

MARIA NILMA RODRIGUES DA COSTA

**INDEXAÇÃO DOCUMENTAL: COMPLEXIDADE DA ESCOLHA DE  
DESCRITORES PARA A REPRESENTAÇÃO E RECUPERAÇÃO DE  
INFORMAÇÃO**

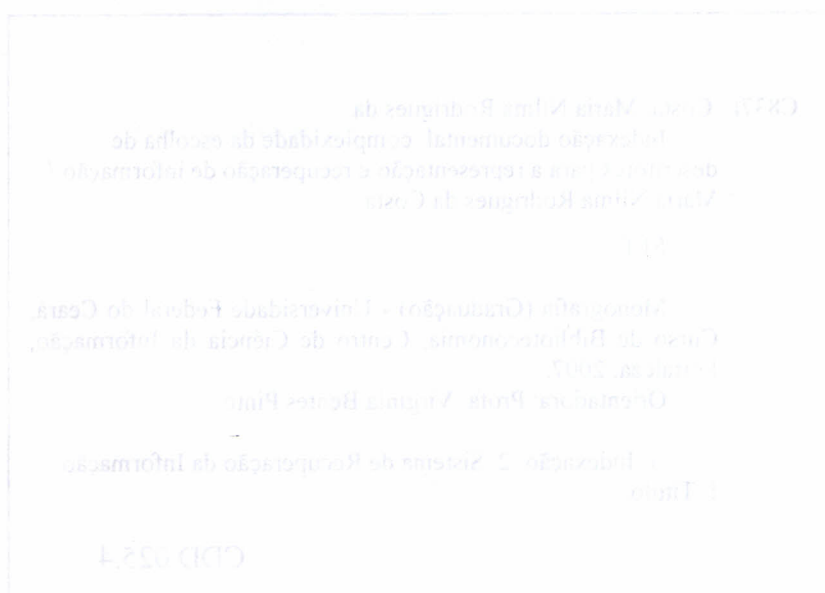
**FORTALEZA  
JULHO-2007**

**MARIA NILMA RODRIGUES DA COSTA**

**INDEXAÇÃO DOCUMENTAL: COMPLEXIDADE DA ESCOLHA DE  
DESCRITORES PARA A REPRESENTAÇÃO E RECUPERAÇÃO DE  
INFORMAÇÃO**

**Monografia apresentada ao Curso de Biblioteconomia, da  
Universidade Federal do Ceará para obtenção do grau de  
Bacharel em Biblioteconomia.**

**Orientadora: Profa. Dra. Virgínia Bentes Pinto**



**Fortaleza – CE  
2007**

## AGRADECIMENTOS

A Deus, que me deu a vida, inteligência, saúde, paz, persistência para que eu concluísse a monografia e me dá força para continuar a caminhada atrás dos meus objetivos.

À amiga Alice que tanto me apoiou, dando-me forças quando não estava com vontade para estudar, por ter paciência e me dado suporte para que eu pudesse concluir o curso e a monografia, e que me fez ver a grande oportunidade que seria entrar na UFC.

A meus pais que mesmo longe me deram apoio e uma educação que agradeço ter recebido.

À professora e orientadora Virgínia Bentes Pinto, pelos ensinamentos, colaboração, por ter disponibilizado livro e texto que muito me ajudaram e que me serviram de referência neste campo do conhecimento.

À todos os meus amigos de curso que conheci e que me acompanharam durante todo o percurso.

A todos os bibliotecários que colaboraram na construção desta monografia.

A todos que me ajudaram direta ou indiretamente nesta caminhada.

## RESUMO

Apresenta os resultados de uma pesquisa exploratória realizada junto aos bibliotecários indexadores de Bibliotecas Universitárias de Fortaleza. O objetivo geral da pesquisa é identificar as dificuldades enfrentadas pelo indexador na extração de termos em um Sistema de Recuperação da Informação (SRI), a fim de verificar as falhas que ocorrem na execução da taxa de indexação. Trata-se de uma investigação sobre as tomadas de decisões dos indexadores na extração de termos de indexação bem como o conhecimento que estes atores possuem sobre a indexação e as suas etapas. Esta preocupação se justifica ao se perceber a importância de uma indexação de qualidade ou eficaz, capaz de recuperar o que é desejado e de evitar recuperação de informações com ruído excessivo. Para a coleta de dados utilizou-se o questionário. Os resultados da pesquisa mostram que as maiores dificuldades no exercício da atividade de indexação são: subjetividade, a exaustividade, a especificidade, a falta de uma política de indexação e as linguagens. Além do mais, percebeu-se que muitos dos profissionais não têm o entendimento do que seja a indexação.

**PALAVRAS-CHAVE:** Indexação; Sistema de Recuperação da Informação.

## ABSTRACT

It presents the results of a exploratory research carried through next to the librarians of University Libraries of Fortaleza. The general objetctive of the research is to identify the difficulties faced for the they in the extraction of terms in a Information Retrieval System (IRS), in order to verify the imperfections that occur in the execution of terms of indexing as well as the knowledge that these actors possess on the indexing and its stages. This concern if justifies to if perceiving the importance of an indexation of quality or efficient, capable to recoup what it is desired and to prevent recovery of information with extreme noise. For the collection of data the questionnaire was used. The results of the research show that the biggest difficulties in the exercise of the activity of indexing are: subjectivity, the exhaustificy, the specificity, the lack of one indexation politics and the languages. In addition, one perceived that many of the professional do not have the agreement of what it is the indexing.

**KEY-WORDS:** Indexing; Information Retrieval System.

## SUMÁRIO

<b>1 INTRODUÇÃO</b> .....	07
<b>1.1 Justificativa</b> .....	08
<b>1.2 Objetivos</b> .....	08
<i>1.2.1 Objetivo Geral</i> .....	08
<i>1.2.2 Objetivos Específicos</i> .....	08
<b>1.3 Aspectos Metodológicos</b> .....	09
<b>1.4 Estrutura do trabalho</b> .....	09
<b>2 A COMPLEXIDADE NA ESCOLHA DOS DESCRITORES NO SRI</b> .....	10
<b>2.1 Sistemas de Recuperação de Informação</b> .....	10
<b>2.2 Considerações sobre a representação indexal</b> .....	16
<i>2.2.1 O que significa representação?</i> .....	16
<i>2.2.2 Indexação documentária como forma de representação</i> .....	17
2.2.2.1 A árdua tarefa na escolha das “pistas” de recuperação informação.....	25
<b>3 METODOLOGIA</b> .....	30
<b>4 INTERPRETAÇÃO DOS DADOS E DISCUSSÃO DOS RESULTADOS</b> .....	31
4.1 Concepção que os indexadores possuem sobre a atividade de indexação e suas etapas .....	32
4.2 Requisitos para o indexador fazer uma indexação adequada .....	34
4.3 Procedimentos a serem seguidos na extração de termos de um documento.....	36
4.4 Fatores que interferem no trabalho do indexador.....	37
4.5 Dificuldades encontradas na escolha de termos e em que essas dificuldades influenciam no SRI.....	40
<b>5 CONCLUSÃO</b> .....	44
<b>REFERÊNCIAS</b> .....	46
<b>APÊNDICE</b> .....	49

## 1 INTRODUÇÃO

A partir da década de 50, uma massa crítica de cientistas, engenheiros e empreendedores começaram a trabalhar um problema que começara a surgir: o aumento da produção científica e de seus registros, particularmente em ciência e tecnologia. Em razão de que os sistemas tradicionais de recuperação da informação já não davam conta de tantos documentos que, a cada dia cresciam muito mais, Calvin Mooers, ainda na década de 50, propõe alternativas e cunha o conceito de Sistema de Recuperação de Informação (SRI). Na realidade o SRI surgiu para atender às necessidades específicas dos usuários, permitindo o acesso a informações relevantes, em meio a um grande volume de documentos disponíveis.

O tema Recuperação de Informação é bastante abrangente, sendo trabalhado em diversas áreas: teoria da informação, canais de comunicação, o usuário da informação, seleção e aquisição de documentos, linguagens de indexação, formação de base de dados, armazenamento da informação, planejamento e avaliação de sistemas de informação. Contudo, pouquíssimas pesquisas têm dado ênfase ao trabalho realizado, pelo indexador, na extração de termos<sup>1</sup>, antes da entrada de documentos nos SRI's, como bem corrobora Cesarino (1985).

A eficiência de um SRI depende da qualidade da análise conceitual, no processo de indexação. Fazendo-se, pois, necessário, um estudo sobre as falhas que ocorrem na recepção da informação, que se deve a erros ou omissões na interpretação do conteúdo dos documentos. Lucas (1996), quando escreveu sobre a responsabilidade do profissional da informação na recuperação de informação em bases de dados, disse que o bibliotecário não é apenas um organizador da informação, mas sim, um profissional que atua como um filtro refinador de informações. Neste sentido é relevante que o indexador compreenda a linguagem do usuário, para que o sistema interaja com o mesmo. Pois sem isso, não há como se fazer uma indexação correta, para que assim o usuário possa obter, pelo menos o mínimo de informações para a sua pesquisa ou de seu interesse.

---

<sup>1</sup> Nesta monografia, à maneira de (GOMES, 1994, p.57), entendemos que o **termo** (grifo nosso) é a “expressão verbal de um conceito”. Por exemplo, os termos representados pelas palavras livro, homem, banana, fruta, dentre outras, constituem significantes. GOMES, Hagar Espanha. **Tesouro e normalização terminológica**: termo como base para intercâmbio de informações. Disponível em: <<http://dgzero.org/dez04/Art.02.htm>>. Acesso em: 14 jun. 2007.

## **1.1 Justificativa**

O interesse pelo tema surgiu na leitura do texto de Cesarino (1985) onde são abordados: o conceito, as funções e os componentes de um SRI. A autora chama a atenção para as falhas que ocorrem em um SRI, que acredita serem oriundas da geração e assimilação da informação por parte do bibliotecário ou especialista do domínio no qual essa atividade está sendo realizada. As observações dessa autora nos motivaram a empreender uma pesquisa que contemplasse a área de indexação e recuperação de informações, especificamente em que concerne a falhas ocorridas durante esse processo. Para estudar o tema, partimos das seguintes questões: sendo a eficiência de um SRI dependente da qualidade da indexação, quais os erros ou omissões que ocorrem nesta atividade? Quais influências o indexador sofre ao indexar? Existem falhas na recepção da informação por parte do indexador?

## **1.2 Objetivos**

### ***1.2.1 Objetivo Geral***

O presente trabalho tem por objetivo geral identificar as dificuldades enfrentadas pelo indexador na extração de termos de indexação a fim de analisar as falhas que ocorrem na execução dessa taxa em Sistemas de Recuperação da Informação (SRI).

### ***1.2.2 Objetivos Específicos***

A partir do objetivo geral decorrem os seguintes objetivos específicos:

- a) conhecer a partir da identificação das dificuldades, os problemas ocasionados em um SRI;
- b) analisar o conhecimento do indexador sobre indexação e as suas etapas;
- c) identificar os erros, falhas ou omissões que ocorrem na taxa de indexação.



### **1.3 Aspectos Metodológicos**

A metodologia norteadora desta pesquisa contempla os estudos exploratórios iniciando-se com a pesquisa bibliográfica e documental para aprofundarmos nossos conhecimentos sobre o tema estudado.

A pesquisa foi realizada junto às Bibliotecas Universitárias de Fortaleza-CE, e, para a coleta dos dados utilizamos o questionário com questões abertas e fechadas.

Entendemos que nossa abordagem de pesquisa contempla a visão denominada de pesquisa “quali-quantitativa”, uma vez que coletamos tanto dados quantitativos, através de questões fechadas (oferecendo um rol de questões para que os indexadores apontassem aquelas que melhor lhe conviessem), como qualitativos onde os participantes emitiram suas opiniões acerca de suas compreensões sobre a atividade de indexação e as dificuldades para eleger as “pistas” visando à recuperação de informação.

### **1.4 Estrutura do trabalho**

Esta monografia está estruturada em cinco (5) capítulos. No primeiro, encontra-se a introdução onde constam a problematização, os objetivos (geral e específicos), a justificativa e a estrutura do trabalho.

O segundo capítulo contempla uma revisão de literatura sobre os Sistemas de Recuperação da Informação, seus aspectos históricos e conceituais, onde são realizadas considerações sobre representação indexal; indexação documentária como forma de representação e reflexões acerca da escolha dos descritores.

A metodologia utilizada na monografia encontra-se no terceiro capítulo e, no quarto capítulo é apresentada a interpretação dos dados e discussão dos resultados. As reflexões conclusivas estão discutidas no quinto capítulo.

