

# Ferramentas de busca na Internet: para quê, por quê e como utilizá-las\*

*Silvana Ap. B. Gregorio Vidotti*

Departamento de Ciência da Informação  
Faculdade de Filosofia e Ciências  
Universidade Estadual Paulista - UNESP  
17525-900 Marília / SP - Brasil  
Tel: (014) 4211270 / 4211296  
E-mail: vidotti@marilia.unesp.br

*Márcia Correa Bueno*

Seção Técnica de Aquisição e Tratamento da Informação  
Serviço de Biblioteca e Documentação  
Universidade Estadual Paulista - UNESP  
13506-900 Rio Claro / SP - Brasil  
Tel: (019) 526-2836 / 526-2817  
E-mail: mcbueno@rc.unesp.br

## RESUMO

A evolução das tecnologias de informática e de comunicação aplicadas à informação tornou possível a indexação, recuperação e disseminação automática da informação, em especial na rede Internet com a utilização das ferramentas de busca - programas que manipulam base de dados com informações sobre documentos disponíveis na rede. A Internet é uma grande fonte eletrônica de informação que, por meio da World Wide Web, coloca à disposição documentos hipertextuais dos mais variados assuntos e de diferentes arquiteturas de informações textuais, sonoras e imagéticas. Para que esse universo possa ser acessado por pessoas de diferentes culturas e territórios, softwares percorrem essa rede indexando as páginas informacionais disponíveis pelos/para os usuários. As ferramentas de busca por meio das estratégias fornecidas pelos usuários, consultam bases de dados com o objetivo de fornecer endereços de sites ou páginas pertinentes ao assunto solicitado. Elas se diferenciam na forma de estruturar as bases de dados, nos recursos referentes às estratégias de busca, nos níveis de busca e na apresentação dos resultados. Como objetivos deste trabalho de pesquisa temos a delimitação das características fundamentais dos diversos tipos de ferramentas de busca e avaliação das respostas oferecidas mediante as estratégias de busca e níveis de busca pré-estabelecidos. Como resultados, apontamos algumas diretrizes para a otimização do uso estratégico das ferramentas de busca com relação aos tipos de consulta e assuntos desejados. Concluímos que existe um campo de pesquisa amplo para os profissionais da Biblioteconomia, com relação aos tratamentos descritivo e temático dos documentos disponíveis na Internet que são catalogados em bases de dados pelas ferramentas de busca de forma automática ou manual, para que esses documentos possam ser recuperados de maneira a atender as expectativas do usuário numa relação eficaz de relevância x pertinência x tempo de busca.

**PALAVRAS-CHAVE:** Ferramentas de busca, Search Engines, Acesso à informação – Internet

Com a evolução das tecnologias de informática e comunicação e a aplicação destas em Unidades de Informação, as informações bibliográficas e catalográficas passaram a ser armazenadas, manipuladas

e recuperadas em vários locais e por diferentes pessoas, sem a limitação de tempo e espaço.

As Bibliotecas, como Unidades de Informação, passaram a implantar softwares integrados para automação/informatização de seus serviços como aquisição, circulação, representação descritiva e temática e recuperação dos documentos pertencentes ao acervo local ou a acervos de acesso remoto.

A introdução destas tecnologias em bibliotecas nos faz lembrar de Vannevar Bush que, em 1945, baseado na estrutura convencional de uma biblioteca, antecipando a explosão da informação e motivado a desenvolver suas idéias pela necessidade de suportar formas naturais de indexação e recuperação de informações, idealizou o sistema MEMEX, que seria, segundo o autor

"um dispositivo no qual um indivíduo armazenaria todos os seus livros, registros e comunicações, e seria mecanizado de tal forma que pudesse ser consultado com alta flexibilidade e velocidade ... um suplemento da própria memória do indivíduo." [1]

O sistema de recuperação proposto por Bush era o de associação de idéias, objetos ou itens, semelhante ao que ocorre no cérebro humano:

"a mente humana ... opera por associação. Com um item enfocado, ela pula instantaneamente, para o próximo item sugerido pela associação de pensamentos, de acordo com alguma teia intrincada de caminhos formada pelas células do cérebro. A mente tem naturalmente outras características: trilhas que não são seguidas freqüentemente podem ser apagadas, os itens não são totalmente permanentes, a memória é transitória. Ainda assim, a velocidade de ação, a complexidade das trilhas, os detalhes das imagens mentais são mais espantosos que qualquer coisa na natureza ... o homem não pode esperar que esse processo mental seja completamente reproduzido artificialmente, mas deve ser capaz de aprender com isto. Ninguém pode esperar igualar a velocidade e flexibilidade com a qual a mente humana segue um caminho associativo, mas deve ser possível vencer decisivamente a mente no que diz respeito à permanência e a clareza dos itens

recuperados do armazenamento". [1]

O armazenamento das informações, no sistema proposto por Bush, seria feito em microformas e o acesso a essas informações seria mecânico e através de índices. Dois itens quaisquer seriam codificados para associação através de uma trilha, que poderia ser gerada e/ou manipulada pelo sistema.

Apesar do sistema MEMEX ter sido projetado em detalhes, a tecnologia adequada para a sua implantação apareceu apenas vinte anos depois com o surgimento dos computadores digitais.

As atuais tecnologias de informática e os novos suportes de informação possibilitaram novos processos de organização, análise, recuperação e disseminação da informação, baseados na estrutura humana de associação de idéias, objetos ou itens, apontada por Bush, com a vantagem de que as informações contidas em uma biblioteca possam figurar simultaneamente em tantos locais quantos forem necessários, via Internet e/ou Intranet, e em ambientes informacionais hipertextuais e multisensoriais, nos quais o usuário é um gerenciador ativo do processo de armazenamento e principalmente de recuperação das informações inter-relacionadas por meio do multidimensionamento dos pontos de acesso informacionais.

Lancaster [2, p.9], alertava para a mudança de filosofia proveniente deste novo tempo para a Biblioteconomia: a importância do acesso à informação ao invés da propriedade da informação, ou seja,

“... os bibliotecários têm reconhecido que eles não mais controlam tudo o que o usuário necessita, e nem deveriam. Ao contrário, o seu papel é fornecer o acesso aos recursos, em quaisquer formas que eles se apresentam, à medida que a necessidade por eles aparecer.”

Com esta nova postura, as mudanças exigidas na administração e planejamento de biblioteca são certas, pois os investimentos são voltados para equipamentos, acessos, catálogos cooperativos, convênios como por exemplo a SciELO (Scientific Electronic Library Online) biblioteca virtual que abrange uma coleção selecionada de periódicos científicos brasileiros e o ProBE (Programa Biblioteca Eletrônica) que oferece para a comunidade científica, acadêmica e administrativa das instituições consorciadas a consulta ágil e atualizada, por meio eletrônico, a textos completos de revistas científicas internacionais através da Rede ANSP (Academic Network at São Paulo).

Se para Lancaster [2], no início da década de 90, o investimento em acervo físico se tornava uma questão a ser amplamente discutida, hoje ela já se torna uma realidade principalmente com relação à formação da coleção de periódicos, visto que os periódicos eletrônicos estão sendo amplamente difundidos e utilizados no meio acadêmico, ainda que a versão em papel não tenha sido encerrada.

Diante deste quadro atual, o atendimento às necessidades

informacionais dos usuários torna-se um procedimento mais dinâmico, considerando-se o tempo de acesso e de atualização da informação eletrônica, o formato eletrônico do documento que pode ser facilmente copiado e/ou transferido (resguardados os direitos autorais), a comunicação entre pessoas feita pelo correio eletrônico, as listas de discussão, o crescente número de bases/bancos de dados on-line com textos completos, a facilidade em se trabalhar com imagens devido ao aumento de memória das máquinas e das unidades de armazenamento, os novos softwares gráficos, a transmissão de dados via rede de computadores local e/ou remota, e as novas interfaces gráficas cada vez mais amigáveis.

Também, diante da automação/informatização de serviços, das novas formas de documentos, novos suportes e acesso eletrônico à informação, mudanças na execução de atividades técnicas das Bibliotecas passam a ser fundamentais.

