

# **A INTERNET E SEU IMPACTO NOS PROCESSOS DE RECUPERAÇÃO DA INFORMAÇÃO**

**Cenidalva Miranda de Sousa Teixeira  
Ulrich Schiel**

## **Resumo**

Mostramos como as tecnologias da informação, em particular a Internet, estão provocando uma globalização das atividades de recuperação de informação e pesquisa bibliográfica. Apresentamos as diversas formas de interação na Internet e os principais serviços existentes atualmente referentes à recuperação bibliográfica.

## **Palavras-chave**

Internet; Recuperação da informação

## **INTRODUÇÃO**

Com a necessidade de colocar a informação ao alcance de qualquer pessoa, quando e onde for necessário e, por outro lado, com o aumento global do volume de documentos publicados, torna-se difícil gerenciar, disseminar e recuperar a informação em tempo hábil.

Considerando a informática como fator fundamental no processo de recuperação de informação, percebemos que a cada momento surgem tecnologias cada vez mais sofisticadas que permitem uma interação direta entre os usuários e os sistemas.

Atualmente, a arte de disseminar e recuperar informações bibliográficas dispõe, além dos recursos usuais de recuperação, dos serviços a nível mundial, como WWW (World Wide Web ou Web), baseado nos conceitos de hipertexto e hipermídia.

Este trabalho encontra-se estruturado da seguinte forma: inicialmente abordaremos os fundamentos da recuperação da informação e os respectivos serviços, em seguida faremos algumas considerações sobre a Internet e alguns dos seus serviços disponíveis, abordaremos também os catálogos Web e as ferramentas de buscas mais conhecidas e usadas na Internet.

## **A RECUPERAÇÃO DA INFORMAÇÃO**

### **Fundamentos da Recuperação da Informação**

A explosão documentária aumentou significadamente a dificuldade de recuperar informação em sistemas manuais.

O processo de recuperação de informação compreende basicamente três etapas: indexar, armazenar e recuperar.

Com o advento da informática, indexar, armazenar e recuperar informações tornaram-se uma tarefa mais simples e eficiente, por haver recursos que permitem maior rapidez nestes três processos.

Para proporcionar a distribuição de dados e a prestação de uma variedade de serviços com documentos eletrônicos, é imprescindível o uso do computador e das telecomunicações.

Inicialmente, o computador foi usado para armazenar informações bibliográficas e gerar índices impressos que poderiam ser consultados da mesma maneira que se consultam os índices impressos produzidos manualmente. Hoje, os sistemas informatizados oferecem uma infinidade de recursos de busca, como as consultas utilizando uma combinação complexa de

termos, feitas em linha a bases de dados, que não podemos comparar aos dos sistemas manuais. A cada dia nos deparamos com novas tecnologias de recuperação de informação, a telemática surge para reduzir drasticamente as distâncias por intermédio das redes de teletransmissão de dados e a multiplicação, não só da potência e velocidade de processamento, mas também dos recursos lógicos por meio de redes locais e remotas, que se integram nos chamados sistemas abertos, equipamentos informáticos de tipos mais variados [Cun-94].

Um outro fator relevante diz respeito às interfaces dos sistemas de recuperação de informação. Elas se baseavam em comandos, tornando-se confusas, de difícil entendimento e impenetráveis pelo usuário inexperiente. Era necessário que o mesmo aprendesse várias formas e expressões sintáticas para que pudesse realizar as suas tarefas rotineiras. Para vencer essas dificuldades, os sistemas de recuperação adotaram as interfaces baseadas em *menus*, para, dessa forma, torná-los acessíveis aos usuários inexperientes ou ocasionais.

Atualmente é cada vez mais freqüente o uso de interfaces gráficas nesses sistemas, que têm como principal benefício, tornar as atividades do usuário final, diante da máquina, bastante comuns, disponibilizando recursos gráficos que simulem as atividades do cotidiano do usuário, quando este deseja realizar determinada tarefa. Isto se resume por uma fácil memorização das tarefas a realizar [Shn-92].

Visando a recuperar informação de forma eficiente e eficaz, os mais diversos centros de documentação, serviços de informação e bibliotecas construíram suas bases de dados, que são fontes de informação automatizada que podem ser pesquisadas de diversos modos. Elas podem ser armazenadas em meios magnéticos ou ópticos e acessadas local ou remotamente.

