

Recuperação de informações em bases de dados: usos de *tesauro*¹

Information retrieval in databases: thesaurus uses

Leonardo Fernandes SOUTO²

RESUMO

Uma das maiores dificuldades enfrentadas na obtenção de informações relevantes é quanto à filtragem de informações. Nesse sentido é bom destacar que o profissional da informação pode ser considerado como o primeiro filtro do sistema. O uso de instrumentos de linguagem documentária se faz necessário e contribui de forma substancial para o sucesso do serviço oferecido. Esse trabalho defende a necessidade de um profissional da informação, na contribuição ao gerenciamento de sistemas de recuperação da informação, e tem como foco, o uso de tesouros em bases de dados, analisando-os sobre dois aspectos: como uma ferramenta de trabalho do indexador e como um poderoso recurso facilitador para a busca de informações.

Palavras-chave: bibliotecário, profissional da informação, tesouros, recuperação de informação, linguagem documentária, bases de dados, filtros de informação.

ABSTRACT

One of the biggest difficulties faced by researchers is how to retrieve relevant information and how to filter it. In this way, it is good to emphasize that the

¹ Trabalho de conclusão da Disciplina Fundamentação Linguística no Tratamento da Informação, sob a orientação da Profa. Dra. Else Benetti Marques Válio. Curso de Mestrado em Biblioteconomia e Ciência da Informação. O trabalho foi apresentado no Seminário em Ciências da Informação (Londrina, 23-25 ago., 2001).

² Bacharel em Biblioteconomia e Documentação pela ESBI/FUOM; Mestrando em Biblioteconomia e Ciência da Informação pela PUC-Campinas. Bolsista do CNPq. Bibliotecário de Referência, Biblioteca Central/DSP, Universidade Estadual de Campinas. Cidade Universitária "Zeferino Vaz", Caixa Postal 6136, 13081-970, Campinas, SP, Brasil. E-mail: lfsouto@unicamp.br

Recebido para publicação em 2002 e aceito em 27/8/2003.

information expert can be considered the first filter of the system. The use of documentary language instruments is necessary and gives a substantial contribution for the success of the service offered. This paper defends the necessity of an information professional to contribute in the coordination of information retrieval systems and also focus on the use of thesaurus, in databases, both as tools of work of the indexes and as a powerful tool to help the process of information retrieval.

Key words: librarians, information professional, thesaurus, information retrieval, documentary language, databases, information filters.

INTRODUÇÃO

Com o surgimento do computador, passamos da era industrial para a era da informação de forma muito rápida. O acesso/posse da informação tornou-se uma necessidade vital para a “sobrevivência” de pessoas e empresas. O uso de meios magnéticos para o registro de informações significou um avanço incrível em seus processos de produção, tratamento, disseminação e recuperação.

Levando-se em conta que informação é matéria-prima para informação, ou que conhecimento gera conhecimento, com o uso dos computadores a literatura científica aumentou nos últimos anos. As facilidades de acesso permitidas pela *Internet* contribuíram para o aumento dessa produção científica. Um documento pode ser acessado, em qualquer lugar do mundo, em questão de segundos, a qualquer hora, sem a possibilidade de estar emprestado ou em processo de restauração. Com tantas facilidades para os pesquisadores, realmente a única consequência esperada seria a “explosão bibliográfica”.

É evidente que a explosão bibliográfica não tem como única causa a facilidade de acesso às informações. A necessidade de registrar/produzir conhecimentos é inerente ao “ser humano”. O que acontece é que a partir de um conhecimento estruturado, já testado cientificamente e sistematicamente organizado, fica bem mais fácil produzir um “novo conhecimento”.

A estruturação do conhecimento é tão importante que diante de tantas informações tornou-se necessário organizá-las em bases de dados. Pode-se entender as bases em seu sentido mais amplo como um “arquivo de informações ou ‘dados’ ou uma coleção de tais arquivos de dados” (BROONER, 1983, p.13). É importante levar em consideração que as informações arquivadas em uma base de dados são organizadas, observando uma sistemática que permite sua posterior recuperação.

