

INDEXAÇÃO DOCUMENTÁRIA: UMA FORMA DE REPRESENTAÇÃO DO CONHECIMENTO REGISTRADO¹

Virgínia Bentes Pinto²

Resumo

A indexação documentária é uma das áreas da Ciência da Informação que mais evoluiu nestes últimos anos. Esta evolução está estritamente ligada às mudanças de paradigmas que ocorrem na nossa sociedade, independente do domínio do conhecimento. Ela faz parte de um sistema global: o Sistema de Recuperação de Informação (SRI) o qual é constituído por um conjunto de atividades que contemplam desde o processo de seleção e aquisição até a recuperação da informação ou de documentos nas Unidades de Documentação³. A indexação tem por objetivo teórico expressar de maneira o mais fiel possível, a representação dos elementos que pertencem ao conteúdo de um documento - seja ele real ou virtual- a fim de que o mesmo possa ser recuperado posteriormente. Por se ocupar do tratamento da informação visando a representação mínima do conteúdo do documento, esta atividade possui uma forte carga cognitiva. Neste estudo é apresentado o estado da arte da indexação documentária, os seus fundamentos teóricos e a indexação manual.

Palavras-chave: Indexação documentária; indexação manual; sociedade da informação.

Abstract

One domain of Information Science that has most developed in the last 30 years is the documentary indexing. This evolution is strictly linked to the changes of paradigms

that happen in our society, independent of the domain of the knowledge. This article is concerning the new changes of indexing and it is the first part of a sequence of two papers concerning indexing.

Key words: Indexing documentary; indexing manual; society of the information.

1 INTRODUÇÃO

O desenvolvimento científico e tecnológico favorece o aumento da produção do conhecimento, de um lado, e a sua fragmentação de outro, em função do aparecimento de novos campos do saber. Essa fragmentação não implica em uma individualização das ciências, muito menos da tecnologia, muito pelo contrário, ambas buscam apoio intra e/ou entre elas e em outros domínios do conhecimento, a fim de serem melhor compreendidas nesse novo paradigma da sociedade dita da informação (Information Society), do conhecimento (Knowledge Society) ou do “aprendizado” (Learning Society).

Neste contexto interdisciplinar, observamos que a produção do saber se apresenta formalmente estruturada sobre várias formas, como por exemplo sobre a forma impressa (livros, periódicos, folhetos, patentes, relatórios técnicos, normas técnicas, etc.), sobre a forma de imagem (fotografias, desenhos, etc.), sobre a forma de multimídia (combina-

¹ Artigo baseado no capítulo II da tese de doutorado “La representatio des connaissances dans le contexte de la documentation technique: proposition d’un modèle d’indexation.

² Doutora em Ciência d Informação e da Comunicação-Institut de Communication et des Médias-Université Stendhal Grenoble 3-França. Professôra do Departamento de Comunicação Social e Biblioteconomia-UFC.

³ Em todo o decorrer deste trabalho nós consideramos “Unidades de Documentação” como sinônimo de Bibliotecas, Centros de Documentação e outros do gênero.

ção de textos, de imagens, de sons e outros dados). Além destas, existem ainda os conhecimentos tácitos, os materializados nos próprios produtos, como é o caso das máquinas e equipamentos, e também aqueles apresentados informalmente sobre a forma oral em feiras, exposições e outros do gênero, sem contar com o conhecimento tácito.

Ora, se a forma de apresentação do conhecimento mudou, seu suporte de estocagem também mudou e, hoje, encontramos os conhecimentos registrados tanto em suportes tradicionais como o papel, quanto em suportes eletrônicos, ópticos e magnéticos.

Esse aumento do conhecimento se traduz pela disponibilidade de uma quantidade enorme de informação, recurso que está sendo considerado como o capital mais importante de nossa sociedade, posto que ele representa um alto valor agregado para o setor produtivo industrial e empresarial. Ora, se nesta sociedade a informação tem um papel fundamental para o avanço da ciência e da tecnologia, ela tem igualmente um lugar importante na vida cotidiana dos cidadãos, que precisam estar bem informados para que possam exercer dignamente a sua cidadania.

Foi à partir destas observações que produzimos este artigo, tecendo comentários, inicialmente, a respeito do acesso a informação na sociedade do conhecimento, em seguida trabalhamos a indexação em seu estado da arte e finalmente a indexação manual como forma de representação do conhecimento registrado visando a sua recuperação posterior.