## 2 A COMPLEXIDADE NA ESCOLHA DOS DESCRITORES NO SRI

### 2.1 Sistemas de Recuperação de Informação

Embora o conceito de recuperação de informação tenha surgido no século XX, a sua semântica remonta às primeiras iniciativas de construção e organização dos indexes a partir do final do século XII, principalmente na França, e cuja função era facilitar, ao leitor bíblico, o acesso às informações. Porém, não se pode negar que desde a grande Biblioteca de Alexandria ficou evidente a necessidade de se desenvolver métodos e técnicas que, não somente permitissem o controle e conservação das coleções, mas, a identificação de assuntos de um documento assim como também cada um de seus exemplares, com o objetivo de recuperá-los corretamente. Ora, com isto observa-se que Unidades de Documentação<sup>2</sup> trazem embutidas em seu conceito o sentido de possibilidades de acesso a informações, porém, tal qual ocorre na Sociedade Contemporânea, isto nem sempre era possível, pois à medida que os acervos aumentavam as dificuldades para o acesso às informações também cresciam junto. Então, as técnicas de tratamento, organização e recuperação de informação foram se aperfeiçoando, e cada vez mais os bibliotecários se empenhavam em propor métodos e técnicas que viessem contribuir para melhorar o acesso à informação e atender de forma mais eficaz as demandas de seus usuários. Entretanto, mesmo com todas essas invenções, o conceito de sistema de recuperação de informação não era enunciado nesses espaços.

Percebendo que a informação cada vez mais se tornava complexa e se constituía em objetos de estudos de vários campos, outros especialistas começaram também a contemplá-la em suas pesquisas buscando apresentar soluções ao desafio da explosão documentária, principalmente após a Segunda Guerra Mundial. Diante deste fato, outras possibilidades de tratamento e organização da informação foram desenvolvidas e resultaram em um conjunto de atividades se constituindo no sistema de recuperação de informação (SRI) ou *information retrieval system* (IRS) ou simplesmente recuperação da informação (*information retrieval-IR*)<sup>3</sup>.

O termo recuperação da informação foi introduzido na literatura biblioteconômica e documentária por Calvin Mooers (1951, 1999 apud MILANÉS; VELÁSQUEZ, 2006, p.5)

---

<sup>2</sup> Ao longo deste trabalho, adotamos o termo Unidade de Documentação para nos referirmos a bibliotecas, arquivos, museus, centros de informação, ou outro do gênero.

<sup>3</sup> Ao longo deste trabalho será utilizado tanto o termo recuperação da informação (RI) como também sistema de recuperação da informação como sinônimos.

para quem a *Information Retrieval* (IR) "engloba os aspectos intelectuais da descrição de informações e suas especificidades para a busca, além de quaisquer sistemas, técnicas ou máquinas empregadas para o desempenho da operação". O que o autor buscava na sua proposta era encontrar soluções que fossem capazes de resolver os seguintes problemas: "como descrever intelectualmente a informação?; como especificar intelectualmente a busca?; que sistemas, técnicas ou máquinas devem ser empregados?". Embora tenham surgido outros problemas, esses três continuam fundamentais e, ainda hoje se encontram sem solução. É bom lembrar também que, mesmo parecendo ter sido introduzido um novo conceito, na realidade seu sentido já vinha sendo utilizado na prática dos antigos bibliotecários uma vez que na preocupação do campo biblioteconômico os aspectos de tratamento e recuperação da informação sempre estiveram presentes no cotidiano das atividades desenvolvidas por esses profissionais.

Tomando por base a proposta de Mooers, foram aparecendo outros conceitos de SRI, destacando-se aquele apresentado por Cendón (2005, p. 62), para quem a recuperação da informação "consiste na busca de uma coleção de documentos para identificar aqueles que satisfazem uma determinada necessidade de informação e os sistemas de recuperação de informação são aqueles sistemas criados para facilitar essa busca". Saracevic (1999 *apud* MILANÉS; VELÁSQUEZ, 2006, p. 2), defende o sistema de recuperação de informação como sendo um espaço que "*abarca los aspectos intelectuales de descripción de información y sus especificaciones para la búsqueda, además de cualquier sistema, técnica o instrumento que se utilice en la operación*".

A partir destes conceitos, entendemos que a recuperação da informação engloba os aspectos intelectuais de representação de informações e suas especificidades para a busca, além de técnicas e máquinas empregadas para o desempenho dessa atividade. Ora, todos esses aspectos são muito complexos e, portanto, trazem embutidas as dificuldades de conceituação do que seja um Sistema de Recuperação de Informação (SRI). Para Souza (2006), essas dificuldades advêm, tanto da ambigüidade dos conceitos de sistema, como também do conceito de informação em si. Nos SRI's, costuma-se evidenciar o conceito de informação como sendo registros de conhecimentos em documentos, em detrimento de outras definições e contextos que levem em consideração os aspectos cognitivos tanto dos envolvidos nas atividades de tratamento de informação quanto dos usuários. Além, disto, Korfhage (1997 *apud* SOUZA, 2006) ressalta o caráter pessoal da informação, e aponta o fato de que sistemas de recuperação de informações armazenam dados, distinguindo as

informações que foram armazenadas por um usuário das que serão absorvidas por outro. Os SRI's seriam os intermediários nesse processo mediado de trocas de informações, ou seja, conforme afirma Lancaster (2004, p.7), os SRI's são uma espécie de interface entre uma coleção de recursos de informação, em meio impresso ou não, e uma população de usuários; e desempenham as seguintes tarefas: aquisição e armazenamento de documentos; organização e controle desses documentos; e distribuição aos usuários.

Mesmo com todas as características aqui apresentadas, Lancaster (1968 *apud* SOUZA, 2006) chama atenção para o fato de que os SRI's não informam o usuário, no sentido de mudar seu conhecimento sobre o objeto de sua questão, mas, apenas o informam sobre a possível existência de documentos referentes à questão, além dos dados de representação descritiva desses documentos. Baeza-Yates e Ribeiro-Neto (1999 *apud* SOUZA, 2006), entendem os SRI's como sistemas que lidam com as tarefas de representação, armazenamento, organização e acesso aos itens de informação.

Partindo das definições citadas e nos apoiando em Souza (2006, p.165), assumimos que nos SRI's se organizam e se viabilizam o acesso a informações uma vez que nestes sistemas são efetuadas as seguintes tarefas:

- a) representações temáticas e descritivas das informações contidas nos documentos, usualmente através dos processos de indexação e de catalogação dos mesmos (tratamento técnico da informação);
- b) armazenamento e gestão física e/ou lógica desses documentos e de suas representações;
- c) recuperação das informações representadas e dos próprios documentos armazenados, visando a satisfação das necessidades de informação dos usuários. Para isso é necessário que haja uma interface na qual os usuários possam descrever suas necessidades e questões, e através da qual possam também examinar os documentos atinentes recuperados e/ou suas representações.

Rowley (1994) argumenta que, geralmente, SRI são constituídos ao menos por três etapas: a indexação, o armazenamento e a recuperação da informação. Porém, não podemos esquecer que estas etapas somente serão efetuadas se antes de qualquer coisa se efetuarem a seleção e coleta (entrada), e o *feedback* a fim de avaliar os resultados dos serviços oferecidos. Nesse processo, a entrada tanto diz respeito ao documento em si mesmo, como a demanda de

informação feita pelo usuário. Tomando por base os estudos de Lancaster, a professora Maria Augusta Nóbrega Cesarino estruturou o modelo teórico dos componentes do SRI, de acordo com a figura 1, abaixo.

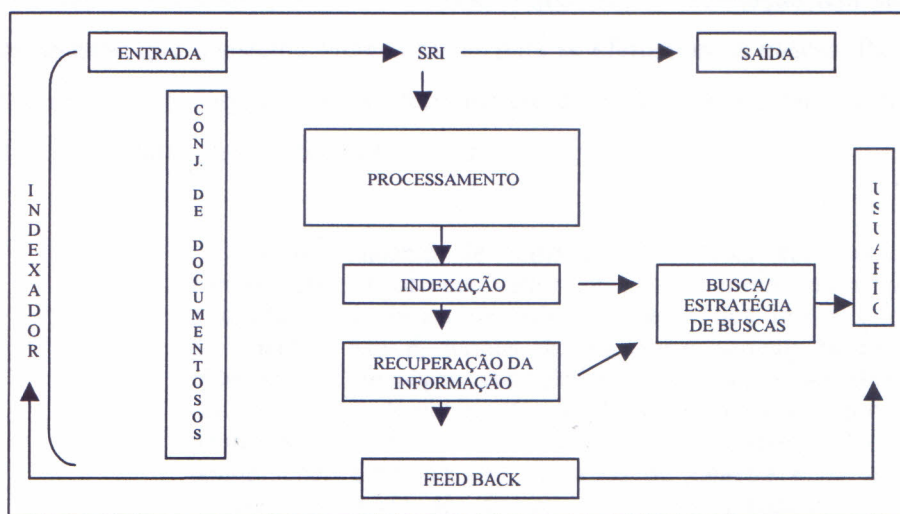


Figura 1 – Componentes dos SRI's  
Fonte: Cesarino (1985, p.159).

Conforme observado na figura 1, a constituição de um SRI demanda inúmeros elementos e ações para que a informação possa ser recuperada. A autora mostra que o início das ações do SRI se efetiva pela formação de um “conjunto de documentos previamente selecionados e adquiridos dentro dos critérios estabelecidos pela instituição que o mantém e de acordo com os objetivos a que ele se propõe atingir”. Essa etapa é fundamental para o bom desempenho de qualquer sistema. Porém, essa seleção não se efetua de forma linear, muito pelo contrário, a formação de coleções se estrutura levando-se em conta os objetivos da organização que abriga as Unidades de Documentação e também o conhecimento dos usuários. Diferentemente dos SRI's das antigas Unidades de Documentação, a estruturação de um SRI, na sociedade contemporânea, demanda que sejam observados aspectos que antes não se dava atenção, ou seja, os vários suportes e formas de documentos. Corroborando, Rowley (2002, p.161) aponta vários destes suportes e formas de documentos incluindo “Cd-Rom, Internet, serviços de consulta em linha, sistemas de gerenciamento de documentos e catálogos em linha de acesso público”. Outro aspecto a ser considerado é que, inicialmente, os SRI's automatizados se estruturaram de uma forma muito simples, sem levar em conta as várias linguagens de comando, que já existiam à época. Porém, atualmente, a maioria dos sistemas funciona com uma interface gráfica que inclui: manipulação direta; seleção de menus;

linguagens de comandos; preenchimento de formulários; e perguntas e respostas.

Além desses fatos, os estudiosos dos SRI's também apontam as estratégias de buscas como um desafio para a área, pois, cada usuário que busca informação tem suas próprias estratégias, mesmo que o sistema ofereça pistas para se efetuarem as buscas. Para a solução destes entraves, foram surgindo outras ferramentas de estratégias de buscas como aponta Lopes (2002) ressaltando que os SRI's favorecem:

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

Mesmo com todas estas observações é importante salientar que a formulação de estratégia de busca depende do tema de pesquisa, da linguagem de busca dos diversos bancos de dados manuais ou automatizados, do domínio dessas estratégias por parte dos responsáveis pelas buscas delegadas e também do conhecimento que os usuários possuem sobre as referidas estratégias. Para tanto, Lopes (2002, p.60) argumenta ser necessário um conhecimento minucioso dessas estratégias e uma preparação adequada tanto para a busca computadorizada quanto para a busca manual. Para a autora são necessárias orientações organizadas em etapas, que podem ajudar o intermediário da busca e o pesquisador a definir o problema, contribuindo para acentuar a qualidade desse serviço. Assim, define como fundamental neste processo sete etapas, a saber: a) discussão do tópico geral da pesquisa; b) conhecimentos básicos sobre os instrumentos de busca; c) formulação 'provisória' da estratégia de busca; d) compreensão da lógica dos conjuntos de termos; e) interdisciplinaridade; f) eliminação de termos indesejados; g) especificação dos parâmetros relevantes para a execução da busca.

Conforme defende Cardoso (2003, p. 5), o campo da RI tem sido influenciado por dois paradigmas ou tendências em suas práticas, concepções e investigações, quais sejam: o enfoque centrado nos sistemas e o enfoque centrado nos usuários. No paradigma do enfoque centrado no sistema "o problema consiste principalmente em construir índices eficientes,

processar consultas de usuários com o melhor desempenho possível, e desenvolver algoritmos de ordenação que possam trazer qualidade à resposta do usuário” (CARDOSO, 2003, p. 5). Porém, é aí que se percebe o grande problema, pois não é somente a construção de índices apoiados em modelos que lhe confere a eficácia, muito pelo contrário, é aí que se verifica o grande impasse, pois, quem constrói os índices são os indexadores (sejam eles humanos ou não-humanos) e quem os utiliza para a busca são os usuários, portanto, a coincidência entre os índices nem sempre acontece. Muitas vezes a dificuldade do usuário está em dizer através de uma expressão de busca, a palavra ou expressão que foram usadas para representar os documentos que irá satisfazer sua necessidade (FERNEDA, 2003). A figura 2 apresenta a RI centrada no sistema.