Se compararmos as palavras de Foskett (1973, p.3), escritas na década de 70, com a realidade que vivenciamos hoje, veremos que a preocupação com o aumento da informação já é objeto de estudo há muito tempo, e que se substituirmos as palavras biblioteca/bibliotecas por bases de dados perceberemos que elas se enquadram perfeitamente à nossa realidade e que poderiam perfeitamente ser escritas por qualquer pesquisador contemporâneo.

As bibliotecas constituem uma parte essencial da corrente da comunicação humana. Antes do conhecimento ser registrado (e mesmo ainda hoje em sociedades muito primitivas), eram os próprios indivíduos que constituíam o repositório do conhecimento, a ponte entre gerações sucessivas e entre os que criavam novas informações e os que precisavam utilizá-las. O volume de informações que pode

ser assim transmitido é limitado, tendo a sociedade passado a progredir quando informações de vários tipos começaram a ser registradas em formas relativamente permanentes que podiam servir como substituto da pessoa do “sábio”.

Atualmente, a quantidade de novas informações produzidas é de tal ordem, que nenhum indivíduo pode alimentar a esperança de estar ao corrente delas, mesmo que seja de uma ínfima parcela. E o problema que temos de enfrentar é o de possibilitar às pessoas que precisam de informações a sua obtenção com o mínimo de gastos (de tempo e de dinheiro) e sem que sejam assoberbadas por grandes quantidades de material irrelevante. Ao invés de depósito individual de conhecimentos, temos o propósito coletivo: a biblioteca; ao invés da memória individual, temos a memória coletiva: os catálogos de bibliotecas e instrumentos bibliográficos. E, assim como o indivíduo cuja memória falha não pode transmitir as informações necessárias quando solicitadas, também a biblioteca cuja memória coletiva for inadequada fracassará em seus propósitos.

Justamente para garantir que a memória coletiva não falhe é fundamental que o processo de construção de bases de dados seja coordenado por um profissional da informação, qualificado e que possua conhecimentos específicos de linguagens documentárias.

A responsabilidade deste profissional que procede a entrada de dados vai muito mais além do que simplesmente indexar os documentos. De forma geral, ele preocupa-se com a seleção

dos documentos a serem incluídos, com a entrada de dados (indexação) e com a recuperação da informação.

Uma das maiores dificuldades enfrentadas, atualmente, em relação à obtenção de informações relevantes é quanto à filtragem de informações. Nesse sentido é bom destacar que o profissional da informação pode ser considerado como o primeiro filtro do sistema. A filtragem inicia-se na seleção dos materiais e continua por todo o processo de organização do sistema. Mas, é no momento da indexação que a filtragem faz-se mais evidente e necessária. O uso de instrumentos de linguagem documentária faz-se necessário, e contribui de forma substancial para o sucesso do serviço oferecido.

A partir das premissas acima, é objetivo desse trabalho, defender a necessidade de um profissional da informação contribuir no gerenciamento de sistemas de informação (neste caso, restringindo-se às bases de dados), identificando sua responsabilidade quanto ao uso de tesouros, um eficiente instrumento de indexação como “ferramenta de trabalho” do indexador (responsável pela filtragem inicial de informações) e também, destacar a necessidade de se disponibilizar o tesouro para o pesquisador, oferecendo desta forma mais uma opção de filtro, visto que a hierarquização dos cabeçalhos permite ao pesquisador definir a melhor estratégia de busca.

Bases de Dados

O grande desenvolvimento tecnológico e da produção literária científica levou à inserção das bases de dados em nossa sociedade, e isto, podemos comprovar nas palavras de Rosetto (1997, p.136).

O volume crescente de informações e os vários meios de armazenagem fizeram com que as organizações responsáveis pelo tratamento/armazenamento/recuperação criassem mecanis-

mos para possibilitar o uso dessa grande “massa de dados”, acoplando as tecnologias de automação e propiciando, dessa forma, as **bases de dados** com acesso *on-line* e/ou em CD-ROM. A tecnologia vem subsidiando também o progresso das **redes de comunicação de dados**, o que possibilita o acesso significativo às informações disponíveis em nível mundial e atende aos variados requisitos da comunidade usuária.

Nos últimos anos tornou-se evidente a crescente aceitação das bases de dados nas atividades/serviços de bibliotecas/centros de informação, enfim em qualquer unidade de informação. Em muitos casos, quando tais unidades de informação disponibilizam as bases para consulta, os próprios pesquisadores elaboram as estratégias de busca e consultam as bases de dados objetivando encontrar informações relevantes e atualizadas.