A indexação, a recuperação e a disseminação da informação, que são processos totalmente inter-relacionados, sofrem grandes mudanças com a inserção de tecnologias de informática. A atividade de indexação, se realizada manualmente, leva tempo, além de ser subjetiva, ou seja, de depender do tipo de documento, e da capacidade e experiência do indexador. Além disto, existem vários tipos de indexações, que dependem dos tipos de documentos e de usuários.

Em oposição ao processo de indexação humana, a forma automática ganha rapidez, porém exige maiores cuidados, como em alguns casos de sinônimos e hierarquia de assuntos, onde, a partir do documento indexado, cria-se um índice geral indicando a quais documentos estão relacionados cada termo extraído da indexação.

A recuperação automática da informação é possível a partir do uso deste índice por meio das estratégias de busca e de acordo com a necessidade informacional do usuário.

A disseminação da informação objetiva divulgar as novas fontes e/ou informações existentes a partir da inserção de novos documentos e da atualização dos índices de indexação, criando uma dinâmica constante. Um grande facilitador na recuperação automática da informação é a interface gráfica dos novos sistemas, embutindo as estratégias de busca em opções gráficas para o usuário, principalmente para o leigo, tornando mais agradável e descomplicada a sua interação com o sistema automatizado.

Atualmente, o acesso à informação eletrônica é o ponto alto das tecnologias de informática aplicadas às Bibliotecas, pois, com a tecnologia das redes eletrônicas, torna-se possível o surgimento de novos documentos e produtos e, por consequência, a criação de novos serviços, como a orientação aos usuários na utilização de seus recursos, o desenvolvimento de home-pages, o agendamento e o atendimento de novos serviços on-line, como a comutação, o empréstimo entre bibliotecas, a disseminação da informação e o catálogo.

Como fonte eletrônica de informações, a Internet tornou possível o acesso a instituições como museus, órgãos públicos, empresas, bibliotecas e aos mais variados tipos de documentos eletrônicos. Podemos citar, receitas culinárias, artigos e livros científico, home-pages pessoais e institucionais, catálogos de bibliotecas tradicionais, bibliotecas digitais e/ou virtuais. Assim, a procura pela informação desejada, mesmo que não se saiba exatamente o que se quer e nem onde a informação se encontra armazenada, ganha uma dimensão maior principalmente em tempo de procura-resposta.

A World Wide Web, também conhecida como WWW ou Web, é a forma mais comum de acesso a Internet e corresponde à sua parte gráfica, com documentos hipertextuais dos mais variados assuntos e de diferentes arquiteturas de informações textuais, sonoras e de imagens.

O acesso à informação disponível na Web se dá por diferentes pessoas de diferentes idades, culturas, territórios e necessidades, sejam cientistas, profissionais liberais, estudantes ou professores, que possuem também uma necessidade de informação diversificada, ou seja, profissional, cultural, lazer ou entretenimento.

Podemos pensar na Internet como uma grande biblioteca, na qual os usuários são agentes ativos do processo de armazenamento, indexação, recuperação e disseminação de documentos eletrônicos hipertextuais. Uma biblioteca auto-organizada em permanente mutação.

Atualmente, o crescimento exponencial do número de documentos e informações dos mais diversos assuntos disponíveis na WWW, se dá pela facilidade de se elaborar um documento eletrônico em forma de home-page, da maneira que se deseja e disponibilizá-lo ao mundo em um servidor/provedor de acesso a Internet, muitas vezes gratuito.

Como esses documentos inseridos na rede Internet diariamente são descritos e indexados? Como os usuários têm acesso a esses documentos?

As atividades de indexação e recuperação dos documentos são feitas pelas Ferramentas de Busca, que possibilitam aos usuários a busca de informações disponíveis na rede Internet.

As Ferramentas de Busca, conhecidas também como Máquinas de Busca ou Search Engines, são programas computacionais desenvolvidos com o objetivo de indexar informações descritivas e temáticas das páginas e/ou sites da Internet em bases de dados, com a finalidade de possibilitar a recuperação de documentos solicitados, pelos usuários da Internet, segundo as estratégias de busca e os critérios adotados.

