

BIBLIOTECA DIGITAL PAULO FREIRE: APLICABILIDADE DO *THESAURUS* BRASED NA INDEXAÇÃO

Fabiana da Silva França
Maria Elizabeth B. C. de Albuquerque

Resumo

A busca de informações, em meios digitais, requer dos usuários certo conhecimento da linguagem utilizada pelos sistemas de informação. Isso faz com que algumas instituições disseminadoras de informações tenham uma atenção especial no tratamento dessas informações. Para este estudo, cujo objetivo é o estudo da utilização do *Thesaurus* Brased na Indexação da Biblioteca Digital Paulo Freire (BDPF), enfocou-se a indexação de documentos na padronização da linguagem, a partir do Tesouro. Para responder a este objetivo, realizou-se uma explanação sobre os temas: Sistemas de Recuperação da Informação (incluindo Bibliotecas Digitais, Indexação e Linguagem Documentária), o Tesouro e sua utilização na indexação de documentos, inclusive o *Thesaurus* Brased. Conclui-se que existe certa rigidez no processo de indexação da BDPF e que a padronização da linguagem de uma área específica do saber, facilita a busca de informações pelos usuários.

Palavras-chave

Sistema de Recuperação de Informação; Indexação; Tesouro; *Thesaurus* Brased

1 INTRODUÇÃO

A chamada sociedade da informação vem provocando um turbulento crescimento da informação e do uso das tecnologias de informação e comunicação (TICs), prin-

cipalmente se direcionamos nosso olhar para a busca de informações digitais.

A busca de informações requer dos usuários uma relação de proximidade entre **o quê** efetivamente se quer, **onde** e **como** procurar.

Trabalhando fatores que envolvem o tratamento da informação, esperamos diminuir as dificuldades de busca e recuperação da informação para aqueles usuários que não possuem o domínio da terminologia específica, principalmente, quando se trata de uma área do saber, em questão.

Percebemos que, em face à crescente demanda de informações eletrônicas, novos tipos de instituições estão sendo construídas no ciberespaço, a exemplo das bibliotecas que, deste modo, vêm diversificando seus serviços, a fim de suplantar o modelo tradicional.

Toda a evolução que as bibliotecas vêm sofrendo com o avanço das TICs, também é marcada pela transferência de informação para os novos suportes, no caso do formato digital, pelo suporte denominado hipertexto, em que o acesso acontece de maneira aleatória, não seqüencial, fazendo com que o usuário obtenha informações em diferentes formatos através do simples clicar em *links*.

Apesar da diversidade de formas de acesso à informação, as bibliotecas recorrem ao uso de instrumentos como listas de cabeçalhos, tesouros e outros, com o intuito de controlar o vocabulário e facilitar a recuperação da informação, como é o caso da Biblioteca Digital Paulo Freire (BDPF) que, entre muitas bibliotecas digitais já desenvolvidas, merece destaque visto que é nosso campo de estudo.

A Biblioteca Digital Paulo Freire (BDPF) está sendo desenvolvida por um grupo multidisciplinar de professores e alunos do Departamento de Informática, do Departamento de Biblioteconomia e Documentação e do Programa de Pós-graduação em Educação, além de parcerias com a Universidade Federal de Pernambuco e com o Centro Paulo Freire - Estudos e Pesquisa, do Estado de Pernambuco.

A política de indexação da BDPF, elaborada por professores do Departamento de Biblioteconomia e Documentação, estabeleceu a utilização do *Thesaurus* Brased do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira – INEP para indexação e padronização dos termos, por se tratar de uma biblioteca cujos conteúdos abrangem mais diretamente a área de Educação.

O trabalho de definição de termos e de manuseio dos instrumentos de padronização e recuperação de informação, estimulou o aprofundamento de estudos sobre a utilização do *Thesaurus* Brased na indexação da Biblioteca Digital Paulo Freire (BDPF).

2 BIBLIOTECA DIGITAL PAULO FREIRE

Para implementar o uso das tecnologias de informação e comunicação na área acadêmica, a Universidade Federal da Pa-

raíba (UFPB) inseriu sistemas de informação para disponibilizar documentos digitais na Internet, com objetivo de construir um canal direcionado a atender toda a comunidade científica.

Uma das iniciativas da UFPB foi o projeto “Pólo de produção e capacitação em conteúdos digitais multimídias da Paraíba”, financiado pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). Esse projeto objetivou o processo de aquisição e consolidação de competência para conceber, implementar e avaliar serviços de recuperação de informação baseados em bibliotecas digitais multimídia. A proposta apresentou a implementação de três bibliotecas digitais temáticas: a Biblioteca Digital da Organização Não-Governamental Para’iwa – Cultura, Imagem e Ação (OnG Para’iwa) a Biblioteca Digital do Núcleo de Documentação e Informação Histórica Regional (NDHIR); e a Biblioteca Digital Paulo Freire (BDPF).

Este projeto se configura, então, como iniciativa de fortalecimento de infra-estrutura de aplicações da tecnologia digital na educação e na pesquisa, com vistas a minimizar problemas de geração e difusão de inovações (BEZERRA; BRENNAND, 2000, p. 2)

Corroborando com esse pensamento Aquino (2001, p. 4) ressalta que,

A aplicação da tecnologia (das redes eletrônicas) no campo educacional, sem dúvida, trará mudanças no campo dos processos de ensino-aprendizagem, na postura dos professores, na participação dos alunos, no trato com a informação, enfim na mudança do paradigma educacional.

Nessa perspectiva, a BDPF tem como principal objetivo “disponibilizar pressupostos filosóficos, sociológicos e pedagógicos do pensamento freireano, para dar suporte a ações educativas democráticas” (BRENNAND, 2000, p. 7).

Bezerra (2003, p.36) ressalta,

[...] a idéia de concepção e implementação da Biblioteca Digital Paulo Freire – biblioteca digital de personalidade –, articula esta perspectiva aos objetivos do Programa Sociedade da Informação em âmbito nacional, decorrendo na disponibilização (*on-line*) de uma ferramenta para a divulgação de informações a respeito de uma das mais ilustres personalidades do cenário da educação mundial: Paulo Freire.

Para disponibilização e tratamento das informações nessa biblioteca, fez-se necessário a criação de equipes específicas e especializadas.

