

ELZA MARIA DE FARIA

LEI DO DIREITO AUTORAL

Todos os direitos reservados e protegidos
pela Lei 9.610/1998

Este trabalho não pode ser reproduzido ou
transmitido sejam quais forem os meios
empregados: eletrônicos, mecânicos,
fotográficos ou quaisquer outros.

**ANÁLISE DE ASSUNTO NOS CONTEXTOS
TRADICIONAL E DIGITAL**

Belo Horizonte

Escola de Ciência da Informação da UFMG

2006

N.Cham. M025 F224a 2006

Autor: Faria, Elza Maria de.

Título: Análise de assunto nos contexto



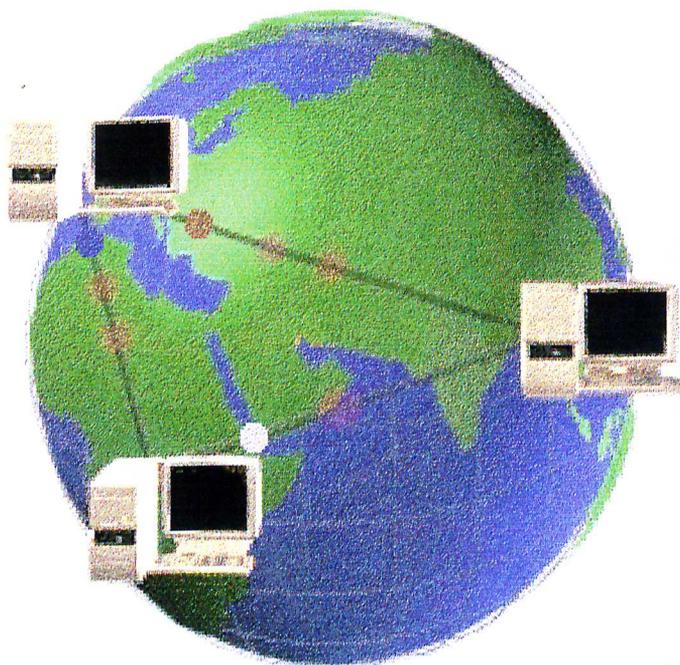
348430806

Ac. 469491

ELZA MARIA DE FARIA

LEIDO DIREITO AUTORAL
Todos os direitos reservados e protegidos
pela Lei 9.610/1998.
Este arquivo não pode ser reproduzido ou
transmitido sejam quais forem os meios
empregados: eletrônicos, mecânicos,
fotográficos ou quaisquer outros.

**ANÁLISE DE ASSUNTO NOS CONTEXTOS
TRADICIONAL E DIGITAL**



*lido
madalena m. farias
18/12/06
Aprovado 95/100*

Belo Horizonte

Escola de Ciência da Informação da UFMG

2006

Udc. 469991

ELZA MARIA DE FARIA

LEI DO DIREITO AUTORAL
Todos os direitos reservados e protegidos
pela Lei 9.610/1998.
Este arquivo não pode ser reproduzido ou
transmitido sejam quais forem os meios
empregados: eletrônicos, mecânicos,
fotográficos ou quaisquer outros.

ANÁLISE DE ASSUNTO NOS CONTEXTOS TRADICIONAL E DIGITAL

**Monografia apresentada ao Curso de
Especialização em Organização da
Informação em Contexto Digitais, da
Escola de Ciência da Informação da
Universidade Federal de Minas Gerais,
como requisito parcial à obtenção do
título de especialista em OICD.**

Orientadora: Madalena Martins L. Naves

Belo Horizonte

Escola de Ciência da Informação da UFMG

2006

U.F.M.G. - BIBLIOTECA UNIVERSITÁRIA



348430806

NÃO DANIFIQUE ESTA ETIQUETA

BIBLIOTECA UNIVERSITARIA

24 / 09 / 08

3484308-06

LEI DO DIREITO AUTORAL

Todos os direitos reservados e protegidos
pela Lei 9.610/1998.

Este arquivo não pode ser reproduzido ou
transmitido sejam quais forem os meios
empregados: eletrônicos, mecânicos,
fotográficos ou quaisquer outros.

043
F224i

Faria, Elza Maria de
Indexação de assunto nos contextos tradicional e digital [manuscrito] / Elza
Maria de Faria. – 2006.
73 f., enc.

Orientadora: Madalena Martins L. Naves.

Monografia apresentada ao Curso de Especialização em Organização
da Informação em Contextos Digitais, da Escola de Ciência da
Informação da Universidade Federal de Minas Gerais, como requisito
parcial à obtenção do título de especialista em OICD.

1. Análise de assunto. 2. Indexação. 3. Biblioteca tradicional.
Biblioteca digital. I. Naves, Madalena Martins L. II. Universidade
Federal de Minas Gerais. Escola de Ciência da Informação. III. Título.

CDU: 025.4

AGRADECIMENTOS

LEI DO DIREITO AUTORAL
Todos os direitos reservados e protegidos
pela Lei 9.610/1998.
Este arquivo não pode ser reproduzido ou
transmitido sejam quais forem os meios
empregados: eletrônicos, mecânicos,
fotográficos ou quaisquer outros.

À Nossa Senhora de Fátima, pela presença em minha vida.

À Professora Madalena Martins Lopes Naves, pela excelente orientação e apoio recibo.

À Professora Gercina Ângela Borém de O. Lima, como coordenadora do curso.

Aos colegas da Biblioteca do ICB, pela colaboração e incentivo.

À minha família, pelo incentivo, amor e carinho, em tempo integral.

Aos meus dois denguinhos “maiores presentes” de minha vida, Marianna e Letícia, pelo amor e carinho “incondicional”.

RESUMO

No domínio da Ciência da Informação, o processo de indexação de assunto é importante num Sistema de Recuperação da Informação, que envolve duas etapas: a análise de assunto e a tradução para linguagens de indexação. A análise de assunto, ou análise conceitual, é a primeira etapa do processo de indexação, a qual é dividida em três fases: leitura do texto; extração de conceitos e determinação da afinidade. Vários fatores interferem neste processo, tais como: fatores ligados ao indexador, como experiência, grau de compreensão, conhecimento do assunto; fatores ligados ao vocabulário, como a especificidade, sintaxe, qualidade do vocabulário, fatores ligados ao documento, como complexidade, língua; fatores ligados ao processo, como o tipo de indexação, produtividade, exaustividade; fatores ambientais, ruído, e iluminação. O objetivo geral é identificar as semelhanças e diferenças nos processos de tratamento da informação, quanto à análise de assunto e tradução dos conceitos, para linguagens de indexação, nos contextos tradicional e digital. A análise de assunto, no contexto tradicional, é aquela etapa que o indexador extrai os conceitos dos documentos, no sentido de determinar os seus assuntos, tendo como recurso a cognição, a metacognição, a língua e sua estrutura, e em seguida, usam-se instrumentos convencionais para a tradução dos conceitos para uma linguagem de indexação. Já a análise de assunto, no contexto digital, é parecida com a tradicional, utilizam-se as mesmas técnicas e instrumentos. Com o avanço das tecnologias, surgem novas ferramentas e padrões, como os metadados, as ontologias, para representar melhor os conceitos extraídos dos documentos. Conclui-se que a grande maioria das bibliotecas digitais usa os mesmos processos e instrumentos de tratamento da informação que as bibliotecas tradicionais. Percebe-se que estes dois contextos possuem mais semelhanças que diferenças, e que os bibliotecários ou indexadores estão no caminho certo e têm muito campo pela frente, com o advento da WEB e da Internet.

PALAVRAS-CHAVE

Análise de assunto - Análise conceitual - Análise documentária - Biblioteca tradicional - Biblioteca digital - Contexto digital - Contexto tradicional - Indexação - Tratamento da Informação.

ABSTRACT

In the domain of Information Science, subject indexing process plays an important role in the Information Retrieval System which involves two phases: subject analysis and translation into indexing languages. Subject analysis, or conceptual analysis, is the first phase of indexing process, which is divided in three stages: text reading; concept extracting and determining aboutness. Several factors interfere in this process such as: indexer-related factors (experience, understanding degree, knowledge of the subject); vocabulary-related factors (specificity, syntax, quality of vocabulary); document-related factors (complexity, language); process-related factors (type of indexing, productivity, exhaustivity); environmental factors, noise and illumination. The general purpose is to identify similarities and differences in information management processes, as for subject analysis and concept translation into indexing languages, in traditional and digital contexts. Subject analysis in the traditional context is that stage when indexer extracts concepts from documents in order to determine their subjects, with the use of cognition, metacognition, language and its structure and then making use of conventional tools to translate concepts into an indexing language. As for subject analysis, in the digital context, it is similar to the traditional one, using the same techniques and tools. With the advance of technologies, new tools and standards arise, such as metadata, ontologies, to better represent concepts extracted from documents. It is concluded that most digital libraries use the same processes and tools to manage information as traditional libraries. These two contexts are observed to have more similarities than differences, and that librarians and indexers are on the right way and have a lot ahead with the arrival of the WEB and the Internet.

Keywords

Subject analysis – concept analysis - document analysis – traditional library – digital library – digital context – traditional context - Indexing – information management.

SUMÁRIO

1 INTRODUÇÃO	7
2 AS BIBLIOTECAS	10
2.1 Bibliotecas tradicionais.....	12
2.2 Bibliotecas digitais.....	13
3 PROCESSO DE INDEXAÇÃO	17
3.1 Política de indexação.....	22
4 ANÁLISE DE ASSUNTO	25
4.1 Análise de assunto no contexto tradicional	27
4.2 Análise de assunto no contexto digital	37
5 TRADUÇÃO DE CONCEITO: PARA LINGUAGENS	42
DE INDEXAÇÃO	42
5.1 Tradução de conceitos: no contexto tradicional.....	45
5.2 Tradução de conceitos: no contexto digital.....	49
6 CONSIDERAÇÕES FINAIS.....	63
REFERÊNCIAS.....	67

LEI DO DIREITO AUTORAL
Todos os direitos reservados e protegidos
pela Lei 9.610/1998.
Este arquivo não pode ser reproduzido ou
transmitido sem que sejam empregados meios
mecânicos, mecânicos, mecânicos,
fotográficos ou quaisquer outros.

1 INTRODUÇÃO

A escolha do tema deste trabalho se justifica pela necessidade de conhecer e averiguar como acontece o processo de indexação no contexto digital, ou seja, saber se esse processo é parecido com o processo de indexação das bibliotecas tradicionais.

Assim sendo, o objeto deste estudo é a indexação e a análise de assunto nos contextos tradicional e digital. Acredita-se que o processo de indexação possua semelhanças e diferenças, entre esses dois contextos.

Diante disso, **os pressupostos** que direcionam o trabalho são:

- 1) A indexação e a análise de assunto, no contexto digital, podem ter semelhanças e diferenças daquelas do contexto tradicional.
- 2) Os fatores que determinam semelhanças ou diferenças são a política vigente, os objetivos e a tipologia da instituição.

O objetivo geral desta monografia é: identificar as semelhanças e diferenças nos processos de tratamento da informação, no que diz respeito à indexação e à análise de assunto, nos contextos tradicional e digital.

E como objetivos específicos, pretende-se: averiguar e conhecer como é feita a análise de assunto nos contextos tradicional e digital; identificar as etapas do processo de indexação nos contextos tradicional e digital, demonstrando semelhanças e diferenças; identificar a importância da política de indexação nos dois contextos.

Diante da grande massa de informação e dos novos avanços tecnológicos existentes no mundo, faz-se necessário estudar as técnicas de organizar e tratar

as informações, no sentido de verificar se “o fazer” do indexador, do contexto tradicional, está prestes a ser extinto, ou se no domínio digital o indexador terá espaço de atuação.

Nota-se que o paradigma da biblioteca tradicional é o de necessitar de um espaço físico, e de material bibliográfico concreto presente no local, para ser disseminado. Já o paradigma da biblioteca digital é o contrário, não precisar de uma localização física e nem de materiais propriamente ditos para armazenar, somente um conjunto de meios eletrônicos para facilitar e localizar as necessidades de informação dos usuários, indiferentes da presença física e da localização dos materiais. De modo que, a falta da presença física, e da localidade, no meio digital é explicada pela rapidez da recuperação da informação desejada, ou seja, a recuperação é em tempo real. Percebe-se então que, a informação não se fixa em um em um único lugar, e que esse novo formato digital é o desvinculo da forma simbólica e tradicional, da organização do conhecimento.

Fala-se muito da organização do conhecimento no meio digital, da virtualidade das linguagens e do meio, do espaço de leitura, atualizado como espaço de escrita, e vice-versa, ou leitura em outras leituras, com uso da força virtual, ou seja, com potência de atualização e multiplicação das oportunidades do mundo real. De modo que a virtualidade é o texto em movimento, vetorizado, mais próximo do movimento do pensamento, ou da imagem que se tem dele. Isto é explicado pelo surgimento de uma estrutura de blocos de textos não-lineares e não seqüenciais conectados por *links* eletrônicos, que oferecem múltiplos caminhos de leitura ao usuário, com linguagem e meio virtual, sendo diferente de texto comum, linear, impresso e estático.

Fica evidente, também, o surgimento de novidades quanto a: instrumentos, padrões, linguagens, tecnologias, suportes e tipologia de informações na área de Ciência da Informação. Os metadados e as ontologias são padrões internacionais

que propiciam o intercâmbio de dados. A instituição que tiver um certo grau de conhecimento a respeito dessas novas tecnologias da informação, terá meios de proporcionar o diferencial em organização do conhecimento, usando e participando de redes, como um facilitador para os serviços técnicos.

Percebe-se, neste novo cenário, que a grande oferta de informação encontrada e disponibilizada no meio eletrônico traz muita dificuldade para o usuário final recuperar exatamente aquilo que lhe é relevante, devido à falta de critérios e padrões na organização da informação na WEB¹. Assim, cabe à Ciência da Informação encontrar, urgentemente, mecanismos de organização e recuperação da informação para a WEB, antes que outros profissionais o façam.

Estas considerações estão abordadas ao longo do trabalho, e para que se possa compreendê-las melhor, é necessário e imprescindível contextualizar as bibliotecas. O segundo capítulo trata de um breve histórico das bibliotecas, e do conceito de bibliotecas tradicionais e digitais.

O terceiro capítulo trata de pequenas considerações históricas, terminológicas e conceituais, sobre o processo de indexação, como também a respeito da política de indexação. O quarto capítulo aborda os conceitos do processo da análise de assunto, nos contextos tradicional e digital. O quinto capítulo trata dos conceitos do processo de tradução dos termos selecionados para uma linguagem de indexação, e especifica também como acontece a tradução nos contextos tradicional e digital.

E o sexto e último capítulo trata das considerações finais após as leituras feitas ao longo do trabalho, e mostra se a monografia alcançou, ou não, os objetivos propostos.

¹ WEB é o ambiente de World Wide Web, e abreviado como www. que emprega a Internet como meio de transmissão.

2 AS BIBLIOTECAS

A história das bibliotecas é marcada por fatos históricos e de resistência do conhecimento. A biblioteca da antiguidade sofre e supera ações de guerras e censuras da igreja católica. Na Idade Média, acontecem até tentativas de destruição das bibliotecas por parte da igreja católica, que mostra a dominação de classes e poder. Porém, hoje, com a democratização da informação, percebe-se que essa conotação está deixando de existir. A informação não é mais disseminada, e sim, disponibilizada. Pode ter acesso à informação quem precisar e quiser, tanto nas bibliotecas, como na WEB.