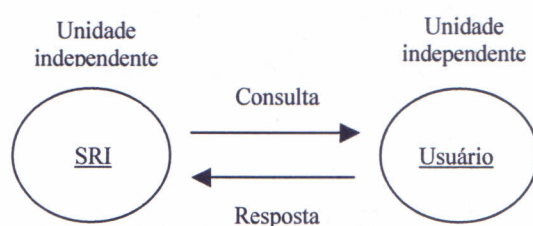


Figura 2: Recuperação da informação no enfoque centrado no sistema  
Fonte: Milanés; Velásquez (2006, p.7)

No paradigma centrado no usuário, “o problema principalmente consiste no estudo da conduta do usuário, no entendimento de suas principais necessidades, e em como este entendimento afeta a organização e as operações do SRI”. (CARDOSO, 2003, p.5). Neste enfoque o usuário passa a ser um elemento a mais no SRI, incorporando-o ao sistema, conforme pode ser observado na figura 3.

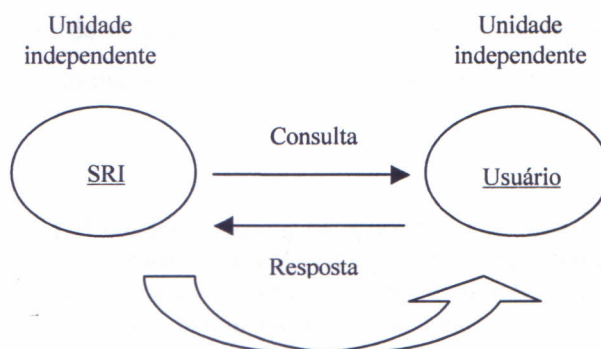


Figura 3: Recuperação da informação com enfoque centrado no usuário  
Fonte: Milanés; Velásquez (2006, p.9)

Observando este modelo tem-se a impressão de que um SRI é mais eficaz e eficiente, se a informação for previamente contida nele e organizada conforme as exigências e características de seus usuários. Neste modelo, o usuário desempenha um papel-chave, uma vez que o usuário pode contribuir interagindo com o sistema a fim de afinar ou restringir suas buscas e, assim, contribuir para a satisfação dos usuários.

Pelo que foi aqui apresentado e embasado em diversos autores, observamos que um SRI é estruturado por inúmeras etapas e ações e que em todas elas o ser humano está presente e traz consigo suas subjetividades, aí a razão das dificuldades e dos problemas inerentes ao próprio sistema. Destaca-se entre eles a complexidade da indexação, principalmente no que diz respeito às tomadas de decisões dos indexadores humanos na escolha dos descritores para a construção dos índices a serem oferecidos como “pistas” para a busca e recuperação de informações, aspectos que serão apresentados a seguir.

## **2.2 Considerações sobre a representação indexal**

### *2.2.1 O que significa representação?*

Ao longo do tempo o homem foi produzindo conhecimentos e desenvolveu técnicas para o seu registro visando não apenas a preservação, porém, a sua transmissão entre as gerações, por meio de representações materializadas nos ideogramas da escrita cuneiforme dos sumérios e hieróglifos dos antigos egípcios. Mas efetivamente, o que significa representação?

Se procurarmos a etnologia da palavra representação, percebemos que ela é originária do latim *repraesentatio*, de *repraesentarer* ou tornar presente. De acordo com Bentes Pinto (2006), o conceito de representação, independentemente do domínio ao qual ele se refira, sempre:

[...] nos leva a pensar na preexistência de qualquer coisa, porque o prefixo “re” evoca a idéia de retorno, de regresso, entre outros. Então, podemos admitir que, o cenário da representação é estruturado através de três elementos fundantes: o representante (objeto ou coisa), o sujeito que se apropria deste objeto, através de sua representação e o representado (resultado da representação do objeto ou coisa representado). Sendo assim, compreendemos que a representação se estrutura como sendo uma forma de sistema simbólica de interpretação dos sujeitos sobre a imagem ou a coisa



(objeto) que ele percebe do mundo, construindo assim sua representação. Percebida assim, a representação se assemelha a um processo de comunicação com a entrada do objeto (a percepção, o tratamento - decodificação desse objeto percebido) e a saída (a imagem do objeto construída). Na realidade, o objeto, enquanto signo emite uma espécie de reflexo e o receptor ativo se apropria da realidade "refletida" e constrói em seu espírito a representação deste. (BENTES PINTO 2006, p. 1).

No Dicionário Larousse (2005, p.841) esta palavra está ligada à filosofia, "a representação é o meio pelo qual um objeto se faz presente ao espírito", à psicologia, "é uma percepção, uma imagem mental cujo conteúdo refere-se a um objeto, uma situação, a uma cena (etc.) do mundo no qual vive o sujeito". Pensada de outro modo, a representação "é a ação de tornar sensível qualquer coisa através de uma figura, de um símbolo, de um signo". (DUBOIS; MITTERAND; DAUZAT, 1998, *apud* BENTES PINTO, 2006, p.3).

Independentemente de qualquer contexto, no conceito de representação está implícita a idéia de "estar no lugar de...". No campo da Biblioteconomia e da Ciência da Informação ela também contempla esta idéia, e perpassa em vários momentos, uma vez que estes campos de saberes têm como objetos de estudo a informação, o documento e o usuário e em todos estes objetos está presente o sujeito humano com as suas subjetividades simbólicas, ou seja, com as suas representações, daí a complexidade de se tratar este conceito nestes campos. Contudo, objetivamente falando, é possível dizer que a representação nestes campos se estrutura nas atividades de representação descritiva (catalogação) e representação temática (indexação) e em ambas estão inseridas outras formas de representações, inclusive social, uma vez que ao se tratar informações visa-se uma comunidade e que em todos os dois casos os sujeitos possuem suas próprias representações. Nesta monografia interessou-se pela representação temática, conhecida popularmente como indexação.

### *2.2.2 Indexação documentária como forma de representação*

A informação e a importância de sua utilização têm sido muito contempladas em trabalhos e estudos dos campos da Biblioteconomia e da Ciência da Informação nas últimas décadas, principalmente devido a fatores como o aumento de informações em circulação, a diversidade de suportes disponíveis para seu registro e tratamento e, em especial, o desenvolvimento das tecnologias da informação e de comunicação. Estas tecnologias criaram um ambiente propício para o experimento de novos métodos de indexação e de recuperação

de informação. Mas, conforme argumenta Bentes Pinto (2001) a indexação não é privilégio da Biblioteconomia e muito menos da Ciência da Informação, ela pode ser aplicada, por exemplo, à economia, à demografia, ao comércio, à estatística, entre outros.

Independentemente do campo onde a indexação é empregada, o senso desse conceito remete para a construção de indexes, popularmente conhecidos como índices. Porém, no contexto da Biblioteconomia e da Ciência da Informação, a indexação é percebida como uma forma de representação do conhecimento registrado e visa oferecer “pistas” para a recuperação da informação em um momento posterior.

Naves e Kuramoto (2006, p.104) entendem a indexação como sendo um processo intelectual que envolve atividades cognitivas na compreensão do texto e na composição da representação do documento. Para Lancaster (2004, p.6), a indexação é uma atividade que visa estabelecer uma representação do conteúdo temático dos documentos. Wooster (1964, *apud* LANCASTER 2004, p.15) se refere à indexação como a atribuição de termos “provavelmente relacionados de alguma forma com o conteúdo do documento original, para ajudar você a encontrá-lo quando precisar”. Do ponto de vista de Rowley (2002, p.163) a indexação é percebida como o processo de atribuir termos ou códigos a um registro de informação ou documento, os quais serão úteis posteriormente na recuperação do documento ou de registros informacionais. Ampliando o conceito, Gardin (1976, *apud* BENTES PINTO, 2001) diz que a indexação documentária é entendida como:

Um conjunto de atividades que consiste em identificar, nos documentos, os seus Traços descritivos (TD's) ou macro-proposições e, em seguida, extrair os elementos (palavras-chave, termos, descritores (sintagmas)) indicadores do seu conteúdo, visando à sua recuperação posterior. Esses elementos vão se constituir na representação do conteúdo do documento e, não na sua representação, pois esta só pode ser pelo próprio documento.(GARDIN, 1976, *apud* BENTES PINTO, 2001, p.226).

A partir dos teóricos citados, entendemos que a representação indexal é um conjunto de atividade e ações que visam captar o conteúdo dos documentos, sejam eles verbais ou não-verbais e expressá-los através de palavras e/ou imagens que serão oferecidas como “pistas” para a recuperação dos referidos documentos. Sendo assim, o principal objetivo da atividade de indexação é possibilitar a recuperação dos documentos ou informações no momento em que o usuário busca um assunto ou um item conhecido (autor, título, editora) em um sistema

de recuperação de informações. Em sendo assim, a indexação é realizada visando atender aos usuários de uma determinada unidade de documentação ou de uma publicação específica.

Do ponto de vista de como se efetua a indexação, ela pode ser realizada ao menos de três maneiras:

a) manualmente (também conhecida como intelectualmente); na qual o indexador é o responsável pela construção dos índices. Esta maneira de indexar é bastante difícil, pois, enquanto sujeito ativo, o indexador possui suas subjetividades e isto, certamente que interfere no processo, mesmo que alguns autores defendam a neutralidade deste sujeito. Desde que se tornaram inviáveis em alguns contextos os processos manuais de indexação de documentos, se buscam alternativas eficazes que possibilitem a representação automática dos assuntos principais desses documentos. Os processos mais comuns de indexação automática descrevem os documentos através de uma lógica simplista advinda da análise de frequência de ocorrência e co-ocorrência das palavras que neles ocorrem;

b) semi-automática; na qual a representação é estabelecida pelos *softwares*, porém, a decisão final das “pistas” representativas dos conteúdos é efetuada pelo indexador;

c) automaticamente (também conhecida como mecanicamente); na qual as representações são construídas pelos *softwares*, sem interferência alguma do indexador humano. Os procedimentos de indexação automática têm se revelado segundo Kuramoto (2006) pouco eficazes no sentido de possibilitar o acesso preciso à informação. Apesar da evolução da tecnologia da informação, os modelos ainda hoje utilizados pelos sistemas de recuperação de informação baseiam-se no uso das palavras como forma de acesso à informação. A questão do tratamento e da indexação da informação está ligada à recuperação da informação. A indexação automática, apesar de idealizada há muito tempo, somente agora vem ampliando a sua utilização, que abrange os conteúdos de publicações eletrônicas e bibliotecas digitais, que têm como base a disseminação de informação em texto integral. A indexação automática é baseada na extração pura e simples das palavras contidas nos documentos. Portanto, a representação de um documento, nesse tipo de procedimento, é dada pelo conjunto de palavras extraídas do referido documento, complementada por indicativos de sua relevância calculados com base na sua frequência nos documentos e/ou coleção. Portanto, o resultado da indexação automática é um conjunto de termos extraídos do conteúdo dos

documentos, acrescido de sua relevância, baseada na ocorrência desses termos no texto.

Em que concerne à extensão com a qual a indexação pode ser realizada, destacamos que ela pode ser exaustiva e específica. A indexação seletiva apresenta uma indicação muito geral do que se trata o artigo ou documento, equivalente neste caso ao título, e um nível de acesso muito limitado. A indexação mais exaustiva proporciona uma indexação muito melhor do assunto específico de que trata o documento, bem como possibilita muito mais pontos de acesso.

De acordo com Lima (2006, p. 104) a indexação de assuntos envolve três etapas: a **análise do documento**, a **identificação e seleção dos principais conceitos**<sup>4</sup> do documento e a **tradução destes conceitos** em termos de uma linguagem de indexação. Bentes Pinto (2001) acrescenta uma quarta etapa que denomina de controle de qualidade.

#### a) **Análise do documento**

Na análise do documento se busca identificar de que trata um documento ou qual o assunto nele tratado. A efetivação desta atividade demanda a leitura dos documentos, a identificação dos termos ou dos conceitos e a seleção daqueles considerados importantes para representar o(s) conteúdo(s) tratado(s). Bentes Pinto (2006, p.7) chama a atenção para esta ação, mostrando que, embora a leitura de um documento pareça uma atividade simples, na realidade se trata de algo bastante complexo, visto que, em sua trama, “estão envolvidas atividades de análise e síntese para a construção representacional dos conteúdos documentários e isto não pode ser dissociado das subjetividades do sujeito leitor, mesmo que, ingenuamente se deseje”. No processo de análise, Fujita e Rubi, (2006) chamam a atenção afirmando que muitas vezes a indexação é percebida, somente como um processo técnico que não necessita de procedimentos sistematizados para identificação de assuntos, porém, somente de uma leitura rápida para identificar e extrair os termos para representar o conteúdo.

