
WALTER MOREIRA

BIBLIOTECA TRADICIONAL X BIBLIOTECA VIRTUAL:
MODELOS DE RECUPERAÇÃO DA INFORMAÇÃO

DISSERTAÇÃO APRESENTADA AO DEPARTAMENTO
DE PÓS-GRADUAÇÃO EM BIBLIOTECONOMIA DA
PONTIFÍCIA UNIVERSIDADE CATÓLICA DE
CAMPINAS COMO REQUISITO PARCIAL PARA
OBTENÇÃO DO TÍTULO DE MESTRE EM
BIBLIOTECONOMIA.

ORIENTADORA: PROFA. DRA. SOLANGE PUNTEL
MOSTAFA

CAMPINAS, 1998

AGRADECIMENTOS

Agradeço,

- a Deus, pelo dom da vida;
- à Lourdes, pelo primeiro estímulo, pela amizade verdadeira e por ter me apresentado Balzac;
- à Ângela, por ter me ensinado a pensar a Biblioteconomia;
- às novas amigas : Marisa e Lourdinha, pelo apoio e pela lisonja;
- à Irmã Olga, pela raiz e pelas asas;
- aos meus pais e meus irmãos, os quais, muitas vezes sem mesmo o saber, amenizaram minhas angústias;
- à Professora Solange, pelo respeito e pelo entusiasmo com que me orientou;
- à Capes, pela bolsa concedida;
- ao Departamento de Pós-Graduação em Biblioteconomia, pela formação e pela informação;
- à Raíssa, minha querida filha, por não questionar o tempo que lhe roubei e
- à Marilda, esposa amada, que me acompanhou e incentivou incansavelmente durante todo o caminho e, nos momentos finais da redação deste meu "filho", prepara um outro, nosso, de muito maior importância, em seu ventre.

SUMÁRIO

Agradecimentos	2
Sumário	3
RESUMO	4
ABSTRACT	5
Apresentação	6
O Problema	8
Capítulo 1 - Tecnologias da Inteligência	11
Capítulo II - O virtual	19
Capítulo III - Sistema de Recuperação da Informação	23
Subsistema de Recuperação da Informação : Desvio Histórico	26
A Biblioteca Virtual	27
Objetivos	33
Objetivo Geral	33
Objetivos Específicos	33
Capítulo IV - Método	35
Capítulo V - Categorias de Análise	36
Revocação / Precisão e o Conceito de Navegação	37
Os Metadados	39
Sistemas ou Serviços de Busca (Search Engines)	43
Modelos de Recuperação da Informação: Lógica Booleana X Lógica Nebulosa	52
Interfaces e Ambientes de Recuperação da Informação	58
Teoria da Caça de Sustento Informacional	70
Recuperação da Informação Hipermídia	73
(In)conclusões	80
Referências	83

RESUMO

Inserindo as bibliotecas no conjunto da "tecnologias da inteligência" o autor desenvolve uma reflexão onde a biblioteca tradicional é contraposta/justaposta à sua contraparte virtual, ressaltando as interpenetrações que lhes caracterizam. Fundamenta a contraposição/justaposição entre os modelos de biblioteca na discussão sobre o conceito de virtualidade e sua falsa oposição ao real. Os modelos de recuperação da informação são discutidos em ambas as bibliotecas e novos conceitos como "navegação" ou "surfing" são comparados com aqueles presentes na recuperação tradicional da informação, como "matching". Discute, ainda, alguns outros conceitos em oposição, como: a lógica nebulosa e a lógica booleana aplicadas às técnicas de recuperação da informação; a inter-relação entre a base da tecnologia da informação e o desenvolvimento de ferramentas de recuperação. Por último, discute a especificidade do ambiente hipermídia em função das novas questões que apresenta ao desenvolvimento da Teoria da Recuperação da Informação. As diferenças radicais percebidas entre as bibliotecas tradicionais e as bibliotecas virtuais não levam o autor a uma polarização quando considera o futuro. Embora a digitalização seja uma tendência irreversível relacionada ao mundo da informação, o autor conclui pela complementaridade dos dois modelos de bibliotecas.

MOREIRA, Wálter **Tradicional library X Virtual library:** models of information retrieval (Mastership Dissertation) Pontifícia Universidade Católica de Campinas, 1998

ABSTRACT

Taking libraries as "technologies of intelligence" the author develops a reflection where traditional library is "opposed" to its virtual counterpart. Models of information retrieval are discussed within the two mentioned models of libraries. New virtual concepts as "navigation" or "surfing" are compared with those present in more traditional information retrieval systems, as "the best match". Here are some other opposite concepts discussed: fuzzy set theory opposed to boolean logic; the interrelation between the basis of information technology and the development of information retrieval tools are also under scrutiny. At last, it discusses the specificity of hypermedia environment which poses new questions to the development of Information Retrieval Theory. The radical differences perceived between traditional and virtual libraries does not lead the author to a polarization when considering the future. Although digitalization is an irreversible tendency related to information world, the author concludes by the complementarity of the two models of libraries.

APRESENTAÇÃO

No meu projeto original (anteprojeto) apresentado como intenção de pesquisa na seleção para o Mestrado, manifestei uma dúvida que era objeto de minhas preocupações em torno da automação de bibliotecas. O questionamento básico estava em identificar até que ponto a propalada automação de bibliotecas havia resolvido ou estaria resolvendo os problemas relacionados à transferência de informação.

Em torno desta pergunta, respondendo-a ou questionando-a, fui reconstruindo ou refazendo a forma de encarar o problema, de maneira que a pergunta na sua forma e intenções iniciais cresceu em complexidade. Implícitas nesta pergunta havia várias generalizações que faziam parte do meu repertório de representações antes de iniciar o Mestrado. Por exemplo, a idéia de que os computadores, como também qualquer outra ferramenta, eram meros instrumentos de trabalho; simples meios. Retirar esses “meros” da minha representação para poder de fato entender o instrumento de trabalho como constitutivo do homem custou-me discussão, reflexão e leituras. Leituras que incluíram principalmente Marx (1988), Levy (1993) e Lojkin (1995), as quais se revelavam essenciais no tratamento do objeto.

Das bibliotecas do projeto original caminhei em direção às tecnologias da inteligência, termo cunhado por Levy (1993) em livro homônimo onde quer mostrar que não há informática em geral mas sim uma floração de novas tecnologias intelectuais — posteriores à fala e à escrita, as quais também recebem este enquadramento — abertas, conflituosas e indeterminadas. Encontrei no mesmo Levy apoio à decisão que me levou a alterar meu foco de discussão: “quanto valeria um pensamento que nunca fosse transformado por seu objeto?”.

A resposta à indagação inicial sobre a resolução ou não dos problemas da biblioteca pelos processos de automação viria mais facilmente através de estudos de caso, opção que rejeitei por acreditar que o assunto merecia ser tratado teoricamente e por identificar que permeando este questionamento estava a recuperação da informação, ou seja, não há determinismo na relação recuperação da informação e tecnologia, nem razão diretamente proporcional.

Tomo, portanto, a discussão das tecnologias da inteligência como aporte teórico, a qual me permite, inserindo a biblioteca neste contexto, compreendê-la na sua especificidade. A questão que se apresenta atualmente refere-se, portanto, à qualidade da recuperação da informação nas bibliotecas virtuais, ou nas *bibliotecas infinitas*, como muitos a chamam, numa alusão ao fato de que a biblioteca virtual, tendo a Internet como *habitat* natural, pode prover infinitas e impensáveis ligações (*links*) a partir do conceito de navegação por hiperdocumentos.

Discutir bibliotecas virtuais e, conseqüentemente, alguns aspectos da Internet, provou-se logo um empecilho a minha (ingênua?) visão de preservação da língua portuguesa. Muito a contragosto, não se me apontou outra alternativa que não fosse traduzir somente os termos para os quais julgasse ter a necessária competência. Assim, com o perdão do leitor e do poeta que queria cultivar "inculta e bela" a última flor do Lácio, algumas palavras e/ou expressões permanecem como as encontrei na literatura de língua inglesa.

Outra questão nascida com o desenvolvimento da dissertação deu-se pela desconfiança/insegurança de que o que eu poderia dizer — e o modo como poderia dizê-lo — seria demasiado complicado para quem dela se aproximasse buscando uma dissertação sobre recuperação da informação, *domínio* da Ciência da Informação; e por demais próximo da Ciência da Informação (*lato sensu*) para aqueles dispostos a encontrar textos mais complexos sobre este *domínio* da Informática. Espero ter sido bem sucedido na busca do necessário equilíbrio.

No geral o corpo do trabalho será dividido em cinco capítulos: no **primeiro**, de caráter introdutório, realizar-se-á uma discussão sobre o conceito de tecnologias da inteligência (fala, escrita e informática), buscando identificar seu grau de influência na configuração da sociedade e na determinação do pensamento humano. Tal como em Levy (1993), ou a partir dele, as técnicas são encaradas não como meros instrumentos do pensamento, mas sim seus sujeitos. O **segundo capítulo** complementa o primeiro quando leva a discussão para a questão/conceituação de virtualidade/realidade. Discutem-se aqui os elementos qualificadores da biblioteca virtual contrapondo-os com o modelo convencional. Apresenta-se no **terceiro capítulo**, em caráter exploratório, a atual configuração do subsistema de recuperação da informação nas bibliotecas

convencionais. A apresentação do estado-da-arte das bibliotecas virtuais e dos objetivos gerais e específicos desta dissertação finalizam este capítulo.

A metodologia utilizada para a consecução do trabalho compõe o **quarto capítulo** da dissertação.

No **quinto capítulo** são propostas e discutidas as categorias de análise da recuperação da informação na biblioteca virtual, quais sejam:

- os conceitos de revocação e precisão e de navegação;
- a lógica booleana e a lógica nebulosa como modelos de recuperação da informação;
- as interfaces e os ambientes de recuperação da informação e
- a recuperação da informação hipermídia.

Finalmente, mas sem a pretensão de firmar o ponto final, são apresentadas as considerações finais.

O PROBLEMA

As tarefas básicas de organizar, administrar e disseminar o conhecimento humano, fazem eco — como aliás não poderia deixar de ser — à forma como a sociedade irá legitimar e justificar o sistema de informação. Em linhas gerais, o quadro evolutivo dos sistemas de informação é pontuado por três fases distintas:

- ênfase no acervo — onde o mais importante é a coleção, considerada, aliás, como um símbolo de status; a biblioteca é encarada como depositária do conhecimento humano;
- ênfase no atendimento - as necessidades do usuário determinam todos os serviços da biblioteca; a biblioteca é interpretada principalmente pela sua utilidade prática e pela sua competência em prestar bom atendimento, passa a ser denominada serviço ou sistema de informação. Não haveria nada de errado nisto não fosse o fato do bibliotecário idealizar o usuário e

- ênfase na informação — a noção de que o importante nas unidades de informação — a nova denominação de que a literatura sugere a partir de então — é a informação contida num documento; a constatação de que o usuário não busca algo físico, não busca suporte informacional, quando quer resolver um dado problema, quando quer reduzir sua incerteza. Quando se dirige à biblioteca quer, antes, algo intangível, incomensurável, quer a informação mesma. Não importa mais saber o que está sendo transmitido, não é a mensagem em si — e nem o meio — que importa, mas sim a intensidade das mudanças que ela — a informação, a mensagem — promoverá.

O ambiente eletrônico, o novo paradigma imposto pela tecnologia informática, que inaugura uma nova fase no tratamento informacional, cujo centro é a questão do acesso à informação/documento e não mais sua posse, derruba velhas teorias biblioteconômicas como as referentes à formação e desenvolvimento de coleções, por exemplo, e coloca em cheque várias fórmulas de controle informacional.

