instrumento facilitador da comunicação em contextos documentários específicos, como é o caso dos sistemas automatizados.

Pode-se verificar que o fator de maior importância na organização da informação é a recuperação por parte do usuário. É importante analisar o que os usuários procuram? Como procuram? Qual a terminologia utilizada por eles? Essa aproximação com o usuário se faz necessária para o conhecimento de suas reais necessidades. O bibliotecário deve conhecer a linguagem das pessoas com o qual o sistema se interage, para melhor caracterizá-lo, e também torná-lo auto-suficiente na busca da informação, por processos automatizados.

Esse comportamento aponta para o surgimento de um novo usuário e de um novo bibliotecário. Usuário mais exigente quanto à qualidade dos serviços prestados pela biblioteca, exigindo do bibliotecário novas responsabilidades que o qualifiquem cada vez mais, passando a ser um educador no sentido de orientar para o uso das tecnologias, e no envolvimento no planejamento de *softwares* interativos de acesso à informação.

Cabe ao bibliotecário da informação facilitar e agilizar o uso da informação armazenada no contexto digital, auxiliando o usuário, através de treinamento na elaboração da melhor estratégia de busca, ajudando-o a definir o seu perfil de busca, de forma a refletir seu interesse corrente, tornando o usuário mais independente para a realização de suas próprias buscas, implementando o uso efetivo de sistemas de recuperação da informação.

Com relação aos objetivos propostos neste estudo, pode-se chegar às seguintes conclusões: o objetivo geral foi mostrar a importância da utilização do vocabulário controlado na indexação de assuntos em sistemas automatizados que visam agilizar e ampliar a recuperação da informação em publicações técnico-científicas. Pode-se afirmar, com base na literatura estudada, que existe um consenso entre os autores, no que se refere à necessidade de controle de vocabulário, na representação de conteúdos de forma a viabilizar um fluxo informativo efetivo em sistemas de informação automatizados.

Em relação aos objetivos específicos:

- Verificar a necessidade do controle de vocabulário, e a função de sua aplicação nos sistemas que tratam e organizam a informação para ser recuperada.

Com base nos textos lidos, pôde-se verificar que existe uma necessidade do controle de vocabulário, nos sistemas de informação, que precisam viabilizar os pontos de acesso à informação. O processo de organização persegue o uso de um controle de vocabulário de forma a permitir a comunicação numa comunidade específica ou entre comunidades para o compartilhamento do conhecimento técnico-científico.

- Mostrar os tipos de linguagens como a natural e artificial, utilizadas como parte integrante da atividade de controle da terminologia no processo de indexação.

Existe uma concordância entre os autores estudados, com relação ao uso da linguagem na representação do conhecimento, como um controle terminológico necessário para a transferência do conhecimento. A utilização de linguagens de indexação é uma forma de melhor representar e recuperar a informação, e esta deve estar de acordo com as necessidades dos usuários. Existe uma tendência, entre alguns autores, em não se utilizar simplesmente a linguagem indexação, e sim a LN e a LI ao mesmo tempo, para se obter uma melhor recuperação da informação. A linguagem excessivamente artificial não é capaz de acompanhar o desenvolvimento do conhecimento científico.

- Analisar os instrumentos utilizados para controle de vocabulário em sistemas de informação, que visam padronizar a indexação e a recuperação, tendo em vista os objetivos do sistema para o qual ela será criada.

Pela literatura lida, percebeu-se que todos os instrumentos (verbais e simbólicos), são necessários, no processo de organização e transferência da informação, e devem ser construídos de forma a atender os objetivos da instituição para a qual foram criados. Existe uma preferência por um determinado instrumento de indexação em relação ao outro, apesar de todos os instrumentos deverem ser utilizados. Nos sistemas automatizados, o tesauro é o mais adotado, por trabalhar com termos, num sistema pós-coordenado, e ser desenvolvido para uma área específica do conhecimento.

BOCCATO, Vera Regina.

- Mostrar a influência do controle de vocabulário na recuperação eficiente da produção científica.

2, p.267-271, 1984, pp. 271.

Pode-se constatar nos textos que o uso de vocabulário controlado interfere na recuperação da informação, proporcionando diversificados pontos de acesso. Além disso, permite uma busca mais refinada ao tema pesquisado, facilitando a comunicação em contextos documentários específicos.

Conclui-se que o tema estudado ainda deve ser muito discutido entre especialistas da área de Ciência da Informação. Deve-se incentivar a realização de pesquisas que possam trazer soluções para os problemas que ainda persistem, apesar das facilidades advindas do surgimento dos sistemas de informação automatizados.

CAMPOS, Maria Lúcia

Construção de Tesauro

História, v. 11, n. 1, p.

1984, pp. 1984.

COMIRA JR., Carlos

Biologias Inform.

4, p.283-285, 1984.