Podemos visualizar na Figura 1, a página de entrada da BDPF, com acesso pelo endereço eletrônico <http://www.paulofreire.ufpb.br>, cuja operacionalização localiza-se, atualmente, no Pólo Digital da UFPB, Campus I.



Figura 1 – Biblioteca Digital Paulo Freire (BDPF)
Fonte: BDPF (2005)

A BDPF reúne, em formato digital, um acervo com uma diversidade de documentos relacionados à vida e à obra do educador Paulo Freire. Esses documentos referem-se a todas as obras impressas e não impressas, independentemente do formato. A transferência de documentos, do formato impresso para o eletrônico, efetiva-se com o processo de digitalização.

2.1 Acervo

O acervo da BDPF é formado de produções nacionais e internacionais relacionadas à vida e à obra de Paulo Freire, e está estruturado da seguinte forma:

a) A obra – Produção do próprio educador Paulo Freire (15 livros);

b) A crítica – Produção de autores que escreveram e escrevem sobre a vida e a obra do educador Paulo Freire (9 artigos de jornal, 34 artigos de revista, 1 correspondência, 16 resumos de dissertação, 6 livros, 1 palestra, 10 resenhas, 15 resu-

mos, 89 seminários, 1 texto didático e 28 outros documentos);

c) Multimídia: áudio, vídeo e imagem de Paulo Freire, dividida em duas categorias:

- **A Obra** com áudio, imagem e vídeo do educador Paulo Freire (6 vídeos, 20 áudios, 38 imagens) e;
- **A Crítica** com áudios e vídeos, depoimentos relacionados ao educador Paulo Freire (2 áudios, 14 vídeos).

2.2 Índice de cabeçalhos de assunto

Para fazer a busca de informações por assunto, a BDPF coloca à disposição dos usuários um índice alfabético, ou seja, uma relação de termos que representa o conteúdo dos documentos, facilitando assim o acesso ao documento de forma rápida e precisa. Este índice é formado de cabeçalhos de assunto e suas respectivas remissivas.

2.3 Glossário

O glossário (Figura 2) elaborado pelo grupo de Indexação da BDPF tem como

principal finalidade a recuperação da informação, sobretudo para os usuários que se interessam pela temática e não dominam a terminologia utilizada.

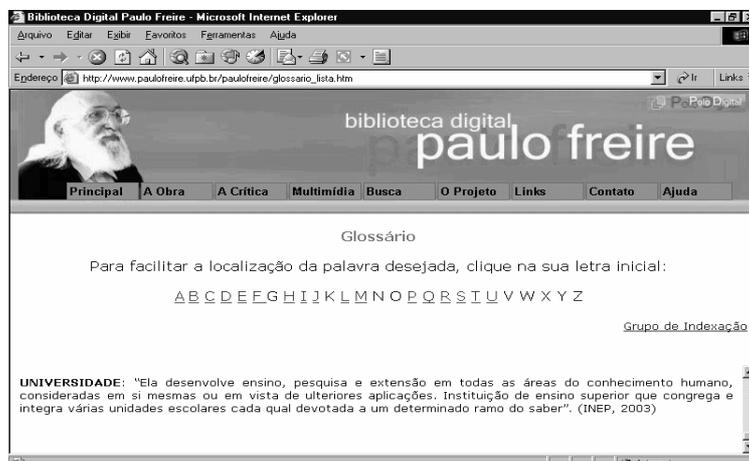


Figura 2 – Glossário da BDPF
Fonte: BDPF, 2005

O glossário faz referência a cada cabeçalho que se encontra no índice dessa biblioteca, acompanhado de seu significado, buscando, dentro do possível, expressar as idéias de Paulo Freire.

Para definição dos termos do glossário, o Grupo de Indexação recorreu ao *Thesaurus* Brased do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira – INEP, também a outros autores como Moacir Gadotti em sua obra intitulada “Paulo Freire: uma biobibliografia”, aos textos de Paulo Freire e aos de especialistas que escrevem ou escreveram sobre o referido educador.

Algumas definições sobre o assunto estão descritas tal como se apresentam no documento e outras foram, sempre que necessário, adaptadas ou elaboradas pelo

Grupo de Indexação com a revisão de um especialista da área de Educação.

2.4 Política de indexação da Biblioteca Digital Paulo Freire

A política de indexação é uma decisão administrativa e, de acordo com Carneiro (1985, p. 221), ela deve servir de guia para tomadas de decisões, levando em conta os seguintes fatores:

[...] características e objetivos da organização determinantes do tipo de serviço a ser oferecido; identificação dos usuários, para atendimento de suas necessidades de informação e recursos humanos, materiais e financeiros, que delimitam o funcionamento de um sistema de recuperação de informações.

Para Lancaster (2004), todas as decisões referentes à política de indexação devem ser tomadas pelos gestores do ser-

viço de informação, ficando, deste modo, fora do controle do indexador individual.

Carneiro (1985), citado por Rubi (2004), destaca os elementos que compõem uma política de indexação:

a) Cobertura de assuntos – assuntos cobertos pelo sistema;

b) Seleção e aquisição dos documentos-fonte – cobertura do sistema em áreas de assunto de seu interesse e qualidade dos documentos;

c) Processo de indexação – refere-se aos níveis de exaustividade e especificidade requeridos pelo sistema, escolha da linguagem, capacidade de revocação e precisão do sistema;

d) Estratégia de busca – decisão entre a busca delegada ou não;

e) Tempo de resposta do sistema – identificação do tempo permitido para ser consumido no momento da recuperação;

f) Forma de saída – definição da forma em que os resultados de busca são apresentados;

g) Avaliação do sistema – determinará até que ponto o sistema satisfaz as necessidades dos usuários.

No ano 2003 foi elaborada a Política de indexação da Biblioteca Digital Paulo Freire, com os seguintes objetivos: a) Recuperar conteúdos através da identificação

de termos (cabeçalhos de assunto), identificando o assunto; b) Produzir um índice com os cabeçalhos atribuídos; e c) Produzir um glossário, utilizando os termos atribuídos aos cabeçalhos, em consonância com as idéias de Paulo Freire (ALBUQUERQUE, 2004).