A partir da maior abrangência alcançada pela Internet, com a consolidação da WWW (World Wide Web) em 1992 e do *browser* Mosaic em 1993, alguns conceitos basilares da Biblioteconomia, referentes ao controle absoluto sobre a informação, são colocados em questão. Em termos práticos isto se reflete na quase ausência de profissionais bibliotecários atuando na administração ou manutenção de grandes bases de dados antes do ambiente de Rede, ou ocupando postos de *webmaster* nos diversos *sites* que aglomeram fontes de informação das mais diversas naturezas, de imagens a arquivos sonoros, de páginas pessoais a bases de dados de periódicos eletrônicos, ficando este mercado reservado em sua maior parte aos profissionais de informática.

O determinismo que norteou a profissão exemplificado de modo combativo por Mostafa (1981) pela fórmula “se A então B”, perdeu seu chão. Senão, veja-se. Se existe uma biblioteca o bibliotecário pode e deve exercer suas funções nela. A metáfora mais representativa hoje do conceito de biblioteca, responsável pelos maiores desenvolvimentos no campo da ciência hoje, contudo, não possui bibliotecários e ao mesmo tempo possui vários. Paradoxo? Contra-senso? Não, esta é simplesmente a opinião emitida recentemente pela revista Informática Exame (1996). A poderosa

biblioteca referida é a Internet; os bibliotecários ausentes são os mesmos que se associam, defendem seus direitos sobre a prioridade no tratamento da informação, freqüentam congressos e escolas de Biblioteconomia etc.. Os *bibliotecários* atuantes são os serviços de busca disponíveis hoje na Rede. O texto explora ironicamente, mas não de forma intencionalmente inconseqüente, a situação, identificando quatro tipos de *bibliotecários*, cada qual preparado para responder determinado tipo de necessidade informacional:

- os diretórios ou catálogos : onde as informações são organizadas hierarquicamente, do tópico mais geral ao mais específico, como o Yahoo! ou o brasileiro Cadê;
- os sistemas de busca (*search engines*) : ferramentas baseadas em grandes bases de dados alimentados por *softwares* especialmente desenvolvidos para rastrear a Rede, como o Alta Vista, o Excite, o Hotbot, o Infoseek ou o Lycos;
- as ferramentas de pesquisa como o Dejanews, que catalogam dados provenientes de grupos de discussão (*newsgroups*);
- os metaformulários, como o Metacrawler, os quais não dispendo de bases de dados próprias, funcionam como interface e realizam a consulta a diversos sistemas de busca. Estes seriam os *meta-bibliotecários*, ou os *bibliotecários* responsáveis por compilar o trabalho de busca de vários outros *bibliotecários* e oferecê-los num único pacote.

Claro que a metáfora é discutível. Causou, por sua própria ingenuidade ou inconseqüência não intencionais, reações de diversas naturezas; não deixa de pré-configurar, contudo, o problema com o qual a Ciência da Informação deverá atracar-se pelas próximas décadas.

CAPÍTULO 1 - TECNOLOGIAS DA INTELIGÊNCIA

Sabe-se, desde Marx, que o trabalho é uma atividade essencialmente humana, que somente o homem possui os elementos condicionantes à sua realização: o uso da razão e a intencionalidade. O trabalho, então, pode ser definido, de forma bastante sintetizada, como uma ação intencional cujo fim seja transformar a realidade. Sabe-se também que os meios de trabalho (o conjunto de instrumentos e técnicas) de que dispõe o trabalhador são determinantes e determinados pela configuração social onde se inserem. Não é o produto que caracteriza uma época econômica, adverte Marx (1988), mas sim os meios com que se dá a sua realização; os meios de trabalho ganham, então, status de “indicadores das condições sociais nas quais se trabalha”.

Pode-se afirmar, então, que o aparato técnico colocado à disposição de uma sociedade possui relação direta com sua forma de pensar / agir. Esta linha de pensamento, ampliada e atualizada para discutir as questões referentes às técnicas eletrônicas, é adotada por Levy (1993). Para este autor, o conceito de técnica está de tal forma calcado no pensamento humano, constitui-o tão amplamente que adquire contornos culturais; assim, por esta mesma razão, as novas técnicas são comumente “denunciadas como bárbaras e contrárias à vida”.

Os processos informáticos, uma das maiores inovações tecnológicas de nosso tempo, não poderiam deixar de gerar também, como de resto todas as mudanças, sua resistência. Se tal resistência ainda hoje se manifesta, que se dirá de sua força em 1993, data da obra de Levy. Por isto um de seus principais objetivos é conscientizar a sociedade da irreversibilidade das mudanças ocasionadas no ambiente informacional (conceito vastíssimo!) a partir do momento em que o computador invade as casas e possibilita contatos antes impensáveis, com o desenvolvimento das redes de comunicação de dados e da Internet em particular. Disso decorre o desabafo de Levy (1993): “alguém que condena a informática não pensaria nunca em criticar a impressão e menos ainda a escrita. Isto porque a impressão e a escrita (que são técnicas!) o constituem em demasia para que ele pense em apontá-las como estrangeiras”.

Em verdade, não são novas as discussões em torno do princípio da exclusão entre as tecnologias da informação, entre aqueles que conseguem enxergar o caráter peremptório das novas aplicações tecnológicas e aqueles que se limitam a rejeitá-las. Novas tecnologias e seus conseqüentes resultados que, de praxe, não escapam da dicotomia útil/inútil, obrigam a discuti-las, a formular questões para as quais o próprio evoluir da humanidade trará as respostas. Assim se deram os questionamentos sobre a exclusão da pintura em favor da fotografia, do teatro em função do cinema, deste pela televisão etc..

O que o pensamento de Levy traz de contributivo é justamente a abordagem das tecnologias da inteligência como marcadas por interpenetrações durante o seu longo processo evolutivo, isto é, a oralidade, a escrita e a informática não são, em absoluto, excludentes. Em Franco (1997) isto aparece bem exemplificado; o autor considera que “o ‘antigo’ costume de escrever cartas sofreu um brusco encolhimento com o uso da telefonia à distância. O uso do correio eletrônico estabelece uma reversão neste processo, provocando um verdadeiro renascimento do hábito de escrever cartas e de exercitar a escrita”. Por vezes exige-se que a escrita seja mesmo ampliada, como ocorre quando se usam os chamados *emoticons*, determinados pictogramas que reforçam a frase escrita em nível semântico. A frase “Amanhã teremos aula”, por exemplo, transmite muito menos que “Amanhã teremos aula :-)”.

Não são negadas, portanto, a presença das tecnologias que não são novas, como a oralidade e a escrita, nos tempos modernos. Ninguém em juízo perfeito pode caracterizá-las como subprodutos enterrados pelo processo informático; ao contrário continuam a exigir discussões. O jornal, por exemplo, levanta-se como um exemplo típico. Este veículo de comunicação ocupa posições privilegiadas tanto na cultura oral, por meio dos radiojornais ou telejornais, quanto na cultura escrita, um número enorme de publicações diárias espalhadas pelo país, ou na cultura informática, meio onde, no Brasil, ainda despontam aquelas empresas jornalísticas já consolidadas no meio escrito, como é o caso da Folha de São Paulo e seu Universo *Online*¹ — na verdade algo mais próximo de uma biblioteca digital do que de um informativo diário — do Jornal do Brasil², d'O Globo³ ou do Estado de São Paulo⁴.

¹ <http://www.uol.com.br>

² <http://www.jb.com.br>

³ <http://www.oglobo.com.br>

⁴ <http://www.estado.com.br>

Com o fim de preservar a clareza necessária ao texto, julga-se necessário distinguir aqui técnica de tecnologia. A idéia de técnica é normalmente associada à sua forma mais evidente, à máquina. No caso das tecnologias da informação, aos computadores, equipamentos de telecomunicação, de impressão e aos demais equipamentos. Lalande (1951) distingue, com a eloquência que a simplicidade consegue dar aos enunciados, **técnica**, como sendo um conjunto de procedimentos bem definidos e transmissíveis, destinados a produzir um certo resultado julgado útil, de **tecnologia**, caracterizada como o conjunto de estudos dos procedimentos técnicos nos seus aspectos gerais e nos seus relacionamentos com o desenvolvimento da civilização.

A amplitude do conceito de tecnologia está, portanto, nos efeitos que o desenvolvimento tecnológico impinge à sociedade. Rodrigues (1994) apresenta o quadro atual como sendo o de uma “modernidade tecno-lógica”. Espera-se da técnica a solução para os problemas humanos hoje, assim como outros períodos foram mitológicos — quando o homem dependia da benevolência dos deuses — ou teo-lógicos — quando os problemas eram solucionados pela mediação da Igreja. Neste mesmo sentido, Levy (1993) entende que se vive um desses momentos onde a partir de uma nova configuração técnica, um novo estilo de humanidade é inventado.

Esta discussão tem invadido várias outras disciplinas, não sendo portanto domínio específico da Ciência da Informação *strictu sensu*, onde tem se tornado assunto preferencial em congressos, seminários e outros eventos. Os Cadernos de Problemas Brasileiros (1997) relatam a recente palestra promovida pelo Conselho de Economia, Sociologia e Política da Federação do Comércio do Estado de São Paulo realizada pelo jornalista Jayme Sirotsky, atual presidente da Fiej (Associação Mundial de Jornais). Falando a seus pares, Sirotsky aborda a questão das tecnologias da informação de forma apaixonada e ao mesmo tempo bastante cautelosa. Juntamente com uma perspectiva luminosa apresenta uma outra que denomina de escuridão: “Esse fantástico mundo das comunicações do limiar do novo milênio convive de forma quase inacreditável com um percentual de quase 50% da população terrestre que ainda não deu seu primeiro telefonema. (...) O que joga a favor? As tecnologias estão evoluindo de tal maneira que não será possível por muito tempo preservar a escuridão”. Ou seja, entende o quadro

como irreversível; o conjunto das tecnologias colocados à disposição da humanidade exige-lhe conhecimento e acomodação a esta nova ordem.

Em Levy (1993) o conceito de tecnologias da inteligência relaciona-se às “técnicas fundamentais de comunicação”, as quais permitem “classificar as culturas em algumas categorias gerais”. Assim, em face do uso das tecnologias da inteligência, as sociedades são classificadas em categorias de :

- **oralidade primária**, a qual “remete ao papel da palavra antes que a sociedade tenha adotado a escrita”;
- **oralidade secundária**, “relacionada a um estatuto da palavra que é complementar ao da escrita, tal como o conhecemos hoje”;
- **escrita**, um dos pilares, com certeza o mais importante, da herança cultural da humanidade e
- **informática**, tecnologia marcadamente moderna.

São recortes realizados, obviamente, para fins de clareza do quanto deseja expor, pois se a informática é marca característica da contemporaneidade, as outras tecnologias aparecem em maior ou menor grau de uso nas demais épocas, inclusive nesta.

A oralidade, por exemplo, desempenhou importante papel na construção do conhecimento humano. Quando Machado (1995) discorre sobre a cultura do manuscrito adverte que “não podemos nos esquecer de que até o séc. XV, toda literatura existia, antes de mais nada, para ser recitada em público e o manuscrito era apenas um instrumento acessório dessa vasta e influente ‘cultura oral’, que nos deu pensadores como Pitágoras, Sócrates e Demócrito e poetas como Homero e os trovadores medievais”. Certamente disto decorre a associação que Levy (1993) faz da inteligência com a memória dos indivíduos — sobretudo a auditiva —, nessa sociedade que denomina de oral primária.