Suas origens datam de 1994, por iniciativa das Universidades Norte-Americanas e, atualmente, as iniciativas particulares e/ou privadas estão também atuando nesta área, disponibilizando suas próprias ferramentas, muitas delas com acesso gratuito.

Preocupados com a construção de ferramentas de busca na Internet para usuários diversificados, Rosenfeld e Morville [3, p.102-103] indicam alguns fatores a serem considerados com relação às necessidades informacionais dos usuários:

- alguns usuários têm claramente definido o tipo de informação que precisam e onde ela pode ser localizada;
- alguns usuários sabem as informações que querem, mas não sabem exatamente onde existem ou mesmo se existem;
- alguns usuários não sabem exatamente o que esperam encontrar, pois não sabem exatamente o que existe sobre o assunto;
- alguns usuários querem tudo sobre um assunto específico.

A navegação pela Internet torna-se um grande problema quando não há tempo disponível para se visitar todos os sites interessantes conhecidos e, principalmente, avaliar a qualidade do conteúdo de cada um, e quando não se conhece o que há disponível na Internet sobre o assunto desejado. Porém, recuperar informações na WWW sem uma estratégia e um instrumento adequado significa obter milhares de documentos irrelevantes. Portanto, é imprescindível conhecer os recursos disponíveis pela própria WWW para se ter a resposta desejada.

Quando se utiliza qualquer uma das Ferramentas de Busca, na verdade o que está sendo consultado não é a Internet propriamente dita, e sim uma Base de Dados Referencial dos documentos existentes na Internet, que retorna como resposta os endereços das páginas relevantes à consulta.

As Ferramentas de Busca procuram ser de fácil utilização, através de uma interface amigável, e a busca é concretizada em segundos com as respostas apresentadas diretamente pelos links das páginas ou por categorias de assuntos ou, ainda, pela forma de exibição dos resultados. Apresentam roteiro de ajuda e exemplos de estratégias de busca.

Quanto ao tipos de Ferramentas de Busca, podemos classificá-las em Catálogos, Índices e Metapesquisadores. Os Catálogos são organizados a partir de páginas cadastradas por seus criadores, que informam dados como título, resumo ou descrição, palavras-chave e endereço de localização da página (URL). A partir de seu cadastro, a página é analisada e recebe um tratamento manual, sendo classificada por assunto ou categoria. A partir de então, são disponibilizadas as informações descritivas e temáticas da página, tornando o seu acesso recuperável na Base de Dados.

Os catálogos frequentemente possibilitam ao usuário a procura por sites de interesse segundo categorias e permitem a solicitação de inclusão de novos documentos e/ou categorias. Têm como grande característica o tratamento temático manual das páginas cadastradas criando categorias e sub-categorias de assunto. Alguns exemplos de catálogos são: Cadê?, Excite, Onde ir?, Surf e Yahoo!.

Os Índices são, geralmente, criados automaticamente a partir de uma busca na Internet realizada por “robôs de busca” ou também conhecidos como spiders (aranhas).

Os robôs de busca percorrem a Internet, efetuando uma “varredura” periódica, procurando páginas e criando, automaticamente, bases de dados com informações recuperáveis sobre a página. Estas informações compreendem título, texto, URL, porém não recebem nenhuma classificação temática manual. Têm como grande característica a constante atualização de seus dados. Alguns índices são: AltaVista, Bookmark, Infoseek, RadarUOL, Todobr e o Netscopio.

Algumas Ferramentas de Busca já possuem os chamados *sites regionais* para um determinado país, como por exemplo o Yahoo! e o AltaVista para o Brasil.

Numa terceira categoria, tem-se os Metapesquisadores, que são Ferramentas de Busca que não possuem sua própria Base de Dados, mas que acionam as Bases de outras Ferramentas de Busca. Uma série de Ferramentas de Busca são arroladas ao usuário e ele define se a busca percorrerá todas ou apenas algumas. Como exemplos temos o MetaCrawler dos Estados Unidos da América e o Metaminer do Brasil.

Hoje, o MetaCrawler faz pesquisas, simultaneamente, no AltaVista, Infoseek, WebCrawler, Thunderstone, Excite, Google, Lycos, LookSmart, GoTo DirectHit e RealNames, enquanto que o Metaminer utiliza as Ferramentas de Busca AltaVista, AOLNetfind, Excite, Lycos, WebCrawler, Yahoo! e Fast e as nacionais: RadarUOL, Zeek, Achei, Cadê?, Yahoo! Brasil.