As primeiras bibliotecas de que se tem notícia são aquelas chamadas de “minerais”, devido aos tabletas de argila, e a biblioteca do rei Assurbanipal, do século VII a.C. a mais importante. Depois, aparecem as bibliotecas vegetais e animais, cujos materiais são rolos de papiros e pergaminhos, tal como a famosa Biblioteca de Alexandria, no Egito, com 700 mil volumes, que foi destruída por motivos religiosos. Mais tarde, com o aparecimento do papel, as bibliotecas com acervos formados de papel vão aparecendo e, logo depois as de livro. A partir do século XVI, evidencia-se uma transformação com a democratização da informação, e assim, aparece um grande número de bibliotecas (Enciclopédia Wikipédia).

De acordo com Morigi (2005) e Martins (2006), a história da biblioteca é marcada por transformações e fatos sociais. A biblioteca do passado tem caráter bem restrito quanto aos seus usuários. O administrador e usuário deste tipo de biblioteca, quase sempre membros da igreja, procuravam garantir *status* e poder com o uso da biblioteca. Mais tarde, a biblioteca perde esse caráter e amplia a categoria e o número de usuários. Assim, a biblioteca da antiguidade tinha o objetivo de agregar *status* e poder aos imperadores, e a biblioteca da Idade Média, apresentava-se como sinônimo de *status* social e poder econômico da alta

sociedade. Mas, com o aparecimento da imprensa, a biblioteca começa a crescer cada vez mais, e perder um pouco desta conotação de *status* e poder. No final do século XX, vários tipos de bibliotecas aparecem, as bibliotecas eletrônicas, as virtuais e as digitais.

A biblioteca ideal para o século XXI, afirma Martins (2006), sendo utopia ou não, deve conter materiais e produtos tradicionais, eletrônicos, digitais, e virtuais, cujo nome deverá ser BIBLIOTECA UNIVERSAL, ou seja, este modelo deve englobar todo tipo de biblioteca, com a união de acervos reais e virtuais, com objetivo de disseminar acervo impresso, virtual, visual, audiovisual, oral, e multimídia. A informação encontrada neste tipo de biblioteca pode estar alocada tanto num espaço físico, quanto na rede, com documentos reais ou virtuais.

Nota-se na história que as bibliotecas passam por várias fases e denominações, tais como: biblioteca fechada (acorrentada a determinadas instituições de poder, para restringir o uso e divulgação da informação), biblioteca tradicional, eletrônica, virtual, interativa, digital, biblioteca sem paredes, e por último, biblioteca universal. Percebe-se que, tradicionalmente, as bibliotecas desempenhavam papel de dominadores da informação, e, hoje, desempenham papel de disseminadores da informação, e acredita-se que isso deve perpetuar por muito tempo, mesmo com a desmaterialização das obras no meio digital.

Após este breve histórico sobre as bibliotecas, há necessidade de se relacionar alguns conceitos de autores de épocas diferentes, sobre bibliotecas tradicionais e digitais, assunto tratado no próximo sub item desta monografia, cujo interesse é perceber e analisar melhor os processos de indexação nestes dois contextos.

2.1 Bibliotecas tradicionais

De acordo com a Enciclopédia Wikipedia, a palavra Biblioteca é de origem grega “Biblioteke”, que significa ambiente de livros, *biblion* (livro) *teke* (caixa, depósito). Nota-se, que até os dias atuais, essa idéia de guarda, de ambiente de livro ainda permanece.

Segundo Aurélio (1999), biblioteca é o lugar onde se guardam os livros, estante ou coleção pública ou privada de livros e documentos organizados por estudo, leitura ou consulta.

De acordo com Alvarenga (2001), por biblioteca tradicional entende-se que seja aquela que tem, por objetivo, adquirir, tratar, armazenar e divulgar documentos, cujas origens sejam a materialidade e a preservação convencional, ou seja, cuja origem seja livro, periódico, CD-ROM, gravuras, e outros. Esses materiais são caracterizados por conhecimentos acumulados, publicados e preservados como memória do conhecimento, disseminados inicialmente através de catálogos em fichas, e depois através de sistemas automatizados de recuperação de informações.

De acordo com Milanese (2002), as bibliotecas eram instituições voltadas para a produção e a organização do produto a ser preservado. Atualmente, o autor acha difícil definir o que é biblioteca, e sua polêmica gira em torno de “o que será” a biblioteca.

Já Alencar (2004), Morigi e Souto (2005) e Martins (2006) afirmam que biblioteca tradicional é aquela cujo acervo é praticamente constituído de papel, seus serviços normalmente não são automatizados e também tem a figura do bibliotecário sempre presente na biblioteca.

Após as definições anteriores, observa-se posição semelhante de autores quanto ao conceito de bibliotecas tradicionais. Normalmente, são constituídas de papel com objetivo de adquirir, tratar, armazenar e divulgar seus acervos. De um modo geral, são aquelas que possuem acervos, cuja materialidade deve ser preservada.

Depois de ter visto os conceitos de bibliotecas tradicionais e o parecer de autores sobre o assunto, será relacionado na próxima parte, o *conceito de bibliotecas digitais*, no sentido de acompanhar suas idéias e possíveis evoluções.

2.2 Bibliotecas digitais

De acordo com a Enciclopédia Wikipedia (2006), biblioteca digital “é a biblioteca constituída por documentos primários, que são digitalizados quer sob a forma material (disquetes, CD-ROM, DVD), quer em linha através da Internet, permitindo o acesso à distância”.

Este conceito inclui, também, a idéia de organização composta por serviços e recursos, cujo objetivo é selecionar, organizar e distribuir a informação, conservando a integridade dos documentos digitalizados.

Cunha (2000) define bibliotecas digitais como conjunto de mecanismos eletrônicos, com objetivo de facilitar a demanda de informações, interligando recursos e usuários.

Alvarenga (2001) afirma que biblioteca digital pode ser definida como um conjunto de objetos digitais, com uso de instrumentos eletrônicos tanto para construção, quanto para recuperação, com o objetivo de registrar e comunicar pensamentos, idéias, imagens e sons, disponíveis a um número ilimitado de pessoas, na plataforma www.

Segundo Alencar (2004), biblioteca digital é aquela na plataforma WEB ou em redes locais, em que os profissionais realizam busca, recuperação, tratamento, indexação e digitalização de acervos em diversos formatos (vídeo, áudio, imagem e texto), combinados com os serviços da biblioteca tradicional, tais como indexação e organização da informação. Alencar afirma que as bibliotecas digitais são aquelas construídas por e para determinados tipos de comunidades de usuários, os quais procuram manter e otimizar seus serviços. De modo que as bibliotecas digitais são um conjunto de técnicas e meios eletrônicos para a criação, busca e uso da informação, com função básica de colecionar, reunir, preservar, usar e armazenar coleções.

Já Morigi e Souto, (2005), afirma que bibliotecas digitais são aquelas que possuem técnicas e práticas modernas, com uso de tecnologias de informação e comunicação. Procedimentos baseados em processos virtuais sustentados pela cibercultura.

Também sobre bibliotecas digitais, Martins (2006) define que são aquelas que estão em formato digital, ou seja, em disquetes, winchester, CDs e outros. Dispõem de todos os recursos de uma biblioteca eletrônica, oferecendo pesquisa e documentos com textos completos (*full text*, vídeo etc.) tanto local, quanto por meio de redes de computadores.

Após a definição desses autores, é importante se conhecer a opinião de Fujita (2005) sobre biblioteca digital, eletrônica e virtual. Para ela, entre um documento impresso e digital, existe a possibilidade de transformação mútua, o impresso pode ser digitalizado, e através deste, pode-se obter uma cópia. E, para que se possa compreender isto, a autora descreve a diferença entre Biblioteca eletrônica, virtual e digital.

Biblioteca Eletrônica é aquela que organiza e funciona como *entidade informativa individual*, com automação de rotinas básicas.

A *Biblioteca Virtual* é aquela com coleção de documentos eletrônicas na WEB sem local físico, disponível à distância. Já a *Biblioteca Digital* é aquela com acesso através de tecnologia digital.

De acordo com a autora, a biblioteca eletrônica ainda vai perdurar no Brasil, em função da existência das coleções impressas, tais como teses, dissertações, que ainda não são geradas em formato digitalizado. Pesquisadores e pós-graduações ainda não estão gerando conhecimentos em documentos digitais e em ambientes digitais, principalmente nos meios acadêmicos universitários.

Verifica-se que biblioteca digital é aquela que está na WEB, com novas tecnologias, cujo foco de acesso não é o documento físico, mas os objetos digitalizados e organizados na www, com textos completos. Só Martins (2006) faz referência a textos completos, e isso é fundamental numa biblioteca digital acreditando-se que seja o que faz a diferença.

Pode-se afirmar então, que a Biblioteca Digital está alcançando grande destaque na WEB, tanto na forma de organizar, como buscar e visualizar qualquer tipo de informação que se necessita. Percebe-se, atualmente, que o mundo todo está voltado e preocupado com o desenvolvimento de novos suportes de comunicação e processamento da informação na Internet. Existem, hoje, vários grupos de estudos interdisciplinares trabalhando em conjunto, tais como da Ciência da Computação, com a Ciência da Informação, e de outras áreas mais, no sentido de aprimorar este domínio. Além disso, a produção da informação no domínio digital está mais ligado a áreas de pesquisa e inovações tecnológicas, àquelas que percebem como vantajoso o compartilhamento de recursos da informação.

Observa-se que a biblioteca digital apresenta diversas vantagens, principalmente para os meios acadêmicos, tais como: beneficiar a humanidade e a ciência, facilitar a execução das pesquisas e buscas textuais, com economia de tempo e dinheiro. Além disso, ela possibilita armazenar, em um único lugar, grande volume de dados, com a compactação dos conteúdos digitais. Outra vantagem percebida é a democratização da informação, o uso simultâneo por vários usuários, devido ao livre acesso. São vantagens bastante louváveis na era atual.

Nota-se também, após a leitura dos textos, que a biblioteca digital, por enquanto, ainda não substituirá a biblioteca tradicional, devendo ambas conviver com suas vantagens e desvantagens, tendo sempre em vista o objetivo que é atender as necessidades de informação de seus usuários. Fica evidente que os profissionais da Ciência da Informação, aprimorando seus serviços, estudando e inovando, com categoria e *staff*, terão condições de concorrer, "igualmente" com as demais profissões que estão avançando cada vez mais sobre a área da informação.

Após se ter abordado e conhecido as diferenças entre biblioteca tradicional e digital, deve-se conhecer o processo de indexação, por ser o objeto deste estudo.

O próximo capítulo aborda pequenas considerações históricas, conceituais e as políticas do processo de indexação.

3 PROCESSO DE INDEXAÇÃO

Para se conhecer o processo de indexação, aborda-se, a seguir, um breve relato histórico, terminológico e conceitual, sobre a indexação. Nos sub itens seguintes, são citadas algumas considerações metodológicas da indexação.

Silva e Fujita (2004), fizeram uma breve explanação do contexto histórico da indexação. De acordo com essas autoras, o surgimento da indexação de deu devido à necessidade de se elaborar índices, para melhor armazenar e recuperar as informações. Esta necessidade surge na Mesopotâmia, no século II, antes de Cristo, com a grafia de resumos temáticos, sobre os documentos da época, em tabletes de argila. Depois, no século XIV, aparecem nos mosteiros os inventários de catálogos e os sumários de livros. Surgem, também, a escrita dos copistas, nas laterais dos livros, com algumas palavras ou sentenças representando os conteúdos. Isto vem provar que, desde essa época, já se fazia indexação, operação condizente com a primeira fase, a da análise de assunto dos documentos.

Ainda de acordo com as autoras citadas, no ano de 1737 surge Alexandre Cruden com o primeiro índice de uma bíblia, grande marco na indexação, que relaciona citações e as devidas localizações no texto para melhora da recuperação. Depois disto, a indexação começa a ser mais difundida.

No ano de 1876, Melvil Dewey² apresenta o índice da Classificação Decimal com expressividade. Mas, somente no século XIX, a indexação teve seu aprimoramento, devido à grande quantidade de informação encontrada. Percebe-se, nesse período, o aparecimento dos termos com seus relacionamentos (Silva e Fujita, 2004).

² CDD- Classificação Decimal de Dewey

Quanto à indexação, no domínio da documentação, também de acordo com as autoras mencionadas, os belgas Paul Otlet e Henri La Fontaine deram a melhor contribuição para a indexação, com a criação da Classificação Decimal Universal - CDU.

Para se entender o conceito de indexação, Silva e Fujita (2004) fazem um breve histórico das terminologias conceituais encontradas no processo de indexação, assim: Gardin (citada por Silva e Fujita p. 136), afirma que a indexação, sob a ótica teórica, é vista como uma operação de representação documentária, com fins de recuperação da informação. Porém, a despeito da conceituação de indexação, existem várias correntes teóricas que explicam a terminologia referente ao processo de indexação: os norte-americanos percebem a indexação, como a própria análise documentária, composta pelas mesmas operações e etapas, para se representar o conteúdo dos documentos e se elaborar índices.

A corrente francesa, de acordo com Silva e Fujita, adota a expressão análise documentária, introduzida por Gardin (1981) seguida depois, por Chaumier, Kobashi, Tálamo, Cintra, Cunha, Guimarães, Fujita, Pinto Molina e outros. De modo que a indexação está inserida dentro de um macro universo, sendo vista como a representação ou o resultado final da fase da análise documentária.

As autoras afirmam ainda que a análise documentária, vista pelos espanhóis, é percebida como descrição da forma física, e de representação dos documentos, ou seja, é a análise descritiva, e também representativa dos conteúdos dos documentos.

Já a corrente inglesa, de acordo com as mesmas autoras, representada por Foscett, Lancaster, Campos, Farrow e outros, utiliza a expressão "indexação" como processo de analisar os documentos e os representar numa linguagem de

indexação, em que a análise de assunto é tida como primeira etapa do processo de indexação, sendo a análise documentária e indexação, processos idênticos.

Assim, percebe-se que, na concepção francesa, a análise documentária se refere somente ao tratamento do conteúdo do documento, fazendo uso da expressão indexação, como processo. Na concepção inglesa, a análise documentária e indexação são processos idênticos, tal como a análise de assunto e a tradução dos conceitos, de modo que a análise, representa a primeira etapa. E na corrente espanhola, a indexação é vista como análise descritiva e representativa dos conteúdos dos documentos.

Essas diferentes correntes explicam as diferentes terminologias existentes, tais como indexação, análise de assunto, análise documentária, leitura documentária, análise de conteúdos documentários, análise de informação.

Após esta breve explanação do contexto histórico da indexação, e das suas terminologias, torna-se necessário **conceituar indexação**.

Indexação é o processo cuja ação é descrever e identificar os assuntos de um documento de acordo com o UNISIST (1981).

Pinto (1985) define a indexação como indicação do conteúdo informativo dos documentos, através de termos que representarão o assunto do documento.

Campos (1987) define a indexação como processo de captar o conteúdo informacional de um documento, e a tradução desse conteúdo para uma linguagem que sirva de intermediária entre o usuário e o documento, num Sistema

de Recuperação da Informação³, em uma ordem cronológica de operações, e que o processo de captação do conteúdo, vem em primeiro plano, e depois a tradução.