A tecnologia do Compact Disc/Read Only Memory — CD-ROM — discos compactos de memória apenas para leitura — suporta as mais criativas bases de dados em multimídia e tem como vantagens: a grande capacidade de armazenamento, pois pode conter milhões de páginas de texto, permitindo o uso de som e imagens gráficas de fácil utilização [Row-94]. Esta tecnologia foi utilizada inicialmente para distribuir bases de dados de modo a não sobrecarregar as bases centralizadas e, atualmente, vem substituindo documentos impressos, tais como, coleções de revistas, enciclopédias, anais etc.

Atualmente, o surgimento da Internet como fonte de informação, disponibilizando os seus mais diversos serviços, possibilitou o acesso a uma enorme quantidade de bases de dados, como STN International, Dedalus etc., e também aos acervos das grandes bibliotecas, como Library of Congress, Biblioteca Nacional, Public Library etc., de maneira rápida e eficiente.

### **Serviços de Recuperação**

O avanço tecnológico, que possibilita a disseminação da informação por computadores, CD-ROM e outros meios de armazenamento, torna-nos cada dia mais exigentes, buscando meios mais eficientes de armazenar e de principalmente recuperar a informação.

As bibliotecas e os demais serviços de informação construíram suas bases de dados que podem ser acessadas por meio das redes de telecomunicações em nível nacional e internacional, e também mediante a atuação dos servidores em linha que são responsáveis pela integração de uma variedade de bases de dados produzidas por si mesmas ou por outras instituições, bem como pelos mecanismos necessários para que essas bases sejam consultadas remotamente.

Segundo Cuadra [Cua-89], no seu diretório de bases de dados *on-line* até 1988 constam mais de 500 desses servidores e a tendência é aumentar cada vez mais. Até então, estas bases de dados estavam disponíveis por meio de comunicação em linha. Em 1990, praticamente inexistiam servidores disponíveis na Internet. Em 1994, as instituições produtoras começaram a disponibilizar seus serviços na Internet e atualmente projeta-se um total de oito milhões de servidores ao final de 1996.

Citamos, a seguir, alguns desses servidores em linha nacionais e internacionais que acoplam várias bases de dados sobre os mais diversos assuntos:

- AMERICA ONLINE(Estados Unidos) – com 50 bases de dados sobre jogos, correio eletrônico, teleconferência, anúncios, finanças, noticiário ( TV, teatro etc.)
- ARUANDA/SERPRO(Brasil) – com mais de 11 bases de dados sobre cadastros industriais, marcas e patentes.
- BIREME(Brasil) – com mais de quatro bases de dados sobre área médica.
- BRS Information Technologies(Nova Iorque) – com mais de 100 bases de dados sobre medicina, química, comércio, ciências sociais e humanidades.
- CENAGRI(Brasil) – com mais de cinco bases de dados sobre ciências agrícolas
- CIN/CNEN(Brasil) – com mais de oito bases de dados sobre energia nuclear, física, eletrônica e energia elétrica.
- COMPUSERVE(Estados Unidos) – com 80 bases de dados sobre finanças, telecompras, jogos, ciência e medicina.
- DATA-STAR(Suíça) – com mais de 250 bases de dados sobre comércio, ciências biomédicas, química, engenharia e notícias.
- DIALOG Information Retrieval Services(Palo Alto, Califórnia) – com 380 bases de dados, praticamente sobre todas as áreas.
- DIMDI(Alemanha) – com mais de cinco bases de dados sobre ciências biomédicas.
- Dow-Jones(Estados Unidos) – com mais de cinco bases de dados sobre comércio e apoio à investimentos.
- ECHO (Comunidade Européia) – com 30 bases de dados, praticamente sobre todas as áreas.
- European Space Agency's Information Retrieval Service(Frascati, Itália) – com mais de 130 bases de dados sobre ciência e tecnologia, comércio e finanças, informações sobre empresas, saúde e segurança ocupacional, patentes e notícias.
- FGV(Brasil) – com duas bases de dados sobre dados econômicos-estatísticos, catálogo de livros e dados.
- IBICT(Brasil) – com seis bases de dados em ciência da informação, catálogo coletivo de periódicos, teses.
- JOIS(Japão) – com mais de 10 bases de dados sobre diversas áreas temáticas.
- ORBIT Search Service(Santa Mônica, Califórnia) – com mais de 100 bases de dados sobre informações científicas, técnicas e sobre patentes.
- Pergamon Financial Data Service(Londres) – com mais de 30 bases de dados sobre serviços destinados à comunidade comercial européia, incluindo *marketing* e prospecção de vendas, finanças e proteção ao crédito e informações sobre empresas.
- PRODASEN(Brasil) – com mais de 12 bases de dados nas áreas de direito e jurisprudência.
- QUESTEL(França) – com mais de 60 bases de dados sobre patentes, química, marcas registradas, medicina, ciência e tecnologia, comércio e notícias.
- STN International(Alemanha) – com 132 bases de dados em ciência e tecnologia.