Sob o ponto de vista do sistema de recuperação da informação, segundo Fujita e Rubi

---

<sup>4</sup> Entende-se por conceito o uso de termos em seus respectivos contextos, ou seja, corresponde aos seus significados. Na terminologia moderna, Rey (1989, *apud* BENTES PINTO, 2006, p. 7) apresenta a palavra conceito para “designar o ato do pensamento e o objeto do pensamento (idéia), enquanto abstrato e geral. Ainda segundo esse autor, o conceito [...] corresponde a uma unidade do pensamento delimitado”.

(2006, p.49) a indexação é reconhecida como sua parte mais importante dentro dos procedimentos realizados para o tratamento da informação, pois condiciona os resultados das estratégias de busca. Nesse contexto, o indexador tem como função compreender o documento ao realizar uma análise conceitual que represente adequadamente seu conteúdo, de modo que ocorra correspondência entre o índice e o assunto pesquisado pelo usuário.

No processo de identificação dos principais conceitos do documento, o indexador evoca seus conhecimentos prévios sobre o assunto tratado, e isto se constitui, tanto quanto a leitura, em um processo cognitivo. Corroborando, Pinheiro (1978) ressalta que a atividade de indexação é, um processo subjetivo que:

[...] envolve julgamento e, conseqüentemente, oscila muito no seu nível de concordância e apresenta discrepâncias. Pode-se dizer que o que coloca em risco a coerência da indexação é que indexadores percebem de forma diferente:

- o conteúdo real do documento;
- a parte desse conteúdo que será suscetível de responder realmente às necessidades dos usuários;
- os conceitos importantes que devem ser conservados para representar este conteúdo;
- os descritores definidos para representar esses conceitos. (PINHEIRO, 1978, p.109).

Segundo Naves (2006) a identificação de conceitos é um dos problemas recorrentes nos sistemas de representação e recuperação, uma vez que ela deve refletir a estrutura organizacional do assunto e o entendimento dos mesmos pelos indexadores. Esta etapa é uma das mais importantes e complexas no processo de indexação, pois é nela que as decisões serão tomadas quanto à escolha dos termos ou conceitos que serão adotados como “pistas” para a recuperação da informação.

Para a identificação dos conceitos, faz-se uso da língua e da linguagem natural com o objetivo de representar os conteúdos dos documentos em “etiquetas” ou “pistas” indexais capazes de nomear os objetos do mundo e com vistas à localização posterior de informação. A linguagem natural (verbal ou não-verbal) é utilizada cotidianamente pelo ser humano de maneira livremente ou espontaneamente no tratamento da informação e no processo de comunicação<sup>5</sup>. Por exemplo, quando buscamos informações nas unidades de documentação,

---

<sup>5</sup> O termo tratamento da informação visando ao processo da comunicação, no decorrer deste trabalho, refere-se tanto às atividades desenvolvidas por especialistas da área em estudo, quanto pelos diferentes sujeitos em suas

na Internet, na *Web*, ou ainda, em outras fontes, utilizamos palavras, ou imagens do nosso cotidiano de comunicação, arquivados em nosso repertório cognitivo. Neste caso, estamos adotando a linguagem natural. E, é aí que se situa o grande problema da representação indexal, pois, como diz Bentes Pinto (2006) na identificação de termos e conceitos visando à construção de representação indexal de único documento, entram em cena:

[...] ao menos três atores: o autor do documento, o indexador (humano ou não-humano) e o usuário, sendo que cada um destes tem percepções diferentes sobre os temas e conceitos adotados nos documentos. Portanto, são aspectos de tamanha complexidade, principalmente para os indexadores que funcionarão como mídia entre os autores, o sistema e os usuários. (BENTES PINTO, 2006, p.7).

Todas as colocações da autora são pertinentes, pois, nesta etapa se trabalha com a linguagem natural e na perspectiva de uma indexação por extração (Lancaster, 2004), na qual os termos ou conceitos para representar os conteúdos dos documentos, são extraídos do próprio documento. Então se formos analisar o preceito de coerência da indexação, vê-se que a linguagem natural dificulta a concordância entre os termos escolhidos pelos indexadores, deixando margem para que um mesmo assunto seja representado de formas diferentes e, com isso, possibilita que as informações sejam perdidas mediante determinada solicitação de busca em um sistema. Por isso, a coincidência na identificação e seleção destes termos ou conceitos visando traduzir os temas tratados nos documentos, provavelmente não ocorrerá, afinal as palavras são polissêmicas pela sua própria natureza e os sujeitos constroem os sentidos a partir de suas culturas, então o que uma palavra significa conceitualmente para um, certamente que não terá qualquer significado para outro. Além do mais com o avanço da ciência e da tecnologia, a produção documental também expandiu e cada vez foi ficando mais difícil de se trabalhar com a linguagem natural, afinal as terminologias também começaram a se sobrepor e, com as propostas de indexação mecânica, foram construídas as linguagens documentárias que, supostamente viriam minimizar os problemas de polissemia na indexação e recuperação da informação.

#### **b) Tradução dos conceitos em termos de uma linguagem de indexação**

A tradução envolve a conversão do resultado da análise conceitual do documento para

---

experiências cotidianas, independente de níveis ou domínios de conhecimento especializados. Haja vista que a condição existencial do ser humano pressupõe a capacidade de percepção e produção de sentido diante do universo informacional em que se encontra.

a linguagem documentária. As linguagens documentárias são linguagens convencionais cujo objetivo é descrever o conteúdo dos documentos visando à recuperação da informação. As linguagens documentárias podem ser alfabéticas, com vocabulário controlado (tesauros, listas de autoridades, cabeçalhos de assuntos) e alfanumérica, com vocabulário controlado e índices (sistemas de classificação). Estas linguagens podem ser consideradas como tendo estruturas hierárquicas e combinatórias. As linguagens com estruturas hierárquicas seguem o princípio da arborescência e possuem relações hierárquicas entre os termos. Também podem ser construídas de forma sistemática, com os termos estruturados do geral ao particular e cujos termos são combinados no momento do processamento da informação (pré-coordenação). Exemplo, Classificação Decimal de Dewey e Cabeçalhos de assuntos. As linguagens combinatórias seguem os princípios da interseção; onde um conceito pode ser representado por vários descritores e combinam os termos no momento da busca (pós-coordenação). Exemplo, tesauros.

Segundo Lancaster (2004), na etapa de tradução a indexação é feita por atribuição fazendo-se uso dos descritores, recolhidos de repertórios terminológicos construídos a partir das linguagens documentárias (tesauros, cabeçalhos de assuntos, listas de autoridades ou outro do gênero), trata-se de indexação por atribuição. Quer dizer, esta maneira de indexar se estrutura nos descritores que, segundo Bentes Pinto (2006, p.7) são palavras ou grupos de palavras de uma “linguagem documentária ou terminológica, escolhidos para construir a representação indexal dos temas ou assuntos tratados nos documentos, sejam eles verbais ou não verbais, e com fins de criar possibilidades de recuperação da informação”, portanto, são percebidos como vocabulários controlados e cujos sentidos dizem respeito ao contexto de uso. As linguagens controladas, normalmente são estruturadas seguindo classes indutivas, ou seja, partem do geral para o particular. Logo, os procedimentos de indexação e recuperação da informação em linguagem controlada são ligados à manipulação de classes. Quando indexamos um documento, estamos colocando-o em uma classe determinada. Para facilitar o processo, cada classe recebe um nome, que é chamado termo indexador. Ao conjunto de termos indexadores se denominam de linguagens de recuperação de informações ou linguagens de indexação. Neste mesmo sentido, Lancaster (2004, p.19) destaca que, “quando um sistema adota algum tipo de controle sobre os termos utilizados, estamos empregando uma linguagem artificial, através de um vocabulário controlado”. Ao descrever a finalidade dos vocabulários controlados, pode-se, segundo Lancaster (2004, p.19) determinar que a estrutura semântica deste tipo de instrumento de indexação destina-se, especialmente, a: “Controlar

sinônimos, optando por uma única forma padronizada, com remissivas para todas as outras; Diferençar homógrafos; Reunir ou ligar termos cujos significados apresentam uma relação mais estreita entre si”.

A proposta dos vocabulários controlados seria possibilitar convergência entre as linguagens utilizadas por indexadores e usuários – premissa fundamental para comunicação de informações dentro de um sistema. Nesta proposta, os possíveis termos de indexação seriam atribuídos a partir de uma linguagem documentária, tanto em sistemas manuais como, semi-automatizados ou automatizados, como por exemplo, os tesouros, os cabeçalhos de assuntos etc. No caso de sistemas automatizados, os tesouros seriam abrigados no sistema e a indexação seria feita com base na ocorrência e co-ocorrência de palavras num registro ou documento. Uma opção pode ser os computadores também utilizados para pôr em ordem termos de indexação atribuídos por seres humanos. Esses termos serão selecionados e atribuídos por indexadores com base no julgamento subjetivo que fazem acerca do conteúdo do documento, ou escolhem termos que tenham a probabilidade de virem a ser procurado por um usuário no futuro. Muitos sistemas incorporam elementos tanto de linguagens de indexação controlada quanto da linguagem natural.

Uma função importante do indexador é contribuir para o aperfeiçoamento do vocabulário controlado, comunicando suas deficiências aos responsáveis por sua manutenção. Lancaster (2004, p.24) aponta a grande quantidade de itens por dia a serem indexados como um entrave para se ler a obra completa, fazendo com que o indexador passe os olhos rapidamente nas partes essenciais de um livro. Chama a atenção que o indexador deve levar em conta o documento inteiro e que os termos atribuídos precisam refletir o todo.

Lancaster (2004, p.83) defende de um modo muito pragmático a adequada indexação como aquela que permite recuperar “itens de uma base de dados durante buscas para as quais eles sejam respostas úteis, e que impede que sejam recuperados quando não sejam respostas úteis”. Contudo, Bentes Pinto (2006) argumenta que, mesmo com todas as possibilidades do tratamento da informação, que a cada dia tenta se aproximar das estratégias de buscas dos usuários:

[...] nenhum tratamento da informação pode ser entendido como uma expressão entre x e y, muito pelo contrário, ela deve ser entendida como uma expressão entre x, y e z. De tal modo que x leia o texto de y e



estabeleça uma representação para que z possa recuperar a informação, quando precisar. Por outro lado, x, y e z, possuem as suas próprias representações de mundo, onde estão envolvidos inúmeros aspectos simbólicos, por exemplo, histórico, social, psicológico, lingüístico, antropológico, legal, dentre outros. (BENTES PINTO, 2006, p. 12).

### 2.2.2.1 A árdua tarefa na escolha das “pistas” de recuperação da informação

A grande quantidade de produção científica, juntamente com a interdisciplinaridade entre as áreas do conhecimento que vem ocorrendo, principalmente, a partir da “explosão bibliográfica”<sup>6</sup> e mais recentemente, da “explosão documentária” nas últimas décadas, tem tornado o trabalho de organização e tratamento da informação cada vez mais complexo.

No campo da Ciência da Informação, a ação de identificar e descrever o conteúdo de um documento ocorre na indexação ou representação temática. A indexação é considerada como processo básico na recuperação da informação e, dentro desse processo, ocorre a fase de escolha das “pistas” que possam representar o conteúdo temático do documento, em que são selecionados termos em linguagem natural, e depois traduzidos para uma linguagem de indexação, passam a ser chamado de termos de indexação ou descritores.

Sob o ponto de vista do sistema de recuperação da informação, a indexação é reconhecida como sua parte mais importante dentro dos procedimentos realizados para o tratamento da informação, pois condiciona os resultados das estratégias de busca. Nesse contexto, o indexador tem como função compreender o documento ao realizar uma análise conceitual que represente adequadamente seu conteúdo, de modo que ocorra correspondência entre o índice e o assunto pesquisado pelo usuário. Fujita (2006) aponta a utilização dos manuais de indexação que devem refletir a política de indexação do sistema de informação e a realidade de trabalho do indexador.

Lancaster (2004, p.25) ressalta que, o motivo para se examinar o documento é a decisão sobre o que incluir na indexação. E para fazer isso de forma eficiente, o indexador precisa conhecer muito bem o interesse da comunidade a ser atendida. Além do usuário,

---

<sup>6</sup> O termo “explosão bibliográfica” foi utilizado, pela primeira vez, em 1948 na *Royal Society Conference of Scientific Information* e entendido como o surgimento em grande número dos periódicos especializados e as resenhas. Com relação à expressão “explosão documentária”, ela apareceu com aumento global do volume de documentos publicados.

entendemos ser necessário que ele também conheça a área de conhecimento na qual atua, as ferramentas terminológicas de indexação e também que exista uma política de indexação que possa nortear esta atividade. É preciso que as unidades de documentação percebam a importância da indexação em todo o ciclo documentário, considerando-a como parte da administração, compreendendo as tomadas de decisões relacionadas a esta atividade, daí a importância de que sejam estabelecidos parâmetros mínimos que possam guiar os indexadores.