Vieira (1988) define a indexação como atividade-meio que, possibilita ao usuário o acesso adequado aos conteúdos dos documentos.

Naves (2000 p. 10) define a indexação “como ação de identificar e descrever o conteúdo de um documento, a qual ocorre na classificação ou catalogação de assunto”.

Pinto (2001) define indexação como atividade para o estabelecimento e representação do conhecimento registrado.

Fujita (2003) define a indexação também como uma ação de construir índices.

Silva e Fujita (2004) conceituam a indexação como operação do tratamento temático dos documentos, ligados à análise, ou leitura documentária, e de síntese e representação.

Por outro lado, Lancaster (2004) define a indexação como um processo, para identificar do que trata o documento, e de rotulá-lo, de acordo com um esquema de classificação, tesauro ou uma lista de cabeçalho de assunto.

Percebe-se que o termo indexação tanto no contexto histórico, como no tratamento da informação, apresenta dois sentidos. Um no sentido de criar índices, de autores, de títulos ou de assuntos. E outro, no sentido de indexar,

³ Sistema de Recuperação da Informação é a interface entre uma coleção de recursos de informação em meios impressos ou não, e uma população de usuários, com a organização, representação e acesso aos itens das informações (Lancaster, 1993).

classificar e categorizar assuntos tratados nos documentos. Todos os autores consultados pensam da mesma forma.

Esse segundo sentido percebido, indexação de assunto de um documento é aquela etapa que consiste em analisar, selecionar e representar o conteúdo informacional de um documento, de acordo com uma linguagem de indexação.

Nota-se que a indexação possui dois estágios, de acordo com Fujita (2003): o analítico e a tradução, em que o primeiro é composto pela compreensão do texto, identificação e seleção de conceitos pertinentes, e o segundo estágio é composto pela tradução, da linguagem de indexação.

Já Vickery (citado por Fujita 2003), afirma que, além desses estágios, pode existir mais um, entre o estágio analítico e a tradução; a sumarização. Ele afirma que o estágio analítico pode ser dividido em dois: análise e sumarização, de modo que a sumarização pode ser explicada como síntese.

Logo, percebe-se que o estágio de determinação de assunto, ou análise de assunto é a mesma coisa que síntese ou sumarização, tal como nos princípios de indexação. E pode-se concluir, então, que o processo de indexação, no domínio da Ciência da Informação, percebido no campo de tratamento da informação, é visto como uma operação para descrever e identificar o assunto dos documentos, ou seja, é uma operação que efetua o tratamento temático, através da análise, da síntese e da representação desse conteúdo temático. É visto, também, como uma operação para a construção de índices, para serem representados em catálogos ou bancos de dados.

Antes de se verificarem as semelhanças e diferenças, entre a indexação nos contextos tradicional e digital, é necessário se abordar sobre a importância do estabelecimento de políticas de indexação.

3.1 Política de indexação

Uma política de indexação é basicamente um guia, cujos componentes são princípios, critérios e variáveis que possivelmente poderão assegurar que as informações e demandas sejam ofertadas, de forma eficiente e econômica, em tempo hábil. Para que um serviço de indexação possa apresentar qualidades e eficiência, faz-se necessário que haja alguns requisitos, tais como: estabelecimento de uma política de indexação bem definida, na qual deve estar contida a exaustividade e a especificidade⁴ da indexação, definição clara dos objetivos da organização, e de identificação dos usuários, e determinação dos serviços condizentes com usuários e recursos humanos. Todos esses requisitos devem estar interagidos, para não afetar a recuperação da informação, de acordo com Carneiro (1985) e Fujita (2004).

Fujita e Rubi (2006a) afirmam que a qualidade e eficiência da política de indexação deve ser composta por estratégias que permitam alcançar os objetivos de um Sistema de Recuperação de Informação e, não somente, como uma lista de procedimentos e elementos a seguir. A indexação não pode ser vista somente como uma operação técnica, mas como uma tarefa da organização. E para que a indexação tenha eficiência e eficácia, é essencial que se tenha uma política de indexação com elementos bem definidos.

A autora Strehl (1998) enumera os elementos necessários numa política de indexação:

- 1) Cobertura de assunto: definição de áreas que receberão tratamento mais específico ou geral.
- 2) Processo de indexação: definição de todo o processo de indexação.

⁴ Exaustividade é a capacidade do sistema que permite indexar um documento em profundidade e a especificidade é a capacidade do sistema de permitir ser bastante preciso na especificação do assunto de um documento.

- 3) Estratégia de busca: definição da realização da busca.
- 4) Tempo de resposta do sistema: definição do tempo entre a busca e a recuperação.
- 5) Forma de saída: definição do formato de saída da recuperação.
- 6) Avaliação do sistema: identificação da forma de avaliação a ser usada no sistema.

LEI DO DIREITO AUTORAL
Todos os direitos reservados e protegidos
pela Lei 9.610/1998.
Este arquivo não pode ser reproduzido ou
transmitido sejam quais forem os meios
empregados: eletrônicos, mecânicos,
fotográficos ou quaisquer outros.

Ainda quanto aos elementos da política de indexação, também Fujita e Rubi (2006a) afirmam que estes são valores dos sistemas de informação, que deverão estar relacionados nos manuais de indexação, em que expressam como proceder dentro do ambiente organizacional. Os mesmos devem fazer parte da cultura da organização. Desse modo, os manuais têm o propósito de uniformizar os procedimentos de indexação. Devem ser um documento oficial com etapas, de análise de assunto, diretrizes e procedimentos, para o indexador, em ordem lógica. Estes manuais devem ser de três tipos:

- 1) Manual de operação ou procedimentos: descrição das atividades da indexação.
- 2) Manual de política: descrição de maneira geral e filosófica as políticas a serem seguidas pelo indexador.
- 3) Manual de organização: descrição de relatos, experiências acumuladas, para servir como manual de consulta.

A falta de uma política de indexação, de acordo com Strehl (1998), e de um vocabulário controlado, podem acarretar numa grande interferência na recuperação da informação, uma vez que a análise de assunto é um processo que envolve julgamentos e discrepâncias. Isto é percebido quando distintos indexadores, em momentos diferentes, percebem de forma diferente, o mesmo conteúdo de um documento. Assim, a coerência é um dos quesitos para uma boa indexação, ela pode ser vista tanto na concordância, quanto nos termos usados

para indexar um mesmo assunto, em diferentes documentos. A premissa da coerência é: todos documentos, que tratam de um mesmo assunto, devem ser indexados da mesma forma; a política de indexação deve estar inserida dentro do contexto do Sistema de Recuperação da Informação em questão; a definição da clientela deve ser estruturada, como também, o levantamento de recursos materiais e humanos; a definição dos objetivos da organização deve ser observada também com atenção.

Nota-se que a política de indexação é muito importante no processo de indexação, dela podendo depender o tipo de catalogação, classificação a eficiência e eficácia da indexação, enfim, todo o processo de representação da informação.

É importante, então neste trabalho, separar as duas etapas do processo de indexação, a análise de assunto e a tradução, para que haja uma melhor percepção, quanto ao "fazer" de cada etapa. Pretende-se verificar, também, na literatura, se o processo de indexação, é uma operação que se faz em um único momento, ou em momentos distintos.

Assim, o próximo capítulo traz o conceito de análise de assunto, nos contextos tradicional e digital.

4 ANÁLISE DE ASSUNTO

Este capítulo trata da primeira etapa da indexação, a análise de assunto, nos contextos tradicional e digital, em um Sistema de Recuperação da Informação. O processo tem, como objetivo, mostrar os assuntos principais de um documento, através de palavras-chaves ou enunciados, para uma futura recuperação.

Faz-se necessário conhecer o significado da palavra assunto, que pode ser definido como o conteúdo, ou a essência de um documento. Após conhecer esse significado, citam-se alguns conceitos de análise de assunto.

Segundo Cesarino e Pinto (1980), a análise de assunto é uma operação fundamental para todo o procedimento de recuperação de informações.

Pinto (1985) define análise de assunto como processo mental para a determinação do conteúdo informativo dos documentos.

Strehl (1988) define a análise de assunto como análise conceitual, atividade de definição dos assuntos dos documentos.

Vieira (1998) define análise de assunto, como a técnica de condensação do documento, mediante a atribuição de descritores, cujo objetivo é aumento da capacidade de armazenamento e o desempenho da recuperação.

Conforme Naves (2000), a análise de assunto é o ato de ler um documento e extrair deste, os conceitos que traduzam a sua essência. Conhecida também como análise temática, análise documentária, ou análise conceitual, ou ainda como análise de conteúdo. Ainda segundo Naves, análise de assunto pode ser definida como uma ação, dentro do processo de indexação, com a finalidade de

identificar e descrever o conteúdo de um documento, através da compreensão e da interpretação temática contidas nos documentos.

De acordo com Vickery (citado por Fujita 2003 p. 356), análise de assunto é vista pelo seu produto, ou seja:

“Análise de informação significa derivar de um documento o conjunto de palavras que servem como uma representação condensada desse documento. Esta representação pode ser usada para identificar o documento, para prover pontos de acesso na busca, para indicar seu conteúdo, ou como substituto para o documento”.

Já Lancaster (2004) afirma que a análise conceitual procura decidir do que trata um documento, decidir qual é o seu assunto.

Dias, (2004) define a análise de assunto, como a etapa do tratamento temático da informação.

Percebe-se, após os conceitos dos autores citados, que a análise de assunto é a etapa mais importante do processo de indexação, no domínio da Ciência da Informação. É um processo que requer compreensão, interpretação, e análise para identificação dos conteúdos dos documentos, no sentido de prover pontos de acesso para os documentos. Todos os autores têm a mesma corrente de pensamento, embora haja terminologia diferente para designar o mesmo “fazer”.

A próxima parte trata das considerações metodológicas e das análises de assunto, no contexto tradicional, e dos fatores que interferem nesse processo.

4.1 *Análise de assunto no contexto tradicional*

Este sub item trata da primeira etapa do processo de indexação, a análise de assunto, no contexto tradicional, no sentido de averiguar e conhecer quais são suas fases, bem como, saber como elas ocorrem, e quais são os fatores que interferem nesse processo.

Tal como visto anteriormente, análise de assunto é o ato de ler e descrever os assuntos essências de um documento; aqui direcionado ao contexto tradicional, ou seja, àquele contexto, cujo material pode ser manuseado. E depois do conceito desta terminologia, torna-se necessário conhecer as suas fases.

De acordo com Naves (2000, 2001), o indexador, ao executar o processo de análise de assunto, passa pelas seguintes fases: fase de leitura do texto, fase de extração dos conceitos, e fase da atinência. Essas fases sofrem interferência de fatores cognitivos, lingüísticos e lógicos. E, em qualquer operação com relação à leitura de textos, deve-se considerar esses três fatores. Acredita-se que são esses fatores citados que legitimam a interdisciplinaridade da Ciência da Informação, os quais estão explicados, um pouco mais à frente.

Naves afirma que a primeira fase da análise de assunto é a *fase de leitura do texto*, fase em que se identifica a tipologia informacional do texto e se faz análise condensada através de palavras-chaves, verificando a textualidade e a intertextualidade do documento. É uma fase que apresenta subjetividade, onde a cognição e a lingüística entram em ação.

Observa-se que, de acordo com Fujita, Nardi e Santos (1998) e os autores revisados, a maior dificuldade na análise de assunto está centrada justamente, na primeira fase, na leitura para extração dos conceitos dos documentos. E sob esta ótica de leitura, para fins de análise de assunto, e tratamento da informação, o

profissional indexador-leitor capta as idéias principais do texto e levanta hipóteses para se compreender a temática no todo. Testam-se as partes do documento, introdução, argumentos com a conclusão, usando estratégias ascendentes como “*bottom up*” e descendentes como, *top-down*. A estratégia ascendente é explicada pela etapa em que o leitor fica na dependência do contexto escrito, para extração dos significados, também chamado de movimentos das partes para o todo; e a estratégia descendente, pela etapa em que predomina o conhecimento prévio do leitor, fazendo uso da memória, e formulações de hipóteses.

Ainda sobre leitura, no processo de análise de assunto, Fujita, Nardi e Santos afirmam que o processo de leitura para identificação, seleção e síntese de conceitos, sempre resulta em enunciados, para uma futura recuperação da informação. Assim, é um processo que elas chamam de desestruturação de um texto, para construção de um outro texto, ou seja, o texto é desmanchado e transformado em conceitos pelo indexador, os quais são selecionados e depois transformados em outro texto, que traduz o conteúdo do texto analisado.

É também preocupação de Naves (2000), o processo de leitura de um texto, para fins de indexação. Ela afirma que é uma fase muito importante, ela é o meio que se permite a comunicação das idéias entre o sujeito que cria e usa a informação, ou seja, entre o emissor e o receptor, na qual a subjetividade costuma imperar. A leitura de textos para apresentar sentido e interpretabilidade, deve ser coesa e coerente. Ela é coesa quando a relação semântica entre os elementos do texto lido, e os elementos usados para sua interpretação, apresentam uma relação de “sentido”, numa ordem seqüencial lógica. É coerente quando os sentidos do texto, integram os fatores lógicos, cognitivos, e interacional, de forma adequada.

E de acordo com a autora, essa coerência é entendida, como princípio da interpretabilidade. Ela afirma que a coesão é responsável pela estruturação da seqüência do texto, a qual permite uma análise mais profunda do texto, de modo

que o leitor consiga extrair do texto, toda a sua essência. Desse modo, a análise de assunto de textos, interessa mais ao estudo da coerência textual, pois ela lida mais com a estruturação dos sentidos dos textos; elemento fundamental na análise de assunto.

Já de acordo com Neves (2004), o ato de ler se apresenta como de múltiplas facetas, em que o leitor e autor ficam sempre em constante interação. A leitura de um texto apresenta uma construção dos sentidos, quando o autor direciona a obra e espera que o leitor recupere e extraia dela o sentido que ele quis retratar. Porém, o leitor ao ler um texto, traz consigo toda a bagagem pessoal, suas experiências e vivências que irão confrontar, ou conjugar com as idéias do leitor. A autora afirma que a relação entre leitor-autor é vista tanto como conflituosa, como harmônica. O indexador tem a tarefa de decodificar o que está escrito, e extrair conceitos dos textos dos autores para torná-los acessíveis àqueles que necessitam de informações. Considera esta tarefa bastante difícil e subjetiva, pois as pessoas possuem habilidades, vivências e culturas distintas. Algumas pessoas desenvolvem habilidades metacognitivas e outras não.