2 ACESSO À INFORMAÇÃO NA SOCIEDADE DA INFORMAÇÃO

A enorme quantidade de informações que hoje estão disponíveis favorece a que se tenha a ilusão de que:

“jamais nous n’avons été mieux informés, ce qui ne veut pas dire que nous savons traiter et intégrer ces données qui littéralement nous submergent. Trop d’information tue peut-être l’information, suscite des évasions imaginaires, des refus de savoir, et se heurte de toute façon à la « clôture informationnelle » de chacun (un organisme n’utilise qu’une infime partie des signaux qui traversent son environnement)” (BOUGNOUX, 1993, p.11).

Corroborando, Pierre LÉVY (1997, p.17), afirma que:

« nous ne disposons que de très peu d’instruments pour filtrer l’information pertinente, pour opérer des rapprochements selon des significations et des besoins toujours subjectifs, pour nous repérer dans le flux informationnel.” (LÉVY, 1997, p.24)

Estas duas reflexões mostram claramente o paradoxo desta sociedade, a qual pode ser caracterizada, entre outras, como:

- 1) uma sociedade grande produtora e consumidora de informações, portanto a sua matéria-prima é a informação;
- 2) uma sociedade que, mesmo sendo produtora e consumidora de uma quantidade enorme de informações, se depara com inúmeras dificuldades para acessar esta avalanche de informações que nos submerge quotidianamente;
- 3) uma sociedade produtora e utilizadora das ferramentas de tratamento, estocagem e recuperação da informação, propiciadas pelas Tecnologias da Informação e da Comunicação (TIC’s);
- 4) uma sociedade onde as TIC’s possibilitam, por um lado, alto alcance, predominam as lógicas das redes e a suposta flexibilidade. Por outro lado convivemos com a chamada infoexclusão e com inúmeras perdas;
- 5) uma sociedade na qual o ser humano necessita, ansiosamente, estar atualizado para poder acompanhar as transformações que se processam, quotidianamente, em uma velocidade infreável;

Dentre estas variáveis, as correspondente ao ser humano, são infinitamente atingida pelas outras, pois este sujeito vive submerso pelas informações, necessita estar atualizado para não ficar a reboque das mudanças que se processam a todo momento e, se depara com as dificuldades para acessar tais informações, mesmo com a existência das ferramentas das NTIC’s. Essas barreiras de acesso a informação são explicadas por KURAMOTO (1999), quando afirma que para navegar sobre o “espaço do saber” é preciso conhecer este espaço, saber utilizar as ferramentas das NTIC’s e igualmente conhecer as estratégias de busca e recuperação de informação, ou ainda ter a ajuda de um especialista neste domínio para que as necessidades de informação dos usuários possam ser atendidas de maneira eficaz e eficiente. Nesta perspectiva, torna-se imprescindível que saibamos o que fazer com as NTIC’s e quais as conseqüências de sua má utilização o que pode, certamente, comprometer tanto a tecnologia (produto/serviço) quanto os que estão sendo seus usuários.

Além destas variáveis, consideramos ainda; a ausência e/ou deficiência no processo de indexação. Esse fato é notório, tanto quando buscamos informações nas Unidades de Documentação, ou ainda sobre a rede Internet que, quase sempre nos fornece como respostas um calhamaço de dados que muitas vezes não condizem com o que foi demandado. Por exemplo, um dia havia necessidade de uma informação sobre a minha conta no “Banco do Brasil” em Fortaleza. Estava na França e, o acesso podia ser feito pela Internet. Com a estratégia de busca “Banco do Brasil” & “Fortaleza”, procurei no “Alta Vista” e no “Yahoo”. Tive 193 respostas compostas pelas palavras “banco”, “Brasil”, “Fortaleza” e “do”, mas nenhuma resposta com “Banco do Brasil”. Problemas desta natureza são enfrentados por inúmeros usuá-

rios da rede e, certamente, pode ter sido decorrente de uma indexação baseada em unitermos, que considera cada palavra individualmente, e não os grupos sintagmáticos, daí, o fornecimento quase sempre de respostas insatisfatórias.