É possível que uma determinada base de dados se encontre acoplada a diversos servidores e que podem conter informações similares entre si.

Como podemos notar, existem muitos serviços específicos para recuperação de informação bibliográfica dificultando que uma pessoa que necessite de uma informação, conheça e acesse o serviço adequado.

Um levantamento feito por Cecília Pontes[Pon-94] sobre os recursos e serviços de informação disponíveis nos institutos de pesquisa da Associação Brasileira das Instituições de Pesquisa Tecnológica Industrial (Abipti), nos itens bases de dados adquiridas pela instituição e acesso a bancos de dados e redes de informação nacionais e internacionais, constatou-se que somente cinco institutos de pesquisa fazem acesso regularmente a banco de dados *on-line* internacionais. Os serviços mais utilizados foram Dialog, STN International, Pergamon, Telesystems e Bases em CD-ROM.

Para minimizar essas dificuldades, segundo Jennifer [Row-94], algumas características devem ser consideradas quando da avaliação dessas bases de dados, das quais citamos:

- Cobertura temática – identificada através dos guias de bases de dados.
- Tipos de dados – se inclui referências, imagens, figuras, textos integrais etc.
- Atualidade – qual a frequência de atualização dos arquivos e o período de abrangência.
- Facilidade de uso – se a interface é amigável e se inclui treinamento.
- Tipo de saída – se disponibiliza diversos formatos de impressão e permite importação de dados para geração de arquivos.
- Linguagem de Indexação – se controlada ou natural, qual o grau de exaustividade da indexação.
- Custo – custo inicial e adicional caso venha a ter, a relação custo benefício.
- Documentação e instrumentos auxiliares de busca - meios de ajuda, *help's*.
- Viés – verificar se existe ênfase em informações regionais, européias, norte-americanas ou brasileiras.
- Cobertura cronológica – o período de cobertura da base de dados desde quando se acha disponível em linha.
- Servidor – o servidor em linha através do qual a base de dados se acha disponível.

Muitos desses servidores em linha colocaram suas bases de dados disponíveis via Internet , utilizando-se dos serviços telnet, como o DIALOG (estritamente comercial) e-mail, ftp,archie, gopher, verônica, WAIS e WWW.

## **A INTERNET E A RECUPERAÇÃO DA INFORMAÇÃO**

### **Considerações Gerais sobre a Internet**

A cada dia surgem novos lançamentos que permitem dinamizar a recuperação de informação, como é o caso da Internet, que nasceu em 1969, como Arpanet, patrocinada pelo Departamento de Defesa Norte Americano – Defense Advanced Research Projects Agency (Darp) – com o objetivo de permitir que engenheiros e cientistas que trabalhavam em projetos militares em toda a América pudessem compartilhar computadores caros e outros recursos. A Arpanet original expandiu-se e suas ramificações formam a espinha dorsal do que chamamos hoje Internet.

A Internet é um mundo novo, colorido, quase sem limites, que simplificada pode ser definida como uma rede mundial de redes de computadores, interligando todos os continentes, alcançando mais ou menos 150 países. Significa dizer que a Internet tornou-se uma biblioteca cibernética universal, com vários bibliotecários, onde cada um utiliza um determinado serviço de pesquisa, para encontrar o que deseja na rede. Na década de 80, a rede proliferou no meio acadêmico.

No Brasil, os primeiros passos visando à conexão às redes internacionais foram dados pelo Laboratório Nacional de Computação Científica (LNCC), ao se conectar com a University of Maryland em setembro de 1988. No mês de novembro do mesmo ano, foi a vez da Fundação de Amparo à Pesquisa do Estado de São Paulo (Fapesp) com o Fermi National Laboratory (Fermilab) em Chicago. Posteriormente, em maio de 1989, a Universidade Federal do Rio de Janeiro (UFRJ) conecta-se à University of California at Los Angeles (Ucla), por intermédio da rede Bitnet<sup>1</sup>, visando à comunicação com pesquisadores de universidades e centros de pesquisa no exterior.

Em 1989, foi implementada a Rede Nacional de Pesquisa (RNP) que se tornou a "espinha dorsal" (*backbone*) dessa grande rede de computadores, rompendo paradigmas e nos colocando em plena revolução "informacional". O principal objetivo era promover e incentivar a troca de informações entre cientistas brasileiros e estrangeiros, permitindo um intercâmbio global de conhecimentos.

A Internet integra os esforços em redes acadêmicas no país. Ela independe da plataforma, isto é, pode se conectar a qualquer tipo de computador, inclusive PCs.