Essa cultura da oralidade perpassa o tempo e alcança a contemporaneidade, seja nos círculos da vida cotidiana, pois “a representação e as maneiras de ser continuam a transmitir-se independentemente dos circuitos da escrita e dos meios eletrônicos”

(Levy, 1993), seja nos círculos acadêmicos, porquanto “alguns de nossos mais importantes pensadores foram essencialmente pensadores orais”, conforme Machado (1995), o qual cita como exemplos Ferdinand de Saussure e Jacques Lacan, intelectuais cujo legado considerado importantíssimo para a configuração do pensamento contemporâneo, têm sua acessibilidade condicionada às “compilações realizadas por seus alunos, com base em anotações de aulas”, por terem deixado “pouquíssimos escritos de próprio punho”.

Não há mais, a rigor, separação entre oralidade, escrita ou informática nos processos de transferência da informação. Megaprojetos de digitalização de textos impressos como o Gutenberg⁵, por exemplo, trazem à cena novamente a cultura oral da qual falou Machado (1995) ou Levy (1993). Um texto digitalizado pode converter-se em voz ou mesmo reverter-se num texto escrito novamente, sem que perca sua natureza digital. Levacov (1997) esclarece que muitas pessoas que se utilizam do Projeto Gutenberg são pessoas cegas que, depois de realizada a transferência do arquivo de interesse para sua máquina, "utilizam sintetizadores de voz para 'ouvir' o texto", além, é claro, daqueles que mesmo com a visão perfeita ainda fazem muitas restrições à legibilidade da tela e optam por esse meio.

A discussão criada em torno dessas tecnologias interessa de perto à Ciência da Informação. Às bibliotecas, consideradas aqui como representantes de todo e qualquer sistema de recuperação da informação, sempre coube o papel de organizar e transferir informações. Isso sem falar na função sociabilizadora que se lhes atribui, de informar o maior número de pessoas possível. O raio de alcance das novas formas de transferência da informação promovido pela tecnologia digital é, então, o primeiro elemento sedutor apresentado às bibliotecas tradicionais. Ao permitir a transposição do conhecimento para um suporte mais amigável e potencialmente acessível a um maior número de pessoas, a telemática constitui-se, neste quadro, como um dispositivo fundamental para a construção do homem, em última instância.

Existe atualmente uma multiplicidade de textos impressos e de outros materiais sendo digitalizados para integrar o ambiente eletrônico, o ciberespaço. Projetos como o citado Gutenberg (biblioteca digital composta de obras de referência e da literatura

⁵ <http://www.promo.net/pg/>

universal, consideradas de domínio público, como Shakespeare, em texto completo) iniciado em 1971; o Perseus⁶ (uma grande biblioteca digital de cultura grega e romana clássicas), criado em 1987 na Universidade de Harvard; o Documentary Images⁷, da Biblioteca Nacional da Austrália, que visa a digitalização de aquarelas, pinturas e fotografias; o canadense Music Library of the Future, da McGill University⁸, cujo objetivo é a criação de uma biblioteca virtual na área de música, com ênfase em música canadense e muitos outros, podem ser arrolados como exemplos⁹.

Essa cultura da transposição de suportes do conhecimento corrobora o que sempre esteve patente, que o homem, enquanto produtor de conhecimento, é figura superior às tecnologias, ainda que elas o constituam e possam determinar o seu pensamento. Ou seja, a despeito de pensar com os recursos que a oralidade permitia, Sócrates, por exemplo, estabeleceu conceitos e metodizou de tal maneira o conhecimento, que ainda hoje é respeitado por suas idéias e considerado um marco divisório dentro da história da filosofia. Assim como Shakespeare ou as tragédias gregas levantam-se como criações representativas e como pontos de referência da cultura em todos os tempos. Não há que se falar, portanto, em sobreposição de tecnologias neste caso.

A tecnologia escrita, principalmente após os avanços conquistados pela imprensa, marcou profundamente a construção do conhecimento, tornou-se mesmo sua memória, fato possibilitado pelo enquadramento da palavra, efêmera e fugaz por natureza. Reforçando a idéia de que as tecnologias da inteligência constituem o pensamento humano, Levy (1993) encara a escrita não como “um simples modo de expressão da filosofia, das ciências, da história ou do direito” mas como constitutivas dessas disciplinas, “já que estes domínios do conhecimento, tal como os conhecemos hoje, não preexistem a ela. Sem escrita, não há datas nem arquivos, não há listas de observações, tabelas de números, não há códigos legislativos, nem sistemas filosóficos e muito menos críticas destes sistemas”. Esse quadro causou mudanças tão marcantes na configuração social que o termo cultura precisou ser reconceitualizado, pelo menos fora dos círculos acadêmicos, e associado ao novo conceito informacional promovido pelo

⁶ <http://www.perseus.tufts.edu>

⁷ <http://www2.nla.gov.au/imagecole/>

⁸ <http://lecaine.music.mcgill.ca/MF-Project/Html/MLF-Home.html>

⁹ Para uma lista mais detalhada ver

CUNHA, Murilo Bastos da. Biblioteca digital: bibliografia internacional anotada. **Ciência da Informação**, Brasília, v. 26, n. 2, p. 195-213, mai./ago. 1997

livro impresso; cultura, nesta acepção, passa a significar cultura escrita e, por extensão, cultura livresca.

Apesar de conviver com o analfabetismo, ainda em larga escala, o que se entende por civilização não mais o aceita como um fato natural, justamente porque a escrita torna-se o fundamento da sociedade; por meio dela é que o homem se sociabiliza. Este estado de coisas engendrou a institucionalização desta tecnologia, o que se deu pela criação de escolas e redes de ensino pelo Estado moderno (Franco, 1997). Como é certo que ainda existe o analfabetismo é certo que as pessoas que não dominam o uso da tecnologia escrita vivem à margem da sociedade. Este mesmo processo, ainda incipiente, já invade o domínio da informática, os anúncios classificados de empregos e a invasão dos *cursinhos de informática* estão aí para confirmar esta realidade.

A escrita, enquanto técnica utilizada para representar o conhecimento, sofre um vertiginoso processo evolutivo a partir dos estudos de Bush (1945) sobre a não-linearidade da estrutura do pensamento humano em detrimento da estrutura linear do texto escrito. O autor acreditava que a artificialidade dos sistemas de indexação, fosse por meio de métodos alfabéticos ou numéricos, obrigava o usuário a percorrer linearmente a representação classe por classe; quando encontrava algo precisava emergir do sistema e reiniciar sua busca por um novo item. A isto contrapõe a velocidade de associação não linear de idéias da mente humana. Nasce com este estudo os primeiros conceitos da tecnologia hipertexto; termo cunhado mais tarde por Theodor Holm Nelson, um continuador da obra de Bush, idealizador do Projeto Xanadu, "um lugar mágico da memória literária onde nada seria esquecido".

A idéia básica do hipertexto, a de ligação remissiva entre partes de um mesmo documento ou para outras fontes que desenvolvam de forma ampliada o tema focado em um texto, não constitui — mesmo naquela época — absoluta novidade. As citações, notas de rodapé, referenciação de trabalhos e outros recursos auxiliares que possuem a mesma intenção final, têm sido largamente utilizados desde há muito (Vilan Filho, 1994); compõem o que Levy (1996) chama de "aparelhagem de leitura artificial". O que há de novidade é o fato de que não há mais o desconforto físico ou o gasto de tempo ocasionado pelo ato de deslocar-se em busca do documento referenciado no texto original. O hiperdocumento e seu conjunto de *links referenciais* (ligações entre partes

do mesmo documento) e *links semânticos* (ponteiros para outros hiperdocumentos) inaugura um novo paradigma nas técnicas biblioteconômicas de disseminação da informação. Conforme aponta Levy (1996), o leitor não percorre mais com seus passos a biblioteca "virando páginas, transportando pesados volumes", pois tem à sua frente um texto que dobra-se e desdobra-se a sua vontade. Isso leva à proposição de um novo modelo interacional onde não é mais o usuário que se encontra distante da biblioteca e sim a biblioteca, ausente da Rede, que se distancia do usuário, quando priva-lhe de sua coleção e de seus serviços.

Uma das vantagens que o sistema hipertexto possui sobre os sistemas tradicionais de recuperação da informação é permitir estratégias de busca realmente personalizadas, criando condições para que o processo de busca tenha como foco principal o contexto que o motivou. Contexto este que não é estável, mas algo que está em jogo, sendo sempre reconstruído (Levy, 1993), ou seja, a realização de uma busca não obedece a uma estrutura hierárquica, tal como as que são utilizadas hoje para representar tematicamente os documentos nas bibliotecas tradicionais, mas a uma rede onde as idéias se associam livremente a outras, como sonhou Bush (1945).

Isto representa um avanço que não pode ser negligenciado por quantos estejam envolvidos com as técnicas de representação e recuperação da informação. A possibilidade real do hipercatálogo causa uma revolução sem precedentes no modo de representar o conhecimento e, por conseguinte, na elaboração de estratégias de busca e, como se tentará demonstrar nos processos que envolvem a recuperação como um todo.

O processo hipertexto e seu produto, o hiperdocumento, viabilizaram o desenvolvimento da tecnologia hipermídia, elemento fundamental para o vertiginoso crescimento e a larga aceitação do ambiente WWW (World Wide Web). Uma explicação simplista para este alcance e esta aceitação é dada pela facilidade de se trabalhar com a linguagem HTML (Hypertext Markup Language), uma linguagem de programação voltada para a marcação de elementos de ligação (*links*) no texto que não exige grandes conhecimentos de informática ou de programação de computadores de quem a manipula. A linguagem HTML gerencia por si mesma o controle de *links* referenciais; para o controle dos *links* semânticos (dinâmicos) estabeleceu-se um

Protocolo de Transferência de Hipertexto (HTTP ou Hypertext Transfer Protocol), o qual dá garantias de uma mesma linguagem quando dois computadores se comunicam.

Em outras palavras, a tecnologia hipermídia criou um ambiente informático interacional que não se restringe aos informatas, dada sua relativa simplicidade de elaboração e de uso e sua interface amigável. Se os procedimentos de transferência da informação inaugurados pelo FTP (File Transfer Protocol) ou pelo Gopher já eram bem-vindos, a WWW, num mundo governado pela imagem, por permitir uma interface gráfica extremamente amigável (*user friendly*), representou o salto que faltava para que a Internet invadisse todas as áreas da atividade humana. Serviços básicos como os de saúde, bancos, correios, bibliotecas estão sendo transformados em suas essências, em suas especificidades. São redesenhados a partir de uma nova configuração técnica. Com o uso dessas tecnologias inventa-se um novo estilo de organizar e disseminar a informação, o que inclui, é claro, as bibliotecas.

Interessa aos propósitos deste trabalho identificar como universos tão diferentes (e tão próximos ao mesmo tempo) como o da biblioteca tradicional, cujo acervo é composto de átomos, e o da biblioteca digital, cujo acervo se compõe de bits, — para utilizar os conceitos com os quais Negroponte (1995) encara a dualidade desta realidade — estão se relacionando.

CAPÍTULO II - O VIRTUAL

Certamente o conceito de virtualidade que habita as mentes das pessoas alfabetizadas em assuntos informáticos hoje, ainda que não se possa dizer que é ponto pacífico, está muito longe do apresentado por Aurélio Buarque de Holanda Ferreira em seu famoso dicionário: "o que existe como faculdade, porém sem exercício ou efeito atual"; "suscetível de realizar-se; potencial", ou, numa abordagem mais filosófica: "o que está predeterminado e contém todas as condições essenciais à sua realização".

Não é preciso uma grande ginástica mental para entender-se que para uma obra de onze anos, ainda que fruto de um intelectual, era totalmente descabida e ainda freqüentava as obras de ficção a idéia de ciberespaço, ou seja, de um espaço onde a barreira entre o virtual e o real começam a dissolver-se; onde o virtual e o real interpenetram-se constantemente e fixam desta forma uma nova ordem.