Neves afirma que, para se entender o que significa habilidades metacognitivas e metacognição, se faz necessário conhecer cognição. A cognição pode ser explicada pelos processos mentais, inconscientes de um ser humano. Já a metacognição, é a gestão, ou monitoramento dos processos cognitivos da pessoa. Assim, ter habilidade metacognitiva, é saber distinguir aparências falsas do que é real, fazer associações da nova informação lida, com as anteriores, integrar o passado com o presente. Mas a autora afirma que experiências e tempo são fatores fundamentais para se desenvolver habilidades metacognitivas, e que leitores e indexadores proficientes, antes de lerem um texto, planejam a meta que irão atingir, verificarem o grau de dificuldade, e tamanho do texto que irão ler, e o que irão ignorar. Depois, procuram ativar a memória a respeito do assunto em

questão, e constroem uma hipótese do assunto abordado. Ao longo da leitura, os leitores executam várias atividades tais, como:

- 1) Extraem as partes importantes.
- 2) Reformulam suas idéias, tendo como referência àquelas adquiridas anteriormente.
- 3) Elaboram hipóteses.
- 4) Interpretam o texto.
- 5) Avaliam a qualidade do texto.
- 6) Testam as hipóteses

Fujita (2004) conclui também em seu texto, que a leitura documentária, fase importante da análise de assunto, é um ato social devido à interação existente entre leitor e autor, levando-se em consideração, o contexto de cada um, em que o entendimento, e a cognição de quem lê, determina os objetivos da leitura, tal como Naves, e Neves percebem o processo de leitura.

Percebe-se, também, que Fujita e Rubi (2006b) dividem a análise de assunto em três fases, as quais elas denominam de “suporte”. Quanto ao primeiro suporte, elas afirmam que, para se conhecer a *estrutura textual*, e *identificação de conceitos*, ou seja, para saber qual é o tema, ou idéia principal de um texto, é sempre bom fazer o seguinte questionamento, usando o “Protocolo Verbal”, falando alto, categorizando, tematicamente o texto:

- 1) O que?
- 2) Onde?
- 3) Como?

Ainda de acordo com as autoras, o primeiro questionamento faz parte da categoria essencial: O texto fala de que? As duas últimas, são categorias acessórias, as

quais vão complementar a categoria essencial. Nessa primeira parte, de exploração do conhecimento da estrutura textual, o indexador deve observar as partes que compõem o documento, os sinalizadores, os destaques, tais como: títulos, autoria, resumos, introdução, objetivos, materiais e métodos, resultados, figuras, discussão dos resultados, conclusões e referências bibliográficas. Analisar na introdução o tema principal do trabalho, conferir e analisar os objetivos, com os resultados propostos, e conclusão.

Quanto à segunda fase, de acordo com esses mesmos autores, e Naves (2000, 2001), pode ser explicada da seguinte forma: a análise de assunto, é a fase da *extração de conceitos*, aquela em que são extraídos os conceitos, ou os assuntos contidos nos documentos, e verificação do contexto, no qual o documento está inserido.

De acordo com Fujita e Rubi (2006b), a análise de assunto, quanto ao segundo suporte, pode ser explicada assim: *identifica-se os conceitos dos documentos*, quanto à *abordagem sistemática e análise conceitual*, ou seja, sistematiza-se os conceitos e os extrai do texto, utilizando a norma ABNT 12.676 (citado por Fujita e Rubi, 2006b), na qual fazem o seguinte questionamento:

- 1) O documento apresenta no seu contexto, um objeto com efeito de uma atividade?
- 2) O assunto do documento apresenta uma ação, uma operação, um processo?
- 3) O objeto é influenciado pela atividade identificada?
- 4) O documento apresenta um agente que tenha praticado esta ação?
- 5) O agente refere-se a modos específicos para realizar a ação, como técnicas, métodos, instrumentos?
- 6) Todos estes fatores são referentes a um mesmo contexto ou ambiente?

- 7) Existem variáveis dependentes ou independentes?
8) Qual o ponto de vista do estudo abordado?

Desse modo, cada tópico verificado deve revelar conceitos, os quais são estruturados em um enunciado de assunto, representando a compreensão do texto de acordo com as autoras. Elas descrevem esse modelo de leitura para indexação de textos científicos, que consiste da combinação das sistemáticas de identificação de conceitos e abordagem sistemática da Norma 12.676, com a localização dos conceitos e estrutura textual. Nessa segunda parte, ou fase, a identificação de conceitos é resultante da compreensão global do texto, da combinação, exploração e do questionamento, feitos anteriormente, em que se formulam idéias por palavras, tendo em mente o objeto, a ação, o agente, métodos, local, ou ambiente e relação causa e efeito, para a extração de conceitos.

Já a terceira fase, da análise de assunto, afirma Naves (2000), é a fase da *atinência* = **“aboutness”**, cujo significado é concernência, sobrecidade, e temacidade, não havendo um consenso de qual o termo melhor, em português, para tradução do termo em inglês. É aquela fase em que se faz a pergunta, sobre o que trata o documento, qual é o seu assunto, quais palavras que melhor representam o documento. Nessa fase, a dificuldade de escolha dos termos fica evidente, pois palavras, ou conceitos extraídos, podem apresentar significados diferentes, dependendo do seu contexto. Assim, a atinência resulta na frase de indexação, elaborada pelo indexador, e depois da frase de indexação, “sobre o que é esse documento”, passa-se para a fase da tradução dos termos, para uma linguagem de indexação, terminando na atinência, o processo de análise de assunto, ainda de acordo com Naves.

Porém, Silva e Fujita (2004), adotam a terminologia “temacidade” para essa terceira fase. De acordo com elas, o termo provém da leitura documentária, cujo

objetivo é expressar o *tema*, ou o conteúdo relevante dos documentos, que ocorre na análise conceitual para identificação dos conceitos do documento. E mediante a estrutura da língua portuguesa, o termo “aboutness” pode ter o significado de “tematicidade”, originado do termo *tema*, *temática* dos documentos. Outros autores, como Lancaster, também adotam o termo “atênência”, tal como citado por Naves, anteriormente.

E quanto ao terceiro suporte da análise de assunto, de acordo com Fujita e Rubi (2006b), é aquela que, de posse das respostas às questões feitas, selecionam-se os conceitos que são mais importantes, que melhor representam os conteúdos dos documentos, para traduzi-los para uma linguagem de indexação.

Pode-se concluir que Fujita e Rubi (2006b) apresentam um estudo com objetivo de se conseguir eficiência e eficácia, na análise de assunto, apontando procedimentos a serem seguidos. Essas autoras afirmam que a identificação de conceitos para fins de indexação, é uma operação difícil, pois envolve interação entre processos cognitivos, conhecimentos do leitor, lingüísticos, textuais, de mundo e profissionais. Assim sendo, é fundamental que indexadores sigam modelos que têm, por base, as estratégias de leitura documentária do indexador, proposta por Fujita (2003), baseada na norma da ABNT 12.676 (citado por Fujita e Rubi 2006b). Esse estudo direciona o foco para o procedimento de identificação de conceitos, mediante o método “Protocolo Verbal, ou “Pensar Alto”, tal como citado. Assim, verifica-se que o estudo é dividido em fases e etapas similares às fases mencionadas pela autora Naves. Logo, vale a pena indexadores fazerem uso desses métodos citados, no dia-a-dia, para um melhor aproveitamento dos recursos existentes, e também para se obter organização e recuperação de informação de qualidade.

Já de acordo com Albretchtsen (citado por Fujita, 2003 p. 221), a análise de assunto vai depender dos objetivos da instituição. E de acordo, com o autor,

existem três tipos de concepções. A concepção simplista é aquela em que os assuntos serão tratados como entidades absolutas, derivada da abstração lingüística do documento, a mais indicada para ser automatizada. A concepção orientada para o conteúdo é aquela que envolve interpretação além da estrutura léxica e gramatical, contidas nos documentos. E a concepção orientada para a demanda é aquela em que o assunto é considerado como instrumento de transferência de conhecimento, de modo que o indexador costuma questionar como tornar aquela indexação, ou conteúdo de um determinado documento, visível para o usuário potencial.

Também, o estudo de Lancaster (2004 p.9) é bastante importante; ele sugere que o indexador, ao fazer análise de assunto de um documento, deve fazer sempre essas três perguntas sobre o mesmo:

- 1) De que trata?
- 2) Porque foi incorporado a nosso acervo?
- 3) Quais de seus aspectos serão de interesse para nossos usuários?

Conforme já mencionado, sabe-se que vários fatores interferem no processo de análise de assunto. De acordo com Naves (2000) ,este sofre interferência de várias ordens, tais como: a política adotada pelo sistema, tipo de vocabulário adotado no sistema, os objetivos da instituição, o perfil dos usuários e os fatores humanos.

Os fatores humanos interferem no momento da análise de assunto, como o ato de pensar, interpretar e definir o assunto do documento. Por ser uma atividade intelectual, o indexador sofre interferência simultânea, originada da Lingüística, da Ciência Cognitiva e da Lógica, conforme afirma a mesma autora.

Os fatores lingüísticos que interferem na análise de assunto, são a sintaxe e a semântica. Apresentam interferência na conversão da linguagem natural para a documentária. A terminologia humana e suas relações não acompanham a terminologia documentária e as mudanças da sociedade. Então, percebe-se que, no processo de análise de assunto, a leitura e a compreensão são de suma importância para que o processo todo se realize com eficácia. Porém, ao se ler um texto, a compreensão deste é reconstruída, devido à agregação de conhecimentos prévios e informações não percebidas pelo indexador, momento em que a cognição entra em ação, de acordo com a autora. A representação de conteúdo de um texto é uma atividade intelectual, em que há identificação de palavras, compreensão do léxico, ou seja, compreensão dos vocábulos de um idioma, elaboração das representações cognitivas percebidas no texto, memorização e agregação aos conhecimentos anteriores.

Já os fatores lógicos, de acordo com Naves (2000), são percebidos através de assuntos novos, que exigem raciocínio, generalização e construção teórica pelo indexador, que foge do bom senso, e do senso comum, em que os processos dedutivo e indutivo entram em ação. O indexador tem que usar a lógica, a dedução e os juízos para estabelecer relacionamentos, causas e conhecimentos próprios para se chegar a conclusões lógicas.

Assim sendo, percebe-se que a análise de assunto, principalmente o processo de leitura de textos, no contexto tradicional é bastante importante. Ela é vista por muitos autores como uma tarefa intelectual e difícil. É considerada a principal dificuldade do profissional, uma vez, que ao ler um documento, deve-se levar em conta, o objetivo do autor, do sistema, e tipos de usuários da instituição. E vários outros fatores interferem neste processo, tais como: grau de conhecimento do indexador, falta de conhecimentos gerais; falta de conhecimento quanto ao assunto tratado; e falta de conhecimentos quanto aos instrumentos de indexação.

Percebe-se, também, que quase todos os autores dividem a análise de assunto, em três fases, como citado por Naves. Esta afirmação pôde ser constatada também, pelos estudos de Lancaster, Fujita e Rubi.

Verifica-se, através das idéias dos autores, e conhecimentos adquiridos, que a análise de assunto, no contexto tradicional, é uma tarefa que envolve raciocínio intelectual, conhecimentos prévios, experiência do indexador e necessidade de áreas interdisciplinares, tais como a lingüística para explicar a semântica, o léxico, e os relacionamentos; a psicologia para explicar e acompanhar a agregação de conhecimento e memorização; a lógica para explicar a intuição, a dedução, e a explicação de novos conhecimentos. Isso mostra que a análise de assunto é uma operação muito subjetiva, e que os fatores humanos têm grande influência na eficácia da recuperação da informação, tal como grau de incerteza, ansiedade e de seriedade, encontrados na prática da análise de assunto, e difíceis de serem controlados.

Fujita, Nardi e Santos (1998), afirmam que o indexador não precisa ser especialista no assunto a ser tratado, devido à sua habilidade de se interagir com o texto, com base na estruturação textual. Acredita-se que isso não seria viável pois, um especialista no assunto, terá muito melhor desempenho que um leigo. Sabe-se que, nos Estados Unidos, os cursos de pós-graduação de Biblioteconomia, todo aluno é especialista numa determinada área, logo o currículo e o desempenho deste profissional será melhor que aquele não especializado, uma vez que ele domina a terminologia, e tem conhecimento da taxonomia das classes de assuntos da área em que está trabalhando.

Após conhecer o que é análise de assunto, no contexto tradicional, suas fases, e sua importância, bem como os fatores interferentes neste processo, necessita-se conhecer a análise de assunto no contexto digital, tema da próxima parte desta monografia.

4.2 Análise de assunto no contexto digital

Este sub item aborda as considerações metodológicas, da primeira etapa do processo de indexação, no contexto digital.

Com base nas definições anteriores, pode-se afirmar que análise de assunto é a operação que procura definir o assunto dos documentos, no domínio digital.

De acordo com Alvarenga (2001), a análise de assunto, os conceitos, e a atenção são componentes invariáveis no tratamento da informação tradicional e digital, ou seja, não sofrem diferenças ou alterações no contexto digital. A autora argumenta que não poderia ser diferente, pois uma máquina não seria capaz de definir conceitos de imagens em movimento, documentos sonoros, imagens, sons e outros tipos de materiais. Afirma, também, que na análise de assunto manual, a extração de conceitos é fundamental para tratar e organizar a informação, para melhor compreender, categorizar e recuperar as informações.

A autora ainda cita experiências na área de Ciência da Informação, com o uso de técnicas e ferramentas do domínio da Biblioteconomia, e também experiências com o uso de técnicas não convencionais, tentativas automatizadas, cujos resultados foram bem limitados. Mostra ferramentas com base na parte da filosofia e metafísica em que trata da natureza do ser, na prática da "engenharia do conhecimento". A autora argumenta que o ideal seria o estudo de uma nova linguagem que pudesse ser entendida pela máquina e pelas pessoas, no sentido de codificar os conceitos discursivos, gráficos e sonoros.

Os processos tradicionais de análise de assunto, de acordo com a autora, são primordiais para o tratamento de informações digitais, e para ela, mudam-se os meios, sofisticam-se os instrumentos e surgem nomes novos para designar coisas velhas, entretanto, a essência das coisas permanece.

O pensamento de Dias (2001) também é parecido com Alvarenga, quando afirma que o tratamento da informação, no contexto digital, acontece tal como no contexto tradicional. Porém, deve-se levar em consideração a natureza, a característica e a tipologia das bibliotecas digitais, e não generalizá-las.

Também Naves (2001), afirma que apesar de a tecnologia avançar cada vez mais, ainda não se conseguiu automatizar a abstração, a percepção, a interpretação e o pensamento da mente humana, no processo de análise de assunto num Sistema de Recuperação da Informação.

Silva (2004) também concorda com essa idéia de que a análise de assunto, no processo de indexação, utiliza-se da linguagem natural da análise de assunto, para se recuperar as informações, tanto no contexto automatizado, quanto no digital.

Dentro da vertente de Alvarenga, Dias, Naves e Silva, estão Fernandes e Azevedo Netto (2005) quando afirmam que a WEB e a Internet estão um caos, falta mão-de-obra de profissionais capacitados, como os indexadores, para solucionar os problemas referentes à recuperação da informação. Afirmam, ainda, que nenhum mecanismo automatizado será eficaz, se não tiver o indexador como especialista para indexar, organizar e tratar as informações.

Verifica-se que o tratamento da informação, na WEB, ou no contexto digital, demanda uma equipe multidisciplinar tal como indexadores, especialistas da área a ser indexada, como também interdisciplinaridade de outras áreas para a análise de assunto, tal como a lingüística, a psicologia e a lógica. Acredita-se que a atividade de análise de assuntos de documentos, no contexto digital, não possa ser feita por máquinas, pois essas não são capazes de definir e compreender os assuntos dos documentos e muito menos de contextualizá-los.