A estrutura tecnológica que permite o funcionamento da Internet são os protocolos<sup>2</sup> — Protocolo de Controle de Transmissão (TCP) e Protocolo Internet (IP).

O modelo de arquitetura dos protocolos da Internet se divide em quatro camadas [Soa-95], conforme mostra a figura 1:

FIGURA 1  
Arquitetura da Internet

<b>Aplicação (1)</b>
<b>Transporte (2)</b>
<b>Inter-rede (3)</b>
<b>Interface de Rede (4)</b>

Onde:

- a camada 1 diz respeito às tecnologias usadas para prover serviços aos usuários, como, Telnet, FTP etc.
- a camada 2 diz respeito às tecnologias usadas para implementar a comunicação entre estações, no caso, o TCP.
- a camada 3 diz respeito às tecnologias usadas para realizar a abstração de inter-rede, no caso o IP.
- a camada 4 diz respeito às tecnologias física e de enlace de dados usadas para transmissão ao nível do meio físico. Define as características mecânicas, elétricas, funcionais e procedurais para ativar, manter e desativar conexões físicas que se destinam a transmitir *bits* entre entidades do nível de enlace [Soa-95]. Temos como exemplos a Ethernet, Token Ring, X.25, ATM etc.

---

<sup>1</sup> Estabelecida em maio de 1981, nos Estados Unidos, como uma rede cooperativa entre a City University of New York e a Yale University, utilizando como padrão os computadores IBM, rodando sob sistema operacional VM/370, sendo o Remote Spooling Communication Subsystem (RSCS) a base do sistema de comunicação. Os serviços básicos disponíveis são correio eletrônico, transferência de arquivo, mensagens interativas, acesso a banco de dados, sistemas de quadro de aviso, sistemas de listas de distribuição etc.

<sup>2</sup> Protocolo é um conjunto de regras e comandos preestabelecidos entre os componentes de uma rede para que se efetuem suas comunicações, ou seja, para que os computadores envolvidos "entendam" às mensagens uns dos outros. Os usuários não precisam conhecê-los em detalhes para que possa deles se utilizar [Kro-93].

A literatura atual fornece informações sobre o mundo maravilhoso da Internet, mas a maioria dos usuários fica se perguntando como conseguir o acesso adequado, pois, dependente desse acesso, as ferramentas e serviços disponíveis variam enormemente.

Os tipos de acesso são:

- Emulação de Terminal – neste caso o computador serve como porta de entrada para outro computador.
- Acesso TCP-IP Completo:

⇒ Acesso Dedicado – neste caso, o computador é sempre um membro da Internet. É a única forma para fornecer serviços na Internet.

⇒ Acesso Discado – neste caso, o computador é um membro da Internet no momento que desejar.

Considerando o acesso TCP-IP Completo, a Internet oferece vários serviços [Cri-94], [Liu-94], [Hah-95], dentre os quais citamos:

- Correio Eletrônico (Electronic Mail ou e-mail) – é uma das ferramentas mais utilizadas que implementa um sistema de troca de mensagens eletrônicas entre pessoas, instituições e empresas.

Cada endereço eletrônico é único, na maior parte do mundo é constituído pela seguinte lógica: usuário@computador.

Exemplo: ceni@dsc.ufpb.br, onde:

- "ceni " – identificação individual do usuário na Internet
- "@" – este símbolo é lido como "at ", que significa "em "
- "dsc.ufpb " – computador no qual a conta do usuário está localizada
- ".br " – significa o país ao qual a máquina pertence

Nos Estados Unidos, os endereços não respeitam esse padrão, utilizam siglas que identificam os tipos de organizações, como "com "- comercial; "edu "- educacional; "gov "- governo; "mil "- militar e "net "- redes.

Por intermédio da tecnologia Multipurpose Internet Mail Extensions (Mime) é possível transmitir documentos multimídia.