A notável difusão das bases de dados, em virtude da explosão bibliográfica, demonstra a necessidade de controle da informação. O fato da informação ter assumido um valor significativo no mercado, chegando a ser considerada como “bem” e sendo seu acesso/posse um fator crítico para o sucesso de pessoas e empresas na Sociedade da Informação, contribuiu para o desenvolvimento desta tecnologia da informação.

As bases de dados são elaboradas com o objetivo de fornecer informação atualizada, precisa e confiável, buscando atender a demanda de uma clientela específica. A produção pode ficar a cargo de pessoas, agências, entidades e até mesmo bibliotecas. As informações são registradas em meio magnético ou em papel, como por exemplo a versão impressa do *Index Medicus*. O acesso pode ser local (CD-ROM e papel) ou remotamente (acesso *online*). Quanto à abrangência podem ser locais, regionais,

nacionais e internacionais. Quanto ao conteúdo podem ser referenciais, de resumos, texto completo, ou representar o documento indicando seu conteúdo de forma mais completa fazendo a junção de mais de uma destas características, como por exemplo *abstract* e texto completo.

Seja qual for a fonte produtora, o suporte de registro, a forma de acesso, a abrangência e a representação do conteúdo, a seleção das bases de dados observa, ainda, os seguintes critérios; – Custo: geralmente o valor gasto para aquisição destes repositórios de informação é muito alto. Em alguns casos é importante fazer-se uma análise dos usuários para verificar a possibilidade de se cobrar pelo uso do serviço, e verificar se a biblioteca ou centro de informação possui verba suficiente para manter a assinatura; – Cobertura: existem bases de dados realmente específicas e outras que embora atendam uma clientela específica, possuem um conteúdo mais generalizado, por atenderem diferentes ciências de uma mesma área; – Atualização: a atualização das bases de dados fica a cargo da produtora. Geralmente é feita em intervalos regulares: diariamente, semanalmente, mensalmente, previamente estabelecidos; – Linguagem de indexação/documentária utilizada: refere-se aos instrumentos, manuais, tesouros, listas de cabeçalhos de assuntos, utilizados para a entrada de dados e sua disponibilidade para busca, e – Estrutura: está diretamente relacionada à forma de entrada de dados. Quantos e quais campos ou pontos de acesso estão disponíveis para consulta?

Em vista dos critérios elencados, neste trabalho, optamos por nos concentrar no penúltimo critério citado: “Linguagem de indexação/documentária utilizada”. A linguagem documentária utilizada para a representação temática do documento é importante para a recuperação das informações, podendo servir como instrumento de consulta tanto para o indexador que inclui o documento na base, quanto para o pesquisador (entende-se aqui

qualquer pessoa que utiliza o sistema), na definição da melhor estratégia de busca.

Linguagens Documentárias

As linguagens documentárias são linguagens “construídas para indexação, armazenamento e recuperação da informação e correspondem a sistemas de símbolos, destinadas a “traduzir” os conteúdos dos documentos” (CINTRA, 1994, p.23).

A linguagem de indexação é composta de três partes: vocabulário (conjunto dos descritores), sintaxe (função/ordem do vocábulo em relação ao outro) e a semântica (sentido do vocábulo). A linguagem de indexação pode ser natural ou artificial. A linguagem natural trabalha com o uso de termos e a representação do assunto está ligada às palavras utilizadas pelo autor, esta linguagem também é conhecida por linguagem livre. A linguagem artificial trabalha com o uso de conceitos e a representação do assunto está ligada aos descritores permitidos pelo instrumento de indexação utilizado pelo sistema.

Se uma linguagem de indexação estiver em constante avaliação, é possível garantir sua atualização. O conhecimento evolui rapidamente e, às vezes, a lista oficial não contém os cabeçalhos desejados pelo indexador para representar o assunto do documento analisado. O ideal é que a linguagem de indexação utilizada seja um sistema aberto e permita a inclusão de novos termos para que se mantenha a confiabilidade do sistema. Entretanto, é fundamental que todos os utilizadores do instrumento de indexação sejam informados da inclusão de um novo cabeçalho na lista. Incluir somente termos que não possuam nenhuma representação na linguagem, pode evitar o risco de utilizar-se a linguagem natural, colocando a perder todo o trabalho já desenvolvido.