Além disso, considera-se que a política de indexação, em tempos de inovações tecnológicas, deixou de ser localizada para se tornar também globalizada. Devemos isso ao fato de que a automação das bibliotecas permitiu que os catálogos, antes locais e restritos a determinada comunidade, se tornassem disponíveis através da Internet, atravessando fronteiras geográficas. Essa visibilidade do catálogo fez com que fosse necessária à adoção de critérios de qualidade para a recuperação da informação e, isto contribui para que as decisões com relação à escolha dos termos indexadores seja cada vez mais complexa, independentemente da unidade de documentação. Fujita (2006) chama atenção para este fato, citando o exemplo de bibliotecas que trabalham com áreas específicas do conhecimento, tais como energia nuclear; odontologia; agricultura, medicina e direito. A autora continua seu raciocínio dizendo que os problemas enfrentados pelos indexadores são os mesmos, independentes do tamanho e da área de assunto coberta pela instituição. Eles dizem respeito principalmente aos procedimentos de indexação; ao manual de indexação, muitas vezes inexistente; à falta de política de indexação; à atualização da linguagem documentária; à utilização eficiente de *software*; à capacitação dos usuários e à falta de um grupo de estudos sobre indexação. Devido a essas dificuldades e à importância da indexação sob o ponto administrativo e gerencial da biblioteca, procuram-se novos caminhos através da cultura e do conhecimento organizacional, uma vez que um sistema de informação não deixa de ser uma organização, e pretende-se transpor essa realidade administrativa para os serviços de recuperação dos sistemas de informação.

Neste contexto Gismondi (20-?) aponta os descritores como orientadores de pesquisas eficientes, por parte do usuário, resultando em um número menor de referências bibliográficas, mais estreitamente relacionadas com o objetivo da pesquisa. Sabe-se que a escolha dos descritores é realizada pelo indexador tomando por base a política de indexação adotada pelo sistema, tipo de vocabulário utilizado, objetivos da instituição, perfil e

necessidades de informação dos usuários o principal fator que interfere diretamente no processo parece ser, assim, o fator humano. No ato de pensar, quando faz abstrações, o indexador interpreta e define o assunto de um documento, sofrendo influência de diversos fatores pertencentes a vários campos do conhecimento.

Bentes Pinto (2001) afirma que a maneira de indexar depende do tipo de documento que se tem em mãos e sugere uma indexação de documentos levando-se em conta as suas estruturas lógicas, pois as mesmas refletem com mais clareza o conteúdo tratado no documento. Desta forma, a indexação pode ser realizada a partir dos capítulos, seções, parágrafos, passagens etc. Outro problema identificado pela autora diz respeito à dificuldade de se escolher, na indexação manual (que não foi extinta até hoje), 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.

Além destes problemas Fidel (1994 apud BENTES PINTO 2001) aponta que outros fatores influenciam nas tomadas de decisões por parte dos indexadores, quais sejam, as fontes de vocabulário, o grau de precisão, o peso dos termos, a tradução dos conceitos e o uso da linguagem popular. Estes aspectos são apresentados a seguir:

- a) em quais fontes de vocabulários de indexação os indexadores podem se apoiar 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?
- b) 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?
- c) o peso relativo dos conceitos de um documento pode ser definido pelo indexador?
- d) 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?
- e) o indexador pode designar os termos de um índice em uma linguagem mais próxima da do usuário? (FIDEL, 1994 *apud* BENTES PINTO 2001, p.230).

A análise do assunto de um documento permite estabelecer relações entre os documentos de uma coleção e o usuário. O usuário exige informações, como aponta Robredo (1994) muito mais precisas na representação do conteúdo dos documentos, para poder chegar

o mais rapidamente possível àqueles que lhes interessam. Na mesma linha de pensamento Tinker (1966 apud LANCASTER, 2004) defende a importância de se fazer uma boa escolha de termos de indexação afirmando que, na atribuição de um descritor:

[...] a um documento, o indexador declara que esse descritor possui um alto grau de relevância para o conteúdo do documento; quer dizer, ele declara que o significado do descritor está fortemente associado a um conceito incorporado ao documento, e que é adequado à área temática do documento. (TINKER, 1966 apud LANCASTER, 2004, p.15).

Embora, o autor defenda a clareza na escolha dos termos indexadores e no alto grau de relevância em relação ao conteúdo tratado nos documentos, é mister que se tenha em conta que esta objetividade não ocorre linearmente como parece na passagem do texto acima. Ora, não existe indexador neutro e muito menos autores e usuários. Somente para se ter uma idéia da complexidade da atribuição de termos indexadores, é só questionarmos sobre o que realmente os autores querem dizer quando utilizam determinados conceitos e também porque os usuários constroem suas próprias maneiras de buscar informações, mesmo que a eles sejam oferecidas possibilidades para tal.

Kuramoto (1996 apud PARREIRAS, 2003) afirma que o uso de palavras-chave é ineficiente, uma vez que a palavra tem como propriedades a polissemia e a sinonímia. Ressalta que essas propriedades podem aumentar as taxas de silêncio e ruído. Destaca a necessidade do uso de descritores contextualizados, que representam a informação sem descaracterizá-la. Contudo, também não podemos esquecer que, mesmo nas linguagens documentárias os descritores também são polissêmicos, afinal trata-se de palavras e, em sendo assim, não são desprovidos dos sentidos de contexto de uso construído pelos sujeitos.

Fujita (2006) ressalta que a indexação é vista, dentro da biblioteca, como um processo técnico, que é preciso somente uma leitura rápida para identificar e extrair os termos para representar o conteúdo. Entretanto, trata-se de uma atividade bastante complexa, portanto, há necessidade de se entender que os indexadores têm suas próprias compreensões de mundo e, certamente que ela se altera também em seus momentos de trabalho o que certamente contribui para as falhas ocorridas em um sistema de recuperação da informação. Cesarino (1985, p.161) aponta como causa das falhas na recuperação da informação a “erros ou omissões na interpretação do conteúdo dos documentos e na percepção da demanda das pessoas a que se destina o sistema”. Por que ocorrem esses erros? Por que tantas omissões na

interpretação do conteúdo dos documentos? Cesarino (1985) lembra ainda a importância de se conhecer a linguagem das pessoas com que o sistema irá interagir, pois sem esse conhecimento as falhas irão aumentar, fazendo com que os usuários percam seu tempo e encontre documentos irrelevantes para suas pesquisas. Este fato é notório em sistemas que utilizam a indexação automática, como por exemplo, nas páginas *webs*, onde as indexações são feitas levando em conta as ocorrências e co-ocorrências das palavras e sem qualquer tratamento visando a minimizar as polissemias das mesmas.

Embora alguns *softwares* tenham sido desenvolvidos para fazer o trabalho do indexador, o elemento humano será sempre imprescindível. De fato, parece que a criatividade decisória do indexador é insubstituível, pois a maioria dos índices requer entradas constantes de palavras e como já abordado ao longo deste documento, elas são polissêmicas pela própria natureza. Logo, sua compreensão está diretamente ligada ao contexto de uso e, o qual não é desvinculado da cultura.

Nos apoiando nos autores aqui apresentados, podemos afirmar que os indexadores humanos estabelecem os índices a partir da compreensão dos assuntos ou temas tratados nos documentos. Porém como ressalta Bentes Pinto (2006, p.8) “essa compreensão tanto pode ser influenciada pela sua bagagem cultural, como também pelo conhecimento que eles possuem sobre os tais assuntos ou temas”. Por outro lado, a autora ainda chama a atenção para o fato de que no momento das buscas os usuários “utilizam esses índices, ou os seus entendimentos enquanto representações que eles têm sobre as coisas ou sobre os objetos do mundo, visando o acesso às informações concernentes àquilo de que necessitam”. Em quaisquer dos casos, nos deparamos com a “presença do conceito mais simples e mais conhecido de representação; ‘estar no lugar de...’ e, é aí que reside a dificuldade das tomadas de decisão dos indexadores quanto a escolha dos descritores, termos ou conceitos para representar conteúdos de documentos, nos SRI, sejam eles em Unidades de Informação tradicionais ou no ciberespaço”.

Porém, independentemente de qualquer coisa, segundo Lancaster (2004) quanto mais especializados os usuários de um determinado centro de informação, maior será a possibilidade de que a indexação possa e deva ser feita sob medida. Uma indexação de assuntos eficiente implica que se tomem decisões não somente quanto ao que é tratado num documento, mas também, por que ele se reveste de provável interesse para determinado grupo de usuários.

### 3 METODOLOGIA

O presente trabalho classifica-se como sendo um estudo exploratório que, segundo Triviños (2006, p.109) “permite ao investigador aumentar sua experiência em torno de determinado problema”. Corroborando, Gil (1999) diz que as pesquisas exploratórias são desenvolvidas com o objetivo de proporcionar visão geral, e tipo aproximativo, acerca de determinado fato. O pesquisador parte de uma hipótese e aprofunda seu estudo nos limites de uma realidade específica, buscando antecedentes, maior conhecimentos para, em seguida, planejar uma pesquisa descritiva ou de tipo experimental.

Quanto à abordagem nosso estudo enquadra-se como uma pesquisa quali-quantitativa. Neste tipo de pesquisa, adotam-se tanto dados qualitativos como quantitativos. Com relação aos dados qualitativos, Cordeiro (1999) argumenta que se enfatizam aspectos sócio psicológicos, tais como opiniões abertas, motivações e questões subjetivas. Neste sentido, buscou-se captar junto aos bibliotecários indexadores suas opiniões sobre o conhecimento referente à atividade de indexação demandando que os mesmos se expressassem verbalmente (por escrito) sobre este fazer. Além disto, também lhes foi apresentado um rol de opções nas quais estes sujeitos poderiam escolher as que melhor lhes conviesse.

O campo de pesquisa foi: as bibliotecas da Universidade Federal do Ceará (UFC), Universidade de Fortaleza (UNIFOR), Faculdades Integradas do Ceará (FIC), Faculdade 7 de Setembro (FA7) e Universidade Estadual do Ceará (UECE). Estas instituições foram escolhidas por se considerar que as mesmas são as mais representativas do Estado do Ceará, tanto em quantidade de cursos quanto pela quantidade e qualidade de seus acervos.

A técnica de coleta de dados utilizada foi o questionário. Antonio Gil (1999) define o questionário como sendo uma técnica de pesquisa composta por um número de questões “apresentadas por escrito às pessoas, tendo por objetivo o conhecimento de opiniões, crenças, sentimentos, interesses, expectativas situações vivenciadas etc”. Para Triviños (2006) o questionário consiste num formulário contendo um elenco de questões em que o pesquisado deverá anotar suas respostas.

O questionário é constituído por sete (7) questões, sendo duas (2) fechadas e cinco (5) abertas. As questões fechadas foram constituídas de múltipla escolha e os participantes

poderiam marcar quantas quisessem.

Para o nosso estudo, inicialmente entrou-se em contato com os bibliotecários responsáveis pelas referidas bibliotecas a fim de expor o objetivo de nossa pesquisa e solicitar a sua colaboração no sentido de permitir a realização da pesquisa com os bibliotecários indexadores. De posse do aval manteve-se contato com estes últimos para efetuarmos a coleta de dados. Foram enviados vinte (20) questionários sendo que somente treze (13) foram devolvidos devidamente respondidos.

Os questionários foram aplicados de duas maneiras:

- a) Através do correio eletrônico, no qual solicitamos a gentileza de que os mesmos fossem devolvidos em uma data determinada;
- b) Aplicação do questionário *in loco*; o pesquisador foi, pessoalmente, as bibliotecas e entregou o questionário aos bibliotecários selecionados para participar da pesquisa e marcou-se uma data para que os mesmos fossem devolvidos. Ressalta-se que, mesmo tendo sido marcado uma data para a devolução, teve-se muita dificuldade no retorno, inclusive com a demora de mais de um mês.

Embora se saiba que um dos inconvenientes da coleta de dados através de questionário seja a dificuldade em sua devolução, como assegura Gil (1999) e Triviños (2006), o fato de termos enviado por *e-mail* nos dava, de certa forma, uma segurança na devolução, entretanto, mais uma vez ficou evidente que a devolução desse instrumento de coleta de dados independe da maneira em que foi enviado (pelo correio tradicional ou eletrônico, uma vez que obtivemos retorno de apenas 62% dos questionários).