- Telnet ou Emulação de Terminal – é uma ferramenta que permite a conexão a partir de um computador a um outro computador na rede. Uma vez conectado, funciona como se você estivesse à frente de uma máquina que pode estar no outro lado do mundo.
- Finger – é um serviço que permite solicitar informações sobre determinado usuário. É um sistema cliente/servidor<sup>3</sup>, pois age pelo usuário e solicita a informação ao computador apropriado. No computador remoto, o servidor Finger fica aguardando a tal solicitação. Segundo Harley & Rick [Hah-95], o Finger oferece três tipos principais de serviço:

a) Exibir informações públicas sobre qualquer usuário de um *host*<sup>4</sup> Internet qualquer, o que varia de *host* para *host*. Muitos *hosts* divulgam apenas as informações básicas, como a informação do *host* que a pessoa usa, o *userid*, o último ou o primeiro nome. O serviço Finger pode exibir toda, ou parte da seguinte informação:

---

<sup>3</sup> É uma arquitetura na qual o cliente (computador pessoal ou estação de trabalho) é a máquina solicitante e o servidor é a máquina fornecedora. Os servidores podem ser qualquer computador. O cliente fornece a interface de usuário e executa parte ou todo o processamento do aplicativo. O servidor mantém os bancos de dados e processa solicitações do cliente para extrair dados ou atualizar o banco de dados [Fre-95]. Todos os serviços Internet utilizam esse relacionamento cliente/servidor.

<sup>4</sup> É o computador principal em um ambiente de processamento distribuído, refere-se a um computador baseado em *timesharing* (tempo compartilhado) ou a um computador central que controla uma rede [Fre-95].

- ⇒ o *userid* da pessoa
- ⇒ o nome completo da pessoa
- ⇒ o *userid* está logado no momento
- ⇒ a última vez que alguém se logou usando aquele *userid*
- ⇒ se o correio já foi lido
- ⇒ um número de telefone
- ⇒ um endereço comercial
- ⇒ informação que a pessoa preparou especialmente para o público em geral ( por exemplo, o horário comercial )

b) Para checar quem está usando o *host* Internet no momento, mostrando toda, ou parte da seguinte informação para cada *userid* logado:

**P** *userid*

- ⇒ nome completo
- ⇒ quando o *userid* se logou
- ⇒ quanto tempo passou desde a última atividade naquele terminal
- ⇒ número de telefone e informação comercial
- ⇒ nome do computador ou terminal no qual a pessoa se logou

c) Para se comunicar com certos *hosts* Internet que foram organizados para oferecer outros tipos de informação.

Basicamente o uso do serviço Finger é para exibir informação pública sobre uma pessoa em particular.

- **Talk** é um serviço que estabelece uma conexão entre seu endereço no computador e um outro usuário ligado no mesmo momento, possibilitando manter uma conversa com uma pessoa que fisicamente pode estar do outro lado do mundo. As duas pessoas podem escrever ao mesmo tempo, cada uma utiliza uma metade da tela. Funciona como um telefone, sendo que ao invés da voz, utiliza-se a escrita.
- Protocolo de Transferência de Arquivos (FTP) é um serviço que consiste em disponibilizar para os usuários da rede, programas e arquivos armazenados em diversos servidores espalhados pelo mundo. É o principal método para transferência de arquivos entre computadores na rede.
- **Archie** é um serviço que permite ao usuário encontrar o que deseja com relativa facilidade em banco de dados mantidos em computadores em lugares diferentes conectados na Internet. O usuário fornece o assunto do seu interesse e ele lista todos os *sites* com os respectivos diretórios onde se encontra algo parecido ou igual ao que você solicitou, que poderá ser copiado via FTP. Funciona como um catálogo geral do FTP.
- **Gopher**, desenvolvido na Universidade de Minnesota, é um serviço de recuperação de documentos que se apresenta mediante sistemas de *menus* hierárquicos, a partir dos quais pode acessar virtualmente qualquer tipo de informação textual.
- **Very Easy Roadent Oriented Net wide Index to Computerized Archives (Veronica)** é um localizador de informações que utiliza palavras-chave para procurar em todos os servidores gopher.
- **Wide Area Information Service (WAIS)** é um serviço que pesquisa vários grupos de dados também chamados de *source* ou fonte, a partir da indicação da palavra-chave e da fonte desejada, que podem ser mais de uma. O WAIS se conectará aos computadores que contêm as fontes e pedirá a eles para conduzir a pesquisa. Assim, toda ocorrência em todos os artigos nos bancos de dados especificados serão pesquisados e, em seguida, será exibida como resultado uma lista de artigos ou citações. Essa lista é apresentada por intermédio de menus com os assuntos mais relevantes no início. O WAIS considera artigos mais relevantes aqueles com maior número de ocorrência da palavras-chave. Após selecionar o item do seu interesse, o WAIS recupera o texto, exibindo-o em seu monitor, ficando disponível para salvar em um arquivo.

O protocolo padrão é o Z39.50, que fornece os procedimentos uniformes para a recuperação de informação bibliográfica remotamente.

- **Internet Relay Chat (IRC)** é conhecido como o local de bate-papos da Internet, possibilita manter uma conversa sobre um determinado assunto com uma pessoa ou um grupo de pessoas espalhados no mundo.