REFERÊNCIAS

AMARAL, Maria Célia. Lista de assunto FFLCH/USP: relato de experiência da elaboração. In: SEMINÁRIO NACIONAL DE BIBLIOTECAS UNIVERSITÁRIAS, 12., 2002, Recife. **Trabalhos técnicos...** Disponível em: <www.ufpe.br/snbu/docs/90.a.pdf

BOCCATO, Vera Regina Casari; FUJITA, Mariângela Spotti Lopes. Estudos de avaliação quantitativa e qualitativa de linguagens documentárias: uma síntese bibliográfica. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 11, n. 2, p.267-281, maio/ago. 2006.

BUFREM, Leilah S.; PEREIRA, Edmeire C. Os profissionais da informação e a gestão de competências. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 9, n. 2, p.170-181, jul./dez. 2004.

CAMPOS, Maria Luiza de Almeida. Linguagens documentárias: núcleo básico de conhecimento para seu estudo. **Revista da Escola de Biblioteconomia da UFMG**, Belo Horizonte, v. 24, n. 1, p. 52-62, jan./jun. 1995.

CAMPOS, Maria Luiza de Almeida et al. Estudo comparativo de softwares de construção de tesouros. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 11, n. 1, p.68-81, jan./abr. 2006.

COIMBRA JR., Carlos E. A. Produção científica em saúde pública e as bases bibliográficas internacionais. **Caderno de Saúde Pública**, Rio de Janeiro, v.15, n. 4, p.883-888, out./dez. 1999,

CARNEIRO, Marília Vidigal. Diretrizes para uma política de indexação. **Revista da Escola de Biblioteconomia da UFMG**, Belo Horizonte, v. 14, n. 2, p. 221-241, set. 1985.

CESARINO, Maria Augusta da Nóbrega. Sistemas de recuperação da informação. **Revista da Escola de Biblioteconomia da UFMG**, Belo Horizonte, v. 14, n. 2, p. 157-168, set. 1985.

CESARINO, Maria Augusta da Nóbrega; PINTO, Maria Cristina Mello Ferreira. **Revista da Escola de Biblioteconomia da UFMG**, Belo Horizonte, v. 7, n. 2, p. 268-288, set. 1978.

CHAUMIER, Jacques. Indexação: conceito, etapas e instrumentos. **Revista Brasileira de Biblioteconomia e Documentação**, São Paulo, v. 21, n. ½, p. 63-79, jan./jun. 1988.

CUENCA, Angela Maria Belloni. O usuário final da busca: avaliação da capacitação no acesso a bases de dados em biblioteca acadêmica. **Ciência da Informação**, Brasília, v. 28, n. 3, p. 293-301, set./dez. 1999.

DIAS, Cláudia. **Usabilidade na web**: criando portais mais acessíveis. Rio de Janeiro: Alta Books, 2003. Cap.2.

DIAS, Eduardo Wense. Contexto digital e tratamento da informação. **Data Gramma Zero** – Revista de Ciência da informação, v. 2, n. 5, out. 2001.

FERREIRA, Aurélio Buarque de Holanda. **Novo dicionário da língua portuguesa**. 2. ed., rev. e aum. Rio de Janeiro: 1986. 1838 p.

FERREIRA, Sueli Mara Soares Pinto. Novos paradigmas e novos usuários de informação. **Ciência da Informação**, Brasília, v. 25, n. 2, p. 217-223, maio/ago. 1996.

FOSKETT, Anthony Charles. **A abordagem temática da informação**. São Paulo, Polígono, 1972. 427 p.

FUJITA, Mariângela Spotti Lopes. A leitura documentária na perspectiva de suas variáveis: leitor – texto - contexto. **Data Grama Zero – Revista de Ciência da informação**, v. 5, n. 4, p. 1-26, ago. 2004.

FUJITA, Mariângela Spotti Lopes. Precis: perspectiva histórica e técnica do seu desenvolvimento e aplicação. **Revista Brasileira de Biblioteconomia e Documentação**, São Paulo, v. 21, n. ½, p. 21-45, jan./jun. 1988.

GIGANTE, Maristela Cid. Os sistemas de classificação bibliográfica como interface bibliotecário/usuário. **Ciência da Informação**, Brasília, v. 25, n. 2, p. 193-196, maio/ago. 1995.

IHADJADENE, Madjid, BOUCHE, Richard, KURAMOTO, Hélio. Navegação nos vocabulários controlados. **Revista de Biblioteconomia de Brasília**, Brasília, v. 22, n. 2, p. 197-210, jul./dez. 1998.

KREMER, Jeannette M. Estratégia de busca. **Revista da Escola de Biblioteconomia da UFMG**, Belo Horizonte, v. 14, n. 2, p. 187-220, set. 1985.

LANCASTER, F. W. **Indexação e resumos**: teoria e prática. Brasília: Briquet de Lemos, 1993.