A nova conceituação de virtualidade começa a invadir o espaço cotidiano e, do que apresenta o senso comum, somente a noção de que o virtual é algo que não está presente desponta como verdadeira. Existe, na verdade, uma oposição fácil e enganosa — nas palavras de Levy (1996) — entre o real e o virtual. Fácil porque dicotomizante e simplista como a maioria das proposições maniqueístas que regem o senso comum; enganosa porque tal oposição não existe de fato. O virtual modifica o real tanto quanto este o modifica.

Com relação às bibliotecas a *oposição fácil e enganosa* surge espontaneamente da questão: como pode a biblioteca virtual não estar presente e ser atuante? A resposta apresenta-se de saída enviesada pela noção que se dá à presença e sua estreita relação com a existência, pois, "[o] fato de não pertencer a nenhum lugar, de freqüentar um espaço não designável (onde ocorre a conversação telefônica?), de ocorrer apenas entre coisas claramente situadas, ou de não estar somente *presente* (como todo ser pensante), nada disso impede a existência" (Levy, 1996).

O conceito de virtualidade com o qual a biblioteca deve dialogar prescinde, pois, dos ambientes de realidade virtual, justamente porque não são os objetos físicos, os livros, os periódicos, que se tornam virtuais; eles são apenas digitalizados; virtualizam-se seus efeitos práticos e as ações dos atores envolvidos neste processo. Inúmeros outros dispositivos já realizaram, conforme Levy (1996) este intento, o computador, tal qual o automóvel, o trem, o avião ou o telefone, por exemplo, aproxima fisicamente cidades ou

regiões separadas. Isto é a virtualização da sociedade; para aqueles que não se utilizam de nenhum desses progressos tecnológicos a separação e a distância continuam válidas. Para além da virtualização da sociedade estes dispositivos virtualizam ainda os sentidos "o telefone para a audição, a televisão para a visão, os sistemas de telemanipulações para o tato e a interação sensório-motora".

Assim como a escrita virtualiza a memória, a hipermídia dá continuidade a um processo já antigo de artificialização da leitura, expresso pela divisão do texto em capítulos, parágrafos, pontuação etc. Com efeito,

"se ler consiste em selecionar, em esquematizar, em construir uma rede de remissões internas ao texto, em associar a outros dados, em integrar as palavras e as imagens a uma memória pessoal em reconstrução permanente, *então os dispositivos hipertextuais constituem de fato uma espécie de objetivação, de exteriorização, de virtualização dos processos de leitura*" (Levy, 1996)

Para além da tecnologia envolvida na criação de um hiperdocumento e na sua forma de apresentação, o hipertexto é uma nova ferramenta que se integra à atividade humana de leitura.

Discutir a virtualização da memória e da leitura implica necessariamente em revisitar a noção de coleção em bibliotecas, enquanto instituições representativas da memória coletiva da humanidade. O desafio que se apresenta à biblioteca virtual é que ela não poderá se restringir a uma coleção digitalizada de documentos; a virtualização inaugura de fato novos processos de produção, de interpretação e de recuperação da informação. "Considerar o computador apenas como um instrumento a mais para produzir textos, sons ou imagens sobre suporte fixo (papel, película, fita magnética) equivale a negar sua fecundidade propriamente cultural, ou seja, o aparecimento de novos gêneros ligados à interatividade" (Levy, 1996). Costa (1996) discute a noção de memória institucional como um composto híbrido "que comporta o **mesmo** e o **outro**". A memória está presente nos dois: no primeiro como repetição (identidade) e representação; no segundo como interrupção e produção do novo, do inteiramente diverso.

As noções de armazenamento, das funções mais tradicionais das bibliotecas, prevalecem, porém, são radicalmente transformadas. Enquanto que uma coleção física de documentos implica em um número considerável de cópias distribuídas estratégica e

geograficamente, uma coleção virtual (hipertextualizada) exige, no limite, apenas uma cópia armazenada do mesmo documento em algum computador conectado à rede. "A partir das home pages e dos hiperdocumentos *on-line*, pode-se seguir os fios de diversos universos subjetivos" (Levy, 1996). Isto faz com que alguns teóricos da Ciência da Informação, como Marchiori (1997), conclamem as bibliotecas, mesmo as virtuais, a assumir seu tradicional papel, pois, nas palavras da autora, "as políticas da Internet não parecem estar comprometidas com a preservação da informação e talvez as bibliotecas sejam instadas a atuar como armazenadoras de arquivos eletrônicos".

Seguindo-se este raciocínio, não há que se opor acervo e acesso, não do modo como vem sendo amplamente representado na fórmula *acervo x acesso*, antes é necessário dialetizar este postulado e identificar os graus de pertinência de cada um dos elementos em situações específicas ou particulares de cada biblioteca. Há situações, sabe-se, onde o acesso é consideravelmente mais caro que o acervo.

Se existe oposição, esta se dá entre acesso e uso efetivo, pois se a aquisição de uma coleção de documentos em papel é comparativa e economicamente mais acessível do que sua manutenção, o mesmo não ocorre com as bases de dados *on-line*, pelas quais paga-se várias vezes pelo mesmo resultado prático; ainda que esta base de dados esteja em CD-ROM pagar-se-á constantemente por suas atualizações. Em áreas de crescimento (ou de mudanças) muito acelerado, como a Informática ou a Engenharia Genética, por exemplo, o ônus dessa atualização pode tornar-se exorbitante, principalmente se levado em conta o fato de serem as bases, em sua maioria, de dados referenciais.

Se a questão da virtualização aparece hoje fortemente vinculada aos processos informacionais ou informáticos, de forma generalizada, outrora esteve igualmente associada ao fascínio despertado pela realidade virtual (não se pretende afirmar aqui que isto caiu em desuso) devido principalmente ao largo apelo da ficção. Ainda é comum as pessoas esperarem de uma biblioteca virtual um espaço interativo onde, com o auxílio de equipamentos especiais de simulação, possam passear pelas estantes e folhear, literalmente, as obras ali dispostas.

Levy (1996) retira a discussão do ambiente informático e alça-a ao da antropologia cultural, considerando que o ser humano se fez alicerçado em três tipos de

virtualização: a linguagem (virtualização do tempo real, inaugurando o passado e o futuro), a técnica (virtualizando a ação) e o contrato (virtualizando a violência).

Para preservar os interesses do presente trabalho retirar-se-á de sua proposição somente o que se refere à técnica. Uma ferramenta, afirma o autor, mais do que uma extensão do corpo é uma virtualização da ação. O martelo, exemplifica, "pode dar a ilusão de um prolongamento do braço; a roda, em troca, evidentemente não é um prolongamento da perna, mas sim a virtualização do andar".

De modo análogo, respeitando-se sempre o apelo prolixo dos direitos autorais, a informação gradativamente virtualiza-se a partir do advento dos sistemas distribuídos de informação eletrônica. Frente a uma base de dados, a presença ou a ausência de um documento (ou fonte de informação) era um fato coercitivo; a capacidade de uma biblioteca virtual de elevar o percentual de revocação a números impensáveis ao mesmo tempo em que lhe confere extremo dinamismo, torna o acesso ao documento uma eventualidade aberta. Em outras palavras, o usuário pode não ter **o documento**, mas dispõe de formas alternativas para adquirir a mesma informação¹⁰. Se num sistema tradicional de informação alta revocação é um fato encarado aprioristicamente, quiçá de forma preconceituosa, como negativo, num sistema eletrônico interativo, com *softwares* cada vez mais *inteligentes* e especializados em trabalhar com o *feedback* de relevância, há que se repensar a problemática.

Ainda é Levy (1996) quem dá a advertência final: "A memória coletiva posta em ato no ciberespaço (dinâmica, emergente, cooperativa, retrabalhada em tempo real por interpretações) deve ser claramente distinguida da transmissão tradicional das narrativas e das competências, bem como dos registros estáticos das bibliotecas".

CAPÍTULO III - SISTEMA DE RECUPERAÇÃO DA INFORMAÇÃO

¹⁰ A medida de relevância, fator da maior importância para a avaliação da eficácia de um sistema de informação, foi propositalmente omitida aqui, pois será abordada posteriormente

Nunca se discutiu tanto a recuperação da informação, ou pelo menos nunca interessou tanto discuti-la. Amparado pelo crescimento da Internet e pelo crescente número de computadores domésticos, funcionando ou não em rede, questões, antes restritas aos estudiosos da área de informação, como as que envolvem a facilidade de acesso e recuperação rápida da informação, vêm se tornando cada vez mais comuns, invadindo mesmo o espaço cotidiano.

A designação SRI surge a partir da maior influência e uso dos processos mecanizados de tratamento e recuperação da informação e sua estreita relação com as necessidades do indivíduo, do usuário visto de modo particular. Traça a partir disto um quadro evolutivo de rápida ascensão, brilhantemente captado por Lesk (1997). Este autor atribui sete idades à recuperação da informação e as compara com as sete idades do homem, dadas por Shakespeare¹¹. Dada sua acuidade e importância histórica apresenta-se aqui uma síntese de suas idéias.

Partindo das premissas estabelecidas por Bush (1945), colocando aí **infância** da recuperação da informação (1945-1955), destaca a necessidade identificada por este último de se criar interfaces individualizadas, com caminhos particularizados para o espaço informacional. Surgem nos anos 50 também os índices de concordância (como o KWIC), tecnologia ainda utilizada em muitos sistemas comerciais de recuperação da informação atualmente.

A década de 60 é apontada como a **fase escolar** da recuperação da informação, característica dada pelas grandes experiências desenvolvidas neste período, como o estabelecimento dos conceitos de revocação e precisão, pelo Projeto Cranfield, do Cranfield College of Aeronautics. Em parceria com os pesquisadores da inteligência artificial surgem também os estudos em torno de questões/respostas em linguagem natural.

A **fase adulta** da recuperação da informação (década de 70) possui ligação íntima com o desenvolvimento das linguagens de programação e dos processadores de texto. Surgem os sistemas de grande aplicação; além do Dialog e do Orbit entre outros, o OCLC (*On line Computer Library Center*), que inaugurou, aliás, a noção de trabalho

¹¹ O referido texto está na comédia *As you like it*, ato 2, cena 7; disponível on-line em <http://the-tech.mit.edu/Shakespeare/Comedy/asyoulikeit/asyoulikeit.2.7.html>

cooperativo na catalogação, utilizava um sistema de busca relativamente simples e limitado. A estratégia de busca constava das quatro primeiras letras do nome do autor e/ou as quatro primeiras do título do documento. Embora a maioria destes sistemas operassem em termos de indexação e resumos, a década assistiu o surgimento dos sistemas de recuperação em textos completos.

O constante crescimento da tecnologia em processamento de texto e o equivalente decréscimo no custo dos discos (rígidos ou flexíveis) e equipamentos de telecomunicações em geral, marcam a **maturidade**, nos anos 80, da recuperação da informação. Isto significou mais informação disponível em formato legível por máquina, tornando comum a informação *on-line*. Dessa década são também os OPACs (*On line Public Access Catalogs*) e o texto completo *on-line*. Como consequência natural, houve um ressurgimento das pesquisas em torno das técnicas de recuperação. Os sistemas comerciais, contudo, absorveram muito pouco dessas técnicas. Uma mudança tecnológica chave foi o amplo uso do CD-ROM; dispor de 650 Mb (um absurdo para a época), com baixo custo de produção, representou a distribuição de grandes bases de dados, inclusive em texto completo.