Este exemplo mostra, explicitamente, a necessidade de se investir no tratamento técnico dos recursos informacionais, assim como na sua organização, de maneira mais conveniente, visando a racionalização de sua estocagem e, naturalmente, a busca e a recuperação de informação de maneira eficaz e eficiente, porque como afirma CHAUMIER(1990 p.278), « rien ne sert d'archiver un document que l'on ne saura pas trouver parce qu'il n'aura pas été indexé » ou ainda porque ele foi indexado de maneira incorreta.

Diante disto perguntamos, o que significa exatamente a atividade de indexação? Como ela deve ser efetuada? Quais são os problemas enfrentados para a sua execução?

3 INDEXAÇÃO: UMA FORMA DE REPRESENTAÇÃO DO CONHECIMENTO

Para se trabalhar a indexação, inicialmente é preciso considerar que este conceito é utilizado em vários domínios do conhecimento, por exemplo na economia, na demografia, no comércio e nas ciências da Informação. Este último é o que nos interessa.

Seguindo o raciocínio do professor Jean-Claude GARDIN (1974), consideramos a **indexação documentária**, como um conjunto de atividades que consiste em identificar, nos documentos, os seus “Traços Descritivos” (TD’s) ou macroproposições e em seguida extrair os elementos/descriptores (sintagmas) indicadores do seu conteúdo visando à sua recuperação posterior. Estes descritores vão se constituir na representação dos elementos indicadores do conteúdo do documento e não a sua representação, pois esta só pode ser pelo documento mesmo.

A representação do conhecimento registrado, tendo em vista a indexação de documentos pode ser realizada tomando-se por base os conceitos /palavras-chave/unitermos ou ainda, em uma visão mais moderna, os sintagmas nominais (proposta apresentada pelo grupo SYDO), ou frases (proposta de Alain F. Smeaton e Paraic Sheridan), ou ainda os sintagmas verbais (proposta de Geneviève Lallich e de Virginia Bentes Pinto). No caso dos conceitos palavras-chave, eles podem ser extraídos⁴ do documento mesmo ou ainda atribuídos à partir de outras fontes, como por exemplo as Linguagens Documentárias (LD’s)⁵. Em contrapartida, os sintagmas ou as frases só podem ser extraídos do documento mesmo, portanto em uma linguagem natural (LN).

A primeira forma de indexar nos parece representativa dos descritores constitutivos dos documentos, ou seja do conjunto de suas unidades lexicais. Aqui é desmontado o discurso do autor, onde as palavras tinham um sentido em função do contexto ditado por seu criador, ou seja elas eram ligadas ao mundo real do documento. Retiradas do seu contexto tais palavras ou conceitos passam a significar apenas propriedades, portanto, seu sentido vai mudar, naturalmente.

Ex. A geração das tensões de alimentação, variáveis em amplitude e Frequência.

Na indexação por palavras-chave/unitermos, teremos o conjunto seguinte: **A, geração, das, tensões, de, alimentação, variáveis, em, amplitude, e, frequência .**

Nesta maneira de indexar, as palavras são retiradas do contexto do logico-semântico onde elas tinham uma significação determinada por este contexto. Elevadas do mundo real, elas designam um conjunto de propriedades, seu sentido muda e se resume a um conjunto de unidades lexicais. Ora, segundo Michel LE GUERN (1991, p.23)

“Le mot de la langue, contrairement à ses occurrences dans le discours, n'a pas de référence extralinguistique. (...) La relation signe/objet, au sens de Peirce, correspond sommairement à la relation saussurienne signifiant/signifié; les signifiés font eux aussi partie de la structure de la langue. Pour que le descripteur remplisse sa fonction, qui est de mettre en relation un objet du monde-une entité extralinguistique-avec le document qui apportera des informations sur cet objet, il faut que le descripteur soit un signe indiciaire.(...) Les mots de la langue, en tant qu'ils sont mots de la langue, ne signifient que des propriétés, jamais des entités: ils signifient des attributs, et non des substances, tant qu'ils ne sont pas mis en oeuvre dans le discours. Le descripteur, quant à lui, signifie une entité, une substance au sens de la philosophie d'Aristote. Le descripteur ne peut donc pas être considéré, à l'instar des mots de la langue, comme un symbole sans référence.”

Ao contrário, se a indexação é realizada tendo como base os sintagmas ou as frases, os índices serão constituídos por passagens do texto portadoras de informação, neste caso pode-se ter uma representação mínima do conteúdo do documento à medida que esses grupos não são isolados do contexto no qual eles são inseridos (onde eles tem um valor referencial).