Os objetivos de uma política são a definição das variáveis que afetam o desempenho do serviço de indexação, o estabelecimento dos princípios e critérios que servirão de decisões para otimização do serviço, a racionalização dos processos e a consistência das operações envolvidas. (STREHL, 1998, p.2)

A elaboração de tal política considerou quatro elementos: os usuários, o acervo, os critérios e a avaliação.

h) Os usuários: estudantes, docentes, pesquisadores e outros indivíduos interessados na temática de Paulo Freire;

i) O acervo: como visto anteriormente, contém todas as informações relacionadas à vida e obra de Paulo Freire;

j) Os critérios: foram definidos critérios de seleção de cabeçalhos, números de palavras por cabeçalho; uso de singular e plural; sinônimos; cabeçalhos compostos; termos homógrafos; período; indicadores geográficos; e divisão de forma;

k) A avaliação: feita através da opinião do usuário considerada como indispensável. Nesse intuito, a BDPF disponibiliza, no seu site, um espaço para os usuários interagirem, através de sugestões ou para tirar dúvidas.

3 SISTEMAS DE RECUPERAÇÃO DA INFORMAÇÃO

O termo Sistema de Recuperação da Informação (SRI), costuma se referir a um sistema de computador, constituído de *hardware*, *software* e base de dados, exercendo várias atividades que podem ser classificadas, conforme suas funções de *input* (entrada) e *output* (saída) de informação, apresentadas em três estágios: indexação, armazenagem e recuperação.

Os sistemas de recuperação da informação foram tradicionalmente desenhados para oferecer acesso à informação a partir de distintos documentos. [...] Mais recentemente, aplicações para recuperação da informação têm fornecido acesso à informação contida nos documentos eletrônicos, especialmente documentos baseados em textos, mas cada vez mais sistemas têm incluído alguns recursos para acesso a documentos multimídia e acesso à informação e objetos com documentos multimídia (tais como fotos e vídeo-clips) (ORTEGA, 2002, p. 59).

Rowley (2002, p. 9) também esclarece que os SRI(s) foram projetados para “proporcionar acesso a informações e não a documentos”. Os SRI(s) devem representar o conteúdo dos documentos, nos quais os usuários, através de uma expressão de busca, obtêm uma rápida seleção dos itens de interesse. No entanto, para que isso ocorra é necessário que estes se familiarizem com os diversos recursos de busca, para otimização da utilização do sistema.

Para uma correta utilização de um SRI, existem alguns elementos que devem ser necessariamente dominados pelos usuários. Kuramoto (1995, p. 2) apresenta esses elementos, como sendo:

- um certo número de comandos para se colocar o sistema em modo de consulta, para se escrever uma expressão de busca dentro de uma sintaxe correta, para visualizar o resultado de uma consulta etc.;
- a indispensável lógica booleana em uma consulta de múltiplos critérios a qual, para oferecer bons resultados, exige a utilização de operadores como intersecção, união, exclusão, comparação, de proximidade etc.;
- a estrutura conceitual da base de dados, os nomes de campos a consultar e as convenções de escrita de cada um destes campos;
- os termos de indexação, os léxicos, os tesouros etc.

Os recursos oferecidos pelos sistemas devem ser aproveitados pelos usuários de forma que lhes permita atingir uma busca bem-sucedida, pois cada usuário tem sua necessidade de informação e deve converter essa necessidade em algum tipo de estratégia de busca (LANCASTER, 1993), ou seja, um conjunto de decisões e ações adotadas durante a busca.

Para planejar uma estratégia de busca com maior nível de complexidade, os SRI(s) necessitam envolver vários conceitos na mesma estratégia como, por exemplo, permitir a busca de palavras por títulos dos documentos, isto é, termos da linguagem natural; buscar termos específicos de linguagens controladas, nos campos de

cabeçalhos de assunto; buscar por autores; por ano de publicação; por classificação; permitir a busca de conceitos compostos ou simples e a possibilidade de truncamento de palavras e de substituição de caracteres no meio dos termos, dentre outros recursos de recuperação (LOPES, 2002).

Definindo-se esses conceitos, a procura de documentos com informações úteis torna-se mais freqüente. A partir daí, o SRI definirá suas principais características de revocação e de precisão. Entende-se por revocação a capacidade de recuperar documentos úteis e por precisão a capacidade de evitar documentos inúteis (LANCASTER, 2004).

Diante do exposto, algumas bibliotecas digitais tentam aperfeiçoar a eficiência e a eficácia de seus SRI(s), através de suas interfaces, deixando-as mais amigáveis e ainda, facilitando os diferentes tipos de busca que seus usuários, sejam eles inexperientes ou experientes, lançam mão. Como exemplo, tomemos a BDPF que disponibiliza menus como: **busca por assunto** (ao fazer a busca por assunto, o usuário deve clicar em algum cabeçalho de assunto, desse modo serão recuperados todos os documentos associados com a palavra clicada); **busca avançada** (quando o usuário opta por realizar uma busca avançada, ele deve preencher pelo menos um dos campos que o sistema disponibiliza – ca-

beçalho de assunto, título, autor, formato, tipos de documento); **busca genérica ou geral** (ao fazer a busca genérica, o usuário deve digitar alguma palavra chave – seja ela o nome de um autor, um cabeçalho de assunto, ou um título de uma obra – desse modo serão recuperados todos os documentos associados com a palavra digitada).

Notamos, claramente, que as tecnologias de informação e comunicação (TICs) atuais influenciam as bibliotecas na mudança de sua concepção histórica de depósito de livros, para instituições voltadas à disseminação de informações mais específicas, de forma dinâmica e eficaz, disponibilizando uma combinação de materiais convencionais e eletrônicos. O que antes era disponível, hoje, deve se tornar acessível, e a Internet é um elemento facilitador, para, através dela, buscar e tratar as informações no sentido de disponibilizá-las, como é o caso das bibliotecas digitais.

3.1 Bibliotecas Digitais

Ao refletirmos sobre a biblioteca digital, temos como precursores Vannevar Bush e Theodore Nelson, idealizadores do hipertexto.

Em 1945, o cientista americano Vannevar Bush, levando em consideração, não somente, o aumento da produção e do registro de informação, mas também seu ar-

mazenamento, consulta e seleção, antecipou a noção de um repositório de informação, ao qual chamou de *MEMEX (Memory Extension)*. Ele idealizou e delineou, com detalhes, no artigo intitulado “*As we may think*”, uma máquina capaz de armazenar informações, de forma fácil e veloz, com acesso através de uma tela de televisão com alto-falantes. O efeito seria de uma extensão da memória humana o que facilitaria a disseminação da informação científica.