Quando tudo parecia bem surge a crise da **meia-idade**. Os anos 90 produziram outra revolução tecnológica: a Internet. Essa rede de vastíssimo alcance (um milhão de assinantes por mês) tem como característica o fato de que não se trata somente de um grande número de pessoas acessando informação, mas também as provendo. A intermediação dos editores (uns poucos enviando informação para muitos) e dos sistemas de informação (organizando o conhecimento) se contrapõe com uma comunidade usuária que inaugura sua própria ordem, gera sua própria informação e também as classifica, estabelecendo *links* com todos os tipos de fontes pelas quais se interessem. O grande desafio desta década repousa no tratamento das imagens. Tecnologias como o *scanner* e o *browser* fizeram com que várias bibliotecas comessem a se preocupar com a transição de seu acervo para um formato digital, sob o risco de, não o fazendo, torná-los obsoletos.

O autor elenca ainda as duas últimas idades da recuperação da informação. Em sua **plenitude** (ano 2000) acredita que a maioria das informações disponíveis hoje em texto já estarão digitalizadas. Encerra seu artigo discutindo a **senilidade**. Nestes dois

últimos enfoques o autor faz mais prospecções do que considerações, assim não se enquadra nos objetivos deste trabalho discuti-los.

SUBSISTEMA DE RECUPERAÇÃO DA INFORMAÇÃO : DESVIO HISTÓRICO

O termo recuperação da informação foi criado em 1951 por Calvin Mooers, como consequência natural de uma mudança de paradigma já estabelecida com relação à convencional biblioteca, que não podia mais satisfazer a “necessidade de armazenar e dar acesso rápido e preciso ao grande número de documentos que vinham tendo crescimento exponencial desde o século XVII e pelo advento do computador, que era visto como a ‘grande solução’ para os problemas de armazenamento e recuperação da informação” (Araújo, 1994).

A consolidação plena do conceito, contudo, só aparece na década de 70, quando muitos autores desenvolveram a teoria dos SRIs. Os *Principles of Information Retrieval* de Kochen data de 1974; Rijsbergen escreve *Information Retrieval* em 1979; Morse publica *Browsing and Search Theory* em 1973 e Lancaster apresenta o seu famoso *Measurement and evaluation of library services* em 1977.

Essa pequena amostra evidencia que a teoria dos SRIs tem em média trinta anos; termos como armazenamento e recuperação, ou revocação e precisão foram pensados dentro de um paradigma já automatizado, mas longe ainda do ambiente de rede dos anos 90 e da revolução informacional promovida pelo incremento no uso de bases de dados cada vez mais especializadas.

Se bem analisado, o SRI foi a designação do próprio sistema de informação, evidenciando a tomada da parte pelo todo; a recuperação constitui umas das operações desempenhadas pelo sistema de informação, ainda que das mais relevantes e em estreita relação com todos os demais subsistemas: recupera-se informações que a rigor deram entrada no sistema. Subsistemas de entrada, saída e avaliação são, óbvio está, partes de um mesmo processo, o qual visa a garantir, no limite, a disseminação da informação. Barreto (1996) considera esta apropriação — da parte pelo todo — uma aberração teórica, pois, "o usuário deslocar-se-ia da ambiência do sistema para ser colocado dentro dele, como um recurso ou elemento **do** sistema".

Aliás, não só *sistema* ou *subsistema* são termos que não se acomodam ao conceito, como também *informação*. Ainda segundo Barreto (1994) "as estruturas significantes armazenadas em bases de dados, bibliotecas, arquivos ou museus possuem a competência para produzir conhecimento, mas que só se efetiva a partir de uma ação de comunicação mutuamente consentida entre a fonte (os estoques) e o receptor". Discutível, portanto, a aceção de que são *informações* os elementos componentes dos estoques, ou dos sistemas de *informação*.

O princípio que governa estes sistemas automatizados baseia-se na melhor coincidência de palavras (*matching*); a realização de uma busca numa base de dados tradicional tem como meta geralmente: a) identificação ou localização de um item conhecido; ou b) condução de uma busca por assunto (Cronin & Hert, 1995). Em outras palavras, entre a necessidade do usuário e o conteúdo do sistema existe uma condição *sine qua non*, qual seja a de que o usuário esteja devidamente condicionado a extrair informações de tais pacotes informacionais pré-formatados. O sucesso de uma busca em termos de sua operacionalidade baseia-se, assim, em expectativas mais ou menos mútuas.

Disto decorre uma extensa literatura em torno dos conceitos de revocação e precisão, encarados como desafio central para a recuperação da informação. Cronin & Hert (1995) atribuem as metáforas armazenamento e recuperação à uma outra que expressa o SRI como um todo, a metáfora do contêiner.

A BIBLIOTECA VIRTUAL

Inúmeras outras metáforas expressam o conceito de biblioteca virtual: biblioteca do futuro, biblioteca sem paredes, biblioteca biônica, são alguns exemplos. Existe, de fato, uma confusão terminológica em torno do conceito. Enquanto no Brasil o termo biblioteca virtual vem se popularizando, a literatura americana elegeu o termo biblioteca digital como preferencial.

Barker (1994) identifica quatro tipos de bibliotecas:

Biblioteca Polimídia : como o próprio nome indica, são bibliotecas onde são armazenadas variadas *mídias*, como livros, CD-ROMs, fitas de vídeo etc.. Este tipo de biblioteca caracteriza-se por não desenvolver processamento central automatizado, sendo realizado de forma manual.

Biblioteca Eletrônica : possui processamento básico realizado eletronicamente. Paralelo ao conceito brasileiro de biblioteca automatizada, este tipo de biblioteca utiliza-se amplamente de computadores para disponibilizar índice *on-line*, construir processos de recuperação e armazenagem da informação.

Biblioteca Digital : tem como característica uma coleção de documentos eminentemente digitais; independentemente se foram criados na forma digital ou digitalizados a partir de documentos impressos, permite, por meio do uso de redes de computadores, compartilhar a informação instantânea e facilmente.

Biblioteca Virtual : conceituada como uma biblioteca cuja existência depende da tecnologia da realidade virtual. Com o uso de técnicas sofisticadas reproduz-se o ambiente físico de uma biblioteca convencional em um ambiente interacional, seu acesso depende, então, do uso de equipamentos adequados para simular efeitos tridimensionais.

Levy & Marshall (1994) ironizam a problemática. Para eles a pergunta sobre o que é uma biblioteca digital equivale a perguntar pela cor do cavalo branco de George Washington. Uma biblioteca digital, definem de forma direta, é uma coleção de materiais digitais, problematizando aí o que entendem por materiais. Sem reservar muita importância aos diversos tipos de biblioteca, interessa-lhes apenas o que todas têm como elemento comum: são repositórios para coleções de artefatos pelos quais o homem fixa ou estabelece uma intenção de comunicação, em outros termos, uma coleção de documentos.

Conforme apontado anteriormente, este autor não concorda com a conceituação de biblioteca virtual apresentada acima. Apesar de suas escorregadias definições, porquanto ainda se configura como um objeto em construção, a biblioteca virtual não passa de um conjunto de *links* organizados que remetem às várias fontes informacionais na Internet, sendo esta também a compreensão do Electronic Information Group (1995).

Em oposição ao *matching* caracterizado anteriormente como princípio fundamental nos subsistemas de recuperação da informação, a biblioteca virtual, organizada por categorias de conhecimento, permite livres e infinitas associações. A partir da construção de buscas e/ou navegação extremamente particulares, cresce a possibilidade de o usuário ser conduzido à informação precisa. É necessário que se considere que o jogo agora é controlado pelo usuário, tem como pano de fundo o seu contexto informacional, isto é, não há mais classificação subjetivada do conhecimento, a não ser quando realizada pelo usuário. Desta mesma opinião compartilha Barreto (1996):

"[o] usuário do SRI deixa sua posição de espectador do processo de recuperação da informação e interage diretamente com os estoques de informação e com documentos (hipertexto) usando linguagens multimídia. Questiona ou, não mais aceita, a **super-organização** dos SRI, que procuram tirá-lo do lugar conquistado, de participante virtual do sistema, para transformá-lo de novo em espectador do processo de transferência da informação, com o distanciamento brechtiano que caracterizou a interação do feedback".

A informação que a biblioteca virtual armazena está dispersa pela ampla teia mundial e a biblioteca virtual não a classifica da mesma forma com que tradicionalmente as bibliotecas o fazem; apenas a categoriza e a reempacota para que o usuário se oriente.

Enquanto a biblioteca tradicional oferece a literatura branca, seja em livros ou periódicos, a biblioteca virtual é limitada nesta função, pois seu ponto forte é a literatura cinzenta, principalmente por questões de ordem jurídica. Não obstante, existe aqui uma contradição, pois, a oferta de determinado(s) livro(s) ou periódico(s) pela biblioteca tradicional não deixa de ser um fator limitante na recuperação em vista da limitação da própria coleção.

Como se vê, a base tecnológica é um fenômeno que não pode ser pensado separadamente das relações sociais de propriedade. Essa é uma das razões pelas quais vive-se ainda uma fase histórica, onde essas duas bibliotecas, a tradicional e a virtual, ainda encontram-se em plena vigência. As bibliotecas virtuais são necessárias tanto para permitir que as bibliotecas tradicionais ofereçam novos serviços quanto para lidar com a crise econômica que afeta seus serviços atuais (Lesk, 1995). Esta situação pode mesmo ser definitiva, não se sabe. A combinação de facilidade de acesso e grande quantidade

de informações disponível tornou a Web a fonte preferencial de informação para um grande número de pessoas. Entretanto, adverte Lagoze (1997), qualquer sugestão de que as bibliotecas virtuais, ou os sistemas de busca na Rede, ocuparão o lugar das bibliotecas tradicionais é tolice, pois ignora o fato óbvio de que a biblioteca não é somente um repositório de informação, mas uma instituição prestadora de serviços incluindo seleção, preservação, classificação, localização e referência.

Entre a biblioteca tradicional e a biblioteca virtual existem aspectos comuns, como a agregação de valor, pois, qualquer índice — mesmo o índice do final do livro ou o que acompanha o último número do periódico no final de um ano — representa um trabalho de síntese, de relacionamento, de indexação, enfim, pois indexar é, no limite, relacionar. Por outro lado, as bibliotecas tradicionais e as bibliotecas virtuais representam realidades diferentes. Se ambas podem ser definidas como coleções organizadas de documentos, a própria noção de documento, em vista de sua virtualidade possível, precisa ser refeita.

Organizar uma coleção de documentos atualmente, criar uma biblioteca, portanto, tornou-se uma atividade corriqueira, facilmente encontrada nos diversos *sites* institucionais dispersos pela Rede. Quer trate-se de entidades empresariais, esportivas, acadêmicas ou governamentais, qualquer *site* acompanhado de algum cuidado em sua elaboração, provê o usuário de informações correlacionadas por meio de sua política de estabelecimento de *links*. Como exemplo cita-se Central Única dos Trabalhadores (CUT)¹² com uma lista de *links* para entidades cutistas na Internet ou a Câmara Internacional de Comércio do Brasil (Camint)¹³, a qual disponibiliza uma coleção de *links* relacionados por meio de um diretório de assuntos dispostos em categorias como economia e negócios, política mundial e estatísticas, entre outros.

Esse conjunto de *links* não deixa de representar uma *biblioteca virtual*, na medida em que reúne *documentos* organizados segundo uma lógica de proximidade de interesses. Na biblioteca tradicional essa lógica de proximidades de interesse é assegurada pelo estabelecimento de políticas de remissivas ou de referências nos catálogos (um usuário que se interessa por ... poderia também interessar-se por ...), o

¹² <http://www.cut.org.br>

¹³ <http://www.camint.com.br>

que, de todo modo, não se diferencia muito da política de apontadores da biblioteca virtual.