No exemplo apresentado anteriormente, se tomarmos como elementos representativos os sintagmas, iremos ter os seguintes:

⁴ O primeiro caso é chamado por LANCASTER(1991) de indexação por extração e o segundo indexação por atribuição. PAIJMANS (1993) os chama de “assigned indexing” e “derivad indexing”.

⁵ No decorrer deste trabalho, consideramos as Linguagens Documentárias (LD), como um conjunto de termos estruturados utilizados como tradutores dos elementos indicadores do conteúdo dos documentos visando a construção de índices para facilitar a recuperação da informação.

Sintagma1= A geração das tensões de alimentação, variáveis em amplitude e frequência

Sintagma2 = as tensões de alimentação, variáveis em amplitude e frequência

Sintagma3= as tensões de alimentação

Sintagma4= a amplitude e frequência

Sintagma5= a frequência

Sintagma6= a amplitude

Neste caso, podemos ter uma representação mínima do conteúdo veiculado no documento pois as suas unidades gramaticais não são retiradas de seu contexto, onde eles tinham um valor referencial. Além do mais, é preciso levar em consideração as características destes sintagmas, pois eles poderão ser portadores seja de um conjunto de informações, seja de um simples fragmento.

A representação em nível de indexação documental perpassa, ao menos, por três etapas:

- a) análise conceptual
- b) tradução
- c) controle de qualidade

Quanto à maneira de indexar, ela pode ser feita através de:

- 1) uma análise manual, chamada igualmente intelectual, que é feita pelos humanos;
- 2) uma análise mecânica, feita pelas ferramentas da informática;
- 3) uma análise que combina as duas: humana e mecânica. Esta última é chamada indexação semi-automática ou assistida pelo computador, sendo realizada da seguinte maneira: inicialmente o sistema faz uma indexação automática dos documentos levando em conta as ocorrências das palavras mais frequentes no texto. Em um segundo momento, o indexador humano refina a lista dos descritores propostos pelo sistema fazendo os ajustes e/ou complementações necessárias.

Com relação à dimensão da indexação, ela foi proposta pelo professor F.W. LANCASTER (1979), e trata-se da extensão com a qual um documento pode ser indexado, ou seja a exaustividade (exhaustivity), e a especificidade (specificity). A indexação exaustiva procura extrair do documento, o maior número de conceitos de forma a cobrir o

seu conteúdo da maneira mais completa possível. É certo que esta maneira de indexar oferece a oportunidade de acesso a um grande número de conceitos, mas ao mesmo tempo ela pode ser responsável pelo ruído⁶ durante a recuperação da informação. Segundo SOERGEL (1994), a exaustividade pode ser vista de dois aspectos: **a exaustividade de pontos de vista e a exaustividade de importância**. A primeira, assegura que as facetas ou os pontos de vista considerados úteis para a representação proposta pelas LD's serão disponíveis para a recuperação da informação. A segunda, determina o nível de importância dos descritores propostos pelas regras de indexação. Outro fator observado nesta maneira de indexar diz respeito ao seu custo-eficácia, pois quanto mais exaustiva a indexação, naturalmente, maior será o seu custo. Portanto, uma decisão neste sentido deve ser fundamentalmente justificada pelo conhecimento do perfil dos usuários e pelos objetivos da instituição (LANCASTER, 1991).

A indexação específica, como o nome o diz, leva em consideração os conceitos específicos em função dos temas tratados no documento. Esta maneira de indexar diz respeito a profundidade com a qual o conteúdo de um documento é tratado. Se de uma parte ela favorece a precisão, de outra, ela contribui para aumentar o silêncio⁷ na recuperação da informação, pois é levado em consideração apenas o conteúdo principal do documento, deixando de fora outros assuntos tratados, mesmo que de maneira não elementar.

Um outro aspecto que deve se levar em conta na atividade de indexação refere-se à definição das Unidades de Informação⁸ que poderão entrar na construção do índice.

A indexação é uma atividade que desmonta o discurso montado pelo autor do documento, à medida que ela faz recortes neste discurso. Assim, ela permite passar de um documento constituído (um documento primário) à sua reconstituição em um novo documento- índice (um documento secundário), o qual é formado não pela representação do conteúdo do documento inicial, mas pela representação dos elementos indicadores do seu conteúdo e que vão se constituir na chave de acesso a recuperação da informação.