Entretanto, foi apenas nos anos sessenta que Theodore Nelson consolidou essa idéia, empregando o termo hipertexto em sistema de informática, para demonstrar a idéia de escrita/leitura não seqüencial, no Projeto denominado XANADU. O autor imaginou que as pessoas poderiam ter acesso a qualquer informação, de qualquer lugar, conectadas a uma grande rede contendo todo o saber literário e científico, onde seriam armazenados os textos completos de documentos (LÉVY, 1993). Essa noção de armazenamento de textos completos em rede e de leitura não seqüencial concretiza-se com as bibliotecas digitais.

As bibliotecas digitais são sistemas de informação que existem somente na forma digital, ou seja, dispõem de todos os recursos de uma biblioteca eletrônica, oferecendo pesquisa e visualização dos documentos (*full text*, imagem, áudio etc) por

meio de redes de computadores (MARTINS, 2005).

Em relação ao conceito de biblioteca digital, percebemos que, na literatura, existe um consenso entre alguns autores como Marchiori (1997), Macedo e Modesto (1999), Pereira e Rutina (1999), Zang et al (2000), no que se refere à existência da informação **apenas na forma digital**, o que é uma característica importante para diferenciá-la das demais.

Discorremos, então, sobre o conceito, apresentado por Marchiori (1997, p. 4):

A biblioteca digital difere-se das demais porque a informação que ela contém existe apenas na forma digital, podendo residir em meios diferentes de armazenamento, como as memórias eletrônicas (discos magnéticos e ópticos). Desta forma, a biblioteca digital não contém livros na forma convencional e a informação pode ser acessada, em locais específicos e remotamente, por meio de redes de computadores. A grande vantagem da informação digitalizada é que ela pode ser compartilhada instantaneamente e facilmente, com um custo relativamente baixo.

A definição aborda aspectos de tempo e espaço, a partir do momento em que o objeto é digitalizado e compartilhado; a busca de informação acontece em qualquer que seja o local de acesso.

A disponibilização e compartilhamento dos documentos contidos na biblioteca digital, são realizados de forma processável por computador, isto é, convertidos por

digitalização ou por reconhecimentos de caracteres óticos, digitalizados por *scanners*.

Suas coleções são formadas por documentos que contêm textos na íntegra, multimídia, imagens digitais, áudio-visuais, entre outros, com conteúdos e arquivos transferíveis quantas vezes forem necessárias, bastando apenas conexão ou autorização através de *LANs (Local Area Network)* ou *WANs (Wide Area Network)*.

Conforme Ortega (2002, p. 53) “Uma vez acessada, materiais da coleção de uma biblioteca digital podem ser visualizados, impressos, baixando ou manipulando de outro modo para satisfazer necessidades particulares de usuários”.

Na construção dessas bibliotecas, existem grupos cada vez mais específicos. Isso significa que quando áreas de vários domínios do saber se concentram para construir uma biblioteca digital, trazem consigo suas terminologias, suas linguagens, com o objetivo de,

[...] selecionar, estruturar, oferecer o acesso intelectual, interpretar, distribuir, preservar a integridade e assegurar a persistência integral das coleções de trabalhos digitais, de modo que estejam prontamente e economicamente disponíveis para o uso por uma comunidade definida ou conjunto de comunidades. (WATERS, 1998)

Para que ocorra essa integração de saberes, as bibliotecas digitais passam a assumir algumas funções citadas por Fox (1998): Suporte de forma colaborativa;

Preservação de documento digital; Gerenciamento de base de dados distribuída; Hipertexto; Filtros de informação; Recuperação de informação; Módulos e instrução; Gerenciamento de direitos autorais; Serviços de informação multimídia; Serviços de referência e respostas às questões enviadas; Busca de recursos; Disseminação seletiva.

Fica evidente, então, a superação de paradigmas que as bibliotecas devem ultrapassar com a utilização de novas tecnologias. Com isso, mudam também os processos, surgindo assim novos desafios, dentre os quais podemos citar os serviços que envolvem os processamentos técnicos, principalmente, a indexação.

3.2 Indexação

A indexação é um subsistema de entrada do sistema de recuperação da informação, resultado da necessidade de construção de índices. Atualmente, essa concepção está relacionada ao processo de tratamento da informação que permite a análise, a síntese e a representação dos documentos.

Antes de falarmos das etapas da indexação, vamos apontar alguns conceitos relacionados a ela. O Sistema Mundial de Informação Científica (UNISIST, 1981, p. 148) define a indexação como a “representação do conteúdo dos documentos, por

meio de símbolos especiais quer retirados do texto original (palavras-chave extraídas do documento), quer escolhidos numa linguagem de informação ou de indexação”.

Para Van Slype (1991) indexação é a “operação que consiste em enumerar os conceitos sobre os quais trata um documento e representá-los por meio de uma linguagem combinatória: lista de descritores livres, lista de autoridades e o *thesaurus* de descritores”.

Esteban Navarro (1999), citado por Silva e Fujita (2004, p. 137), conceitua indexação como

[...] um processo a identificar e descrever ou caracterizar o conteúdo informativo de um documento mediante a seleção das matérias sobre as quais versa (indexação sintética) ou dos conceitos presentes (indexação analítica) para sua expressão da língua natural e sua reunião em índice, com objetivo de permitir posterior recuperação dos documentos pertencentes a uma coleção documental ou conjunto de referências documentais como resposta a uma demanda acerca do tipo de informação que este contém.

Segundo Lancaster (2004), a indexação implica na preparação de uma representação do conteúdo temático dos documentos. O autor divide o processo de indexação em duas principais etapas que ocorrem de modo simultâneo: a análise conceitual e a tradução.

a) **Análise conceitual** – indica sobre o que trata o documento, ou seja, a qual assunto está relacionado;

b) **Tradução** – abrange a conversão da análise conceitual de um documento, num determinado conjunto de termos de indexação para o vocabulário do sistema.

Para Vieira (1998), citado por Paiva (2002, p. 52) na indexação manual “os conceitos são extraídos, através de um processo de análise intelectual”.

Bentes Pinto (2001) afirma que a indexação manual é realizada intelectualmente por um ser humano, baseando-se no julgamento dos indexadores, do texto e das necessidades da comunidade de usuários.