Qualquer biblioteca, seja a tradicional, seja a virtual, implica em organização. A biblioteca virtual, contudo, está visceralmente ligada à rede de comunicação Internet; isto aliás, é o que diferencia a conceituação de uma biblioteca digital de uma biblioteca eletrônica, cujos serviços são também automatizados e distribuídos em redes *on-line* mas sem a perspectiva de encadeamento externo vigorando somente os *links* referenciais, a exemplo das bases de dados em CD-ROM anteriormente à sua disponibilização na Internet.

Assim, se a biblioteca em geral remete à idéia de organização é preciso reconhecer que um conjunto de *links* extrapola uma coleção local, ainda e principalmente que esta coleção seja a da biblioteca virtual. A biblioteca virtual possui também seus *links* referenciais (neste sentido ela é também digital), assim tanto remete aos textos especificamente selecionados para compô-la, como remete para documentos outros na rede, os quais não primam por qualquer organização. Se esses textos soltos na rede fossem por algum meio selecionados, capturados e disponibilizados na coleção da biblioteca *virtual*, como *links* referenciais (internos ao *site* da biblioteca virtual) e este fosse o seu limite, a biblioteca em questão seria apenas uma biblioteca digital.

Se se pergunta o que diferencia uma biblioteca virtual de um *sistema de busca* do ponto de vista do usuário comum, a resposta certamente passa pelo crivo da informação disponibilizada com valor agregado. Os sistemas de busca podem prover acesso, com maior ou menor sucesso, aos itens da rede, mas as bibliotecas virtuais vão um pouco além. Elas são centros de informações com valor agregado onde os *links* devem ser cuidadosamente selecionados, organizados, preservados e disseminados. O propósito ainda é o mesmo do modelo tradicional: se um recurso da rede é de boa qualidade, de interesse da comunidade usuária e relevante para a coleção, ele permanece no catálogo. A idéia é oferecer ao usuário um quadro mais acurado e completo da informação disponível em uma dada área de assunto.

Disso decorreu o desconforto — procedente como se vê — que as colocações da revista *Informática Exame* (1996) causaram no meio bibliotecário. Esta mesma confrontação foi realizada por Kelly & Robbins (1996) ao texto de Saunders (1992). Este autor define a biblioteca virtual como um sistema pelo qual o usuário pode

conectar-se com bibliotecas remotas e bases de dados usando o catálogo de acesso público *on-line* da biblioteca local ou qualquer outro computador em rede como ponte (*gateway*). Isto é, dá ênfase aos *links* por si mesmos. Kelly & Robbins (1996) acrescentam a essa definição o fato de que as bibliotecas virtuais contêm elementos de valor agregado e não simplesmente conexões para *sites* remotos, deixando claro que o elemento humano (agregando valor semântico à informação) é o fator que norteia a ação de qualquer biblioteca.

Na biblioteca virtual o balcão de referência virtualiza-se também. À medida em que as ferramentas comumente utilizadas no balcão de referência tornam-se disponíveis em formatos eletrônicos, o balcão de referência pode migrar para qualquer local onde os usuários tenham acesso ao bibliotecário e ele tenha acesso a um terminal (Kelly & Robbins, 1996). Uma das vantagens da referência por correio eletrônico, por exemplo, está na possibilidade de rapidamente endereçar-se a questão para a pessoa mais apropriada para respondê-la, não importando se é ou não membro da equipe de referência .

O tratamento das novas formas de documento encetadas pela virtualização ou digitalização propicia também o desenvolvimento de novos serviços que lhes são inerentes. Um exemplo é o InfoFile¹⁴, um sistema de localização de informações de referência em governo, política, eleições e legislação dos Estados Unidos que utiliza as FAQs (Frequently Asked Questions) como suporte, desenvolvido por Stalker & Murfin (1996) na Ohio State University. Baseado em menus hipertexto, o InfoFile provê respostas a setecentas questões de referência específicas (dados de 1996), permitindo ao usuário definir o nível de profundidade e especificidade de sua questão, o qual passa então a *negociá-la*.

Esse é um exemplo de inovação tecnológica nas bibliotecas virtuais: o balcão de referência sendo estruturado a partir de perguntas e respostas (compiladas) dos usuários. Não significa dizer que tal serviço não pudesse ser realizado fora do ambiente automatizado e virtual; com efeito, a biblioteca tradicional já o realizava na medida em que as fontes que compõem o setor de referência são as fontes conhecidamente mais

¹⁴ <http://sjcpl.lib.in.us/InfoFileIntro.html>

utilizadas. A inovação está em liberar o pessoal de referência da extenuante tarefa de compilar as chamadas questões de referência pendentes.

Como se pôde perceber ainda não é pacífica a determinação do grau de influência ou de interdependência entre a biblioteca tradicional e a biblioteca virtual, principalmente no que se refere às questões de recuperação da informação.

Num exemplo (impraticável) levado ao extremo, se fosse possível revalidar na biblioteca virtual todo o conjunto de procedimentos que compõem o campo interacional da biblioteca tradicional, as técnicas de recuperação praticadas nestas últimas certamente seriam um entrave, pois, a mudança do ambiente interacional entre sistema e usuário, o nascimento de novas formas documentais e o condicionamento da completa eficácia do sistema de informação ao desenvolvimento de técnicas que acompanhem *pari passu* o crescimento das tecnologias telemáticas, obrigam a caracterização de novos modelos de recuperação da informação.

OBJETIVOS

OBJETIVO GERAL

Diante do quadro exposto é objetivo deste trabalho comparar a tradicional recuperação da informação comumente praticada nas bibliotecas com a forma pela qual vem se apresentando nas bibliotecas virtuais.

OBJETIVOS ESPECÍFICOS

Para a consecução da pesquisa, os seguintes objetivos foram cumpridos:

- discussão sobre o conceito de navegação em oposição à perspectiva da coincidência de palavras (*matching*) presente na recuperação tradicional;

- caracterização dos modelos de recuperação da informação encontrados na biblioteca virtual por meio da abordagem comparativa entre a lógica booleana e a lógica nebulosa;
- identificação da influência que a base tecnológica exerce no desenvolvimento das técnicas de recuperação da informação, traçando-se a evolução das interfaces e ambientes de recuperação;
- discussão da necessidade de adoção de novos modelos para a recuperação de informação hipermídia.

CAPÍTULO IV - MÉTODO

A prática de ciência tradicional pede que um trabalho científico, um relatório de pesquisa, contemple dentre suas partes lógicas um capítulo destinado ao método, onde são relatados os passos para que possa ser eventualmente produzida uma réplica da pesquisa.

O estabelecimento de uma metodologia adequada foi objeto de apreensão após a definição do objetivo da pesquisa. Se de um lado tem-se a biblioteca tradicional, com uma vasta fonte de consultas disponível, onde se pode, inclusive, questionar a realidade por meio de investigações sobre essas bibliotecas; de outro, tem-se os sistemas virtuais de recuperação da informação (esta também virtual, em muitos casos), uma tecnologia da informação incipiente, que ainda exige discussões sobre os efeitos que causa na realidade aparente.

Optou-se pela pesquisa documental por meio da concretude de um roteiro de observações e de categorias de análise pré-definidas que pudessem fornecer subsídios para a construção teórica pretendida. Questionou-se, portanto, a realidade apresentada pelos autores que discutem as questões propostas para esta reflexão, em outros termos, questionou-se a literatura científica dando voz aos especialistas. A preocupação metodológica foi, então, intermediar este tipo de diálogo entre os autores com a realidade observada.

Ainda assim julga-se necessária uma ressalva. Se a literatura científica tradicional, em meio impresso, forneceu os elementos considerados imprescindíveis ao embasamento teórico que conduziu a pesquisa, as informações mais atualizadas referentes ao tratamento que se pretendeu dar ao tema foram encontradas em ambiente virtual, em fontes eletrônicas. Conforme já fora possível verificar em pesquisas de levantamento, a maioria das informações relevantes sobre a Internet só pode ser encontrada na própria Internet, que está constantemente *under construction*.

Há textos que nascem em meio impresso e são convertidos ao formato digital; há aqueles, contudo, que são gerados em formato digital sem a intenção de se assentar num suporte de papel ou em algum meio magnético ou ótico. O que essa literatura tem de

inovador é o fato de que somente os textos julgados pertinentes permanecem por algum tempo na Rede, o que confere um caráter de extremo dinamismo à pesquisa documental e exige-lhe, por outro lado, extrema acuidade na seleção das fontes, tendo em vista a profusão de textos e do caráter algumas vezes duvidoso da confiabilidade da fonte.

Evidente, portanto, que a pesquisa a ser realizada poderá ser refeita, mas jamais reproduzida ou replicada, pois captou um momento único, precisou mesmo desse recorte na história para que pudesse ser concretizada.

CAPÍTULO V - CATEGORIAS DE ANÁLISE

REVOCAÇÃO / PRECISÃO E O CONCEITO DE NAVEGAÇÃO

A tradicional recuperação da informação é comparada por Cronin & Hert (1995) ao modo como o lobo obtém alimento, ou seja perseguindo sua caça, enquanto que o novo modelo proposto pelas bibliotecas virtuais se aproxima mais da estratégia utilizada pela aranha, que arma sua teia e espera que sua presa venha até ela. Por isto esses autores traçam analogias entre o comportamento de busca por alimento animal e o comportamento de busca por informações pelo pesquisador. A metáfora cunhada por estes autores para os sistemas tradicionais é a metáfora do contêiner, calcada na coincidência de palavras (*matching*).

A biblioteca virtual vai lidar com o conceito de navegação, muito bem colocado por Camargo (1996) como “conceito ímpar no todo fragmentado”. Apesar da autora não referir-se aos aspectos da recuperação *stricto-sensu*, tal singularidade do conceito tem conseqüências fundamentais para a busca da informação. Não é por acaso que o já citado Electronic Information Group (1995) argumenta que uma característica distintiva da proposição de uma busca a uma biblioteca virtual é que nesta a ferramenta de busca (ou catálogo) só será necessária numa etapa inicial, para posicionar o usuário numa “vizinhança” adequada do ciberespaço. Uma vez lá a condução da busca tornar-se-á mais promissora, posto que criada pela imaginação humana e não por alguma ferramenta automática.

Dada a natureza hipertexto da Web, o primeiro conjunto de resultados pode não fornecer a resposta, mas pode estabelecer *links* com outros *sites* que o façam. Mesmo a mais estranha página pode fornecer um *link* para uma fonte interessante (Notess, 1997).

É perfeitamente plausível que num futuro muito próximo o elemento humano seja varrido dos processos intermediários da transferência de informação. Não é este, contudo, o quadro que se desenha. Atualmente, as bibliotecas entendem que juntamente com o desenvolvimento tecnológico é necessário uma metodologia de organizar e disponibilizar a informação que inove e inovar é um ato que requer amplo conhecimento das ferramentas. Por isto as organizações modernas (incluídas aqui as

bibliotecas), vêm tendo uma preocupação crescente com o desenvolvimento da qualidade, da criatividade, da reflexão e não mais com o controle e com forma de reprodução do comportamento das pessoas.

O compromisso assumido com a criatividade, em face dos novos desafios, revolve assuntos dados como assentados pela Biblioteconomia, como a problemática da catalogação. Tradicionalmente, pelo menos no caso da Biblioteconomia brasileira, existe uma separação entre catalogação descritiva (representação descritiva) e catalogação de assunto (representação temática). Se os currículos modernos dão ênfase à representação temática, o mesmo não ocorre com a representação descritiva, parte menos *nobre* da profissão, relegada a trabalhos auxiliares, alvo, inclusive, de comentários não muito estimulantes.