- **Mailing List** permite interação com usuários na rede sobre os mais diversos assuntos, onde as discussões são conduzidas por correio eletrônico.

- **Usenet Newsgroups** forma um grupo de discussão com a troca de mensagens. Esses grupos são organizados hierarquicamente, por tópicos e subtópicos, como, por exemplo:

"alt "- grupos alternativos, sem moderador, vale tudo;

"comp "- sobre computadores;

"brasil "- sobre o Brasil;

"net "- sobre redes;

"rec "- entretenimento;

"sci "- discussões científicas etc.

- **World Wide Web ou Web (WWW)** — a Teia Global – é, sem sombra de dúvidas, a "revolução" da Internet. É uma coleção de documentos hipertextos ligados entre si, criando um mundo de informações digitais que envolve texto, imagens e sons, constituindo-se em um dos maiores acervos multimídia que integra as tecnologias de comunicação, transmissão de imagens e sons, criando uma verdadeira rede de difusão de conhecimento [Ins-96]. As páginas multimídias da WWW são manipuladas por programas especiais chamados de *browsers*, ou Cliente Web.

Cada conexão é referenciada por um único nome, chamado de Uniform Resource Locator (URL). Os URLs permitem que os programas *browsers* identifiquem e localizem as referências de cada serviço, constituindo assim as suas ligações que podem estar em outros servidores, como Gopher, Wais, FTP.

O protocolo de comunicação entre *browsers* e servidores Web<sup>5</sup> é o Hiper Text Transfer Protocol (HTTP), que possibilita a transferência de documentos hipertexto, mas os *browsers* Web também se comunicam com outros servidores, e um servidor WWW pode se comunicar com outro servidor por meio de gateway<sup>6</sup>, como mostra a figura 2:

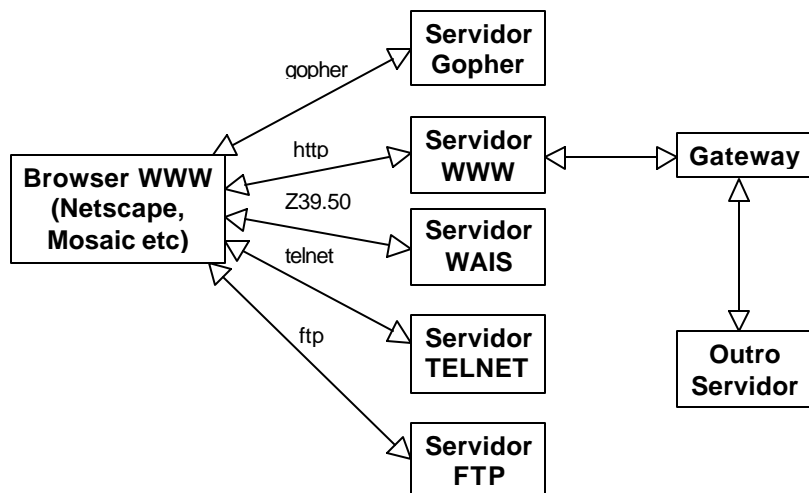
FIGURA 2  
WWW e suas conexões

---

<sup>5</sup> É um computador host que dispõe do serviço Web da Internet atendendo solicitações de outros computadores numa rede.

<sup>6</sup> É um computador que executa conversão de protocolo entre tipos diferentes de redes ou aplicativos. Por exemplo, uma gateway pode conectar uma rede local de computadores pessoais a um mainframe[Fre-95].





### Catálogos Web e Ferramentas de busca

Na Web não existe um diretório oficial universal consolidado de endereços devido ao crescimento constante de informações sobre uma variedade de assuntos. Considerando esse fato, algumas empresas independentes oferecem catálogos Internet e mecanismos de busca que permitem localizar determinada informação sem necessariamente conhecer sua URL. Estas ferramentas permitem a recuperação de informação sobre os mais diversos assuntos.

Atualmente existem vários catálogos, serviços de pesquisa bibliográfica na Web, alguns específicos como STN International, Dedalus, Biblioteca Nacional, Bookpool, e outros genéricos, como Altavista, Yahoo, Lycos e Cadê?.