Na indexação automática, o processo de indexação é realizado com o uso de computador; Guimarães (2000, p. 1) aponta três concepções, levando em consideração como esse uso é realizado:

– a primeira está relacionada pelo uso de programas informáticos que dão suporte ao armazenamento dos termos de indexação obtidos pela análise conceitual;

- a segunda pelo uso dos sistemas que analisam documentos de forma automática com validação dos termos por um profissional (indexação semi-automática);

- e a terceira, refere-se a indexação automática propriamente dita, sendo aquela realizada pelos programas de computador sem nenhum tipo de validação por profissionais.

Os termos criados pela indexação automática, propriamente dita, não são de muita qualidade, no entanto, são mais baratos e apresenta forma muito mais rápida

(obviamente). Isso faz com que algumas instituições adotem esse processo.

O modelo adotado no processo de indexação da BDPF é a indexação realizada por profissionais, usa-se o computador apenas para armazenar os termos indexados em programas específicos. Os profis-

sionais recebem os documentos a serem indexados, para análise, e utilizam programas de computador para armazenar os cabeçalhos indexados, que são disponibilizados ao usuário através de índices de cabeçalhos de assunto e na forma de glossário com seus respectivos conceitos, processo esse representado na Figura 3.



Figura 3 - Processo de indexação realizado pelos indexadores da BDPF
Fonte: Dados da pesquisa

Na segunda etapa da indexação, a **tradução**, Lancaster (2004) faz uma distinção quanto à indexação, dividindo-a em dois tipos: indexação por extração (são retiradas expressões do próprio documento para representar o conteúdo) e indexação por atribuição (são atribuídos termos de uma outra fonte aos documentos, para representar seu conteúdo, a exemplo dos vocabulários controlados).

A indexação é uma operação que lida com idéias a serem transmitidas e envolve vários fatores, dentre eles, o conhecimento do indexador no assunto a ser trabalhado. Isso não quer dizer que o indexador deve ser um especialista na área, contudo deve conhecer a linguagem a ser utilizada. Caso contrário, podem ocorrer falhas de indexação tanto na análise conceitual (deixar de reconhecer algum tópico de interesse do usuário; interpretação errônea do conteúdo

do documento), como na tradução (deixar de utilizar termos específicos da área; empregar termos que não são da área). Dessa forma, faz-se necessário conhecer as linguagens de indexação, item a ser abordado a seguir.

3.3 Linguagem Documentária

Os SRI(s) utilizam instrumentos para que ocorra comunicação com o usuário, instrumentos esses que são chamados de Linguagem Documentária (LD). Essas linguagens recebem várias denominações como: linguagens de indexação (Melton, J.); linguagens descritoras (Vickery, B); codificações documentárias (Grolier, E.); linguagens de informação (Soergel); vocabulários controlados (Lancaster, F.W); listas de assuntos autorizados (Montgomery, C.).

As linguagens documentárias são linguagens “construídas para indexação, armazenamento e recuperação da informação e correspondem a sistemas de símbolos destinados a ‘traduzir’ os conteúdos dos documentos” (CINTRA et al., 2002, p.33).

Gardin (1981), citado por Cintra (2002, p.25), trata a linguagem documentária como “um conjunto de termos providos ou não de regras sintáticas, utilizadas para representar conteúdos de documentos téc-

nico-científicos, com fins de classificação ou busca de informação”.

Essas linguagens precisam de alguns elementos para comunicação necessária com o usuário, ou seja, necessitam de vocabulário, estrutura (elementos sintéticos) e sintaxe.

Para Gomes e Campos (1998), o vocabulário e a estrutura ocorrem no interior da LD e a sintaxe no momento da indexação. Entende-se por vocabulário, os “termos de linguagem natural, incluindo os controles de harmonia e sinonímia”; por estrutura, “a rede de relações hierárquicas entre os termos” e por sintaxe a “ordem que estes termos precisam ter para uma representação adequada dos documentos”.

As linguagens documentárias podem ser classificadas, seguindo dois critérios:

- a) Pela ordenação dos conceitos – como pré ou pós-coordenadas;
- b) Pela forma de apresentação – ordem sistemática ou alfabética.

As linguagens pré-coordenadas combinam ou coordenam os termos no momento da indexação, ou seja, o indexador coordena os assuntos, quando faz o tratamento da informação; como exemplo, temos os cabeçalhos de assunto.

Nas linguagens pós-coordenadas, os usuários coordenam os assuntos no momento em que fazem a busca da informação. Nesse tipo de linguagem, o usuário

tem a possibilidade de realizar todo o processo para chegar à informação desejada, combinando os termos. Um exemplo clássico é o tesouro.

4 TESAURO

Thesaurus é uma expressão inglesa que foi empregada a partir de 1500, para indicar um acervo ordenado de informações e conhecimentos. Uma das primeiras obras a incluir esta expressão no seu título foi o *Thesaurus linguae romanae et britannicae*, publicado em 1565, autoria de Cooper (CAVALCANTI, 1978). Popularizou-se, no ano de 1852, em Londres, com a publicação do *Thesaurus of English Words and Phrases*, de Peter Mark Roget, que utilizava a palavra significando uma coleção de termos organizados, de acordo com as idéias que expressava cada texto, conforme esclarece o subtítulo da obra “palavras classificadas para facilitar idéias e para ajudar na composição literária”.

O *thesaurus* de Peter Mark Roget foi organizado considerando-se a estrutura classificatória, ou seja, um sistema de classificação de idéias e um índice alfabético dos cabeçalhos, dos quais ocorrem as palavras e as frases que representam as idéias. Constituído em seis categorias: as relações abstratas (existência, ordem, número, tempo, etc); o espaço (movimento, mudança de lugar); a matéria (solidez, calor, luz); o intelecto (aquisição, comunica-

ção de idéias); vontade (escolha, ação, intenção); afeições (sentimentos morais e religiosos).