LARA, Marilda Lopes Ginez de. Diferenças conceituais sobre termos e definições e implicações na organização da linguagem documentária. **Ciência da Informação**, Brasília, v. 33, n. 2, p. 91-96, maio/ago. 2004.

LARA, Marilda Lopes Ginez de. Linguagem documentária e terminologia. **Transinformação**, Campinas, v. 16, n. 3, p. 231-240, set./dez. 2004.

LARA, Marilda Lopes Ginez de. O processo de construção da informação documentária e o processo de conhecimento. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 7, n. 2, p.127-139, jul./dez. 2002.

LOPES, Ilza Leite. Estratégia de busca na recuperação da informação: revisão da literatura. **Ciência da Informação**, Brasília, v. 31, n. 2, p. 60-71, maio /ago. 2002.

LOPES, Ilza Leite. Uso das linguagens controlada e natural em bases de dados: revisão da literatura. **Ciência da Informação**, Brasília, v. 31, n. 1, p. 41-52, jan./abr. 2002.

LUCAS, Clarinda Rodrigues. A metalinguagem como lugar da interpretação: terminologia e bases de dados informatizadas. **Delta**, São Paulo, v. 15, n. 1, fev./jul. 1999.

MARCONDES, Carlos Henrique; SAYÃO, Luís Fernandes. Integração e interoperabilidade no acesso a recursos informacionais eletrônicos em C&T: a proposta da Biblioteca Digital Brasileira. **Ciência da Informação**, Brasília, v. 30, n. 3, p. 24-33, set./dez. 2001.

MOURA, Maria Aparecida et al. Linguagem de indexação em contextos cinematográficos: a experiência de elaboração do tesauro eletrônico do cinema

brasileiro. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 10, n. 1, p. 54-69, jan./jun. 2005.

NAVES, Madalena Martins Lopes. Estudo de fatores interferentes no processo de análise de assunto. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 6, n. 2, p.189-204, jul./dez. 2001.

NAVARRO, Sandrelei. Interface entre lingüística e indexação: revisão de literatura. **Revista Brasileira de Biblioteconomia e Documentação**, São Paulo, v. 21, n. ½, p. 46-62, jan./jun. 1988.

NOVELLINO, Maria Salet Ferreira. A linguagem como meio de representação ou de comunicação da informação. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 3, n. 2, p.137-146, jul./dez. 1998.

NOVELLINO, Maria Salet Ferreira. Instrumentos e metodologias de representação da informação. **Informação & Informação**, Londrina, v.1, n. 2, p.37-45, jul./dez. 1996.

OLIVEIRA, Érica Beatriz. Produção científica nacional na área de geociências: análise de critérios de editoração, difusão e indexação em bases de dados. **Ciência da Informação**, Brasília, v. 34, n. 2, p. 34-42, maio/ago. 2005.

PINTO, Maria Cristina Mello Ferreira. Análise e representação de assuntos em sistemas de recuperação da informação: linguagens de indexação. **Revista da Escola de Biblioteconomia da UFMG**, Belo Horizonte, v. 14, n. 2, p. 169-86, set. 1985.

PINTO, Virgínia Bentes. Indexação documentária: uma forma de representação do conhecimento registrado. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 6, n. 2, p.223-234, jul./dez.2001.

RIVIER, Alexis. Construção de linguagens de indexação: aspectos teóricos. **Revista da Escola de Biblioteconomia da UFMG**, Belo Horizonte, v. 21, n. 1, p. 56-99, jan./jun. 1992.

SAYÃO, Luís Fernando. Bases de dados: a metáfora da memória científica. **Ciência da Informação**, Brasília, v. 25, n. 3, p. 314-318, set./dez. 1996.

SOUTO, Leonardo Fernandes. Recuperação de informações em bases de dados: usos de tesouro. **Transinformação**, Campinas, v. 15, n. 1, p. 73-81, jan./abr. 2003.

SOUZA, Renato Rocha. Sistema de recuperação de informações e mecanismos de busca na web: panorama atual e tendências. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 11, n. 2, p.161-173, maio/ago. 2006.

STREHL, Letícia. Avaliação da consistência da indexação realizada em uma biblioteca universitária de artes. **Ciência da Informação**, Brasília, v. 27, n. 3, p. 329-335, set./dez. 1998.

TRISTÃO, Ana Maria Delazari. Sistemas de classificação facetada e tesouros: instrumentos para organização do conhecimento. **Ciência da Informação**, Brasília, v. 33, n. 2, p. 161-171, maio/ago. 2004.

UNISIST. Princípios de indexação. **Revista da Escola de Biblioteconomia da UFMG**, Belo Horizonte, v. 10, n. 1, p. 83-94, mar. 1981.

WANDERLEY, Manoel Adolpho. **Linguagem documentária, acesso ao documento**: aspectos do problema. 1973. 79 f. Dissertação (Mestrado em Biblioteconomia e Documentação) - Escola de Biblioteconomia e Documentação da FEFIEG, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 1973.