“A utilização dos metapesquisadores não elimina a necessidade de conhecer as características individuais dos diversos mecanismos de busca. Quanto mais se conhece sobre as formas de funcionamento das ferramentas que os alimentam, melhor julgamento quanto a confiabilidade dos resultados obtidos. Se, por exemplo, a pesquisa exige determinados refinamentos não processáveis pelas ferramentas que constituem o metapesquisador pode resultar erros e resultados inadequados”. [4]

Bueno e Vidotti [5, p.48] salientam ainda que:

“... no metapesquisador, a restrição do número máximo de resultados obtidos por ferramenta e a delimitação do tempo de busca para a pesquisa são fatores a serem considerados no resultado da busca, pois pode não haver tempo hábil para conectar ou consultar de forma integral/abrangente todas as Ferramentas de Busca selecionadas e ainda o resultado obtido pode ser superior ao delimitado na estratégia.”

Com o aumento constante e exponencial dos documentos da Internet, as Ferramentas de Busca passaram a se especializar quanto ao assunto ou área de abrangência. São exemplos de Ferramentas específicas:

Todobr (<http://www.todobr.com.br/>) para assuntos relacionados ao país ou ainda por região;

Biolinks (<http://www.biolinks.com/>) para artigos científicos;

Cora (<http://www.cora.justresearch.com/>) para Ciência da Computação;

Chemie.De (<http://www.chemie.de/?language=e>) para Química;

Whowhere (<http://www.whowhere.lycos.com/>) para procurar pessoas.

Os conceitos de recuperação da informação estão presentes no uso das Ferramentas de Busca, onde os fatores de pertinência e exaustividade devem estar claros diante de uma estratégia de busca, e onde o resultado deve ser o desejado com relação à quantidade de documentos recuperados e à sua qualidade, ou seja, a relevância de seu conteúdo.

Após a elaboração da estratégia de busca e a consulta a base de dados de uma Ferramenta de Busca é comum obter como resposta milhares ou centenas de informações referenciais dos documentos que contenham ou não as informações solicitadas. Isto acontece por se tratar de um assunto geral ou em função da estrutura organizacional da Base de Dados da Ferramenta de Busca, que pode considerar como respostas, aproximações ao termo consultado, plural, ocorrência do termo no documento, diferenciação ou não de palavras maiúsculas e minúsculas, e que podem levar a obtenção de documentos pertinentes, porém estes nem sempre satisfazem as expectativas e necessidades dos usuários.

Lynch [6] classifica o acesso à informação através do uso das Ferramentas de Busca como altamente democrático devido ao acesso único e igual à todas as informações da rede, porém questiona a irrelevância frequente dos resultados nas consultas realizadas.

As diferenças básicas entre as Ferramentas de Busca decorrem dos critérios adotados na construção de suas Bases de Dados e dos recursos disponíveis para a elaboração de estratégias de busca.

Branski [4, p.2] aponta os seguintes critérios de diferenciação na construção da base de dados:

- “a relação de páginas iniciais a partir do qual o *spider* percorrerá a rede em busca de informação,
- as informações enviadas pelos autores que escolhem as ferramentas onde pedirão a inclusão,
- de como indexa as informações de cada site (se armazena o texto integral, se somente o título e um pequeno resumo algorítmicamente construído do conteúdo, se o título e as primeiras linhas do site, etc.) e
- no caso dos catálogos, os critérios humanos utilizados para a indexação e classificação das informações.”

Para se elaborar estratégias que correspondam às

expectativas informacionais dos usuários, é preciso entender como as Ferramentas de Busca indexam e quais são os recursos disponíveis para a elaboração das estratégias de busca adequadas.

Bueno e Vidotti [5] explicam a utilização de alguns operadores e relacionam alguns recursos, apontando diferenças entre as Ferramentas Alta Vista, Yahoo! Brasil, Cadê?, Netscopio, Metacrawler e Metaminer.