Dias, Naves e Moura, (2001) dão grande importância à análise de assunto. Eles citam vários estudos e pesquisas que apontam a recuperação da informação através de buscas por assunto, e de acordo com eles, esse tipo de busca apresenta o maior índice de fracasso, devido às análises pouco elaboradas, sem levar em conta o usuário. Tendo em vista esse fracasso, a utilização de palavras-chaves nos títulos começam a proliferar cada vez mais, nas recuperações da informação, o que é lamentável, de acordo com eles, pois é uma recuperação limitada. Esses autores percebem o usuário como peça fundamental, na análise de assunto. E acredita-se, que este é o caminho certo, e que a análise de assunto, no contexto digital, não deve ser diferente.

Percebe-se que a análise de assunto é uma operação fundamental para um Sistema da Informação, é a operação em que o profissional indexador faz quando vai determinar o assunto de um documento, para dar entrada num Sistema de Informação, também quando recebe um pedido de informação, ou seja, quando o profissional analisa pedidos de um usuário na tentativa de identificar a necessidade desse usuário sob forma de conceitos, e os traduz para a linguagem adotada pelo sistema. Assim, a primeira tarefa da indexação, de descobrir os assuntos dos documentos, é parecida tanto do contexto tradicional, quanto digital.

Enfim, sabe-se, que bases de dados, que utilizam somente a linguagem natural, num Sistema de Recuperação da Informação, traz falsas recuperações, apresentam alto grau de revocação⁵, e baixo grau de precisão⁶. A autocategorização, também parece não ser muito pertinente, pois conceitos relevantes podem passar despercebidos. Deve-se ter cuidado ao criar Bibliotecas Digitais, tendo sempre em mente que tarefas tais como a análise de assunto, e a

⁵ Revocação é o coeficiente entre o número de documentos inseridos no sistema versus o número de documentos relevantes recuperados pelo usuário.

⁶ Precisão é o coeficiente entre o número de documentos inseridos no sistema e a precisão da busca.

extração de conceitos, não deve ser delegada às máquinas. As máquinas servem como um facilitador, para os profissionais indexadores, mas ainda são se pode afirmar que substituirá a atividade intelectual do ser humano. Por mais instrumentos e padrões que se tenha, não será fácil substituir a tarefa humana de analisar e categorizar os conteúdos dos documentos.

De acordo com Ribeiro e Mendes (2006), fica evidente hoje, que os paradigmas mudaram, eles estão mais voltados para a questão de como se criar pontos de acesso temáticos aos documentos realizados ao vivo.

Porém, argumentam que o maior problema, enfrentado pela biblioteca digital, é a questão da preservação, uma vez que estes materiais têm vida média curta, equipamentos e programas de informática ficam ultrapassados com muita rapidez e isso deve ser repensado.

Outro problema também, é a questão do acesso aos documentos na WEB. Quase sempre, ficam disponíveis os documentos mais recentes. Observa-se, hoje, que um determinado documento possa estar *on line*, mas amanhã, não se sabe.

Essa é uma preocupação de Cunha, (2000) quanto à vida média de um WEBSITES⁷, que é de 70 dias mais ou menos. As bibliotecas e os profissionais da área de informação deveriam se preocupar com isso, uma vez que, no meio virtual, não se percebe essa preocupação. Endereços, URLs⁸ mudam com uma freqüência muito grande, como também arquivos deixam de ser armazenados constantemente. É tarefa dos indexadores estudar sobre preservação e novas e técnicas de indexar, como metadados e marcação de textos e outros mais.

⁷ WEBSITES é o conjunto de páginas da WEB de hipertextos acessíveis pelo protocolo http na Internet.

⁸ URLs (Localizador Universal de Recursos) são endereços de recursos em uma rede.

Diante disso, fica o seguinte questionamento: como fazer, caso as bibliotecas tradicionais deixem o cenário vazio, ou melhor, deixem de existir, permanecendo somente bibliotecas digitais, e houver necessidade, de pesquisadores em fazer prioritariamente, um levantamento bibliográfico retrospectivo?

Sabe-se que hoje, para solucionar essas questões, pesquisadores costumam fazer arquivos em seus próprios computadores, e isso não é a solução, logicamente. Deve-se pensar a solução no nível global, de WEB, e Internet, e não como está acontecendo. E deve-se pensar, com uma certa urgência, pois este transtorno já é percebido, como no caso da Base de dados do Portal Capes de Periódicos Científicos⁹ que na maioria das vezes, seus usuários têm de recorrer aos serviços de comutação bibliográfica¹⁰ para recuperar um assunto, publicado há quatro ou mais anos atrás.

Após fazer a análise de assunto, o profissional indexador passa para a segunda etapa do processo de indexação, ou seja, para a tradução dos termos selecionados, de acordo com linguagens de indexação. O próximo capítulo aborda conceitos e considerações metodológicas dessa segunda etapa.

⁹ Portal Capes de Periódicos Científicos é um portal brasileiro de informação científica, de periódicos nacionais e internacionais de mais de 10.555 títulos de revistas, na WEB.

¹⁰ Comutação bibliográfica é um serviço oferecido pela biblioteca, que visa dar acesso à informação existentes nas bibliotecas brasileiras e na Bristh Library, através do Comut, para obtenção de cópias.

5 TRADUÇÃO DE CONCEITO: PARA LINGUAGENS DE INDEXAÇÃO

Este capítulo trata da conversão da linguagem natural, para a linguagem artificial, ou seja, trata da tradução dos termos selecionados, no processo de leitura, proveniente da análise de assunto, para uma linguagem de indexação. Esta operação tem, como objetivo, representar o conteúdo temático dos documentos, controlar o vocabulário, mostrar os relacionamentos existentes entre os termos, e fazer coincidir o vocabulário do indexador com o do usuário.

Faz-se necessário definir, inicialmente, a palavra conceito, para se entender as expressões extração e representação de conceitos.

Conforme Medeiros (1986), conceitos são unidades de conhecimentos que podem ser identificadas por enunciados, expressados através da linguagem verbal, a respeito de um item. É a representação da idéia de um objeto, representando aquilo que há de permanente, imutável e comum a todos os objetos da mesma espécie.

Dahlberg (1987) define conceito como uma unidade de conhecimento sobre determinado item através de uma forma verbal.

Segundo Naves (2000 p. 55),

“Um conceito é definido como um conjunto de características, que são os elementos dos conceitos e traduzem os atributos das coisas designadas. A característica mais geral é chamada categoria, que é o conceito na sua mais ampla extensão”.

Após a definição dessas autoras, quanto ao termo conceito, pode-se definir a fase de tradução de conceitos, no contexto da indexação.

Pinto (1985) define a etapa da tradução, como operação em que os conceitos selecionados serão traduzidos em termos indexadores, de acordo com algum padrão consistente, ou de acordo com a linguagem de indexação usada no sistema.

Strehl (1998) define a tradução, como atividade de conversão dos conceitos identificados na análise conceitual, para uma linguagem de indexação.

Naves (2000) define a tradução, como aquela etapa que se faz a transferência da frase de indexação, para os termos adotados por uma linguagem de indexação, no sentido de transformar os conceitos selecionados, em termos ou símbolos autorizados, para representá-los em um sistema. É o momento de traduzir a frase de indexação, feita na análise de assunto, para os termos autorizados pelo sistema.

Também Fujita (2003) define a tradução, como operação da representação de conceitos por termos de linguagens de indexação.

Embora a autora Campos (1997) tenha argumentado que existe uma ordem cronológica, na operação de indexação, nas duas etapas, análise de assunto e tradução, ela mesma percebe que as mesmas, têm caráter integrativo, uma **unicidade** das duas operações. Afirma que a indexação é uma única operação, em dois momentos. Porém, é uma operação que carece de suporte teórico, e de cientificidade. Hoje, este processo se apresenta mais como uma arte, e um artesanato, do que propriamente uma técnica. A autora Campos argumenta que o estudo da lingüística, da sintaxe e da semântica, seria um ponto de partida para o alcance dos objetivos necessários para o avanço na indexação.

Após os conceitos, nota-se que a tradução é a etapa da indexação em que se faz transferência de uma terminologia da linguagem natural para uma linguagem

LEI DO DIREITO AUTORAL
Este arquivo não pode ser reproduzido ou
transmitido seja quais forem os meios
empregados: eletrônicos, mecânicos,
fotográficos ou quaisquer outros.

artificial, ou seja, é a troca dos termos ou conceitos encontrados na leitura de um documento, e selecionados, para a conversão em termos indexadores, utilizados pelas linguagens de indexação, de modo a satisfazer usuários e sistemas de informação.

Percebe-se, então, que para representar os conceitos extraídos dos documentos, numa linguagem de indexação, é necessário que o indexador tenha conhecimento dos instrumentos de indexação, como: as linguagens verbais, representadas pelos tesouros¹¹ e listas de cabeçalhos de assunto; as linguagens simbólicas, representadas pelos símbolos de classificação. O indexador deve conhecer estes instrumentos de indexação, pois a qualidade da indexação vai depender da qualificação do indexador e dos instrumentos utilizados. O indexador deve ser, de preferência, um especialista na área a ser indexada, deve entender os termos encontrados e ser imparcial ao indexar. A instituição deve medir a recuperação do seu sistema, pelo cálculo das taxas de revocação e precisão, de acordo com UNISIST (1981).

Então, evidencia-se a necessidade de recorrer à literatura para averiguar e conhecer, como é feita a tradução de conceitos, em termos indexadores, no processo de indexação, no contexto tradicional, assunto tratado na próxima parte deste capítulo.

¹¹ Teseurus é uma lista de palavras com significados semelhantes dentro de um domínio específico do conhecimento.

5.1 Tradução de conceitos: no contexto tradicional

Esta parte trata das considerações metodológicas da segunda etapa de indexação, ou seja, mostra a etapa de representação dos conceitos extraídos dos documentos, para termos indexadores, de acordo com as linguagens de indexação, dentro do contexto tradicional.

Essa segunda etapa do processo de indexação, indexação manual, intelectual ou humana, também chamada de etapa de tradução, é aquela etapa em que os indexadores comparam os conceitos da linguagem natural pré-selecionados, para uma linguagem de indexação, conforme afirma Pinto (2001). Caso haja coincidência dos termos, eles poderão ser escolhidos para representar o conteúdo do documento. Quando esses conceitos não coincidem, haverá necessidade de uma tomada de decisão, e o resultado desta decisão vai depender do grau de conhecimento do indexador quanto à área, à política da instituição, e ao perfil do usuário.

Também Silva e Fujita (2004) acreditam que a indexação, no contexto tradicional, é a indexação manual, realizada no final da representação documentária. É a operação resultante da fase de análise documentária. É a operação que enumera os conceitos extraídos dos documentos e os representa através de uma linguagem de indexação, tendo, como auxílio, as listas de descritores, thesaurus, listas de autoridades, ou outros instrumentos autorizados num Sistema de Recuperação da Informação (Silva e Fujita, 2004).

Ainda segundo as autoras, a indexação foi dividida em indexação alfabética de assunto, indexação coordenada e indexação automática. A indexação alfabética de assunto determina os cabeçalhos de assuntos dos documentos. Também denominada de catalogação de assunto, ela é resultado da operação de análise de assunto. O uso da terminologia cabeçalho de assunto é influência dos Estados

Unidos, por causa das listas de cabeçalhos de assunto utilizadas. Assim, catalogação de assunto significa representação dos conteúdos dos documentos em catálogos de bibliotecas. Porém, as autoras afirmam que é bastante claro, para a comunidade de Ciência da Informação, que a catalogação de assunto é um ramo maior dentro da indexação de assunto. Essas autoras afirmam que os cabeçalhos de assunto surgiram devido à necessidade de representar melhor os assuntos dos documentos, uma vez que os títulos dos mesmos não estavam mais sendo representativos o suficiente. Profissionais da área de Ciência da Informação tinham muita dificuldade em descrever os assuntos compostos e complexos, devido à necessidade de sínteses, recurso difícil de ser usado, por requerer habilidade do profissional.

Dentro da linha de pensamento de Silva e Fujita (2004), a indexação coordenada é caracterizada pela composição de palavras-chave, descritores ou assuntos que representam os conteúdos precisos dos documentos. A indexação coordenada é conhecida como pré e pós-coordenada, ambas utilizadas num sistema de recuperação tradicional. Na pré-coordenada, os termos são combinados no momento da indexação, com o auxílio de listas de cabeçalho de assunto. E na pós-coordenada, os termos são combinados no momento da busca, ou recuperação da informação, na qual os tesauros são os instrumentos mais indicados, por reunirem descritores que representem conceitos simples.

Já a indexação automática, segundo as autoras, é aquela que identifica e seleciona os termos correspondentes aos assuntos, sem a intervenção dos profissionais indexadores. O usuário utiliza termos, palavras ou expressões significativas que descrevem o conteúdo do documento, através do computador. Logo, ela é diferente da indexação automatizada, pois essa última utiliza o computador como meio para recuperar as informações, indexadas por um profissional.

Fujita (2004) argumenta ser de grande importância a etapa de tradução de conceitos para uma linguagem de indexação, devido à necessidade dessa tradução em absorver os objetivos do autor, do leitor e da instituição. A autora argumenta que a linguagem documentária, também conhecida como linguagem de indexação, deve seguir regras semânticas e controle de vocabulário, para representar os assuntos dos documentos durante a indexação, tais como controle de sinônimos, homônimos.

Esta etapa, de acordo com Pinto (2001), é bastante complexa. O indexador deve seguir seriamente alguns padrões e objetivos da instituição, as listas de cabeçalhos de assuntos, os tesouros, ou outras listas autorizadas pela instituição. Estes instrumentos asseguram a desambigüização e a padronização dos termos, mas não garante uma indexação correta, caso o indexador não tenha extraído adequadamente os termos do documento, na primeira etapa de análise de assunto, ou não saiba usar os instrumentos de indexação.

Para que haja padronização, algumas considerações são necessárias para uma boa indexação, conforme afirma Fidel (citada por Pinto, 2001):

- 1) O indexador sabe usar fontes da área como de auxílio para seleção dos termos?
- 2) O indexador conhece o grau de fidelidade da tradução, e da precisão?
- 3) O indexador deve ser mais ou menos específico?
- 4) O indexador conhece qual termo é mais adequado para a sua instituição?
- 5) O indexador conhece a linguagem dos seus usuários e suas necessidades?
- 6) O indexador conhece a política de indexação da sua instituição?

Todos esses questionamentos são necessários e deverão estar contidos na política de indexação traçada pela instituição, tal como já citado anteriormente pois, caso contrário, a recuperação estará comprometida.

Se o indexador for muito abrangente na indexação, utilizando classes mais gerais, o grau de especificidade será baixo, e como conseqüência, o sistema terá alta revocação e baixa precisão. Ao passo que se o indexador for muito específico, terá alto grau de precisão e grau muito baixo de revocação. A escolha, conforme já citado, deverá ser de acordo com a política de indexação do sistema, a tipologia dos usuários e os objetivos da instituição. A eficiência na recuperação também vai depender do grau de conhecimento quanto a novas tecnologias. O profissional indexador tem, por obrigação, estar sempre em interação com os avanços surgidos. Aquele que ficar estagnado no tempo, provavelmente, ficará preso a um depósito de livros, sem muita representatividade para uma comunidade de usuários.