É comum encontrarmos sistemas de linguagem documentária sintéticos que, apesar

de relacionarem cabeçalhos simples e compostos, oferecem regras para a formação de novos cabeçalhos compostos ou especificação dos já existentes.

Em sistemas informatizados o ideal é o uso da pós-coordenação, porque no momento da busca podem relacionar-se os termos a fim de estabelecerem-se as melhores estratégias. Com isso, as possibilidades de pesquisa aumentam significativamente. Quando se decidir por utilizar a pré-coordenação, sugere-se a hifenação e não a inversão, porque o usuário geralmente desconhece a diferença entre hífen e vírgula. Mesmo que o sistema disponibilize um manual contendo as orientações para uso da base, não é óbvio que o pesquisador compreenda e assimile quando o assunto será pesquisado fazendo-se uso da inversão.

Tesouro como Instrumento de Recuperação da Informação

A origem do tesouro, caso este tivesse surgido nos dias de hoje, poderia estar associada à revolução tecnológica, que gerou um volume notável de informações e sua aplicação estaria diretamente associada às bases de dados.

O tesouro documentário surgiu da necessidade de manipular grande quantidade de documentos especializados. Era preciso trabalhar com vocabulário mais específico e com uma estrutura mais depurada do que aquela presente nos cabeçalhos de assunto (remissivas e referências cruzadas tipo ver e ver também). Assim, além da especificidade, cuidou-se de melhorar a estrutura e as referências cruzadas (ver também) deram lugar às relações hierárquicas (vertical) e associativas (horizontal).

Pelo fato desse novo instrumento da documentação possibilitar, através do agrupamento dos termos, o acesso a uma idéia, mesmo sem saber nomeá-la de saída, as novas listas estruturadas de termos passaram a ser chamadas de tesouros... (GOMES, 1990, p.14).

Ainda segundo Gomes (1990, p.16), pode-se considerar o tesouro como “Linguagem documentária dinâmica que contém termos relacionados semântica e logicamente, cobrindo de modo compreensivo um domínio do conhecimento”.

Diante das citações originárias da obra organizada por Gomes (1990), intitulada Manual de Elaboração de Tesouros Monolíngües, já é possível notar a importância dos termos e da relação entre eles, que constituem as duas características de maior destaque dos tesouros. Se nos concentrarmos na observação de que o tesouro possibilita, “através do agrupamento dos termos, o acesso a uma idéia, mesmo sem saber nomeá-la de saída” perceberemos logo que sua utilização, por indexadores e usuários de bases de dados, será um recurso altamente eficaz. Tanto os indexadores poderão utilizá-lo a fim de facilitar seu trabalho, mantendo um controle terminológico, quanto os usuários poderão utilizá-lo como bússola, orientando-se em relação aos termos que representam suas “idéias” e, consecutivamente, localizando os documentos de seu interesse.

O uso do tesouro como instrumento de recuperação da informação é sem dúvida uma estratégia eficaz para a busca de informações em bases de dados. Geralmente, é comum o tesouro ser utilizado somente pelos indexadores como uma ferramenta de trabalho, ficando sua aplicação restrita à indexação, deixando de lado a importância desse instrumento no momento da recuperação. A simples adoção de um te-

sauro como uma linguagem artificial, controlada, já contribui em muito para a diminuição da inconsistência na recuperação da informação em uma base de dados.

Com a visão de uma estrutura hierarquizada dos assuntos de uma base, fica mais fácil a compreensão de como procurar o assunto de interesse. Em alguns casos, o acesso ao tesouro impede que documentos que se encontram indexados deixem de ser localizados. É importante destacar que quando se utiliza uma linguagem controlada, os assuntos são representados tematicamente segundo termos pré-definidos. Assim, na indexação, pode deixar de utilizar-se um termo porque o sistema não o permite, recomendando o uso de outro. Caso o pesquisador faça uma busca pelo termo não aceito, certamente ele não encontrará o documento indexado. Se por ventura o pesquisador consultar o tesouro, ele terá maiores chances de encontrar o documento e ainda poderá visualizar outras formas para pesquisa.