A catalogação, na verdade, é encarada a partir de seu produto, o catálogo, sendo um registro o seu ponto culminante. Ignora-se, dessa maneira, o grande processo intelectual não visível de desenvolvimento de vocabulário, de sistemas de descrição e de classificação presentes no registro como trabalho morto. Toma-se tal processo de forma a-histórica.

O surgimento das formas digitais encerram esta discussão infundada pois, reposicionam o catálogo como responsável pela construção da autoridade literária. Assim como outrora, quando ainda principiava a tecnologia de impressão a prática da catalogação obrigou os editores a definir autores, títulos e outros elementos para as obras publicadas; recentemente bases de dados e o desenvolvimento de estruturas de recuperação da informação calcadas nos títulos dos documentos, como o KWIC, obrigaram novamente os editores modernos a uma preocupação maior com o conteúdo dos títulos.

O desenvolvimento do ambiente sócio-tecnológico obriga, como se vê, a um constante reformular desta ordem. Neste século, o desenvolvimento de tecnologias como a do filme, do vídeo, do microfilme, da fotocópia, todos requisitaram mudanças e acomodações nas regras; na publicação e distribuição, em modelos de consumo e uso (Levy, 1995). O ambiente virtual coloca em cheque novamente estes dispositivos. Isto Levy chama de *digital order*. Já não se tomam mais decisões editoriais que envolvam a criação de textos sem a inclusão de discussões sobre os metadados, posto que se leva

em conta o crescimento do acesso a textos eletrônicos (Butterfield, 1997), realidade que vive a Universidade de Michigan e que pode ser expandida para o contexto editorial de um modo geral.

A catalogação, seja descritiva, seja temática, depara-se com antigos e novos problemas, dentre os questionamentos principais da *digital order* está também o mais antigo e o mais elementar: por que e para quem se descreve as fontes virtuais de informação e qual o nível apropriado para descrevê-las? E ainda, quanto de intervenção humana é necessário e quais os processos que serão realizados por agentes inteligentes?

Da mesma maneira que foi preciso institucionalizar a catalogação convencional para que fosse dado aos homens acesso ao conhecimento registrado em obras impressas (por qualquer meio de impressão), assim também hoje inúmeros organismos debruçam-se sobre a questão da representação das informações digitais, do conhecimento virtualizado.

Inicialmente conhecido por Ohio College Libraries Center por ter sido fundado por este grupo de bibliotecas, em 1971, o hoje famoso OCLC (Online Computer Library Center) ocupa a linha de frente na discussão de novas formas de representação, de estabelecimento de metadados, como são conhecidos estes elementos. O OCLC é o maior catálogo cooperativo do mundo, servindo 17.000 bibliotecas em 52 países, contém cerca de 30 milhões de registros bibliográficos (Borgman, 1997). Sua proposta básica, informa Butterfield (1997), não é substituir sistemas existentes ou ignorar a tradição secular em descrição de elementos bibliográficos, mas continuar a construir esse conhecimento por meio da criação de novos sistemas metainformacionais com o auxílio indispensável do conhecimento especializado desenvolvido durante o processo evolutivo da catalogação tradicional. Busca-se o padrão ideal de metadado: um conjunto de elementos simples o suficiente para autores e editores usarem na descrição de seus próprios documentos quando colocados na Rede, mas útil o suficiente para facilitar a localização e recuperação destes documentos por outros (Caplan, 1995).

OS METADADOS

Segue-se uma descrição das principais iniciativas em metadados:

Text Encoding Initiative (TEI) (Xu, 1997) : usa a SGML como base para codificar e intercambiar textos legíveis por máquina entre comunidades de pesquisadores. O cabeçalho TEI (TEI header), anexado a um texto em conformidade com o TEI ou não, consiste de :

- descrição do arquivo (descrição do objeto eletrônico e suas fontes);
- descrição "codificadora" que mostra como o texto foi codificado durante decisões editoriais feitas durante a marcação (*markup*) do documento;
- descrição de perfil a qual inclui informação bibliográfica não contextual e
- descrição de revisão naqueles documentos de cada mudança feita ao texto eletrônico.

Dublin Core (Xu, 1997; Caplan, 1995) : é um padrão metadado proposto recentemente para a descrição de fontes em rede e auxiliar na sua localização e recuperação. É um simples conjunto de treze elementos de dados cuja flexibilidade permite que esses elementos sejam modificados e expandidos. É tão intuitivo em sua arquitetura que os próprios provedores de informação podem codificar seus próprios documentos no momento da criação. Na verdade este é um dos maiores objetivos dos desenvolvedores.

Conjunto mínimo de elementos do Dublin Core:

- assunto : palavras ou frases indicativas do conteúdo informacional;
- título: título, nome ou pequena descrição do objeto;
- autor: o nome do criador do conteúdo, ordem dos nomes de acordo com a cultura;
- outro agente: o nome de qualquer outra pessoa, tais como tradutores, por exemplo, responsável também pelo conteúdo intelectual do objeto; o sub-elemento papel (*role*) descreve a responsabilidade;
- publicador (editor / publisher): o nome da entidade responsável por disponibilizar o objeto;

- data: data da publicação;
- identificador: uma *string* de caracteres ou um número usado para distinguir esse objeto de outros objetos;
- tipo de objeto : descrição de gênero do objeto, tais como *poema* ou *dicionário*;
- forma : características físicas, lógicas ou de código, como por exemplo um arquivo executável do Windows™;
- relação : relacionamento reconhecidamente importante com outros objetos; o sub-elemento "tipo" descreve a natureza do relacionamento e o sub-elemento "esquema" identifica a notação usada para identificar o objeto relatado;
- língua : língua natural do objeto;
- fonte : objeto, tanto impresso quanto eletrônico, do qual derivou e
- cobertura : caracteriza parâmetros para especificar o público-alvo, ou o tempo ou espaço (extensão).

URC (Uniform Resource Characteristic) (Xu, 1997) : diferentemente dos padrões acima o URC focaliza ostensivamente a garantia de recuperação de fontes eletrônicas. É composto de:

- URN (Uniform Resource Name), uma proposta de identificação única dos objetos de informação presentes na rede, semelhante ao ISBN;
- URL (Uniform Resource Locator), o endereço eletrônico de um recurso Internet;
- URC (Uniform Resource Characteristic), proposto para servir de conexão entre URNs e URLs. Se o URL muda, usuários autorizados podem acessar (abrir) o serviço URC para modificar o mesmo.

O URL pode conter metainformações tais como: autor, título, publicador (editor), assunto e assim por diante, o que pode ajudar na localização e recuperação de fontes. O serviço URC inclui a possibilidade de busca bibliográfica no site de busca do próprio serviço. O objetivo dos desenvolvedores é permitir que no futuro seja possível conectar-se a uma variedade de servidores URC, juntamente com alguns *sites* dedicados tais como OCLC e Library of Congress.

USMARC (Xu, 1997) : atualmente é um complexo conjunto de padrões para a descrição, armazenamento, intercâmbio, manipulação e recuperação de dados bibliográficos legíveis por máquina. Originalmente desenvolvido nos anos 60 para descrição de livros impressos, o USMARC está ainda sendo adaptado para prover descrição, acesso e informação de localização para recursos em rede.

Com a introdução do campo 856, cuja finalidade é viabilizar um *hiperlink* entre o registro MARC e o texto eletrônico que ele descreve, o USMARC tornou-se um padrão possível para a descoberta e recuperação de recursos em rede.

Toda a questão da representação da informação localizada em fontes disponíveis na Rede, por mais completos que possam ser esses metadados, por mais informativo que o registro USMARC seja, por exemplo; os campos que endereçam a busca do usuário, como o 856, neste caso, já são por si contraditórios. O usuário de bibliotecas foi acostumado ao endereçamento *correto* (onde fosse possível). Em outras palavras, a biblioteca, atendendo a uma solicitação, fornece ao usuário uma cópia do documento, coloca em suas mãos um exemplar físico, identificado por meio de seus esquemas de representação e recuperação da informação. Falha na localização da informação é, portanto, falha da biblioteca. Com as bibliotecas virtuais a coisa não se dá assim. Falhas nos *links* relacionados, exceto as relacionadas a *links* inativos, não podem mais ser responsabilidade da biblioteca. Dessa forma, se o usuário encontra alguma informação útil em um *site* hoje e pretende complementar sua pesquisa amanhã, pode ser surpreendido com a já popularizada volatilidade da informação disponível na Internet, ou por várias outras razões, como a indisponibilidade do *site* devido a mal funcionamento ou ao fato de estarem muitos usuários tentando conectar-se ao sistema ao mesmo tempo, causando conflito no tráfego de dados.

Talvez isto possa ser explicado pela ausência — e até pela impossibilidade — de um trabalho realmente bem planejado de indexação na Web. Considerando-se o número de documentos disponíveis na Rede, adverte Woodward (1996), revocação e precisão ainda serão alvos de muitos empreendimentos por muito tempo, se puderem ser alcançadas.

A idéia de *transferir uma parcela da culpa* pelos insucessos nas buscas ao usuário advém da responsabilidade que o mesmo assume no processo de construção da busca. A primeira coisa que o usuário considera relevante quando realiza um *browsing* é aquela que conduz a similaridades em conteúdo intelectual ou produção histórica. Por exemplo, trabalhos sobre o mesmo assunto, trabalhos criados por pessoas da mesma instituição. Além do mais, as categorias que são importantes para as pessoas irão mudar enquanto realizam o *browsing*. Este processo não é nem mesmo exclusivo do ambiente eletrônico, muito embora tenha ganho popularidade no atual ambiente de rede, o termo *browsing* é, na verdade, um conceito adaptado. Cove & Walsh (1988) o definem como tradicionalmente aplicado ao ato de mover-se por entre as estantes de uma biblioteca folheando livros ao acaso, à procura de pedaços (*bits*) de informação de qualquer natureza. O termo, hoje, contudo, está fortemente associado ao uso de navegadores e dos sistemas de busca.

SISTEMAS OU SERVIÇOS DE BUSCA (SEARCH ENGINES)

O termo *sistema de busca* popularizou-se e é fruto de confusão semântica. Sistemas de busca, diretórios, páginas amarelas e *metacrawlers*, a despeito de suas diferenças, são considerados como pertencentes a uma mesma categoria: a dos robôs que rasteiam a rede em busca de URLs. O que, basicamente, diferencia um *sistema de busca* dos demais é que esta ferramenta, sim, utiliza um robô para realizar sua varredura em algum conjunto de dados.

Um *sistema de busca* é, para ser breve, um método para encontrar informações na Internet mantendo-se um índice dos recursos Web, o qual pode ser questionado por meio de um tópico específico. A ação a ser desenvolvida e seu nível de profundidade variam em função do formato e do *software* que o serviço utiliza (*wanderers*, *robots*, *crawlers*, *worms*, *spiders*, *ants*). De modo genérico, indexam o título do documento, o URL, os *links* contidos no hiperdocumento, as primeiras duzentas palavras ou mesmo

todo o documento em alguns casos. Um grande problema advindo disto é que todos usam excessiva largura de banda, criando problemas de tráfego de dados, pois estão continuamente *caçando*. Aliás, essa é uma das razões por que tanto se investe na criação de um padrão de metadados. Não fosse a ineficiência característica dos sistemas de busca, não se justificariam alguns projetos de bibliotecas virtuais, os quais visam, no limite, sanar os problemas daí advindos.

Um sistema de busca compõem-se de: um robô, que coleta automaticamente os *links*, títulos e textos dos *sites*; uma base de dados, onde as fontes são armazenadas e uma ferramenta de busca, propriamente dita, onde o usuário pode pesquisar a base de dados. Já um diretório de assunto, embora permita buscas, são antes categorizações de recursos Internet, próprios para *browsing* (Solock, 1996).