A seguir, faremos uma breve descrição sobre algumas das mais comentadas e usadas ferramentas de busca:

- **Excite**, criado recentemente pela Architext Software, oferece duas ferramentas de navegação em separado: NetSearch, que permite procurar, simultaneamente, os grupos de notícias Usenet via texto simples ou pelo teclado e o endereço Web; NetReviews, que oferece artigos Web organizados por assuntos.
- **Altavista**, desenvolvido em 1995, no Laboratório de Pesquisa Digital, Palo Alto, Califórnia, por uma equipe de técnicos e colocado em uso a partir de 15 de dezembro do mesmo ano. Permite acessar mais de 30 milhões de páginas Web em mais de 275 mil servidores e três milhões de artigos de 14 mil *Usenet news groups*. É uma ferramenta poderosa, conhecida e usada mundialmente por milhões de pessoas, considerada por muitos, a "memória de elefante da Internet" pela sua abrangência, utiliza robôs para manutenção das *sites*<sup>7</sup>, oferece pesquisa simples, por palavra-chave e avançada por frase, elimina pontuação e caracteres especiais como, \$, %, /, \, &, etc.
- **Lycos**, um dos mais abrangentes e poderosos localizador de páginas na Web. Seus recursos exclusivos incluem busca por texto e uma lista dos 250 endereços Web mais populares.
- **WebCrawler**, criado pela America Online, utiliza robô para coletar e indexar as páginas Web, oferece busca rápida e simples a partir de um índice baseado em conteúdos.
- **Whole Internet Catalog**, criado pela Global Network Navigator, é uma lista de 1.200 endereços Web gratuitos. Apesar de muito bem organizado por tópico, não possui um mecanismo de busca.
- **Yahoo!**, criado em abril de 1994, por David Filo e Jerry Yang, na época, estudantes de engenharia elétrica da Universidade de Stanford. Inicialmente começaram a catalogar servidores Web, com o objetivo de ajudar as pessoas a

<sup>7</sup> Corresponde a um endereço Internet na Web. Denota um computador *host* que possui uma página de apresentação Web[Ven-96].

navegarem no imenso volume de informações existentes. Esta ferramenta tornou-se rapidamente popular, por oferecer mecanismo de busca associado a uma lista de *sites* agrupados por assunto. Em fevereiro de 1995, a Netscape Communications criou um *link*<sup>8</sup> de dentro do seu paginador para o Yahoo!, mediante o emprego do botão "Net Directory", colocando à disposição computadores e linhas telefônicas. Incluem como recursos a Web Launch –serve para anunciar ao mundo seu novo endereço Web. Os seus serviços são gratuitos aos usuários.

- **InfoSeek** oferece busca na Internet, tanto gratuita quanto paga. Gratuitamente cobre localidades Web e grupo de notícias, retornando apenas 10 resultados por unidade de cada busca. Encontra-se freqüentemente congestionada. Já o serviço pago oferece diferentes planos de pagamentos e acesso a diversos e peculiares banco de dados de notícias, além de fornecer uma conexão mais rápida.

- **Cadê?**, Índice Geral da Internet em Português – oferece um catálogo dividido em grandes áreas que abrange ciência e tecnologia, compras, cultura, educação, esportes, finanças, governo, indústria e comércio, informática, Internet, lazer, notícias, referência, saúde, serviços, sociedade, os provedores de acesso com suas respectivas tarifas, bem como uma ferramenta que utiliza busca simples por palavra-chave e elimina todos os caracteres especiais da língua portuguesa (não significativos). As palavras pesquisadas devem conter no mínimo dois caracteres, e não se faz distinção entre letras maiúsculas e minúsculas. A pesquisa é realizada no endereço (URL), no nome da página e na sua descrição. Oferece ainda o cadastramento do usuário.

- **Yaih?** O diretório da Internet Brasil –, criado pela Rede Nacional de Pesquisa (RNP), oferece um catálogo dividido em grandes áreas, que abrange artes, ciência e tecnologia, ciências humanas, computação e Internet, educação, entretenimento, governo, negócios e economia, notícias e recreação, referência, regional, saúde e sociedade e cultura e uma ferramenta de busca que utiliza busca simples por palavra chave e composta utilizando os operadores booleanos "and" e "or". Oferece ainda a inclusão de novos URLs.

- **SEI-Bib** Serviço de Informação Bibliográfica –, este serviço é de nossa autoria [Tei-96], está em fase de desenvolvimento na Universidade Federal da Paraíba e atualmente já oferece um catálogo Web especializado em informação bibliográfica, contendo informações sobre bibliotecas, editoras, jornais, livrarias, revistas, sistemas de recuperação de informação e universidades, assim como uma ferramenta de busca que permite ao usuário efetuar pesquisas nos mais diversos serviços de recuperação bibliográfica disponíveis na Web. Posteriormente, serão acopladas novas funções a esse serviço.

## CONSIDERAÇÕES FINAIS

As ferramentas de busca Web são, sem sombras de dúvidas, a melhor opção para efetuar pesquisas na Internet.