Conclui-se que a segunda etapa da indexação, em que o indexador vai traduzir os conceitos extraídos do documento, para uma linguagem de um sistema, no contexto tradicional, é bastante complicada e morosa, pois o ato de indexar um documento, de se pensar em que classe ele poderá pertencer, em que categoria, que termos indexadores serão passíveis de se transformar em uma linguagem de indexação é muito complexo. Essa etapa tem, por objetivo, a transformação da linguagem natural em uma linguagem de indexação utilizada pelo sistema, que faz coincidir a linguagem do indexador, com a linguagem utilizada pelo sistema. É neste momento, que o indexador deverá decidir quanto ao grau de especificidade do sistema, tendo em vista a política do sistema, conforme tratada no terceiro capítulo.

Após se conhecer a tradução no contexto tradicional, pode-se estudá-la no contexto digital, assunto tratado no próximo sub item.

5.2 Tradução de conceitos: no contexto digital

Este item trata das considerações metodológicas da segunda etapa do processo de indexação, da tradução dos conceitos, em termos indexadores, para linguagens de indexação, no contexto digital. Tem, como base, as considerações de autores da área, para uma possível comparação entre a indexação tradicional e a digital.

Percebe-se a existência de duas vertentes de pensamentos; uma que adota os instrumentos tradicionais de indexação, e outra que acha necessário romper com o tradicional, e com a forma hierárquica.

A princípio, serão relacionados os autores *que têm preferência pelos instrumentos tradicionais*, para efetuar o processamento técnico da informação. Depois, aqueles que apóiam as novas tecnologias, e vêem necessidade de mudanças.

O texto de Vieira (1988) descreve a aplicação de uma metodologia para indexação manual e automática de títulos e resumos na língua portuguesa, no domínio da Ciência da Informação, no período de 1972 a 1983, no sentido de comparar essas duas tipologias de indexação. A autora defende que a indexação automática tem uma qualidade tão boa quanto a manual. Nada menciona quanto à indexação no contexto digital, porém acredita-se que seja favorável, uma vez que aprova a indexação automatizada também.

Ainda sobre a indexação automática, Pinto (2001) argumenta que, apesar da evolução dos processos de indexação semi-automática e automática, a indexação manual ainda não foi excluída. Alguns a consideram obsoleta, porém, países do primeiro mundo ainda continuam a utilizá-la, porque a indexação semi-automática e a automática não deram bons resultados na recuperação da informação. O porquê disto se deve à não oferta de respostas satisfatórias na recuperação das

informações. Além disso, de um modo geral, os sistemas de indexação automática ainda não estão sendo muito difundidos.

Já o autor Dias (2001) considera que o nível de detalhes do tratamento da informação vai depender do tipo de sistema e dos instrumentos utilizados. Do ponto de vista temático, em bibliotecas digitais o ideal para o tratamento da informação é o vocabulário controlado e as linguagens documentárias. O autor cita a Net First, da OCLC¹², com processamento manual, no contexto digital, com máquinas de busca iguais a Yahoo, e Altavista¹³, mas com objetivos e seleção bem definidos, e uso de Sistema de Classificação de Dewey para a descrição temática. Pensa ser necessário o uso das linguagens documentárias e de sistema de classificação no contexto digital. No entanto, isso não impossibilita que novos processos e instrumentos venham a ser desenvolvidos. Mas, é evidente que o contexto tradicional ainda permanecerá por bastante tempo, e que as bibliotecas digitais devem aprender com as bibliotecas tradicionais.

Também Alencar (2004), acredita que a biblioteca digital é a biblioteca tradicional presente na Internet, com serviços de bibliotecas convencionais e com suportes digitais e serviços de busca rápida, ou seja, é a biblioteca que faz uso de todos os serviços tradicionais, tratamento, indexação, digitalização de acervos, e organização da informação, com suporte digital.

A seguir, estão relacionados os autores que, de um modo geral, acreditam no **rompimento do "fazer" tradicional, na indexação**, para a organização e a recuperação da informação.

¹² OCLC é Online Computer Library Center, cuja missão é facilitar o acesso à informação mundial através da cooperação

¹³ Yahoo é de uma empresa norte-americana que presta serviços de Internet com missão global de pesquisa. Altavista é de uma empresa Overture Services, com o objetivo de prestar serviços de pesquisa.

Drabenstott e Bruman (1997) pensam que no registro, para fins de localização nas bases de dados, deve-se juntar a programas como o *knowbot*, (um robôzinho que automatiza buscas e coletas de dados em bases de dados partilhadas) Catalogadores de assuntos, que conseguirem romper com os modelos tradicionais, farão análise de conteúdo e indexação, estabelecendo ligações entre materiais digitais, e outros novos formatos e instrumentos que por ventura possam surgir.

Assim também acredita Levacov (1997), e, segundo ela, com as novas tecnologias, as bibliotecas poderão ser chamadas de "bibliotecas sem paredes para livros sem páginas". Afirma que os conceitos de lugar e tempo tomaram-se secundários. A palavra "lugar", na atual conjuntura, significa acesso e confiabilidade da informação, ou seja, quem produziu, identificou, selecionou e disponibilizou, e quem a tornou autêntica. A palavra "tempo" implica em virtualização ou relativização, instantaneidade, velocidade qualitativa do espaço e tempo mutantes.

De acordo com a autora, a construção das bibliotecas virtuais acontecem devagar. Primeiro, a parte chamada de *off-line*, com o controle do inventário, circulação de materiais, catálogos eletrônicos e automação da indexação. Depois, com o aparecimento dos CD-ROMs, com as obras de referências e a seguir bases de dados *on-line* com textos completos. A autora faz o seguinte questionamento: atualmente, a biblioteca deve decidir o que é mais importante, investir na coleção, ou no acesso. Porém, há uma tendência dos profissionais da informação em preferir o acesso global e não local, pois isso permite que as bibliotecas se tomem um setor a mais, justificando investimentos e projetos. E, com isso, profissionais da informação devem rever maneiras de trabalhar e oferecer serviços para que o usuário remoto possa recuperar com eficiência, eficácia e satisfação as informações desejadas.

Ainda de acordo com Levacov, um dos maiores problemas da era digital está em desenvolver políticas que identifiquem quem indexou, compartilhou as informações, como também, quem é responsável pela autenticidade. O maior problema é mudar do paradigma da posse, ou propriedade, para o paradigma do acesso. Faz-se necessário adquirir novas habilidades a respeito do que ela chama de "alfabetização digital". Ela considera que 1% da informação arquivada no mundo, está em formato digital, sendo portanto, um desafio para o futuro dos bibliotecários, tornar os 99% disponíveis no mundo digital.

Conforme Alcaide et al. (2000), a representação automática apresenta melhor consistência que a representação manual, dentro de um contexto de informação documentária. Esta afirmativa tem, como base, o estudo experimental feito com a utilização de um protótipo para fazer o processamento automático, através do método estatístico-morfológico, em que as palavras significativas do texto são contadas através da frequência de repetição dos termos. Faz-se o levantamento dos termos significativos, e estes são comparados em dicionário morfológico. E depois, estes termos significativos vão para a representação automática. O resultado do estudo mostra que indexação automática é mais exaustiva, consistente e pertinente que a indexação manual.

Oliveira (2002) tem uma grande preocupação com a terminologia na WEB, segundo ela a WEB Semântica¹⁴ é uma proposta apresentada como solução para resolver o caos da informação existente na WEB. Assim, a WEB Semântica é uma extensão da WEB. A autora aponta que a proposta da WEB Semântica é tratar da semântica, ou seja, tratar do conteúdo das informações. Pensa que há necessidade de estudos e esforços no sentido de gerar linguagens e mecanismos que possibilitem a conceituação e a representação semântica dos documentos na WEB.

¹⁴ Web semântica é um novo formato de conteúdo que tem significado para computadores, de modo a interligar significados de palavras a idéias.

Dentro dessa mesma corrente de pensamento estão Souza e Alvarenga (2004). Eles consideram que a recuperação da informação, via WEB, deixa muito a desejar, pois a forma de indexação utilizada é baseada em palavras-chave contidas nos documentos. E devido ao caos percebido, neste aspecto, surge a proposta da WEB Semântica, uma vez que esta propõe os estudos do sentido das palavras; do relacionamento de problemas psicológicos dos interlocutores nos processos de comunicação dos signos; dos problemas lógicos, com estabelecimentos dos signos com a realidade na significação; e dos problemas lingüísticos para estabelecimento das funções dos vários sistemas dos signos.

Essa também é a questão levantada por Fernandes, Pinto e Azevedo Netto (2005), ao afirmarem que a representação informacional no ambiente virtual ou digital, a princípio, foi baseado no índice KWIC¹⁵, depois no SMART¹⁶. E, com o advento da Internet, e estruturação da WEB, percebe-se que os unitermos atribuídos pelas máquinas, para recuperar informações, não mais atendiam às necessidades dos usuários. Fica latente uma grande necessidade de se utilizarem novas técnicas de tratamento, nesse ambiente, para organização e recuperação.

Porém, mesmo com boas ferramentas e muitos estudos, é necessário que o indexador tenha conhecimento da indexação, das características da organização e de seu público como um todo.

Verifica-se que uma boa parte dos autores consultados acreditam na indexação puramente manual, tanto para biblioteca tradicional, quanto para biblioteca digital; na necessidade dos serviços de profissionais indexadores, em ambas as etapas da indexação; na necessidade de manuais de serviços com políticas bem traçadas num documento oficial; na mão-de-obra humana.

¹⁵ KWIC é um índice rotado derivado dos títulos dos documentos (Key Word in the Context).

¹⁶ SMART é um índice estatístico e probabilidades de ocorrência e co-ocorrência de palavras (System for Manipulation and Retrieval of Text).

Já outros autores, de um modo geral, preferem romper com o tradicional, havendo também aqueles que querem fazer adaptações quanto aos recursos tradicionais, com novos instrumentos e tecnologias. Quase todos propõem novas formas de catalogar e indexar, que são citadas a seguir, para uma melhor noção desse novo domínio.

Mas, é importante ressaltar que, se numa biblioteca tradicional, com pequeno volume de documentos, existem desafios, imagine-se num universo com grande volume, como no mundo virtual e digital? Percebe-se, também, que a exaustividade e a especificidade são dois elementos que deverão estar bem definidos no sistema de recuperação, pois estes dois elementos vão determinar a eficácia do sistema, e terão influência direta na indexação realizada e na recuperação da informação.

Antes de explicar esse novo contexto, pergunta-se: será que não mudariam somente alguns instrumentos de trabalho? O vocabulário não continuaria a ser controlado? Para solucionar esses questionamentos, faz-se necessária uma breve revisão sobre *metadados*, *XML*, *ontologias*, *ciberespaço* e *hipertexto*, tópicos mais citados nesta monografia, quanto a novos instrumentos, padrões e linguagens de marcação, usados na organização da informação.

Metadados, segundo Souza (1987), são dado sobre dados, a partir de um dicionário digital de dados, usado na geração, recuperação e uso de documentos eletrônicos, para estabelecer padrões de conteúdo, com objetivo de documentar, descrever e organizar os dados, de forma estruturada e padronizada, facilitando a manutenção e a recuperação. Os metadados são estruturados por elementos para a descrição do conteúdo dos dados, os quais podem ser chamados de etiquetas, tais como: nome do campo, descrição do campo, tipo de dado, formato, enfim qualquer dado necessário que se queira recuperar, e o usuário tem liberdade de escolher os campos durante sua busca por informações.

Souza ainda alerta que, no meio eletrônico, é fundamental o uso de padrões para os conteúdos de intercâmbio de dados em metadados. Existem padrões para descrição de dados bibliográficos, formato de intercâmbio de dados bibliográficos, formato de entrada de dados, como o formato MARC¹⁷, e o Dublin Core¹⁸. Este último é um conjunto de quinze elementos de metadados, tais como: autor, título, assuntos, palavras-chaves, resumo, editor, outros colaboradores, data, gênero do recurso, formato, identificação do documento, língua, relação, cobertura e direito autoral. Todos têm, por objetivo, facilitar a recuperação de documentos na WEB, em que o elemento assunto pode ser definido com o uso da Classificação Decimal de Dewey, Classificação Decimal Universal, LCSH¹⁹, Tesouro, até mesmo com um conjunto de palavras, ou seja, utilizando-se da linguagem natural. Os metadados utilizados nas Bibliotecas Digitais são usados com recursos de linguagem de marcação, ou seja, acrescentam-se marcas e etiquetas aos recursos informacionais.

Junto com os metadados são utilizados também recursos das linguagens de marcação. É importante conhecer o que são linguagens de marcação. Almeida (2003) aponta que linguagem de marcação pode ser qualquer informação adicional acrescentada ao texto de um documento, são padrões que servem para transferir informações de um formato para outro. São exemplos de linguagem de marcação: SGML, HTML, XML.

De acordo com o autor, o SGML (Stander Generalized Markup Language) linguagem de marcação, é um padrão simples, não tem elementos pré-definidos,

¹⁷ MARC Machine Readable Catalogin Format – desenvolvido e mantido pela Library of Congress com objetivo de padronizar a representação descritiva automatizada de acervos bibliográficos.

¹⁸ Dublin Core – Padrão de metadados voltado para a descrição e recursos na Internet. Possui vocabulário controlado e quinze elementos que servem para descrever recursos na WEB.

¹⁹ LCSH significa Library of Congress Subject Heandings. É uma lista de cabeçalho de assunto da Biblioteca do Congresso E.U.A..

podendo ser criados de acordo com a necessidade. Esta foi a primeira linguagem de marcação criada que se tem notícia.

HTML (Hiptertext Markup Language) é baseado no *SGML*, e foi criado para documentos via *WEB*. São transportados pelo protocolo *HTTP*²⁰, não permite criação de rótulos. Ele objetiva formatar a estruturação dos documentos digitais, (Almeida, 2003).

Esse mesmo autor aborda a respeito do *XML* (eXtensible Markup Language), que permite criar rótulos, e fornece hiperligações compatíveis com *HTML*. Ele objetiva descrever o objeto digital.

Nota-se nos cursos de Ciência da Informação, e na literatura da área, o surgimento de uma nova ferramenta para auxílio na organização da informação, nos contextos digitais, a ontologia. De acordo com Almeida (2003), essa ferramenta pode ser definida como aquela para uso nas tarefas de organização da informação, como um catálogo de diversas tipologias existentes em um certo domínio. É uma especificação formal e explícita de uma conceitualização compartilhada. As ontologias são organizadas criando taxonomias, sobre conceitos de um domínio, com relações e interações entre os conceitos. Elas possuem um vocabulário de termos e significados que podem ser informais, com o uso da linguagem natural; semi-informais com linguagem natural restrita e semi-formais, com uso de linguagens artificiais, chamadas de semântica formal. Em síntese, a ontologia permite a formalização, o compartilhamento e a definição dos conceitos de um domínio de conhecimento.

²⁰ *HTTP* significa *Hypertext Transfer Protocol* = Protocolo de Transferência de Hipertexto. Conjunto de regras para intercâmbio de informações de textos, imagens e sons.

Faz-se necessário esclarecer também, a semelhança entre ontologia e tesouro, uma vez que estes operam num mesmo nível epistemológico, e mesmo objetivo, ou seja, a ontologia é voltada para o registro de conceitos de um certo domínio automatizado, e o tesouro serve para aprimorar a comunicação entre o usuário e as linguagens de informação. Percebe-se que ambos se preocupam com a estrutura das formas, em que a ontologia se preocupa com as formas da descrição de seus elementos, e o tesouro com as formas de descrição do conhecimento.