A maneira de indexar depende, naturalmente, do tipo de documento à indexar. Se tomarmos como exemplo uma monografia, como devemos indexá-la? Analisando o conteúdo predominante no documento, ou de maneira mais fina levando em conta por exemplo os capítulos, os parágrafos e as seções? No que concerne aos periódicos e aos anais de eventos, a indexação deve se apoiar sobre os artigos, sobre as conferências e comunicações ou a partir de seus títulos?

⁶ Consideramos como barulho, o excesso de documentos propostos pelos organismos de documentação para responder a uma demanda mas na realidade não respondem ao assunto demandado. Ele é medido pela relação entre o número de documentos não pertinentes relacionados e número total de documentos propostos: $B=dn/d$, onde dn =documentos não pertinentes, d =documentos da base

⁷ Neste trabalho, o silêncio corresponde a ausência de documentos que responderiam as necessidades dos usuários, mas na realidade não foram encontrados, mesmo que façam parte da coleção. A taxa de silêncio corresponde a relação entre o número de documentos pertinentes encontrados e o número total de documentos da base= $SI=dp/d$, onde dp =documentos pertinentes, d =documentos da base

⁸ Consideramos como Unidades de Informação, os fragmentos de textos reconhecidos como unidades de indexação

Tratando-se da indexação de documentos técnicos - manuais técnicos, patentes, normas técnicas, bulas de medicamentos, laudos médicos etc., a indexação deve ser conduzida levando-se em conta os títulos, os capítulos, as passagens.....?

Os estudos e experiências, mostram que ainda não temos uma resposta precisa a estas interrogações. No entanto, o que se observa é que a indexação de monografias, de periódicos e de anais de eventos pode ser realizada tanto de maneira específica - levando-se em conta a estrutura lógica das monografias, dos artigos de periódicos e das comunicações de eventos- quanto de maneira bem geral, tomando-se como referência os títulos. Com relação a indexação de documentos técnicos, ela deve ser conduzida no sentido da especificidade, pois os usuários deste tipo de documentos demandam uma informação pontual para responder às suas necessidades que são bem específicas. Assim, talvez a solução seja indexar estes documentos levando-se em conta as suas estruturas lógicas pois, normalmente, elas refletem com mais clareza o conteúdo tratado no documento. Neste contexto a indexação pode ser realizada à partir dos capítulos, seções, parágrafos, passagens, etc.

Estas considerações mostram que não existe uma regra única para a execução da atividade de indexação, a maneira segundo a qual um documento vai ser indexado deve ser estabelecida nas políticas de indexação definidas pelas Unidades de Documentação. Estas políticas devem ser definidas em função dos objetivos e da missão destes organismos, em função do perfil de seus clientes potenciais, e deve estar contida em um manual, de maneira que os indexadores possam tomar conhecimentos das regras estabelecidas e possam segui-las. Esta decisão poderá contribuir para reduzir de certa forma a subjetividade suscitada por esta atividade.

Em que concerne a indexação de documentos audiovisuais, cujas características são o conteúdo (informação), a mídia vetor deste conteúdo e o suporte de estocagem. Isto nos traz problemas, pois neste caso a indexação demanda muito mais detalhes e muito mais informações do gênero: Quem? O quê? Como? Onde, Quando.... ? A maneira de indexar estes tipos de documentos coloca em jogo além das informações visuais, outros tipos de informações percebidas por outros órgãos sensoriais desde que o sujeito conheça o conteúdo (CHELLAPPA, 1995). Segundo a professora Johanna SMIT (1989) as dificuldades para a indexação de documentos audiovisuais resulta da tentativa de passagem da denotação (o que o documento mostra) para a conotação (o que é percebido pelo indexador). Então, como deve ser feita a indexação destes documentos? Levando-se em conta os objetos representados e suas formas? A percepção visual (cor) A cena? O acontecimento?

4 INDEXAÇÃO MANUAL

A indexação manual, chamada igualmente intelectual ou humana, como o próprio nome o diz, é realizada pelos

humanos, sejam eles bibliotecários ou especialistas do (s) domínio (s) no qual (is) as bases de dados são consagradas. Este tipo de indexação se baseia, sobretudo, no julgamento, normalmente intuitivo, dos indexadores, em função do texto e do interesse para a sua comunidade de usuários.