#### **4 INTERPRETAÇÃO DOS DADOS E DISCUSSÃO DOS RESULTADOS**

Após a coleta de dados, passamos a interpretá-los à luz dos problemas e dos objetivos da pesquisa, estruturando-os em cinco (5) categorias: 1) concepção que os indexadores possuíam sobre a atividade de indexação e suas etapas; 2) conhecimentos sobre requisitos para se efetuar a atividade de indexação; 3) procedimentos a serem observados para a escolha dos termos de indexação; 4) fatores que interferem no trabalho do indexador; 5) dificuldades encontradas na escolha de termos e em que essas dificuldades influenciam no SRI. Este item

descreve e discute os principais resultados obtidos em cada parte do questionário. Foram enviados 20 (vinte) questionários, dos quais treze (13) deram retorno. Para melhor análise dos dados optamos por citar cada questionário por Bibliotecário (B).

#### **4.1 Concepção que os indexadores possuíam sobre a atividade de indexação e suas etapas**

A atividade de indexação embora pareça simples, demanda uma compreensão mais acurada acerca da atividade em si e também de suas etapas. Neste sentido buscamos conhecer a concepção que os indexadores possuíam sobre essa atividade, tínhamos um pressuposto de que embora esses profissionais efetuem tal atividade, seu conhecimento teórico sobre a mesma era ainda muito incipiente. Tal pressuposto foi confirmado, pois, quando os bibliotecários foram questionados sobre que compreensão tinha a respeito da indexação, 23% realmente emitiram conceitos demonstrando entendimento, como podem ser observados nos seguintes depoimentos:

“Atividades voltadas em estabelecer a representação temática do conhecimento registrado. Faz parte do SRI, vai desde a seleção até... a recuperação da informação de uma unidade de documentação (Biblioteca)”. (B01)

“É representar, sintetizar o conteúdo de um documento através de termos para que sejam recuperados pelos usuários posteriormente”. (B09)

“Indexação é o processo de descrição do conteúdo de um documento no qual empregam-se termos pertencentes a uma linguagem controlada”. (B07)

Estes dados só vieram a confirmar o nosso pressuposto de que, os conhecimentos de alguns indexadores eram incipientes. Por se tratar de profissionais especializados na área de Biblioteconomia e Ciência da Informação e que atuam cotidianamente executando atividades de indexação, supostamente deveriam ter um entendimento teórico desta atividade. Por outro lado, a falta deste entendimento certamente que compromete a qualidade na recuperação de informação. Por se tratar de uma atividade cognitiva uma vez que contempla representações, logo não é uma atividade neutra, portanto, requer que aqueles envolvidos nesta atividade não



a vejam como algo simplista.

Neste mesmo item, também percebemos que alguns participantes confundem o conceito de indexação com a sua função. Veja-se os depoimentos:

“Que é um processo de suma importância para localização de qualquer suporte, sem ela não podemos ter a informação que queremos em curto prazo. Era bom que pudéssemos fazer também com a Internet”. (B06)

“A indexação é a ponte que liga o leitor e o documento, facilitando a busca do que se procura”. (B11)

“A indexação é o fator principal para recuperação de qualquer documento no momento em que o usuário busca um determinado assunto”. (B02)

Através das respostas se observou que apesar de alguns bibliotecários trabalharem todos os dias com a indexação, os mesmos não conseguem expressar o conceito indexação, percebeu-se que 31% dos questionados confundiram o conceito da indexação com a sua importância e objetivos.

Em que concerne às etapas da indexação observou-se que 38% têm conhecimento sobre este assunto e responderam se tratar da leitura do documento, da identificação dos conceitos, da seleção e tradução, conforme Lancaster, Cesarino e Bentes Pinto. O que chama a atenção foi que 31 % não demonstraram clareza em suas respostas confundindo-as com os componentes do Sistema de Recuperação de Informação. Abaixo algumas falas:

“Entrada, análise, decisão, escolha, armazenamento, saída”. (B04).

“A análise de assunto é a etapa mais importante e compreende a leitura de partes do documento como: introdução, resumo, frases dos capítulos, ilustrações etc.” (B10).

Nas falas aqui apresentadas, percebe-se uma confusão, por parte de alguns bibliotecários, pois entrada, saída e recuperação do documento dizem respeito aos componentes de um SRI, enquanto sistema, que é constituído pelas unidades de entrada

(*input*), tratamento e saída (*output*).

Os outros 31% demonstraram desconhecimento sobre as etapas de indexação: “Então tem que haver uma política de indexação, para que o documento seja analisado, indexado e por fim recuperado”. (B03).

A confusão observada por 31% dos bibliotecários, com relação ao conceito de indexação leva-nos a crer que apesar de trabalharem nesta atividade, os mesmos não conseguem ou não sabem conceituá-la. Isto nos leva a inferir que executam uma tarefa estritamente pragmática, sem que tenham compreensão da complexidade que é a indexação, enquanto uma atividade cognitiva e de representação. Em razão disto perguntamos, se eles não conseguem conceituar a taxa de indexação, como conseguem realizá-la? Essa dissociação do teórico com o prático é responsável também pela falta de compreensão do que seja a área de Biblioteconomia, pois, as atividades práticas são executadas sem uma compreensão dialética, logo, dissociada da práxis.

#### 4.2 Conhecimentos sobre requisitos para se efetuar a atividade de indexação

Nesta pergunta, ofereceu-se um rol de itens, para que os bibliotecários elegeassem dentre várias opções, os requisitos que eles considerassem necessário para efetuar uma indexação adequada ou com qualidade. Nas respostas ficaram evidentes que os requisitos concernentes ao exame do documento e o estabelecimento do assunto de seu conteúdo foram apontados por 85% dos questionados. Lancaster (2004) afirma que muitas vezes não é possível uma leitura completa do documento, mas o indexador deve realizar uma leitura rápida do documento, devendo ler suas partes essenciais: título e subtítulo, resumo, sumário e introdução para ter certeza de que nenhuma informação útil passou despercebida.

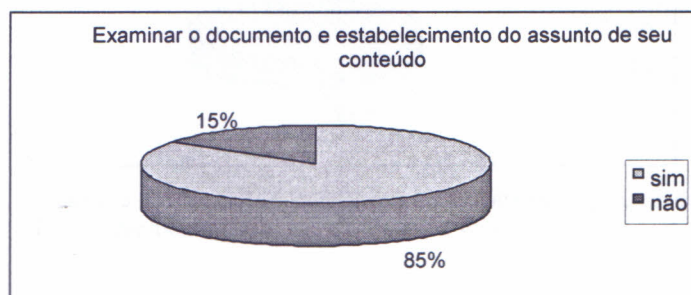


Gráfico 1 - Examinar o documento e estabelecimento do assunto de seu conteúdo

Por outro lado, o que nos surpreendeu foi que 15% dos participantes deixaram de escolher o item referente ao exame do documento. Ora, sabe-se que a atividade de indexação, jamais poderá ser realizada sem que seja feito um exame do documento. Diante disto, questionamos, como será que estes participantes fazem para indexar uma obra, haja vista que examinar o documento é imprescindível para a sua indexação?

Foi colocado como requisito para fazer a indexação, as referências bibliográficas, pois, entendemos que elas também oferecem indícios sobre os temas tratados nos documentos. Contudo, 46% consideram as referências bibliográficas como um requisito para uma indexação adequada. Isto pode ser mais um reflexo de que a indexação feita por esses profissionais assume uma simples prática, sem maiores conseqüências.

No que diz respeito ao requisito tradução dos conceitos em linguagem natural para a linguagem documentária, foi apontado por 92%. Mais uma vez nos surpreende, pois, pelo que sabemos, todas as instituições pesquisadas fazem uso de tesauros, cabeçalhos de assuntos ou listas de autoridades, assim como os sistemas de classificação e todas estas ferramentas são linguagens documentárias. Logo, a segunda etapa da indexação (tradução) se faz presente. Sendo a tradução dos conceitos em termos de uma linguagem de indexação um requisito importante, confirma-se que essa falta de conhecimento por parte dos bibliotecários pode comprometer realmente um SRI, fato defendido por Lancaster (2004). Então, não se justifica que 8% tenham deixado de apontar esse item.

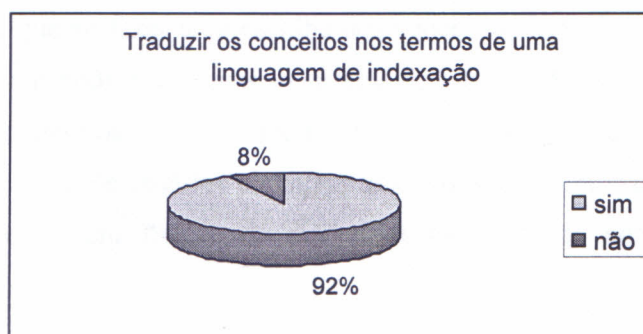


Gráfico 2 - Traduzir os conceitos nos termos de uma linguagem de indexação  
Fonte: Questionário aplicado em fevereiro de 2007

#### 4.3 Procedimentos a serem observados para a escolha dos termos de indexação

Foi solicitado aos bibliotecários que apontassem os procedimentos a serem seguidos na extração de termos de um documento. Esta questão foi colocada para saber se os bibliotecários tinham conhecimento dos procedimentos para se fazer à escolha dos termos de indexação, visto que Lancaster (2004) chama a atenção que a falta de conhecimento pode prejudicar a correta função de um SRI.

Pode-se observar que 92% dos bibliotecários ao extrair termos de um documento, verificam qual é o assunto tratado, sendo que este item é primordial na taxa de indexação. Os demais 8% , não consideram isto como importante o que demonstra a falta de conhecimento por parte destes bibliotecários. Este fato é muito sério, pois fica a impressão de que a indexação é feita aleatoriamente e, isto certamente que compromete a eficácia na recuperação da informação e também o trabalho do bibliotecário.

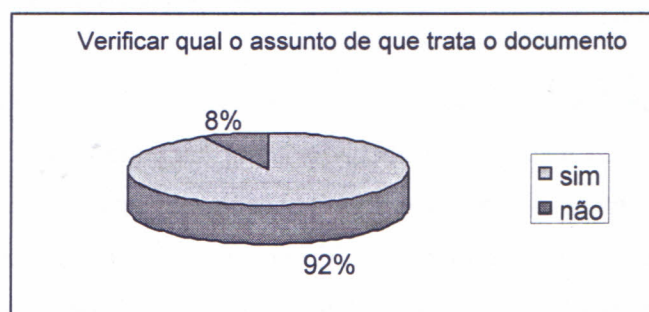


Gráfico 3 - Verificar qual o assunto de que trata o documento  
Fonte: Questionário aplicado em fevereiro de 2007

Percebeu-se que ao fazer uma escolha dos conceitos, 92% dos bibliotecários têm em mente as consultas que podem ser feitas ao sistema de informação. Segundo Lancaster (2004) o indexador deve escolher os conceitos que forem considerados os mais apropriados para uma determinada comunidade de usuários e adaptar tanto os instrumentos de indexação como os próprios procedimentos em função da retroalimentação obtida através dos pedidos de informação.

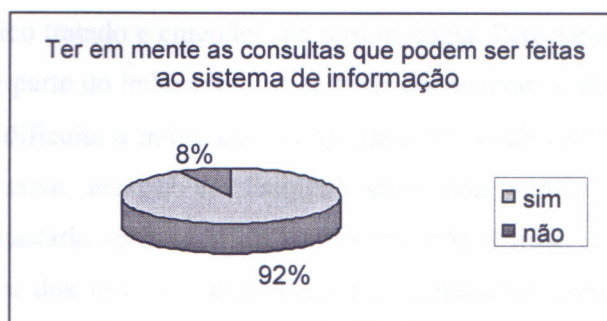


Gráfico 4 - Ter em mente as consultas que podem ser feitas ao sistema de informação  
 Fonte: Questionário aplicado em fevereiro de 2007

#### 4.4 Fatores que interferem no trabalho do indexador

Foi solicitado aos bibliotecários que os mesmos apontassem os fatores que interferem no trabalho do indexador, para se ter conhecimento dos fatores que viessem a prejudicar o seu trabalho no momento em que fizessem a indexação, visto que na literatura brasileira não se encontram textos abordando tal assunto.

O fator mais citado pelos participantes refere-se ao “local”, apontado por 23% dos bibliotecários, pois como executam outras atividades, não podem se concentrar na atividade de indexação, sofrendo interrupções de todo tipo. Em se tratando de uma atividade intelectual, isso certamente consiste em um fator que dificulta o trabalho do indexador. Outros fatores ligados ao ambiente como “refrigeração; iluminação e ruído” são apontados também como interferentes na atividade de indexação, pois condições ambientais desfavoráveis têm um efeito negativo sobre a exatidão dessa tarefa intelectual. Vejam-se alguns depoimentos:

“Local. (se você estiver no balcão de atendimento, você não consegue analisar direito o conteúdo e sendo assim, prejudica a indexação). Ter um espaço reservado para este trabalho, pois é um trabalho minucioso”. (B06).