Para que a ferramenta ontologia possa ser aplicada digitalmente, faz-se necessário conhecer o que é ciberespaço, e hipertextos, palavras que aparecem sempre na literatura ligada aos contextos digitais. Assim, Monteiro (2003) discute filosoficamente sobre as obras no ciberespaço, e como nesta monografia o enfoque é mais técnico, será colocado de forma sintética o parecer final da autora, que representa a parte mais prática. Ela começa definindo a palavra virtual, dentro do contexto da informação, como: "o virtual é o principal atributo do ciberespaço e que melhor o descreve". Depois, afirma que no ciberespaço, há uma desmaterialização das obras e desterritorialização dos signos. A desterritorialização pode ser explicada como movimento, aquilo que não se fixa em um território, vetorizado. E a desmaterialização, pode ser vista como desvinculo das formas simbólicas da organização clássica do conhecimento, do livro impresso, da obra.

Diante disso, a autora descarta a classificação e a catalogação para organizar o conhecimento no ciberespaço, imperando o modelo rizomático, ou seja, modelo que prevê o princípio da heterogeneidade e multiplicidade. Estes princípios representam, respectivamente, conhecimento que se desterritorializa e interpenetram, produzindo novas abordagens e conceitos metodológicos, que não podem ser reduzidos a uma unidade. O rizoma é definido como uma árvore, com imagem do mundo-raiz e árvore-mundo, que pode ser acessada de inúmeros pontos. Um rizoma não começa nem conclui, ele se encontra sempre no meio,

entre as coisas, inter-ser. É um modelo arbóreo de organização do conhecimento, apresentando as taxonomias e classificações das ciências, é como se fosse um caule com crescimento contínuo. Esse modelo vem acabar com o modelo linear, tal como o hierárquico, das classificações, e dos tesouros.

Assim, ainda de acordo com Monteiro, não se pode confinar mais a um fechamento físico de uma forma, tal como o registro material. Pois, com o auxílio do *link*, no hipertexto, faz-se ligação do contexto, com os conteúdos através de "nós" do ciberespaço, representando os conhecimentos, o que vem mostrar que ciberespaço e hipertexto caminham juntos.

Para Monteiro, o hipertexto é o produto do ciberespaço, sendo diferente de texto comum, impresso e estático, pois graças à virtualidade da linguagem, e do meio, produz-se signos que geram outros signos, com movimentos, que têm a tarefa de produzir enunciados e sentido.

Também quanto ao *hipertexto*, Levacov (1997) considera que o ciberespaço necessita de *mudanças no tratamento da informação*, deixando para trás os bancos dos dados hierárquicos para os hipertextuais, que possam acompanhar as constantes mudanças e o dinamismo exigido. O hipertexto é uma alternativa técnico-evolutiva do tratamento da informação, com formas *naturais associativas*.

Já, Ihadjadene, Bouché e Kuramoto (1998) levantam uma série de problemas, como os especificados a seguir. Eles afirmam que um sistema em hipertexto é um conjunto de textos, interligados por elos, cujo funcionamento ou acesso se dá via operadores booleanos, e através da navegação, viabilizada através de vocabulário controlado ou de referência bibliográfica. Mas, de acordo com esses autores, nos sistemas atuais de recuperação de informação, a recuperação funciona centrada somente nas referências bibliográficas. O ideal seria um sistema de recuperação, sob a forma hipertextual, conforme alguns estudos já feitos, cuja navegação é

centralizada nos vocabulários controlados, sem muito esforço mental para decifrá-los.

Ainda de acordo com esses autores, catálogos em linha, cujo acesso se dá através dos índices de classificação, apresentam dois problemas: cada catálogo apresenta forma diferente de acesso; exige um prévio conhecimento do usuário quanto ao índice exato do seu ramo de interesse. E assim, estudos têm testado o uso das classificações como ferramenta de busca e navegação no campo de assunto. Os autores acham que é possível melhorar esse tipo de ferramenta como nos OPACs²¹, após o desenvolvimento do formato USMARC.

Eles citam a classificação de DEWEY, como forma ideal para o hipertexto, devido à estrutura da classificação. As dez classes e as subclasses são os nós, e os elos, as relações entre uma classe e subclasse, permitindo que o usuário navegue de uma classe a uma subclasse. A CDU permite a apresentação dos termos ordenados tal como os índices. Já os tesouros apresentam ordem alfabética, hierárquica e gráfica.

Segundo esses autores, no hipertexto, se um usuário procurar por determinado cabeçalho, ele irá localizar somente se souber da existência do mesmo. Assim, é desejo dos bibliotecários colocar em linha a lista dos assuntos e suas relações, ou seja, seus elos de sinonímia e hierarquia, como auxílio aos usuários, de modo que os cabeçalhos de assuntos representem os nós, e relações entre os cabeçalhos os elos.

Ainda com relação ao hipertexto, Barreto (2004) pensa que o HTML é "um sistema de informação, que fornece *network semântico* não linear, e de múltiplos caminhos, e experiências da informação". E para que haja hipertexto, é necessário

²¹ OPACs significa On-line Public Access Catalogs e USMARC é formato para entrada de dados bibliográficos nos Estados Unidos.

um ambiente navegacional, que crie espaços de informação multidimensionais e sem fronteiras.

Para o autor, o hipertexto é um entrelaçamento de informação, conhecimento, labirintos, espelhos, atalhos e desvios em direção à construção individualizada de conhecimento. O hipertexto trata de uma maneira diferente de leitura, que permite a interatividade, a navegação contígua sem limites de espaço físico. É uma linguagem padrão para os documentos em sistemas WEB (www) que usam o HTTP.

Nota-se que o hipertexto forma associações via *links*, entre as informações ou entre documentos. E a estrutura que resulta disso, é chamada de rede de conhecimentos. A rede e os *links* permitem a flexibilidade e o entrelaçamento de textos, com outros documentos, combinados com multimídia, para formar uma fusão de muitas mídias.

Assim, o hipertexto é um sistema de representação de informação, que apresenta um quadro semântico, não linear e de vários caminhos, e várias opções de informação.

Porém, Ihadjadene, Bouché e Kuramoto (1998) acreditam que hipertexto apresenta uma série de problemas, que desaconselharia a sua aplicação, sobretudo, sem um estudo bem elaborado, pois, a maioria dos catálogos dessa natureza mostra apenas as relações "ver" e "ver também", não mostrando as relações BT, NT, RT (Broader Term = Termo geral, Narrowed Term = Termo específico, Related Term = Termo relacionado). Ao se usarem todas essas relações, surgem problemas tais como sobrecarga de informação vertical devido aos subcabeçalhos. Esses autores acham que a solução para isso seria reagrupar as subdivisões de forma e de assunto em categorias gerais, representando as diferentes facetas do cabeçalho de assunto.

Outro problema, também apontado por eles, é a sobrecarga de informação horizontal, devido ao tamanho dos cabeçalhos e à complexidade dos mesmos. Eles apontam, como solução, a redução das subdivisões, diminuindo o tamanho do cabeçalho de assunto, preferindo o uso de cabeçalhos menos longos. Outro problema seria a remissão, ou seja, cabeçalhos que não remetem a nada, que não apontam a nenhuma referência bibliográfica, e isso é bastante problemático, pois aumenta o tamanho da base, confunde o usuário, e não apresenta uma real utilização. Elas afirmam que o ideal seria eliminá-los.

Assim, os autores chegam à conclusão, que por enquanto, não seria aconselhável a navegação através das relações hierárquicas ou de associações. Seria ideal reorganizar as estruturas tal como a *LCSH*, para que se tenha uma melhor navegação. Também concluem que nem todo vocabulário controlado pode ser representado hipertextualmente. Os melhores são: lista de cabeçalho de assunto, integrada com um sistema de classificação, e enriquecida com a linguagem natural.

Considera-se que, após essa revisão, ficaram esclarecidos questionamentos feitos anteriormente, quanto ao uso dos mesmos tipos de instrumentos, tanto no contexto tradicional, quanto no digital, como também a necessidade de um controle de vocabulário, em ambos os contextos. Percebe-se que, nos contextos digitais houve apenas uma redefinição das maneiras de se trabalhar e organizar a informação. Ficou claro, também, que com metadados, pode-se usar *CDU*, *CDD*, *LCSH*, tesouro, ou linguagem natural, "velhos conhecidos" de todo indexador. Nas ontologias, percebe-se a taxonomia, conceitos de um domínio com relações e interações, sendo necessário o uso de instrumentos de linguagens natural ou artificial. A taxonomia também faz parte da organização da informação, ao longo dos tempos. Assim, os instrumentos são os mesmos nos dois contextos, e

percebe-se que há, de um modo geral, controle de vocabulário, com ou sem a presença de novas tecnologias.

Acredita-se que na tradução de conceitos não se deve abandonar o tradicional, e as formas hierárquicas, pois, pode-se afirmar que não se trata de um rompimento, e sim de novas maneiras de se trabalhar, no contexto digital. Ontologias e metadados são padrões e linguagens para serem utilizadas no trabalho da indexação digital, que os profissionais da informação devem aprender a usar, em integração com os instrumentos tradicionais. Nota-se que os textos referentes a essas novas tecnologias usam a padronização de dados, e o intercâmbio de dados bibliográficos. E tudo isso já não é usado há algum tempo? Usam-se o padrão MARC, há algum tempo, que na verdade é um metadado, usam-se as ontologias, cujas estruturas são taxonômicas, parecidas com as estruturas dos códigos de classificação já utilizados, conforme citado anteriormente por alguns autores.

Assim, pode-se concluir que o tradicional é bastante utilizado. A proposta não é romper com o tradicional, e sim aprimorá-lo.

6 CONSIDERAÇÕES FINAIS

Após, finalizado este estudo, é necessário se voltar ao início deste texto, mais precisamente, aos pressupostos e objetivos, para verificar o alcance dos mesmos. Os dois pressupostos que direcionaram o estudo foram: a indexação e a análise de assunto, no contexto digital, podem ter semelhanças e diferenças daquelas do contexto tradicional; e os fatores que determinam semelhanças ou diferenças são a política vigente, os objetivos e a tipologia da instituição.

Pode-se afirmar, com certeza, que os pressupostos foram alcançados, no sentido de se poder conhecer a indexação e a análise de assunto e suas fases, nos contextos tradicional e digital, bem como conhecer quais são os fatores que interferem no processo de indexação. É possível afirmar que há mais semelhanças do que diferenças, no processo de tratamento da informação, no que diz respeito à indexação e à análise de assunto, nos dois contextos estudados, mesmo com metadados e ontologias, pois o princípio e a filosofia são os mesmos. Como também, pode-se afirmar que os fatores que determinam as semelhanças e diferenças citadas, nos dois contextos, são as políticas, os objetivos e a tipologia da biblioteca.

O estudo teve, como objetivo geral: identificar as semelhanças e diferenças nos processos de tratamento da informação, no que diz respeito à indexação e à análise de assunto, nos contextos tradicional e digital. Assim, pode-se considerar que as semelhanças percebidas estão no uso das ferramentas tradicionais e maneiras de se analisar e extrair os assuntos dos documentos. As diferenças estão no uso de novos instrumentos e metodologias, adaptados a um contexto moderno, mais condizente com as novas tecnologias. Conclui-se que o que ocorre, na realidade, é a interação entre o velho e novo, ou seja, processos antigos convivendo com processos tecnológicos. Ambos estão voltados para um

objetivo maior, que é a organização da informação para sua futura recuperação pelo usuário.

Quanto aos objetivos específicos: averiguar, conhecer a análise de assunto, de identificar as etapas do processo de indexação e de mostrar a importância da política de indexação, nos dois contextos, consideram-se alcançados também, embora o estudo não tenha sido exaustivo, embora tenha procurado conhecer um pouco do histórico do processo de indexação, dos conceitos de biblioteca tradicional e digital e, todo o processo da análise de assunto, nos domínios tradicional e digital.

Conclui-se que a de análise de assunto, primeira etapa do processo de indexação, no contexto digital, passa pelas mesmas etapas do contexto tradicional: a leitura do texto, extração dos conceitos e a atenção. E o fator que irá diminuir a subjetividade é o mesmo nos dois contextos, ou seja, o uso de um vocabulário controlado e da sintaxe, para mostrar as relações existentes entre os termos indexados.

Nessa etapa, a leitura para fins de indexação é de suma importância. O estudo mostra que o indexador, visto como leitor, interage com o texto lido, tendo como suporte seus conhecimentos de mundo, lingüísticos e profissionais, ou seja, o indexador passa, para o texto lido, suas experiências, vivências, ansiedades, expectativas e conhecimentos adquiridos. Assim, diferentes leitores apresentam formas variadas de interpretar e indexar um mesmo documento. Percebe-se que as dificuldades advindas do processo de análise de assunto são basicamente reflexos da interação entre o texto, o leitor e o contexto, na leitura para fins de indexação. Desse modo, pode-se concluir que o indexador deverá ser o mais imparcial possível, ao analisar o assunto de um documento.

Considera-se também que a representação dos termos indexadores, segunda etapa da indexação, no contexto digital, passa pelas mesmas etapas e processos da indexação tradicional. As bibliotecas digitais necessitam do indexador tanto na primeira, como na segunda etapa do processo de indexação, fazendo uso de conhecimentos e instrumentos clássicos, tais como listas de cabeçalhos de assunto, tesouros, códigos de classificação, ou de novos instrumentos tais, como metadados, ontologias, XML, e outros que por ventura possam surgir. Verifica-se que esses últimos citados são formatos que irão facilitar e estruturar melhor a recuperação da informação, principalmente os metadados.

Assim, indexadores devem ter os mesmos cuidados nos dois contextos, quanto a saber definir qual conceito extraído que melhor representará o documento, bem como simular buscas para medir o grau de precisão e eficiência da recuperação da informação no sistema, no sentido de verificar o grau de satisfação dos usuários.

Além desses cuidados, espera-se que os profissionais da informação, principalmente os indexadores, se preocupem com os novos instrumentos de trabalho, procurando estudar e usar os conhecimentos tradicionais adquiridos, com sabedoria, para adaptá-los à nova realidade, tal como o estudo dos autores Ihadjadene, Bouché e Kuramoto. Embora esse estudo não seja atual, ele é bem técnico e eficiente, não tendo sido encontradas publicações mais recentes sobre o tema, cujo enfoque mostre o uso dos instrumentos antigos, e as devidas alterações necessárias a serem feitas para adequá-los à navegação nos sistemas de hipertextos.

A grande vantagem percebida no final da revisão, é de verificar que o uso de metadados e ontologias pode assegurar e otimizar a recuperação final de textos completos e gráficos, principalmente as figuras em movimento. Podem facilitar o intercâmbio de dados bibliográficos, com recuperação e acesso da melhor

qualidade, por meio dos WEBSITES, e isto é o diferencial do momento, na indexação de assunto.

Fica evidente ainda que as relações hierárquicas não são aconselháveis para o hipertexto, e a navegação; e que nem todo vocabulário controlado pode ser representado sob a forma de hipertexto. Seria necessário organizar uma estrutura parecida com a LCSH. Percebe-se que os melhores modelos de organização do conhecimento digital são aqueles com estrutura integrada a uma lista de cabeçalho de assunto, a um sistema de classificação, e à linguagem natural. Alguns autores chegam a descartar a classificação e a catalogação como ferramentas de organização do conhecimento, citando o modelo rizomático, como ideal. Outros preferem descartar os bancos de dados hierárquicos, em troca dos modelos hipertextuais, com alternativa técnico-evolutiva e formas naturais associativas. Já outros apontam que o ideal seria um modelo centrado no vocabulário controlado, com ontologias e metadados compartilhados.