O tesouro apresenta os termos na seguinte estrutura:

Termo	Biblioteconomia
Nota de aplicação	Termo incluído em 18/5/01
Equivalência	UP Documentação
Termo genérico	TG Ciência da Informação
Termo específico	TE Bibliotecário
Termo associado	TA Arquivística

A estrutura de um tesouro segue uma linha hierarquizada de relações entre os termos e sua construção demanda um conhecimento técnico e especializado. Existem alguns profissionais que dominam as técnicas de elaboração de um tesouro, dentre eles, destacam-se os bibliotecários.

Na obra de Gomes (1990, p.16), um comentário apóia a fundamentação desse trabalho, que defende a aplicação do tesouro nas bases de dados.

A estrutura do tesouro é um elemento importante para que ele possa cumprir sua função: ela permite ao usuário (indexador ou consultante) encontrar o (s) termo (s) mais adequado (s), mesmo sem saber, de início, o nome específico para representar a idéia ou o conceito que ele procura. A partir de um termo que o usuário conhece, o tesouro, através de sua estrutura, mostra diversos outros que podem ser tão oportunos ou mais do que aquele que lhe veio à mente.

Diante desse pensamento, consideraremos que as possibilidades de aplicação do tesouro são muito amplas em relação ao que realmente ocorre. De certa forma podemos concluir que ele ainda é um instrumento subutilizado.

Entre algumas bases de dados que disponibilizam seus tesouros para os usuários, a INSPEC é produzida pelo *Institution of Electrical Engineers* (IEE). Ela traz informações sobre física, engenharia elétrica e eletrônica, ciência da computação e tecnologia da informação, oceanografia, engenharia nuclear, geofísica, química, matemática e engenharias no geral.

Indexa mais de 4 300 periódicos e 2 581 Anais de conferências, livros, teses e relatórios técnicos e 1 510 normas totalizando 6 milhões de referências. A taxa anual de inclusão de novos registros é de cerca de 300 mil. Inclui informação especializada e pertinente sobre lasers, semicondutores e supercondutores, astronomia, astrofísica e geofísica, componentes ópticos, engenharia biomédica, inteligência artificial, CAD e análise por computador, equipamento periférico, redes locais e manuseamento de materiais, tecnologia de informação aplicada à gestão, atividade bancária, comercial, seguros e *marketing*.

As referências bibliográficas incluem título, resumo descritivo e detalhes bibliográficos completos.

Responsabilidade do Profissional da Informação na Entrada e Recuperação de Dados

É muito importante que a coordenação do processo de construção de bases de dados esteja sob a responsabilidade de um profissional da informação. Ele pode ser o responsável direto pelo sucesso do serviço. A competência desse profissional está relacionada com o domínio e o conhecimento das técnicas de indexação e de elaboração e uso de linguagens documentárias, sobretudo de tesouros, que é o objeto de estudo nesse trabalho.

Comumente, afirma-se que a biblioteconomia e a documentação têm por objetivos básicos a análise, organização e disseminação da informação. Ratificando tal afirmação, algumas obras de referência associam as metodologias vigentes na área de biblioteconomia e documentação, bem como as funções exercidas por seus profissionais, ao fluxo do tratamento e recuperação da informação (GALVÃO, 1998, p.47).

A afirmativa de que o profissional da informação atua na filtragem inicial é confirmada logo no primeiro momento da indexação, quando o mesmo identifica língua, forma física, nível, público a que se destina, país de publicação e outras características inerentes ao documento. Esta atividade já possibilita o cruzamento futuro de informações, assim, ampliando as possibilidades de filtragem.

A especificidade e a exaustividade na indexação estão diretamente ligadas à precisão e revocação na recuperação. A especificidade e a exaustividade estão relacionadas tanto à quantidade de cabeçalhos quanto ao uso de termos/conceitos gerais e específicos.

É importante destacar que a relevância fica a cargo do usuário. Por mais que se queira

definir padrões de relevância, somente o usuário poderá dizer se os documentos recuperados são realmente relevantes. O máximo que o profissional da informação pode fazer é definir, no momento da seleção, se os documentos a serem incluídos na base de dados podem potencialmente ser de interesse dos usuários do sistema.