Para realizar essa indexação, é preciso, inicialmente, analisar o conteúdo do documento, lendo-o não do início ao fim, mas por partes, ou seja lendo suas estruturas lógicas. Por exemplo, a introdução, os capítulos, as seções, os parágrafos, a conclusão e outras passagens consideradas importantes. Esta análise pode ser estabelecida partindo-se das estruturas fornecidas pelos autores ou pelos editores de documentos ou ainda por uma segmentação proposta pelo indexador. Em resumo, ela comporta a leitura de documentos a compreensão de seu conteúdo, a identificação e a seleção de conceitos para representar os elementos indicativos deste conteúdo. Segundo FIDEL (1994), se faz necessário considerar neste contexto dois aspectos: o documento propriamente, ou seja o seu lado objetivo e a razão ou motivo pelo qual o documento poderá ser utilizado neste caso, os aspectos subjetivos. Aqui a indexação manual enfrenta um grande problema: a dificuldade de escolher os conceitos que podem melhor representar os elementos indicadores do conteúdo do documento e, a subjetividade desta escolha, o que pode explicar os desacordos freqüentes entre os indexadores humanos.

Após esta análise, passa-se a uma segunda etapa, a chamada tradução na qual os indexadores fazem uma comparação entre os conceitos pré selecionados em linguagem natural com os descritores das LD's. Se esses conceitos coincidirem com os das LD's, eles poderão ser escolhidos como representantes dos elementos que fazem parte do conteúdo do documento. Na prática nós sabemos que, se os conceitos selecionados não coincidirem com os descritores das LD's, mesmo assim os indexadores poderão os escolher. Esta decisão vai depender de seu conhecimento sobre o assunto, do seu conhecimento sobre o perfil dos usuários, da política de indexação adotada e, igualmente de sua experiência no domínio da indexação.

A tradução é uma etapa bem complexa, à medida que ela exige o seguimento de regras que foram definidas à priori, como por exemplo, os tesouros, as listas de autoridades, etc. Essas regras são consideradas como uma faca de dois gumes, pois se de um lado, elas parecem assegurar a qualidade da indexação, no que diz respeito à desambiguação das palavras, a organização e normalização dos índices. De outro, elas podem ser responsáveis pelo **silêncio** ou pelo **ruído** no momento da recuperação da informação. Assim, segundo FIDEL (1994), as experiências dos indexadores tem mostrado que quando da etapa de tradução, é necessário levar em consideração alguns questionamentos, entre os quais ela cita:

1) **as fontes dos termos de indexação:** em quais fontes de vocabulários de indexação os indexadores podem se apoi-

- ar para escolher os termos que vão compor os índices ? Existem regras que limitam o indexador aos descritores dos tesouros utilizados pelo sistema, e outras permitem que sejam utilizados os termos da língua natural;
- 2) **a precisão:** que grau de precisão o indexador pode utilizar para traduzir os conceitos em termos de indexação? Os termos selecionados para o índice devem ser tão precisos que substituam o conceito ou eles devem ter um sentido mais geral ?
Ex.: Fruticultura tropical no semi-árido: manga e caju
Este documento deve ser indexado por:
Frutas cítricas?
Manga?
Caju?
- 3) **o peso:** o peso relativo dos conceitos de um documento pode ser definido pelo indexador ? No exemplo anterior, qual conceito terá o peso maior?
- 4) **a fidelidade:** em que medida a tradução deve ser fiel ? Como ser fiel em uma tradução quando o conceito não tem um descritor correspondente? O indexador poderá usar os termos aproximados ?
Ex.: Ciência → Científico
- 5) **a linguagem do usuário:** o indexador pode designar os termos de um índice em uma linguagem mais próxima da do usuário ? Por exemplo, através dos seus perfis é possível estabelecer regras que poderão lhes guiar na escolha dos termos de indexação mais adequados aos seus ?
Ex.: Dor de cabeça → Cefaléia

Estas interrogações são de fundamental importância quando do estabelecimento das políticas de indexação que serão adotadas pelas unidades de documentação. As soluções as questões 1, 2 e 5 são mais fáceis de ser resolvidas, porque elas são ligadas à tomada de decisão operacional. Ao contrário, as soluções às questões 3 e 4 não são fáceis a resolver pois a definição do peso dos conceitos pertencentes a um documento implica em um processo subjetivo, portanto difícil a se colocar em prática. Com relação à fidelidade da tradução dos conceitos que não possuem descritores equivalentes, talvez a solução seja conservar os conceitos, pois os descritores considerados próximos certamente que não possuem o mesmo sentido.