As Ferramentas de Busca, geralmente, trabalham com operadores booleanos, posicionais, truncamento e a combinação destes, além de recursos adicionais como a busca por linguagem natural, obrigatoriedade ou não da ocorrência do termo, diferenciação entre maiúscula e minúscula, diferenciação de acentos e caracteres especiais, além de outros que exigem conhecimento de conceitos prévios e da sintaxe da busca, como busca por data, por domínio, por URL, por título, por outros idiomas, tipo de documento.

A utilização correta dos operadores e dos demais recursos interferem na qualidade das respostas obtidas com relação à sua pertinência e exaustividade, como uma busca por nome próprio que pode ser otimizada levando-se em conta a diferenciação que a Ferramenta faz ou não para esta busca.

Quanto ao nível de busca, geralmente, as Ferramentas de Busca oferecem mais de um nível de busca, conhecidos por busca simples e avançada dependendo do conhecimento e experiência do usuário. A busca avançada fornece maiores opções e recursos para elaboração das estratégias de busca.

As buscas nestes níveis podem variar, ou seja, uma mesma busca efetuada nos níveis simples e avançado pode obter respostas diferentes, ou seja, algumas opções podem estar disponíveis apenas em um nível de pesquisa, como no exemplo a seguir:

Estratégias	Busca Simples		Busca Avançada
	Web Pages Encontradas	Ocorrências de palavras	
frank sinatra	51.352	Frank sinatra: 112557	66.485
Frank Sinatra	56.316	Frank Sinatra: 98521	60.277
+frank +sinatra	33.924	Sinatra: 231330; Frank: 5293791	66.485
frank and sinatra	120.560	Sinatra: 231330; Frank: 5293791 Ignoradas (and): 1623918405	74.224
Frank and Sinatra	107.545	Sinatra: 196221; Frank: 4647454 Ignoradas (and): 1623918405	66.949
"frank sinatra"	51.352	frank sinatra: 112557	66.485
"Frank Sinatra"	56.316	Frank Sinatra: 98521	60.277

frank near sinatra	1.075.220	sinatra: 231330; frank: 5293791; near: 10182866	69.186
frank near(1)sinatra	1.655.679	near 1: about 7000; sinatra: 231330; frank: 5293791	124
frank near(5)sinatra	1.641.360	near 5: about 3000; sinatra: 231330; frank: 5293791	35

**Tabela 1: Busca simples e avançada, sem restrições e respostas apenas das Web Pages**

Este simples exemplo nos informa que devemos conhecer as ferramentas de buscas, seus recursos e operadores para a formulação de estratégias, através de seus textos explicativos e de ajuda. É importante observar que informações são manipuladas de forma diferente pela mesma ferramenta, o que conduz a estudos detalhados da estrutura de armazenamento e indexação da mesma, e como, em geral, estas informações organizacionais não são divulgadas, tornam-se necessários estudos experimentais e deduções por inferências.

Como forma de facilitar a pesquisa, principalmente para o usuário leigo, algumas Ferramentas de Busca fornecem uma interface gráfica mais amigável, embutindo em opções de escolha os operadores que utilizam para busca. Assim, não é exigido do usuário o conhecimento prévio para elaborar a sintaxe da busca.

A forma de apresentação dos resultados também se diferencia, pois algumas Ferramentas de Busca oferecerem opções de melhor visualização e ordenação dos resultados que podem ser parametrizáveis pelo próprio usuário.

Rosenfeld e Morville [3, p.113-115], apontam como fatores para a apresentação dos resultados da consulta de uma ferramenta de busca, o grau de estrutura do conteúdo das Bases de Dados e a escolha do usuário, ou seja, como o usuário deseja ordenada a informação que ele procura? Algumas Ferramentas de Busca permitem que o próprio usuário determine a exibição dos resultados dentre as opções existentes, tais como: quantas e quais informações devem ser exibidas para cada documento recuperado? Quantos documentos devem ser exibidos? Como devem estar ordenados os resultados dentre as opções: cronológica, alfabética ou relevância?