Se houver um maior esforço na entrada de dados, com certeza isso gerará um menor esforço na saída, ou seja, na recuperação de informações. Por isso, é importante que o responsável pela construção e manutenção da base de dados esteja consciente que de seu trabalho inicial depende todo o restante do processo.

CONCLUSÃO

Na visão de Lucas (1996, p.60), e sem dúvida nenhuma, uma visão compartilhada por este pesquisador que se aventurou a escrever sobre a responsabilidade do profissional da informação na recuperação de informações em bases de dados, o bibliotecário não será/é apenas um organizador da informação, mas sim, atuará/atua como um filtro refinador de informações.

As novas tecnologias de informação são inúteis sem os meios de localizar, filtrar, organizar e resumir os seus produtos. Para dar conta destas necessidades um novo profissional está surgindo: o gerente de informações - capaz de combinar as habilidades do cientista da computação, dos bibliotecários, dos editores e especialistas em bases de dados. Estes agentes humanos irão trabalhar com agentes de *software*, cuja especialidade será manipular a informação, fazendo uso das ferramentas que ajudem a navegar na

Internet (WWW, Gopher, VERONICA, etc.).

A função dos bibliotecários evoluirá, sua eficácia não será julgada em termos de quantidade e qualidade da informação fornecida, e sim a partir do tempo economizado para os usuários. O bibliotecário será acima de tudo o “refinador” humano da informação (Les temps des..., v. 31), com a função de criar informação com valor-agregado para serviços específicos - informação é conhecimento com valor agregado.

Pode relacionar-se a responsabilidade do profissional da informação quanto aos instrumentos/mecanismos de controle/refino da informação. No caso de bases de dados, à própria elaboração/utilização do tesouro. Pode considerar-se que a própria estrutura já constitui-se em um filtro para a pesquisa. Com uma estrutura bem elaborada, o tesouro, se disponibilizado para o pesquisador, constitui-se em um excelente recurso de busca. O usuário pode orientar-se quanto à estratégia a ser formulada, definindo o caminho para localizar a informação que melhor atenda às suas expectativas.

É fundamental que os profissionais da informação de forma geral e a sociedade, conscientizem-se de que não há como negar a importância e necessidade deste profissional “refinador humano da informação”, que de acordo com Lucas (1996) será conhecido como o “gerente de informações”. A responsabilidade desse profissional vai muito além do simples desenvolvimento e controle de mecanismos de recuperação da informação. Todas as suas atividades possuem um cunho de responsabilidade social, pois, nunca perdem de vista a totalidade do ser humano, que almeja uma informação rápida, confiável, precisa e com o menor custo e tempo.

REFERÊNCIAS

- BROONER, E.G. Introdução a bases de dados. *In*: BROONER, E. G. *Gerência de bases de dados para microcomputadores*. Rio de Janeiro: Campus, 1983. p.11-24.
- CINTRA, A.M.M. *et al.* Linguagens documentárias. *In*: CINTRA, A.M.M. *Para entender as linguagens documentárias*. São Paulo: Polis, 1994. p.23-34.
- FOSKETT, A.C. Introdução. *In*: FOSKETT, A.C. *A abordagem temática da informação*. São Paulo: Polígono, 1973. p.3-9.
- GALVÃO, M.C.B. Construção de conceitos no campo da Ciência da Informação. *Ciência da Informação*, Brasília, v.27, n.1, p.46-52, 1998.
- GOMES, H.E. (Coord.). *Cabeçalho de assunto unificado: projeto CAU: divisão de forma: instruções de uso*. [S.l.]: IBICT, 1984. p.14.
- GOMES, H.E. *Manual de elaboração de tesouros monolíngües*. Brasília: Programa Nacional de Bibliotecas das Instituições de Ensino Superior, 1990. 78p.
- LUCAS, C.R. A organização do conhecimento e tecnologias da informação 1. *Transinformação*, Campinas, v.8, n.3, p.59-65, 1996.
- ROSETTO, M. Uso do protocolo Z 39.50 para recuperação de informação em redes eletrônicas. *Ciência da Informação*, Brasília, v.26, n.2, p.136-139, 1997.