A cada dia nos deparamos com novas e sofisticadas ferramentas que oferecem diversas formas de realizar as buscas. Conseqüentemente, cresce a dificuldade na escolha da ferramenta adequada que permita ao usuário economizar tempo e dinamizar suas pesquisas.

Com toda essa gama de informações disponíveis, conforme mencionado por Ferreira [Fer-94], 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.

Atualmente, é possível efetuarmos visitas em bibliotecas via Internet, e temos na Web várias publicações disponíveis, tais como, livros, revistas, jornais etc.

## REFERÊNCIAS BIBLIOGRÁFICAS

[Cua-89] *Cuadra Directory Online Databases*. Cuadra/Elsevier, New York, 1989.

---

<sup>8</sup> Corresponde a uma ligação virtual entre dois computadores da Internet, através de um documento Web.

- [Cun-94] CUNHA, Murilo Bastos. As tecnologias de informação e a integração das bibliotecas brasileiras. *Ciência da Informação*, v.23, n.2, p.182-189, maio/ago.1994.
- [Cri-94] CRISPEN, Patrick Douglas. *The roadmap for the information highway:Internet Training Workshop* University of Alabama-Tuscaloosa, 1994.
- [Fer-94] FERREIRA, Sueli Mara Soares Pinto. 'Introdução às redes eletrônicas de comunicação' *Ciência da Informação*, v.23, n.2, p.258-263, maio/ago.1994.
- [Fre-95] FREEDMAN, Alan. *Dicionário de Informática*. São Paulo: Makon Books,1995.
- [Hah-95] HAHN, Harley & STOUT, Rick. *Dominando a Internet*. São Paulo, Makron Books, 1995. 853p.
- [Ins-96] INSTITUTO BRASILEIRO DE PESQUISA EM INFORMÁTICA. *Internet: método rápido*. Rio de Janeiro: Infobook, 1996.
- [Kro-93] KROL, E. *The whole Internet user's guide & catalog*. Sebastopol/CA:O'Reily & Associates, 1993.
- [Liu-94] LIU, Cricket et all. *Managing Internet Information Services*. Sebastopol/CA: O'Reily & Associates, 1994.
- [Pon-94] PONTES, Cecília Carmen Cunha. 'Institutos de pesquisa tecnológica e serviços de informação' *Ciência da Informação*, v.23, n.2, p.165-170, maio/ago.1994.
- [Row-94] ROWLEY, Jennifer. *Informática para bibliotecas*. Briquet de Lemos/Livros: Brasília, 1994.
- [Soa-95] SOARES, Luiz Fernando Gomes et al. *Redes de Computadores: das LANs, MANs e WANs às Redes ATM*. 2. Ed. Rio de Janeiro: Campus, 1995.
- [Shn-92] SHNEIDERMAN, Ben. *Designing the User Interface-Strategies for Effective Human Computer Interaction*. Addison Wesley Publishing Company, 1992.
- [Tei-96] TEIXEIRA, Cenidalva & SCHIEL, Ulrich. *Um serviço de informação bibliográfica via Internet*. XXIX Congresso Nacional da SUCESU, Natal, Nov. de 1996
- [Ven-96] VENETIANER, Tomas. *HTML: desmistificando a linguagem da Internet*. São Paulo: Makron Books, 1996.

### **Abstract**

*We show how information technology, specially the Internet, is promoting a globalization in information resources and retrieval. The different available interactions with the Internet are presented and the most important bibliographic research services are introduced.*

### **Keywords**

Internet; Information retrieval

### **Cenidalva Miranda de Sousa Teixeira**

Professora do Departamento de Biblioteconomia da Universidade Federal do Maranhão, com extensão em análise de sistemas pela PUC-Rio, atualmente mestranda em informática, na Universidade Federal da Paraíba, na área de sistemas de informação e banco de dados sob orientação do professor Doutor Ulrich Schiel.

e-mail: ceni@dsc.ufpb.br

Web: <http://www.dsc.ufpb.br/~ceni>

**Ulrich Schiel**

Professor Titular do Departamento de Sistemas e Computação da Universidade Federal da Paraíba, na área de sistemas de informação e banco de dados. Bacharel em matemática pela Universidade Mackenzie/SP, mestre em informática pela PUC-Rio e Dr. rer. nat. pela Universidade de Stuttgart/Alemanha. Atividades de pesquisa em bancos de dados temporais, bancos de dados textuais e metodologias de projeto de sistemas de informação

e-mail: [ulrich@dsc.ufpb.br](mailto:ulrich@dsc.ufpb.br)

Web: <http://www.dsc.ufpb.br>