Enfim, o estudo demonstra que a indexação não é um processo mecânico e padronizado trata-se de um dinâmico e individualizado, de modo que diferentes documentos exigem análises e procedimentos diferenciados.

Logo, pode-se considerar que a teoria clássica continua a predominar, ficando então como proposta, a indicação de futuros estudos e pesquisas que envolvam novas tecnologias para a indexação, no domínio digital. Essa proposta está fundamentada na escassez, de literatura brasileira a este respeito. Alerta-se também para a constatação de que somente com muito estudo, e pesquisas a Ciência da Informação avançará rumo a um futuro mais promissor.

REFERÊNCIAS

- ABNT. Métodos para análise de documentos – seleção de termos de indexação – NBR 12676. Rio de Janeiro, 1992. apud FUJITA, Mariângela Spotti Lopes, RUBI, Milena Polsinelli. Um modelo de leitura documentária para a indexação de artigos científicos : princípios de elaboração e uso para a formação de indexadores. **Ciência da Informação**, Brasília, v. 7, n.3, jun/ 2006. Disponível em:<DataGramZero>. Acesso em: 10 maio 2006.
- ALBRETCHTSEN, H. Subject analysis and indexing: from automated indexing to domain analysis. *The Indexer*, London, v.18, n.4, p.219-24, 1993. apud FUJITA, Mariângela Spotti Lopes. A identificação de conceitos no processo de análise de assunto para indexação. **Revista Digital de Biblioteconomia e Ciência da Informação**, Campinas, v.1, n.1, p. 60-90, jul./dez. 2003.
- ALENCAR, Anderson Fernandes de. Bibliotecas digitais : uma nova aproximação. **Informação & Sociedade** : Estudos, Universidade Federal da Paraíba, v.14, n.1, 2004. Disponível em:<<http://www.informacaoesociedade.ufpb.br/>> Acesso em: 12-05-2006.
- ALCAIDE, Gabriel Santos. et al. **Análise comparativa e de consistência entre representações e manual de informações documentárias**. 2000. 61 f. (Monografia de conclusão de curso TCC- Biblioteconomia). Fundação Escola de Sociologia e Política de São Paulo, FESP. São Paulo, 2000
- ALMEIDA, Maurício Barcellos. Uma visão geral sobre ontologias: pesquisa sobre definições, tipos, aplicações, métodos de avaliação e de construção. **Ciência da Informação**, Brasília, v.32, n.3, p. 3-20, set/dez. 2003.

LEI DO DIREITO AUTORAL
Todos os direitos reservados e protegidos
pela Lei 9.610/1998.
Este trabalho não pode ser reproduzido ou
transmitido em quaisquer meios
eletrônicos, mecânicos,
fotográficos ou quaisquer outros.

ALVARENGA, Lídia. A teoria do conceito revisitada em conexão com ontologias e metadados no contexto das bibliotecas tradicionais e digitais. **Ciência da Informação**, Brasília, v.2, n.6, dez. 2001. Disponível em: <DataGramZero>. Acesso em: 10 maio 2006.

BARRETO, Aldo de Albuquerque. As palavras voam, a escrita permanece: a aventura do hipertexto. **Ciência da Informação**, Brasília, v.5, n.5, out. 2004. Disponível em: <DataGramZero>. Acesso em: 27 setembro 2006.

CAMPOS, Astério Tavares. A indexação. **Revista de Biblioteconomia de Brasília**, Brasília, v.15, n.1, p. 69-72. 1987.

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, M. A. N.; PINTO, M. C. F. Análise de assunto. **Revista de Biblioteconomia de Brasília**, v.8, n.1, p. 32-43, jan/jun. 1980.

CUNHA, Murilo Bastos da. Construindo o futuro : a biblioteca universitária brasileira em 2010. **Ciência da Informação**, Brasília, v.29, n.1, p. 71-89, jan/abr. 2000.

DAHLBERG, Ingetraut. Teoria do conceito. **Ciência da Informação**, Brasília, v.7, n.2, p. 101-107, 1987.

DIAS, Eduardo Wense. Análise de assunto: percepção do usuário quanto ao conteúdo de documentos. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 9, n.2, p. 146-157, jul./dez. 2004.

DIAS, Eduardo Wense. Contexto digital e tratamento da informação. **Revista de Ciência da Informação**, Brasília, v.2, n.5, out. 2001. Disponível em: <DataGramZero>. Acesso em: 10 maio 2006.

DIAS, Eduardo Wense; NAVES, Madalena Martins Lopes; MOURA, Maria Aparecida. O usuário-pesquisador e a análise de assunto. **Perspectivas em Ciência da Informação**, Belo Horizonte, v.6, n.2, p. 205-221, jul./dez. 2001.

DRABENSTOTT, Karen M.; BURMAN, Celeste M. Revisão analítica da biblioteca do futuro. **Ciência da Informação**, Brasília, v.26, n.2, maio/ago. 1997.

FERNANDES, Joliza Chagas; PINTO, Virgínia Bentes; AZEVEDO NETTO, Carlos Xavier de Azevedo. Representação indexal na WEB : estudo do sintagma "história da Paraíba" nos Sites Alta Vista e Gooble. **Informação & Sociedade : Estudos, Paraíba**, v.15, n.2. 2005. < Disponível em: <http://www.informacaoesociedade.ufpb.br/pdf/IS1520508.pdf>>. Acesso e: 02-05-2006.

FERREIRA, Aurélio Buarque de Holanda. **Novo Aurélio Século XXI** : o dicionário da língua portuguesa. 3. ed. Rio de Janeiro : Nova Fronteira, 1999.

FIDEL, R. User-entered indexing. **JASIS**, v.45, n.8. p.572-576, 1994 apud 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.

FUJITA, Mariângela Spotti Lopes. A biblioteca digital no contexto da gestão de bibliotecas universitárias : análise de aspectos conceituais e evolutivos para a organização da informação. In: ENCONTRO NACIONAL DE CIÊNCIA DA INFORMAÇÃO, 6. , 2005, Salvador, Bahia. **Proceedings**. Salvador, 2005.

LEI DO DIREITO AUTORAL
 Todos os direitos reservados e protegidos
 pela Lei 9.610/1998.
 Este arquivo não pode ser reproduzido ou
 transmitido sem que sejam tomados os meios
 legais: eletrônicos, mecânicos,
 fotográficos ou quaisquer outros.

FUJITA, Mariângela Spotti Lopes. A leitura documentária na perspectiva de suas variáveis : leitor texto-contexto. **Ciência da Informação**, Brasília, v.5, n.4, ago. 2004. Disponível em: <<http://DataGramZero>>. Acesso em 10-05-2006

FUJITA, Mariângela Spotti Lopes. A identificação de conceitos no processo de análise de assunto para indexação. **Revista Digital de Biblioteconomia e Ciência da Informação**, Campinas, v.1, n.1, p. 60-90, jul./dez. 2003.

FUJITA, Mariângela Spotti Lopes, NARDI, Maria Isabel Aspeti, SANTOS, Silvana. A leitura em análise documentária. **Transinformação**, v. 10, n.3, p.13-31, set./dez.1998.

FUJITA, Mariângela Spotti Lopes, RUBI, Milena Polsinelli. O ensino de procedimentos de política de indexação na perspectiva do conhecimento organizacional: uma proposta de programa para a educação à distância. **Perspectivas em Ciência da Informação**, Belo Horizonte, v.11, n.1, p. 48-66, jan/abr. 2006a

FUJITA, Mariângela Spotti Lopes, RUBI, Milena Polsinelli. Um modelo de leitura documentária para a indexação de artigos científicos : princípios de elaboração e uso para a formação de indexadores. **Ciência da Informação**, Brasília, v. 7, n.3, jun. 2006b. Disponível em: <DataGramZero>. Acesso em: 10 maio 2006.

GARDIN, J.C. et al. La logique du plausible: essais d'epistemologie pratique. Paris: Maison de Sciences de L'Homme, 1981. apud SILVA, Maria dos Remédios da;

FUJITA, Mariângela Spotti Lopes. A prática de indexação : análise da evolução de tendências teóricas e metodológicas. **Transinformação**, Campinas, v.16, n.2, p. 133-161, maio/ago.2004.

GONÇALVES, Antonio Cláudio Brasil. Os novos paradigmas da imagem em movimento : em busca de metalinguagens de representação para bases virtuais visando a recuperação de conteúdo semântico. **Ciência da Informação, Brasília**, v.3, n.1, fev. 2002.

HISTÓRIA das bibliotecas: In: **Enciclopédia Wikipedia**. Disponível em <<http://wikipedia.org/wiki/hist>>. Acesso em 16-05-2006.

IHADJADENE, Madjid; BOUCHÉ, Richar; KURAMOTO, Hélio. Navegação nos vocabulários controlados. **Revista de Biblioteconomia de Brasília**, Brasília, v22, n.2, p.197-210, jul./dez. 1998.

LANCASTER, F. Wilfrid (Frederick Wildrid). **Indexação e resumos** : teoria e prática. Brasília : Briquet de Lemos/Livros. 1993. 347 p.

LEVACOV, Marília. Bibliotecas virtuais : ® revolução? **Ciência da Informação**, Brasília, v.26, n.2, maio/ago. 1997. Disponível em <http://dici.ibict.br/archive/00000628/> Acesso em: 10-05-2006.

MARTINS, Robson Dias. Perspectivas para uma biblioteca no futuro : utopia ou realidade? **Revista Informativa On-Line**, Rio de Janeiro: Universidade Estácio de Sá, 2006. Disponível em: <http://biblioteca.estacio.br/artigos/005.htm> Acesso em:23-05-2006.

MEDEIROS, Marisa Brasher Basilo. Terminologia brasileira em ciência da informação. **Ciência da Informação**, Brasília, v.15, n.2, p. 135-142, jul./dez. 1986.

MILANESI, Luís. **Biblioteca**. São Paulo : Ateliê, 2002. 116 p.

MONTEIRO, Silvana Drumond. A organização virtual do conhecimento no ciberespaço. **Ciência da Informação**, Brasília, v.4, n.6, dez./2003. Disponível em:< <http://DataGramZero>>. Acesso em 19-09-2006.

MORIGI, Valdir José; SOUTO, Luzane Ruscher. Entre o passado e o presente: as visões no mundo contemporâneo. **Revista ACB: Biblioteconomia em Santa Catarina**, Florianópolis. V.10, n.2, p. 189-206, jan./dez. 2005. Disponível em: <http://www.cbsc.org.br/revista/ojs/viewarticle.php?id=89> Acesso 22-05-2006.

NAVES, Madalena Martins Lopes. Análise de assunto: concepções. **Revista de Biblioteconomia de Brasília**, Brasília, v.20, n.2, p. 215-226, jul./dez. 1996.

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. 189-203, jul./dez. 2001.

NAVES, Madalena Martins Lopes. **Fatores interferentes no processo de análise de assunto**: estudos de caso de indexadores. 2000. 275 f. Tese (doutorado) – Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2000.

NEVES, Dulce Amélia de Brito. **Aspectos metacognitivos na leitura do indexador**. Belo Horizonte, 2004. 131 f. Tese (doutorado) – Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2004.

OLIVEIRA, Rosa Maria Vivona Bertolini. WEB semântica: novo desafio para os profissionais da informação. In: SEMINÁRIO NACIONAL DE BIBLIOTECAS UNIVERSITÁRIAS, 12., 2002, Pernambuco: UFPE. Disponível em <<http://www.sibi.ufri.br/snbu/snbu2002/main.htm>> Acesso em: 01-06-2006.

PINTO, Maria Cristina Melo 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. 169-186, set. 1985.

PINTO, Virginia 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.

RIBEIRO, Ilda Campos, MENDES, Laurinda. **Biblioteca digital**. Curso de Especialização em Ciências Documentais, Universidade de Beira Interior, Covilhã. Disponível em <http://www.di.ubi.pt~api/digital_library.pdf> Acesso em 18-09-2006.

ROBRETO, Jaime, CUNHA, Murilo Bastos da.. Aplicação de técnicas infométricas para identificar a abrangência do léxico básico que caracteriza os processos de indexação e recuperação da informação. **Ciência da Informação**, Brasília, v.27, n.1, p.11-27, jan./abr. 1998.

SILVA, Maria dos Remédios da; FUJITA, Mariângela Spotti Lopes. A prática de indexação : análise da evolução de tendências teóricas e metodológicas. **TransInformação**, Campinas, v.16, n.2, p. 133-161, maio/ago.2004.

SILVA, Marilée Patta e. **Análise de assunto auxiliada por um sistema de informação, em textos da computação**. 2004. 90 f. Dissertação (mestrado) – Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2004.

SILVA, Neusa C. SÁ, Nysia; FURTADO, Sandra R.S. Bibliotecas digitais : do conceito às práticas. In: SIMPÓSIO INTERNACIONAL DE BIBLIOTECAS DIGITAIS, 2., 2004, Campinas. *Anais...* Campinas : UNICAMP. Disponível em <<http://libdigi.unicamp.br/document/?code=8304>>. Acesso em: 12-05-2006.

SOUZA, Renato Rocha, ALVARENGA, Lídia. A WEB semântica e suas contribuições para a ciência da informação. **Ciência da Informação**, Brasília, v.33, n.1, p.132-141, ago. 2004.

SOUZA, Terezinha Batista de. Metadados : catalogando dados na internet. **Transinformação**, Campinas, v.9, n.2. p.93-105 maio/ago. 1997.

STREHL, Leticia. Avaliação da consistência da indexação realizada em uma biblioteca universitária de artes. **Ciência da Informação**, Brasília, v.27, n.3, set. 1998.

UNISIST. Princípios de indexação. **Revista da Escola de Biblioteconomia da UFMG**, Belo Horizonte, v. 10, n.2, p. 83-94, mar. 1981.

VICKERY, B.C. Analysis of iformation. In: KENT, A.; LANCOUR, H. (Ed.) Encyclopedia of Library and information science. New York : Decker, 1968. v.1 p. 355-384. apud FUJITA, Mariangela Spotti Lopes. A identificação de conceitos no processo de análise de assunto para indexação. **Revista Digital de Biblioteconomia e Ciência da Informação**, Campinas, v.1, n.1, p. 60-90, jul./dez. 2003.

VIEIRA, Simone Bastos. Análise comparativa entre indexação automática e manual da literatura brasileira de ciência da informação. **Revista de Biblioteconomia de Brasília**, Brasília, v.16, n.1, p. 83-94, jan./jun. 1988.

DIREITO AUTORAL
Todos os direitos reservados e protegidos
pela Lei nº 9.610/1998.
Este arquivo não pode ser reproduzido ou
transmitido sejam quais forem os meios
empregados: eletrônicos, mecânicos,
fotográficos ou quaisquer outros.

LEI DO DIREITO AUTORAL

Todos os direitos reservados e protegidos
pela Lei nº 5.130, de 1966.

Este arquivo não pode ser reproduzido ou
transmitido sob qualquer forma, eletrônica,
mecânica, fotográfica ou qualquer outros.