Em 1940, a palavra tesouro foi utilizada em Ciência da Informação, na recuperação de informação. Todavia, foi no ano 1951 que Mortimer Taube consolidou a idéia com o desenvolvimento do Sistema Unitermo, cuja finalidade era representar o assunto por palavras únicas chamadas de unitermos, introduzindo o acesso múltiplo, com arquivos manuais de fichas. As buscas eram feitas com uma comparação entre os números que eram colocados nas fichas. O principal problema desse sistema era não possuir controle de vocabulário, fazendo com que a busca se tornasse exaustiva, pois não se tinha idéia de como aquele documento teria sido representado no texto, problema esse que Lancaster (1993) assevera como “o retorno aos vocabulários controlados e o desenvolvimento do tesouro na recuperação de informação” como solução. Lancaster considera o sistema unitermo responsável pelo aparecimento de uma nova linguagem – o tesouro de recuperação da informação.

A área de Ciência da Informação incorpora vários significados ao tesouro. Na década de 1970, a UNESCO, através do programa UNISIST, define o termo sob dois aspectos: quanto à **estrutura** – vocabulários controlados e dinâmicos, de termos relacionados, semânticos e gerais co-

brindo um domínio específico do saber; e segundo a **função** – instrumentos de controle terminológico usado na tradução da linguagem natural para uma linguagem do sistema mais restrita (CAMPOS, 2001). A preocupação da UNESCO era apresentar conceitos ligados à construção de tesouro, à organização e à recuperação da informação.

O tesouro, quanto à estrutura, é considerado dinâmico, por permitir o registro de novos termos e alterações de termos já existentes, se necessário, além de remissivas.

De acordo com Currás (1995, p. 88), tesouro “é uma linguagem especializada, normalizada, pós-coordenada, usada com fins documentários, onde os elementos lingüísticos que compõem – termos simples ou compostos – encontra-se relacionados entre si sintática e semanticamente”.

4.1 Tesouro na indexação de documentos

O tesouro é o instrumento utilizado no contexto da documentação para indexação e recuperação da informação e é uma linguagem artificial, porque os termos que compõem esta Linguagem Documentária devem possuir um controle rígido do significado.

O uso de tesouros nas tarefas de indexação, permite ao usuário (pesquisador

ou especialista) localizar com maior grau de facilidade o termo adequado para a busca, pois apresenta uma relação lógica e hierárquica dos descritores.

Além da capacidade de organização, o tesouro também tem um valor didático, porque utiliza conceitos específicos de uma área de domínio que contempla e permite, por meio das relações entre os termos, sua melhor compreensão.

Foi com esse entendimento e por se tratar de um instrumento especializado na área de Educação, que foi utilizado o *Thesaurus Brased* na indexação da Biblioteca Digital Paulo Freire.

4.2 *Thesaurus Brased*

O *Thesaurus* Brasileiro de Educação (BRASED) é uma ferramenta que integra e padroniza todas as bases de dados educacionais do INEP e foi desenvolvido no âmbito do CIBEC/ INEP. O Centro de Informação e Biblioteca em Educação (CIBEC) é responsável pela disseminação de informações educacionais e está ligado à Diretoria de Tratamento e Disseminação de Informações Educacionais do Instituto de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP).

a) Histórico do *Thesaurus Brased*

O *Thesaurus* Brasileiro de Educação (BRASED) começou a ser desenvolvido em 1980, sob a coordenação do professor

Gaetano Lo Mônaco. No ano 1989, os analistas do CIBEC elaboraram a primeira versão experimental para uso interno; em 1997, a segunda versão; apenas em 2001, é que foi criado seu protótipo <http://www.inep.gov.br/pesquisa/thesaurus>.

Em princípio o *Thesaurus* BRASED foi elaborado para facilitar a pesquisa em Educação, sendo o tesouro pioneiro desta área no Brasil, diferenciando-se dos outros vocabulários controlados de Educação do país, por trazer seus termos selecionados e estruturados dentro de uma matriz conceitual.

b) Matriz Conceitual do *Thesaurus* Brased

A matriz conceitual foi elaborada a partir de uma análise crítica da realidade educacional e de seu contexto. A base conceitual, na qual o *Thesaurus* Brased foi desenvolvido, considera a educação em seu contexto global e interdisciplinar, o que permite ao estudioso analisá-la e compreendê-la em profundidade (INEP, 2005).

Para conceber o *Thesaurus* Brased partiu-se do princípio que a Educação é o processo pelo qual o ser humano (indivíduo e coletividade) desenvolve seu intelecto, suas potencialidades, sua cultura, satisfaz suas necessidades e se torna agente de sua história, interagindo constantemente com o meio. A matriz conceitual do *Thesaurus* Brased, portanto, coloca o homem

no centro do sistema educacional (INEP, 2005).

O *Thesaurus* Brased insere a Educação no contexto global, sem o qual não é possível compreendê-la. Na definição de seu âmbito temático, foram levadas em consideração áreas que estão relacionadas à Educação. Um *thesaurus* de Educação, dessa forma concebido, atenderá às exigências teóricas e concretas do pensar, fazer e gerir educação dentro de uma sociedade em desenvolvimento (INEP, 2005).

De acordo com a matriz conceitual, o *Thesaurus* Brased é composto por quatro campos (ou subáreas), que delimitam a abrangência da Educação:

- a) Campo 100 – **Contexto da Educação**
- b) Campo 200 – **Escola como instituição social**
- c) Campo 300 – **Fundamentos da Educação**
- d) Campo 400 – **Educação: princípios, conteúdo e processo**

Além desses quatro campos, existe um campo complementar que é o indicador e o especificador de informação: Campo 900 – **Indicadores e especificadores de informações**

c) Estrutura do *Thesaurus* Brased

Na estrutura do *Thesaurus* Brased, os termos se encontram dentro da lógica

da matriz conceitual que contempla quatro campos temáticos e um complementar, cada um desses campos subdivide-se em grupos, facetas e subfacetat. Esses termos são ordenados, de acordo com as relações lógico-ontológicas – de hierarquia, de equivalência e de associações – existentes entre eles.

d) A pesquisa no *Thesaurus* Brased

Para realizar uma pesquisa no *Thesaurus* Brased, encontram-se disponíveis, atualmente, duas formas: através da estrutura clicando sobre os campos 100, 200, 300, 400 e 900, ou em seus grupos ou facetat; ou através do campo termo que está disponível na página principal do *Thesaurus* Brased, digitando o termo ou palavra desejada.

5 RESULTADOS

Observa-se que 351 cabeçalhos de assunto existentes na base de dados da BDPF compõem o índice alfabético e o glossário. Esses cabeçalhos foram indexados pelo Grupo de Indexação da BDPF e cadastrados em seu banco de dados. Nota-se que, na base de dados, existem “campos” específicos, nos quais são cadastrados os cabeçalhos de assunto (termos indexados), e suas definições com suas respectivas fontes de referências.