Em estudo realizado em dezembro de 1997, dois pesquisadores do NEC Research Institute, localizado em Princeton, New Jersey, Estados Unidos da América, Lawrence e Giles [7], avaliaram as respostas oferecidas por seis Ferramentas de Busca que indexam texto completo: Alta Vista, Excite, HotBot, Infoseek, Lycos e Nothern Light. As consultas correspondiam a questões da rotina de trabalho dos pesquisadores do mesmo Instituto e foram avaliadas um total de 575 respostas. Através de uma metodologia previamente determinada, foram considerados:

- somente as resposta cujos documentos pudessem ser

- carregados e/ou acessados;
- número máximo de 600 respostas por consulta (eliminando as duplicações);
- respostas obtidas em até 60 segundos;
- documentos que continham exatamente todos os termos pesquisados, eliminando-se o plural por exemplo, devido à forma diferente de tratamento que as Ferramentas dão a estas particularidades;

e, foram desconsiderados:

- consultas com *stop words* (por exemplo, artigos, preposições), uso de caracteres especiais porque cada Ferramenta de Busca trata-os de forma diferenciada;
- respostas duplicadas entre as Ferramentas, ou seja, documentos que tinham a mesma URL.
- Como resultados, tem-se que:
- HotBot é a Ferramenta mais abrangente nesta avaliação, seguido por Alta Vista, Northern Light, Excite, Infoseek e Lycos.
- dois fatores são importantes para o resultado alcançado: o estado das Bases de Dados de cada Ferramenta no momento da consulta e a forma que cada Ferramenta de Busca indexa as palavras de um documento, uma vez que cada uma possui sua própria metodologia;
- alcance das Ferramentas de Busca é crescente de acordo com o número de Ferramentas utilizadas, assim, o número de documentos retornados é maior tanto quanto maior for o número de ferramentas consultadas, fator que privilegia grandemente os Metapesquisadores;
- *ranking* das Ferramentas de Busca que apresentam nas suas respostas documentos inválidos, ou seja, páginas alteradas ou que não existem mais, estabeleceu-se em: Lycos, 1,6%; Excite 2,0%; Alta Vista, 2,5%; Infoseek, 2,6%; Northern Light, 5,0% e HotBot, 5,3%;
- tamanho da Web indexável também foi calculada em um total de 320 milhões de páginas;

Lawrence e Giles [8] deram continuidade à sua pesquisa com as Ferramentas de Busca, ampliando o número de ferramentas e consultas num total de 11 ferramentas (Alta Vista, EuroSeek, Excite, Google, HotBot, Infoseek, Lycos, Microsoft, Northern Light, Snap e Yahoo) e 1.050 consultas e, calculando o tamanho estimado da Web indexável em 800 milhões de páginas, 180 milhões de imagens, 3 milhões de servidores, onde 83% são comerciais.

As conclusões deste estudo realizado em fevereiro de 1999 foram:

- alcance que as Ferramentas de Busca atingem na Web não é superior a 38,3% pela ferramenta Northern Light, seguido de 37,1% pelas Snap e Alta Vista;
- alcance das Ferramentas de Busca com relação ao tamanho estimado da Web corresponde a 16% também pela ferramenta Northern Light, seguido por 15,5% com Snap e Alta Vista, ou seja, pelo tamanho da Web nenhuma ferramenta consegue cobrir mais que 16% o crescimento da Internet;
- a porcentagem de links inválidos chega a 14% pelo Lycos e a menor porcentagem de 2,2% pelo HotBot;

- a média em dias da atualização de novos documentos chega a 235 pelo Yahoo e, como menor valor, 141, pelo Northern Light.

Diante deste quadro, conclui-se que as Ferramentas indexam apenas uma fração da Web e o seu alcance é significativamente limitado, devido às limitações técnicas que envolvem o desempenho de cada Ferramenta de Busca em particular.

Na literatura científica não há a indicação de uma Ferramenta de Busca ideal ou completa, assim, para se decidir por qual ferramenta escolher, deve-se considerar o assunto desejado, a resposta esperada e as características funcionais e operacionais da mesma, e como estratégia pode-se utilizar mais de uma Ferramenta para a mesma pesquisa pela diversidade dos documentos indexados em suas Bases de Dados.