Além destes entraves, a indexação manual apresenta outras barreiras, entre as quais destacamos:

- ⇒ custo elevado; pois ela exige pessoal especializado e demanda muito tempo. Segundo BERTRAND (1994), o tempo médio de indexação de um documento é de aproximadamente 30 minutos, podendo variar de 10 à 45 minutos, segundo as dificuldades encontradas com relação à compreensão do conteúdo do documento ;
- ⇒ fraca coerência intra e entre os indexadores; a coerência na identificação dos conceitos e na escolha dos termos de indexação, pelo mesmo indexador em

momentos diferentes de indexação, ou por outros indexadores equivale a aproximadamente 30% (CLEVELAND, 1977);

⇒ a dificuldade para escolher uma grande quantidade de conceitos, normalmente, no processo de indexação manual a escolha dos conceitos não ultrapassa de 05 (cinco). Assim, muitos assuntos tratados em um documento poderão não ser percebidos pelos indexadores, o que vai provocar um aumento do silêncio no momento da Recuperação da Informação.

Na indexação manual, observamos que, se de um lado, o indexador tem um grande poder de decisão na definição dos conceitos que representarão os elementos indicadores do conteúdo do documento, de outro lado, encontramos vários inconvenientes os quais causam problemas que são difíceis a administrar por que:

- ⇒ nem sempre o indexador é especialista no domínio do(s) que ele indexa;
- ⇒ nem sempre o indexador é especialista no domínio da indexação ;
- ⇒ existem novos domínios do conhecimento pouco cobertos ou ainda não cobertos pelas LD's especializadas;
- ⇒ as inovações terminologias que se verificam em vários domínios do conhecimento exigem uma constante atualização tanto por parte das LD's, como pelos indexadores. Conteúdo, o que se verifica é que na realidade, nem sempre os indexadores são treinados e/ou reciclados neste sentido, e muito menos as LD's são atualizadas.

Embora a atividade de indexação manual pareça objetiva e neutra, na realidade isto não é verdade, pois esta atividade consiste em um trabalho de análise e síntese e, portanto, tem uma forte carga cognitiva, à medida que ela demanda uma compreensão do conteúdo do documento, a extração dos elementos correspondentes a este conteúdo, a representação destes elementos e a sua organização em forma de índices. Sendo naturalmente uma atividade subjetiva, ela é influenciada pelos conhecimentos sobre o domínio do documento, pelas experiências do indexador, pelo conhecimento da atividade de indexação, pelo meio ambiente onde se realiza a indexação, entre outros. Assim ela não poderá ser desprovida da neutralidade, mesmo que seja desejável.

Outra observação no que diz respeito a indexação manual, é que, apesar da evolução da indexação semi-automática e automática, elas não excluíram esta prática. Assim, mesmo considerada como obsoleta para uns, a indexação manual é ainda utilizada tanto nos países do chamado primeiro mundo como nos do terceiro. Ora, se este tipo de indexação ainda é utilizada nos países grande produtores e consumidores de recursos informáticos é, principalmente, por dois motivos:

- ⇒ porque a indexação semi-automática e automática não oferecem respostas totalmente satisfatórias no momento da recuperação da informação;

⇒ porque os sistemas de indexação automática ainda não atingem 100 % das unidades de documentação desses países.

5 CONCLUSÕES

Nossa problemática, neste trabalho, foi apresentar o estado da arte da indexação, mostrando o seu conceito, assim como a sua prática.

Nós observamos que graças à interdisciplinaridade da ciência da informação, notadamente, com a informática, a estatística, a lingüística, e a psicologia cognitiva, o campo da indexação vem evoluindo bastante. Essa evolução é bem conhecida através da literatura quando observamos que inicialmente, as experiências foram calcadas em uma prática manual simples e de forma intuitiva, cujo o objetivo era fornecer um conjunto de palavras que fossem capazes de oferecer algumas pistas para o usuário encontrar o documento que ele tinha necessidade.