Percebe-se, nesse sentido, que na entrada por assunto, isto é, no índice alfa-

bético dos 351 cabeçalhos de assunto, 190 se encontram com as respectivas definições formando, assim, o glossário. Por esses dados, nota-se que há uma preocupação dessa biblioteca no controle do vocabulário.

Verifica-se que dos 190 cabeçalhos de assunto cadastrados, na base de dados, 61 foram extraídos e definidos a partir do *Thesaurus* Brased para a BDPF. Constatat-se, portanto, que apesar de ser um instrumento de controle terminológico em uma área específica do saber, neste caso, em Educação, foi necessário recorrer a outras fontes, como os próprios documentos indexados (109 cabeçalhos de assunto), quando estes trazem informações com seu significado, e outras obras estabelecidas pela política de indexação.

6 CONSIDERAÇÕES FINAIS

A utilização do *Thesaurus* Brased na indexação da BDPF, proporcionou grande avanço no conhecimento e no aprendizado a respeito do processo de tratamento da informação de uma biblioteca.

As mudanças oriundas das tecnologias de informação e comunicação (TICs), nos convidam a uma reflexão sobre a identificação de um novo modelo de aprendizagem que leva o usuário a uma autonomia na busca de informações.

A indexação dos documentos e a contribuição trazida por um instrumento de controle terminológico, no caso, o tesauro, para a padronização da linguagem de uma área específica do saber, facilita o caminho a ser percorrido pelos usuários para atingir suas necessidades de busca e, tratando-se de uma biblioteca digital, de busca de informações em meios digitais.

Logo, é interessante destacar que o controle de vocabulário utilizado por essa biblioteca foi estabelecido por uma política de indexação, instrumento relevante e necessário, pois influencia na estrutura do Sistema de Recuperação da Informação (SRI), que, de acordo com Lancaster (2004), controla sinônimos, diferencia homógrafos e relaciona termos.

Com relação aos resultados, os dados mostram que existem 351 cabeçalhos de assunto indexados na base de dados da BDPF, e 190 desses encontram-se no glossário, trabalho este que requer pesquisa criteriosa para não fugir do objetivo principal da BDPF: divulgar as idéias, os pressupostos filosóficos, sociológicos e pedagógicos do pensamento freireano, dando assim, suporte a ações educativas.

Pode-se observar, portanto, que há certa rigidez no processo de indexação, escolhendo-se uma linguagem pós-coordenada – o *Thesaurus* Brased, pioneiro na área de Educação – como principal

fonte de padronização da linguagem de Paulo Freire aplicada pela BDPF.

Destaca-se também a relevância de um trabalho interdisciplinar, em que professores e alunos, de várias áreas, em uma ação conjunta, conseguiram construir uma Biblioteca Digital disponibilizando livros, artigos, fotos, vídeos e monografias e outros documentos do educador Paulo Freire, que tanto contribuiu para a Educação mundial.

Pelo exposto, depreende-se, portanto, a relevância da temática deste estudo ser um campo aberto para novas investigações e descobertas, que venham a sistematizar os conhecimentos teóricos e/ou práticos sobre indexação e seu papel fundamental nos Sistemas de Recuperação da Informação.

REFERÊNCIAS

- ALBUQUERQUE, M. E. B. C. de. *Relatório de indexação*. João Pessoa, 2004.
- AQUINO, M. de A. (Coord.). *Recuperação do conteúdo freireano para a construção da Biblioteca Digital Paulo Freire*. João Pessoa : UFPB, 2001. (Projeto de iniciação científica. UFPB. Departamento de Biblioteconomia e Documentação).
- BENTES PINTO, V. Indexação documentária: uma forma de representação do conhecimento registrado. *Perspectivas Ciência da Informação*, Belo Horizonte, v. 6, n. 2, p. 223-234, jul./dez. 2001.
- BEZERRA, E. P. *Digitalizando o virtual: uma análise informacional do processo de*

implementação da Biblioteca Digital Paulo Freire. 2003. 132 f. Dissertação (Mestrado em Ciência da Informação). Universidade Federal da Paraíba, João Pessoa, 2003.

_____.; BRENNAND, E. G. G. *Projeto Implementação do Pólo Produtor de Capacitação em Conteúdos Digitais Multimídia no Estado da Paraíba*. João Pessoa, 2000. (Projeto de Iniciação à Pesquisa)

Biblioteca Digital Paulo Freire (BDPF). Disponível em:
<<http://www.paulofreire.ufpb.br/paulofreire/principal.jsp>> Acesso em: 15 mar. 2005.

BRENNAND, E. G.G. et al. *Concepção e implementação da Biblioteca Digital Paulo Freire*. João Pessoa, 2000. (Projeto de Iniciação à Pesquisa)

CAMPOS, M. L. de A. *Linguagem documental que fundamentam sua elaboração*. Niterói: EdUFF, 2001.

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

CAVALCANTI, C. R. *Indexação & tesouro metodologia & técnicas*. Brasília: ABDF, 1978.

CINTRA, A. M. et al. *Para entender as linguagens documentárias*. 2 ed. São Paulo: Polis, 2002.

CURRÁS, E. *Tesaurus, linguagens terminológicas*. Brasília: IBICT, 1995.

FOX, E. A. *Digital Library courseware*. Blacksburg: Virgínia Tech Department of Computer Science, 1998.

GOMES, H. E.; CAMPOS, M. L. de A. Tesouro e normalização terminológica: o termo como base para intercâmbio de informações. *Datagrama zero: Revista de Ciência da Informação*, v. 5, n. 6, dez. 2004. Disponível em:

<http://www.dgz.org.br/dez04/Art_02.htm>. Acesso em: 23 dez. 2004.

GUIMARÃES, J.A.C. *Recuperação em um contexto de novas tecnologias*. 2000, (Texto didático)

INEP. *Thesaurus brasileiro de educação*. Disponível em:
<<http://www.inep.gov.br/pesquisa/thesaurus>> Acesso em: 15 mar. 2005.

KURAMOTO, H. Uma abordagem alternativa para o tratamento e a recuperação de informação textual: os sintagmas nominais. *Ciência da Informação*, Brasília, v. 25, n. 2, 1995. Disponível em:
<<http://www.ibict.br/cienciadainformacao/include/getdoc.php?id=814&article=476&mode=pdf>> Acesso em: 20 nov. 2004.