“O acompanhamento da evolução das Ferramentas de Busca e das formas de uso de seus operadores são essenciais para uma busca estratégica de informações na Internet, pois o usuário pode através delas usufruir criteriosa e conscientemente do que de melhor a WWW oferece”. [5, p.48].

A cada dia, novas Ferramentas de Busca são criadas e disponibilizadas, porém esta tecnologia não está acompanhando o crescimento exponencial da WWW, principalmente com relação à atualização de suas Base de Dados, o que proporciona um grau de instabilidade (ou insegurança) ao se consultar a Internet com a utilização deste instrumento.

Diante do exposto, e considerando a complexidade da rede Internet, Ferreira [9, p.260] salienta que:

“torna-se necessário o estabelecimento de programas de capacitação do profissional de informação, de modo a torná-lo apto a usar os recursos da Internet e identificar estratégias eficientes para sanar necessidades informacionais de seus clientes, ou ainda para capacitar esses clientes no uso mais adequado daqueles recursos”

As atuais tecnologias de informática e comunicação estão possibilitando uma reorganização das atividades biblioteconômicas com o objetivo principal de atender de forma precisa e rápida as necessidades dos usuários de bibliotecas tradicionais, digitais, virtuais, e da Internet.

O profissional de biblioteconomia, prático ou teórico, não pode ficar à margem deste mundo globalizado e tecnológico, mas sim propor metodologias eficazes de representações descritivas e temáticas automáticas, a fim de permitir a recuperação de documentos eletrônicos relevantes e pertinentes, existentes especialmente nesta imensa Torre de Babel que é a Internet.

Podemos concluir que existe um campo de pesquisa amplo e pouco explorado para os profissionais da Biblioteconomia, com relação aos tratamentos descritivos e temáticos dos documentos disponíveis na Internet, que são catalogados em bases de dados pelas ferramentas de busca de forma automática ou manual,

para que esses documentos possam ser recuperados de maneira a atender as expectativas do usuário numa relação eficaz de relevância x pertinência x tempo de busca.

#### NOTAS

- \* Trabalho Apresentado no XII SEMINÁRIO NACIONAL DE BIBLIOTECAS - SNBU. Florianópolis, Brasil. 2000.
- 1. BUSH, V. - As We May Think. The Atlantic Monthly. [Em linha]. Jul. (1945). Versão eletrônica preparada por Denys Duchier, abril (1994). URL: <http://www.isg.sfu.ca/~duchier/misc/vbush>
- 2. LANCASTER, F.W. - Ameaça ou oportunidade? O futuro dos serviços de biblioteca à luz das inovações tecnológicas. Revista da Escola de Biblioteconomia da UFMG. Belo Horizonte, 23:1 (1994), p.7-27.
- 3. ROSENFELD, Louis, MORVILLE, Peter. - Searching systems. In: \_\_\_\_\_ . Information architecture for the World Wide Web. Sebastopol, CA : O'Reilly, cap.6 (1999), p.99-130.
- 4. BRANSKI, Regina Meyer. - Localização de informação na Internet : características e formas de funcionamento dos mecanismos de busca. Campinas: UNICAMP, 1998
- 5. BUENO, Márcia Correa, VIDOTTI, Silvana Aparecida Borsetti Gregorio. - Uso estratégico das Ferramentas de Busca na Internet. In: SIMPÓSIO INTERNACIONAL "PROF. DR. PAULO TARCÍSIO MAYRINK, 3, 1999, Marília. Anais... Marília : Faculdade de Filosofia e Ciências, 1999. p.39-49
- 6. LYNCH, Clifford. - Searching the Internet. Scientific American. [Em linha]. 276:3 (1997) [Consult. 14 Nov. 1999]. URL: <http://www.sciam.com/0397issue/0397lynch.html>
- 7. LAWRENCE, Steve, GILES, C. Lee. - Searching the World Wide Web. Science. Washington, 280:5360 (1998), p.98-100.
- 8. LAWRENCE, Steve, GILES, C. Lee. - Accessibility of information on the Web. Nature. 400:6740 (1999), p.107-109.
- 9. FERREIRA, Sueli Mara Soares Pinto. - Introdução às redes eletrônicas de comunicação. Ciência da Informação. 23:2, (1994) p.258-263