Salientamos ainda, que a indexação coloca em cena três atores: o autor do documento, o indexador e o usuário. Consequentemente, para cada um a noção de pertinência informacional será percebida diferentemente segundo as suas experiências, os interesses de cada um no momento da produção ou de leitura do texto. Assim a cobertura dos conceitos escolhidos para representar os elementos do conteúdo do documento, não podem certamente chegar a 100%. Pois esta indexação nem sempre é feita de maneira exaustiva, portanto, certas partes do documento não são levadas em consideração. Assim, ela vai provocar a perda de certas informações, o que contribui para aumentar o silêncio. No que concerne ao indexador e ao usuário, é desejável que a taxa de cobertura dos termos seja ótima pois isto vai influenciar a pertinência das respostas fornecidas no momento da recuperação da informação.

Assim, qualquer que seja o método de indexação utilizado, manual, semi-automático e automático, a indexação, através dos componentes do índice, deverá permitir aos clientes o acesso ao documento que contém a informação que ele necessita. Seu resultado se constituirá em um dispositivo chave entre o documento primário a ser lido e compreendido e um documento secundário (índice) a ser constituído, de maneira tal que a representação dos elementos indicadores do conteúdo do primeiro sejam encontrados no segundo documento de maneira mais completa e fiel possível. Pois é este documento índice que, durante a busca de informação, vai oferecer « pistas » para que o usuário possa decidir, sem ver o documento primário, se ele vai considerá-lo ou não, como possível para responder à sua necessidade. Portanto, a atividade de indexação visando a representação dos elementos do conteúdo de documentos é calcada em dois objetivos fundamentais:

⇒ um teórico: estabelecer os mecanismos para a elaboração dos índices ;

⇒ outro operacional: possibilitar a busca e a recuperação da informação.

Para finalizar, lembramos que a indexação documental é uma atividade que pode ser aplicada aos documentos textuais, visuais, sonoros, pictóricos, multimídia, etc.

6 REFERÊNCIAS BIBLIOGRÁFICAS

1. BENTES PINTO, V. (1999) *La representation des connaissances dans le contexte de la documentation technique: proposition d'un modèle d'indexation*. Grenoble, (Thèse de doctorat, Université Stendhal Grenoble-3).
2. BERTRAND, A. (1993) *Comprehension et categorisation dans une activité complexe: indexation de documents scientifiques*. Toulouse, (Thèse de doctorat, Université de Toulouse).
3. BOUGNOUX, D. (1993) *Sciences de l'information et de la communication*. Paris, Larousse, .
4. CHAUMIER, J. (1990) L'indexation documentaire; de l'analyse conceptuelle humaine à l'analyse automatique morphosyntaxique. *Documentaliste*, vol.27, n.6., p.275-284, nov.dec.
5. CHELLAPPA, R. (1995) Human and machine recognition of faces: a survey. *Proc. of. the IEEE*, v.83, n. 5, may.
6. FIDEL, R.. (1994) User-centered indexing. *JASIS*, v.45, n.8, p.572-576.
7. GARDIN, J. C. (1974) *Les analyses des discours*. Neuchatel, Delachaux et Nestlé.
8. KURAMOTO, H. (1999) *Proposition d'un système de recherche d'information assisté par ordinateur, avec application à la langue portugaise*. Lyon. (Thèse de Doctorat l'Université Lumière).
8. LALLICH BOIDIN, G. (1986) Analyse syntaxique automatique du français application à l'indexation automatique. Grenoble. (Thèse de doctorat Université des Sciences Sociales de Grenoble).
9. LANCASTER, F. W. (1991) *Indexing and abstracting in theory and practice*. London, Library Association.
10. _____. (1979) *Information retrieval system characteristics, testing end evolution*. New York, John Wiley, 381p. cap.1, p.1-14.
11. LE GUERN, M. (1991) Un analyseur morpho-syntaxique pour l'indexation automatique. *Le français moderne*. v.59, n.1, p.22-35.
12. LEVY, P. (1997) *L'intelligence collective: pour une anthropologie du cyperspace*. Paris, Découverte/Poche, p.21-24.
13. SOERGEL, D. (1994) Indexing and retrieval performance: the logiciel evidence. *JASIS*, v.45, n.8, p.589-599.