“Local. Distração causada por fatores externos e ambientes não adequado à concentração que o trabalho exige”. (B12).

Outro fator citado que chama a atenção é o “conhecimento do assunto” e “assunto novo” apontado por 15% dos bibliotecários. Lancaster (2004) lembra que os indexadores não precisam ser necessariamente especialistas no assunto, mas devem ter algum conhecimento do

conteúdo temático tratado e entender sua terminologia. Percebe-se que um assunto novo gera insegurança por parte do indexador. “A falta de conhecimento do assunto” apontado por B13 é um fator que dificulta a indexação, visto que o não conhecimento do assunto pode levar à indexação excessiva, incapaz de distinguir entre dois termos, o indexador talvez atribua ambos quando bastaria apenas um ou apenas um seria correto. Fatores como “conhecimento das necessidades dos usuários; experiência; concentração; capacidade; e compreensões de leitura” foram apontadas por B01 e B07 e se assemelham quanto aos fatores apontadas por B08 como “a capacidade de abstração, percepção e interpretação das informações analisadas” se como interferentes no trabalho de indexação. O conhecimento das necessidades dos usuários de um SRI é especialmente importante, porque uma indexação de qualidade deve ser realizada de acordo com as necessidades de determinada comunidade sempre que possível, quando se conhece a necessidade dos usuários pode-se ter em mente quais conceitos ou descritores utilizar ao indexar. Anos de experiência como indexador, também é um fator que influi na atividade de indexação, da mesma forma como a capacidade de se concentrar, a capacidade e compreensão do que está se indexando. Todos os fatores aqui apontados vêm ao encontro das pesquisas de Lancaster (2004) e Fidel (1994, apud Bentes Pinto, 2001), que chamam a atenção para as dificuldades que os indexadores não especialistas dos campos que indexam enfrentam na escolha dos termos de indexação.

O fator “língua” também foi apontado. Evidentemente, o indexador que não souber francês, por exemplo, dificilmente poderá indexar documentos em francês de modo eficiente, a não ser que contenha resumos claros e completos na própria língua do indexador, o que não é usual. Outro entrave apontado foi o “tempo” apontado por 15% dos participantes, o que pode contribuir com a ansiedade e frustração dos indexadores para se dedicarem à indexação, visto que são muitas as atividades desenvolvidas e muitas os documentos a serem indexados, mas pouco o tempo para indexar.

Também ficou evidente a necessidade de uma política de indexação, apontado por 13% dos indexadores e que corrobora com as idéias de Lancaster (2004) e Fujita (2006). Embora este fato tenha chamado a atenção de poucos indexadores ele é de fundamental importância, haja vista que a indexação necessita de parâmetros que guiem os indexadores no momento de tomadas de decisões minimizando subjetividade e incertezas durante o processo de indexação. Como ressalta Carneiro uma política de indexação deve servir como um guia para tomada de decisões.

A “não existência de um vocabulário controlado” (B02) é um fator que interfere na atividade de indexação, pois o mesmo estabelece a forma de representar os assuntos que compõem uma área limitada do conhecimento, tornando possível maior coerência entre os termos indexados. Além da existência de um vocabulário controlado, os participantes mencionaram, “qualidade do vocabulário empregado” (B05) uma vez que ele pode interferir na atividade de indexação. “Não ter acesso aos tesouros, ou instrumentos para ajudar na indexação” (B09). Um vocabulário controla sinônimos; diferencia homógrafos e reuni ou liga termos cujos significados apresentem uma relação mais estreita entre si, devendo, pois ter qualidade, para tornar-se o ponto de convergência entre as linguagens utilizadas por autores, indexadores e pesquisadoras.

Outro elemento apontado pelos indexadores é a necessidade de maior entendimento sobre como e porque indexar: “Falta de suporte técnico (orientação de como e porque indexar)”. (B03). Esta colocação nos parece sem fundamento, afinal, todos os entrevistados são bibliotecários e, portanto, deveriam ter o mínimo de entendimento sobre a atividade e a função da indexação. Ora, se eles não sabem o porquê de indexar, fica um questionamento: este indexador estará fazendo uma indexação de qualidade?

Além dos fatores apontados anteriormente, também foram citados aqueles relacionados ao documento e ao indexador:

“Fatores ligados ao indexador: Conhecimento do assunto; conhecimento das necessidades dos usuários; experiência; concentração; capacidade e compreensão de leitura;

Fatores ligados ao documento: Conteúdo temático; complexidade; língua e linguagem; extensão; apresentação e sumarização.

Fatores ambientais: Refrigeração; iluminação e ruído.” (B07).

Alguns assuntos são de mais difícil compreensão do que outros, e diferentes indexadores criam diferentes figuras ou idéias de uma mesma informação externa, por causa de suas inclinações pessoais e afetivas, que certamente interferem no trabalho por eles desenvolvido. Relacionado a isso, está o grau de correspondência entre o conteúdo temático do documento e o conhecimento ou os interesses do indexador. Com relação à extensão quanto mais informações são apresentadas, mais claramente a representação revela o alcance do documento, tornando-se mais provável que venha a indicar para o indexador se o

documento satisfaz ou não a uma necessidade de informação. A apresentação e sumarização influirão sobre a maior ou menor facilidade que o indexador terá para descobrir de que trata o documento: o título é preciso ou enganador, existe um resumo ou algum outro de tipo de sumarização que reflita integralmente o conteúdo do documento?

O fator que interfere no trabalho do indexador apontado por (B09) é o “próprio indexador ser capaz de identificar os principais conceitos de um documento”, é imprescindível que o indexador saiba identificar os termos do documento, pois sem esta capacidade não será possível realizar a atividade de indexação. Também foram apontados como dificuldades para indexação os assuntos similares tratados em um mesmo documento: “Assuntos similares encontrados no documento” (B11).

Um participante chamou a atenção para as questões da extensão da indexação, como um entrave para o desenvolvimento desta atividade. “De acordo com o tipo de documento, a especificidade e exaustividade dos termos selecionados, levando-se em consideração sua finalidade” (B12). Visto que a indexação exaustiva implica no emprego de grandes quantidades de termos indexadores, de modo bastante completo. Daí porque pode ter sido apontado como um fator que interfira no trabalho do indexador. Com relação à especificidade Lancaster (2004) diz que é aquele segundo o qual um tópico deve ser indexado sob o termo mais específico que o abranja completa mente, e caso o indexador não tenha conhecimento sobre o mesmo, a indexação estará prejudicada. Bentes Pinto (2001) aponta que a questão da especificidade pode ser um entrave na indexação.

#### **4.5 Dificuldades encontradas na escolha de termos e em que essas dificuldades influenciam no SRI**

Estudando a literatura que trata da atividade de indexação verificou-se que é dada bastante ênfase à atividade de indexação propriamente, porém, na literatura brasileira não encontramos qualquer trabalho que focalizasse as dificuldades enfrentadas pelos indexadores durante a escolha de termos de indexação, a partir deste fato evidenciou-se a necessidade de se verificar as dificuldades encontradas na escolha de termos e em que essas dificuldades influenciam no SRI, já que é uma atividade cotidiana dos mesmos.

Foi solicitado que os bibliotecários apontassem as dificuldades encontradas na escolha



de termos e em que essas dificuldades influenciam em um SRI. O fator “tempo” apontado por B01 e B06 são compreensíveis, visto que é grande a quantidade de documentos a serem indexadas e as já citadas outras atividades que o indexador realiza ao mesmo tempo.

Dentre as dificuldades citadas a “falta de coerência nos termos utilizados” apontada por B02. O conceito de coerência é um dos quesitos para uma indexação de qualidade, já que todos os documentos que tratam sobre o mesmo assunto devem estar representados da mesma forma, não havendo perda de informações no momento da recuperação. Um outro diz ser “a necessidade de clareza na política de indexação e que haja padronização nos serviços” (B10). A falta de padrão a ser seguido prejudica, não apenas o bibliotecário, no momento da definição dos termos indexadores de um documento, mas, principalmente, coloca em risco a qualidade da recuperação de informações em um sistema, no momento em que dados pertinentes e relevantes não são identificados em frente de determinada expressão de busca, estando, portanto, comprometido o cumprimento de um SRI considerando-se que as necessidades de informação dos usuários não são adequadamente atendidas. Os indexadores entendem que, a exaustividade e a seleção correta dos descritores são um dos entraves da indexação.

“A escolha em grande quantidade de conceitos” (B01).

“A seleção de um número correto de descritores diante de inúmeras opções” (B05).

Outro item bastante citado é “não representar bem o assunto do documento e ele deixar de ser recuperado, perdendo-se uma fonte de informação que muitas vezes é imprescindível para um pesquisador” que foi colocado por B09; B10; B12. Que vem ao encontro de outra dificuldade apontada por B03 e B10 “documentos não analisados corretamente, não serão recuperados. E documento não recuperado, é documento ‘perdido’ no acervo. Ele não irá cumprir seu verdadeiro papel que é de servir de fonte de informação e pesquisa para o usuário”, o que é interessante observar que os bibliotecários têm em mente a complexidade do que fazem e que qualquer erro deles prejudica tanto o usuário quanto o SRI.

“Escolha em grande quantidade de conceitos e a qualidade do serviço” apontado por B01; a “qualidade do vocabulário controlado” apontado por B07 e “a seleção de um número correto de descritores diante de inúmeras opções” apontadas por B05, o indexador se vê

diante de muitos conceitos e deve escolher o descritor que melhor represente o documento que se está indexando. A qualidade do serviço é o grau de concordância entre os termos atribuídos pelo indexador e um grupo de termos ideais.

No fator “adequação aos interesses do usuário” apontado por B01 percebe-se que o indexador preocupa-se com os usuários, identifica-se com a quem se destina um SRI.

“Não conhecer o conteúdo do suporte (livros, cds, dvds, monografias etc.)” apontado por B06 é um fator que interfere na atividade de indexação, pois se o indexador desconhece o conteúdo do documento não terá como extrair os termos de indexação, prejudicando a recuperação do mesmo.

“Especificidade/sintaxe; ambigüidade ou imprecisão do vocabulário controlado” apontado por B07 e B12, verifica-se que na especificidade somente os assuntos tratados no documento são identificados, de maneira específica, e como às vezes é difícil identificar o assunto do documento dificulta a atividade do indexador. Na ambigüidade ou imprecisão do vocabulário controlado o indexador fica indeciso quanto ao descritor utilizar.

Um outro participante aponta a falta de domínio dos temas a serem indexados como um dos grandes entraves na qualidade da indexação. Assim ele se expressa: “Exige do indexador familiaridade com área que trabalha. Quando se trabalha com todas as áreas do conhecimento faz-se tudo bem generalizado, mas quando passa-se a indexar para uma biblioteca específica, passa-se a ser mais específica também”. (B04). De fato, o trabalho do indexador é complexo, pois, nem sempre ele é especialista do domínio que indexa e isto, certamente que vai influenciar em seu trabalho.

“O contato com os usuários na hora da busca faz perceber o quanto uma indexação muito geral atrapalha”. Dito por B09 relata uma dificuldade enfrentada pelo usuário na hora de recuperar documentos em um SRI.

“Para se obter sucesso neste trabalho é necessário que a política de indexação seja clara e haja padronização nos serviços”, apontado por B10 é visto como necessária, pois em uma política de indexação tem-se a cobertura de assuntos da SRI em que se trabalha; extensão da cobertura do sistema em áreas de interesse da comunidade; a qualidade dos documentos

incluídos no sistema; processo de indexação; processo de busca e a avaliação do sistema.

“A dificuldade de tradução para uma linguagem de indexação é uma das dificuldades mais comumente encontrada, e a similaridade dos assuntos”, apontadas por B12 verifica-se a indecisão que o indexador enfrenta ao ter que escolher os descritores diante dos termos de indexação.

“A falta de um cabeçalho controlado dos descritores utilizados se torna difícil na hora da recuperação da informação por haver termos indexados de várias maneiras”. Dito por B02 revela a falta de instrumentos que auxiliem o indexador na sua atividade, ocasionando indexação excessiva de um mesmo termo.

Nos seguintes relatos percebe-se a preocupação do indexador com o usuário:

“Quando, na busca, há um número muito grande de termos a serem consultados pelo usuário”. (B05)

“Os usuários não encontrarem o que procuram”. (B06; B13).

“Influenciarão negativamente na recuperação da informação se tais dificuldades não forem superadas, isso poderá ocasionar uma inconsistência na busca da informação”. (B13)

Essa preocupação revela também que um controle de qualidade se faz necessário para que o SRI possa atingir o seu objetivo que é atender ao usuário de maneira satisfatória, que o usuário possa encontrar entre tantos documentos o que lhe é relevante.