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

_____. *Indexação e resumos: teoria e prática*. 2 ed. rev. e atual. Brasília: Briquet de Lemos, 2004.

LÉVY, P. *As tecnologias da inteligência: o futuro do pensamento na era da informática*. Rio de Janeiro: Ed. 34, 1993.

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

MACEDO, N. D. de; MODESTO, F. Equivalências: do serviço de referência convencional a novos ambientes de redes digitais em bibliotecas. *Revista Brasileira de Biblioteconomia e Documentação: Nova Série*, São Paulo, v. 1, n. 1, p. 38-54, 1999.

MARCHIORI, P. Z. Ciberteca ou biblioteca virtual: uma perspectiva de gerenciamento de recursos de informação. *Ciência da Informação*, Brasília, v. 26, n. 2, p. 115-124, maio/ago. 1997.

MARTINS, R. D. *Perspectivas para uma biblioteca no futuro: utopia ou realidade?* Disponível em: <<http://biblioteca.estacio.br/artigos/005.htm>> Acesso em: 01 fev. 2005.

ORTEGA, C. D. *Informática documentária: estado da arte*. 2002. 259 f. Dissertação (Mestrado em Ciência da Comunicação). Universidade de São Paulo, São Paulo, 2002.

PAIVA, E. B. *Entre as normas e os desejos: a indexação de periódicos na Biblioteca Central da Universidade Federal da Paraíba*. 2002. 156 f. Dissertação (Mestrado em Ciência da Informação). Universidade Federal da Paraíba, João Pessoa, 2002.

PEREIRA, E. C.; RUTINA, R. O século XXI e o sonho da biblioteca universal: quase seis mil anos de evolução na produção, registro e socialização do conhecimento. *Perspectivas Ciência da Informação*, Belo Horizonte, v. 4, n. 1, p. 5-19, jan./jun. 1999.

ROWLEY, J. *A biblioteca eletrônica*. 2. ed. Brasília: Briquet de Lemos, 2002.

RUBI, M. P. *A política de indexação na perspectiva do conhecimento organizacional*. 2004. 135f. Dissertação (Mestrado em Ciência da Informação). Universidade Estadual Paulista, Marília, 2004.

SILVA, M. dos R.; FUJITA, M. S. L. A prática de indexação: análise da evolução de tendências teóricas e metodológicas. *Tran-sinformação*, Campinas, v.16, n.2, p. 133-161, maio/ago. 2004.

STREHL, L. Avaliação da consistência da indexação realizada em uma biblioteca universitária de artes. *Ciência da informação*, Brasília, v.27, n.3, p.329-335, set. 1998. Disponível em: <www.ibict.br/cienciainformacao/include/

[getdoc.php?id=668&article=354&mode=pdf](#)> Acesso em: 05 dez. 2003.

UNISIST. Princípios de indexação. *R. Esc. Bibliotecon. UFMG*, Belo Horizonte, v.10, n.1, p.83-94, 1981.

VAN SLYPE, G. *Los lenguajes de indización: concepción, construcción y utilización en los sistemas documentales*. Madrid: Fundación Germán Sánchez Ruipérez; Pirâmide, 1991.

WATERS, D. J. What Are Digital Libraries? *Digital Library Information Resources in Berkeley Digital Library SunSite. CLIR Issues*, n. 4, Jul./Aug. 1998. Disponível em: <<http://www.clir.org/pubs/issues/issues04.html>> Acesso em: 15 out. 2004.

ZANG, N. et al. Biblioteca virtual: conceito, metodologia e implantação. *Revista de Pesquisa e Pós-Graduação*, Erechim, v. 1, n. 1, p. 217-236, 2000. Disponível em: <<http://www.uri.br/publicacoes/revistappg/ano1n1/>> Acesso em: 20 jan. 2005.

Fabiana da Silva França

Bibliotecária – Graduada pela UFPB
fabiana21franca@yahoo.com.br

Maria Elizabeth B. C. de Albuquerque

Docente do Departamento de Biblioteconomia e Documentação (UFPB)
ebaltar@openline.com.br

Title

The Digital Library Paulo Freire: the Brased *Thesaurus* application in the indexation process

Abstract

The information search, in the digital media, requires some user's knowledge about the language applied to information systems. In this sense, some institutions that disseminate information must give a special attention to the technical treatment of this information. This study aims to analyze the use of the Brased *Thesaurus*, in the indexation of the Paulo Freire's Digital Library, focused on the documents' indexation in the language standardization, from the *Thesaurus*. In order to achieve this objective, some explanation about the terms Information retrieval systems (including Digital Libraries, Indexation and documentary Language), about the *Thesaurus* and its use in the documents' indexation, and also about the Brased *Thesaurus*, was given. It was possible to conclude that there is some inflexibility in the studied library's indexation process and that the standardization of the language, from a specific area of knowledge, makes user's information searches easier.

Keywords

Information Retrieval Systems; Indexation; *Thesaurus*; Brased *Thesaurus*

Título

Biblioteca Digital Paulo Freire aplicabilidad del *Thesaurus* Brased en la indexación

Resumen

La búsqueda de informaciones, en medios digitales, requiere de los usuarios cierto conocimiento del lenguaje utilizado por los sistemas de información. Hace con que algunas instituciones diseminadoras de informaciones tengan una atención especial en el tratamiento de esas informaciones. Para este estudio, cuyo objetivo es el análisis de la utilización del *Thesaurus* Brased en la Indexación de la Biblioteca Digital Paulo Freire (BDPF), la indexación de documentos en la padronización del lenguaje, a partir del *Tesaurus*. Para responder a este objetivo, se realizó una explicación sobre los temas: Sistemas de Recuperación de la Información (incluyendo Bibliotecas Digitales, Indexación y Lenguaje Documentário), el *Tesaurus* y su utilización en la indexación de documentos, incluso el *Thesaurus* Brased. Concluye que existe cierta rigidez en el proceso de indexación de la BDPF y que la padronización del lenguaje de un área específico del saber, facilita la búsqueda de informaciones por los usuarios.

Palabras clave

Sistema de Recuperación de Información; Indexación; *Tesaurus*; *Thesaurus* Brased