A partir das falas dos participantes da pesquisa, percebe-se claramente que, a despeito do que possa parecer para muitos, o trabalho de indexação é uma atividade complexa, haja vista que envolve o ser humano. Portanto, é uma atividade que não pode ser feita tendo por parâmetros a linearidade, afinal, o ser humano, por sua própria natureza, é dinâmico e, traz consigo suas experiências culturais que certamente influenciarão no seu cotidiano.

## 5 CONCLUSÃO

Após as análises dos resultados obtidos e levando em consideração os problemas e os objetivos apresentados ao longo deste trabalho, é possível apontar algumas reflexões conclusivas com relação às dificuldades dos indexadores para as escolhas dos termos de indexação. Neste sentido, é, pois, necessário que se façam algumas reflexões sobre as questões básicas levantadas pelos bibliotecários, tais como:

- 1 Verifica-se que alguns fatores como tempo e local interferem na indexação de documentos, pois a falta de concentração, ocasionada por interrupções, podem afetar a execução de uma atividade essencialmente intelectual, que exige reflexões, inferências, deduções, raciocínio, busca de informações na memória e utilização do conhecimento prévio;
- 2 Fica evidente a necessidade de se conhecer a área em que se está atuando, e esse fator é apontado por alguns bibliotecários como um dos que influencia na hora de se realizar a indexação.
- 3 Também ficou evidente a confusão de alguns bibliotecários sobre a indexação e suas etapas, que os mesmos confundiram com os componentes de um SRI.
- 4 A necessidade de clareza na política de indexação e padronização nos serviços em algumas bibliotecas é necessária para que a falta de um padrão a ser seguido não prejudique não só o bibliotecário, mas principalmente a qualidade da recuperação de informações em um SRI.
- 5 A escolha em grande quantidade de conceitos e a seleção de um número correto de descritores diante de inúmeras opções, pode sobrecarregar a base de dados com termos pouco informativos sobre o assunto dos documentos.

Diante dos resultados obtidos é preciso considerar que nem todos os bibliotecários souberam identificar os requisitos, para uma extração de termos de um documento, o que pode comprometer um SRI, ocasionando ruídos.

Percebe-se que os problemas enfrentados pelos bibliotecários são os mesmos,

independe do tamanho e da área de assunto coberta pela instituição. Eles dizem respeito principalmente aos procedimentos de indexação; ao manual de indexação, à falta de política de indexação, muitas vezes inexistente; à atualização da linguagem documentária; à utilização eficiente de *software* e à capacitação do usuário.

Acredita-se que o conhecimento dos usuários relacionada à identificação de conceitos poderia oferecer uma abordagem melhor sustentada para oferecer agilidade a uma SRI como também assegurar uma uniformidade de procedimentos ao tratamento temático de informações.

Infere-se ainda que, apesar de os bibliotecários trabalharem com a indexação, alguns sentem dificuldades na extração de descritores, e que sofrem inúmeras interferências como: o local em que se trabalha, interrupções de todo tipo, pouco tempo para indexar muitos documentos, a falta de uma política de indexação, um vocabulário controlado inadequado, o que pode levar a não ser realizada uma indexação adequada.

Finalmente, entendemos que nossos objetivos foram atingidos, e nossos resultados poderão ser utilizados para trabalhos futuros, uma vez que apontam dados importantes em relação à complexidade da indexação, pois se trata de uma atividade subjetiva e, portanto, não é isenta das inferências culturais e cognitivas dos sujeitos nela envolvidos.

## REFERÊNCIAS

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

\_\_\_\_\_. Requisitando o conceito de representação. *In: Relatório de estágio pós-doutoral*. Montreal: [s.n.], 2006.

CARDOSO, Olinda Nogueira Paes. **Bancos de Dados Objetos-Relacionais em Oracle**. Lavras: UFLA, 2003. (Apostila de curso).

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

CENDÓN, Beatriz Valadares. Sistemas e redes de informação. *In: Oliveira, Marlene de. (Coord.). Ciência da informação e biblioteconomia: novos conteúdos e espaço de atuação*. Belo Horizonte: Editora UFMG, 2005. p. 61-95.

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

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

CORDEIRO, Darcy. **Ciência, pesquisa e trabalho científico: uma abordagem metodológica**. 2.ed.rev.aument. Goiânia: UGC, 1999.

DICIONÁRIO Larousse da Língua Portuguesa. São Paulo: Larousse do Brasil, 2005.

FERNEDA, E. **Recuperação de informação: análise sobre a contribuição da Ciência da Computação para a Ciência da Informação**. [S.l: s.n.], [20-?]. Disponível em: <<http://ferneda.sites.uol.com.br/Trabalhos>>. Acesso em: 14 jun. 2007.

FOSKETT, A.C. **A abordagem temática da informação**. São Paulo, Polígono: Brasília, Ed. da UnB, 1973.

FUJITA, Mariângela Spotti Lopes; RUBI, Milena Polsinelli. O ensino de procedimento de política de indexação na perspectiva do conhecimento organizacional: uma proposta de programa para a educação à distância do bibliotecário. **Perspectivas em Ciência da Informação**, Belo Horizonte, v.11, p.48-66, jan/abr.2006. Disponível em: <<http://www.eci.ufmg.br/pcionline/>>. Acesso em: 10 out. 2006.

GIL, Antônio Carlos. **Métodos e técnicas de pesquisa social**. 5.ed. São Paulo: Atlas, 1999.

GISMONDI, Ronaldo Curi. **Utilizando descritores e qualificadores em pesquisas bibliográficas via Internet**. [S.l: s.n.], [20-?]. Disponível em: <[http://www.lampada.uerj.br/lampada/ementas/aulas/info\\_med/DescritoresPesquisaBiblio\\_Internet.pdf](http://www.lampada.uerj.br/lampada/ementas/aulas/info_med/DescritoresPesquisaBiblio_Internet.pdf)> Acesso em: 16 out. 2006.

LANCASTER, F. W. **Indexação e resumos: teoria e prática**. Tradução de Antonio Agenor Briquet de Lemos. 2. ed. rev. atual. Brasília: Briquet de Lemos/Livros, 2004.

LE COADIC, Yves-François. **A Ciência da Informação**. 2ª ed. Brasília, DF: Briquet de Lemos/Livros, 2004.

LIMA, Gercina, Ângela Borém. Organização da informação para sistemas de hipertextos. *In*: NAVES, Madalena Martins Lopes; KURAMOTO, Hélio (Org.) **Organização da informação: princípios e tendências**. Brasília/DF: Briquet de Lemos/Livros, 2006.

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

LUCAS, C. R. A organização do conhecimento e tecnologia da informação. **Transinformação**, v. 8, n. 1, 1996.

MILANÉS, Franklin Marin; VELÁSQUEZ, Alexeis Torres. *Uma mirada em torno al desarrollo de la recuperación de información*. Data Grama Zero - **Ciência da Informação**, v.7 n.4 ago/2006. Disponível em: <<http://www.dgz.or.br/>>. Acesso em: 12 set. 2006.

NAVES, Madalena Martins Lopes; KURAMOTO, Hélio (Org.) **Organização da informação: princípios e tendências**. Brasília/DF: Briquet de Lemos/Livros, 2006.

PARREIRAS, Fernando. O uso de sintagmas nominais como fonte de descritores para textos de periódicos científicos. **Escola de Ciência da Informação**. Belo Horizonte, 2003. Disponível em: <<http://www.fernando.parreiras.nom.br/publicacoes/sn.pdf>>. Acesso em: 12 set. 2006.

PINHEIRO, Lena Vânia Ribeiro. Medidas de consistência da indexação: interconsistência. **Ciência da Informação**, Rio de Janeiro, v.7, n.2, p.109-114, 1978.

ROBREDO, Jaime; CUNHA, Murilo B. da. Representação do conteúdo dos documentos. *In*: \_\_\_\_\_. **Documentação de hoje e de amanhã: uma abordagem informatizada da Biblioteconomia e dos sistemas de informação**. São Paulo: Global, 1994, p.201-263.

ROWLEY, Jennifer. Fundamentos da recuperação da informação. *In*: \_\_\_\_\_. **A biblioteca**

eletrônica. Brasília: Briquet de Lemos/Livros, 2002, p.161-186.

SOUZA, Renato Rocha. Sistema de Recuperação de Informações e mecanismos de busca na *web*: panorama atual e tendências. **Perspectivas em Ciência da Informação**, Belo Horizonte, v.11, n.2, p.161-173, maio/ago.2006. Disponível em: <<http://www.eci.ufmg.br/pcionline/>>. Acesso em: 10 dez. 2006.

\_\_\_\_\_. **Uma proposta de metodologia para escolha automática de descritores utilizando sintagmas nominais**. 2005. 214p. Tese (Doutorado em Ciência da Informação) - Escola de Ciência da Informação da UFMG, Belo Horizonte.

TRIVIÑOS, Augusto N. S. **Introdução à pesquisa em Ciências Sociais: a pesquisa qualitativa em Educação**. 1. ed. 14 reimp. São Paulo: Atlas, 2006.



Senhores (as),

Em razão da necessidade de elaborar uma monografia de conclusão de curso de Biblioteconomia da Universidade Federal do Ceará, sob a orientação da Profa. Dra. Luciene Mendes Pinna, estou realizando esta pesquisa cujo objetivo principal é verificar se há uma mudança de decisão de compra de livros entre os pesquisadores do Instituto de Física, em função da abrangência de assuntos pesquisados.

Neste sentido, gostaria de solicitar a sua colaboração para que seja possível obter respostas para as questões propostas a seguir.

Desde já peço a sua atenção e agradeço desde já a sua colaboração.

QUESTIONÁRIO

1. Qual é o seu cargo atual e suas etapas?
2. Mas, antes de responder as questões, gostaria de saber se você já fez uma indicação bibliográfica.

APÊNDICE

1. Qual o seu cargo atual e suas etapas?
2. Mas, antes de responder as questões, gostaria de saber se você já fez uma indicação bibliográfica.
3. Quando você fez a indicação bibliográfica, qual o assunto que você indicou?
4. Qual o tipo de livro que você indicou?
5. Qual o nível de leitura que você indicou?
6. Qual o tipo de livro que você indicou?
7. Quando você fez a indicação bibliográfica, qual o assunto que você indicou?
8. Qual o tipo de livro que você indicou?
9. Qual o nível de leitura que você indicou?
10. Qual o tipo de livro que você indicou?

Prezados (as) Senhores (as),

Em razão da necessidade de elaborar uma monografia de conclusão do Curso de Biblioteconomia da Universidade Federal do Ceará, sob a orientação da Profa. Dra. Virgínia Bentes Pinto, estamos realizando uma pesquisa cujo objetivo geral é analisar as formas de tomada de decisão do indexador na escolha de descritores para os Sistemas de Recuperação de Informação em bibliotecas universitárias.

Neste sentido, gostaríamos de contar com a sua valiosa colaboração, assegurando-lhe que suas respostas serão utilizadas somente para este fim.

Sem mais para o momento, agradecemos pela sua atenção.

Cordialmente,

Maria Nilma Rodrigues da Costa

### QUESTIONÁRIO

1. Qual a sua concepção sobre indexação e suas etapas?
  
2. Marque os requisitos (pode marcar mais de um) para o indexador fazer uma indexação adequada.
  - ( ) Examinar o documento e estabelecimento do assunto de seu conteúdo.
  - ( ) Considerar o título e subtítulo do documento.
  - ( ) Identificar os conceitos presentes no assunto.
  - ( ) Traduzir os conceitos nos termos de uma linguagem de indexação.
  - ( ) Levar em consideração o resumo, se houver.
  - ( ) Ler a introdução, prefácio, orelha, sumário, apresentação do documento.
  - ( ) Considerar as referências bibliográficas.
  
3. Marque os procedimentos a serem seguidos na extração de termos de indexação.
  - ( ) Verificar qual o assunto de que trata o documento.
  - ( ) Considerar se o documento se refere a métodos, técnicas e instrumentos especiais.
  - ( ) Ao fazer a escolha dos conceitos, o indexador deve ter em mente as consultas que podem ser feitas ao sistema de informação.
  - ( ) O indexador não precisa necessariamente representar com termos de indexação todos os conceitos identificados durante o exame do documento.
  - ( ) Escolher os conceitos que forem considerados os mais apropriados para uma

determinada comunidade de usuários.

( ) Na escolha dos termos ou descritores, levar em conta a política de indexação.

4. Quais os fatores que interferem no trabalho do indexador?
5. Quais as dificuldades encontradas na escolha de termos?
6. Em que as dificuldades enfrentadas pelo indexador influenciam no sistema de recuperação da informação?
7. Caso julgue necessário, por favor, teça algum comentário.