O robô utilizado pelos sistemas de busca é um programa de computador que percorre de forma automática a estrutura hipertexto da Web recuperando hiperdocumentos e todos os hiperdocumentos que são referenciados, pelo uso de *links*, no primeiro. Existem ferramentas que realizam a mesma operação e que não podem ser qualificadas especificamente como robôs, porquanto operadas por mãos humanas e não recuperarem automaticamente os hiperdocumentos referenciados.

Será ainda possível para as bibliotecas trabalhar com esquemas de classificação produzidos aprioristicamente? Numa das passagens mais contundentes de seu recente livro Levy (1996) pontua : "Não me interessa mais pelo que pensou um autor inencontrável, peço ao texto para me fazer pensar, aqui e agora. A virtualidade do texto aumenta minha inteligência em ato". O que leva o autor a essa proposição é o incremento em termos de revocação que a navegação propicia. Com o crescimento exponencial da Internet, contudo, tornar-se-á cada vez mais difícil gerenciar a massa documental indiscriminada de resultados de busca. Outro inconveniente é que as bases de dados criadas pelos sistemas de busca, que também pululam na Rede, podem, potencialmente, tornar-se maiores do que os próprios recursos Internet (Xu, 1997).

As desvantagens do uso de robôs, apresentadas a seguir, foram apontadas por Desai (1997):

- as repetidas tentativas dos robôs de encontrar novas fontes aumenta o tráfego na rede. O número de robôs vem crescendo e os administradores de sistemas irão provavelmente desabilitar visitas de robôs;
- se a rede cresce para o modo pague-e-use (*fee-for-use*) será difícil justificar a ação dos robôs;
- tipo de dado gerado pelo robô não é útil porque é muito simples para amparar descobertas. Eles ainda são muito usados por falta de ferramentas melhores;
- sistemas baseados em linguagem natural não são avançados o suficiente para extrair significado de uma fonte;
- é mais difícil gerar identificação automática de aspectos e conceitos de fontes tais como códigos de programas, imagens digitais e sistemas complexos;

Muitas pesquisas discutindo a eficiência/eficácia e até mesmo a validade dos sistemas de busca como ferramentas para localização e recuperação de informações na Rede têm sido produzidas e disponibilizadas *on-line*, na própria Internet, mais do que na literatura convencional.

Assim que o ambiente Web começa a tornar-se popular (muito embora seja difícil estabelecer limites cronológicos, isto se dá por volta de 1993, com a criação do browser Mosaic) e a Internet passa a ser encarada como uma vasta fonte alternativa aos tradicionais modelos já desgastados de obtenção da informação, sempre mediados, surge a preocupação, inicialmente de estudantes de computação e posteriormente de grandes corporações, de colocar alguma *ordem* no caos documentário da Rede.

Em 1995, conforme dados fornecidos por Shaw num estudo publicado neste mesmo ano, o volume de textos na Internet está em torno de 35 Gb, o que equivale a um milhão e duzentos mil de livros de cinquenta mil palavras cada um. Para se ter uma idéia do que isso significa e tendo-se sempre em mente que os dados apresentados já contam três anos, a presente dissertação tem aproximadamente vinte e quatro mil palavras.

A excessiva preocupação das entidades que desenvolvem os sistemas de busca e sua luta em aperfeiçoá-los cada vez mais dá mostras de que um grande *mercado informacional* se desenvolve, pois, ainda que tais sistemas caminhem para o modo pague-e-use (*fee-for-use*), é difícil imaginar a localização de informações na Internet sem o suporte tecnológico que ferramentas como o Alta Vista ou diretórios como o Yahoo! desenvolveram. Mesmo e principalmente as bibliotecas virtuais dependem visceralmente do conluio destes sistemas de busca pois, de outro modo, corre-se o risco de tornar estática sua *coleção*, decretando-se assim o fracasso do modelo.

Em razão de sua popularidade e de sua abrangência, a literatura elegeu cinco serviços de busca e dois diretórios (ou base de dados seletivas de recursos Web) como modelos, são eles o Alta Vista, o Excite, o Hotbot, o Infoseek e o Lycos, representantes dos sistemas de busca; o Yahoo! e o Magellan, como representantes dos diretórios de assunto. Feldman (1997) atribui-lhes a metáfora de supermercados dos serviços de busca. Todos cobrem cerca de 50 milhões de *sites* em suas bases de dados.

Descrevem-se abaixo alguns destes estudos com o fim de fornecer uma visão panorâmica do que se vem pesquisando em termos de recuperação de fontes informacionais disponíveis na Internet desde então. Em sua maioria são estudos comparativos que correlacionam a qualidade do conjunto de respostas de cada serviço em termos de revocação e precisão para determinadas questões propostas¹⁵.

As pesquisas realizadas, com as honrosas exceções de sempre, ainda não alcançaram o que se pode chamar de padrão científico adequado, pois não apresentam metodologias capazes de lidar com a diversidade. Principalmente devido à grande diferença de performance de tais serviços, estabelecer um *ranking* de resultados para tais testes, preocupação de grande parte destes estudos, não é adequado, pondera Koch (1996).

Dos primeiros sistemas de busca criados, o WebCrawler ainda hoje é sinônimo de busca na Internet. Pinkerton (1994) o desenvolveu a partir de noções genéricas do comportamento de navegação das pessoas pela Rede; o WebCrawler, afirma quando ilustra o funcionamento do sistema àquela época, imita a maneira como as pessoas

buscam informações: a partir de um conjunto de documentos descobre outros, examina os *links* contidos nestes e segue cada um desses *links* que o conduzem a novos documentos, realiza-se então, como num *moto-contínuo*, a repetição de todo o processo.

Em vista de limitações de equipamentos (hardware) e mesmo do software, o WebCrawler não realizava, então, busca em textos completos. A localização da informação a partir da questão de busca passava pelo filtro (ou pela censura) de uma lista de palavras não permitidas (*stopword list*). Ainda assim não deixa de surpreender a taxa de transmissão do tempo de resposta neste período. Em média o servidor de buscas levava um oitavo de segundo para oferecer um conjunto de *links* como resposta a uma dada questão.

As limitações impostas pelo custo de recursos em termos de equipamentos, rede e largura de banda disponível, tem levado os sistemas de busca a se mudarem de seus lares .edu, afirma Shaw num alerta de 1995 reforçado posteriormente por Koch (1996), para as entidades .com, isto é, deixam as universidades para integrarem-se aos serviços comerciais. Em suas próprias palavras: o tempo dos sistemas de busca gratuitos, atualizados todas as manhãs nos laboratórios de computação das universidades por alguns estudantes bem intencionados está chegando ao fim. O Yahoo e o WebCrawler, por exemplo, confirmando sua previsão, foram recentemente adquiridos pela AOL (America Online).

Pinkerton (1994) aprendeu, com o desenvolvimento e manutenção do WebCrawler, valiosas lições, as quais com certeza foram incorporadas às decisões envolvendo os sistemas de busca criados subsequentelemente. Duas das mais significativas são a necessidade da (discutível) indexação por texto completo (para eliminar a censura) e a necessidade imperiosa de um mínimo de padronização nos elementos que compõem o *site*, antecipando, assim, a discussão em torno dos metadados.

Um problema ainda muito comum no universo livresco é identificado por Pinkerton (1994) em referência às páginas Web: a infidelidade do título em relação ao

¹⁵ Assim como são produzidos os artigos, também as bibliografias (webliografias) estão disponíveis *on-line*. Um bom exemplo, com 186 trabalhos elencados e *linkados*, pode ser encontrado no seguinte URL : <http://www.ub2.lu.se/desire/radar/lit-about-search-services.html>

conteúdo do documento. O aprendizado no caso dos livros e periódicos dá-se por força de necessidade; a partir do momento em que as publicações não padronizadas (conforme o modo como se concebe hoje uma folha de rosto e seus elementos mínimos ou as notas informativas de um periódico científico) começam a ficar à margem da cena científica, por impossibilidade, muitas vezes, de representá-las e, conseqüentemente, localizá-las e recuperá-las nas bases de dados, surge um esforço natural no sentido de adequar o documento à sua representação e vice-versa. Koch (1996) manifesta também a preocupação com o estabelecimento de alguns padrões. Dada a impossibilidade — até mesmo por razões de ordem econômica — de indexar todo o conteúdo da Internet em texto completo, julga necessário definir com mais precisão o conceito de documento e dos elementos que o representam, com o fim de possibilitar buscas alvo-orientadas, com maiores possibilidades de sucesso.

O ano de 1996 foi o que mais produziu pesquisas de comparação de sistemas de busca, o que já são indícios de certa maturidade destes sistemas. Alguns estudos significantes são os de Stanley; Webster & Paul; Koch; Solock e o de Eagan & Bender.

A premissa básica dos referidos estudos é fornecer aos interessados informações sobre o desempenho comparado dos maiores e mais populares serviços de busca disponíveis na Rede. Embora em muitos aspectos tais estudos tenham se tornado prisioneiros de seu próprio tempo — à mesma velocidade com que cresce a Internet os sistemas de busca revitalizam seus serviços — tecem muitas considerações que ainda são verdadeiras. Funcionam como manuais de orientação ou, refletindo um gosto tão em voga naquele ano, como *livros de auto-ajuda* para os internautas pesquisadores, sem, por certo, a superficialidade deste tipo de publicação.

O mesmo fator que provocou o surgimento dos sistemas de busca na Rede, a grande massa documental desorganizada, contraditoriamente, os limitam. O Alta Vista, por exemplo, comparado ao Lycos no estudo realizado por Stanley (1996), tem uma base de dados muito maior do que este último e, como conseqüência, mais desatualizada, pois demanda um tempo maior para que seu robô percorra a Internet verificando os *links* atuais e acrescentando novos. Também em função do tamanho de sua base o Alta Vista exige estratégia de busca mais elaborada, com mais qualidade, para o fornecimento de informações relevantes.

Um conselho extremamente prático é o que dá Webster & Paul (1996) aos bibliotecários, o qual serve, inclusive, como justificativa para estudos como o que realizaram: é preciso aprender a avaliar a adequabilidade de um sistema de busca ou de uma árvore de assuntos à especificidade da questão e saber incorporá-los ao trabalho cotidiano da biblioteca, de preferência enquanto são *gratuitos*.

As diferentes concepções, tanto entre os diretórios e os sistemas de busca, quanto entre estes últimos, contribuem para que conjuntos de resultados muito diferentes sejam fornecidos em resposta às mesmas questões propostas a diferentes serviços (Webster & Paul, 1996), além do fato de cada sistema de busca possuir sua própria lista de palavras não permitidas (*stopword list*).

As (poucas) vantagens de índices coletados manualmente, baseados em listas ou formulários (diretórios de assunto) estão na seleção das fontes: somente são incluídos os *links* que um autor ou usuário julgaram como sendo relevantes para outros e de qualidade suficiente. Por outro lado, a maior vantagem dos serviços baseados em robôs (sistemas de busca) é sua relativa completeza. Uma vez que as partes mais importantes da fonte original ou mesmo o documento todo é indexado, uma quantidade substancialmente maior de informação pode ser recuperada (Koch, 1996).

Os sistemas de busca, comparados aos catálogos de acesso público *on-line* e índices de periódicos *on-line*, ainda são relativamente limitados em alguns aspectos, concluem Webster & Paul (1996). Ainda não desenvolveram recursos suficientes para combinar resultados de busca ou de limitá-las a campos específicos, como o campo de autor, por exemplo, com a mesma qualidade dos primeiros.

Koch (1996) propõe uma taxonomia para a área de pesquisa sobre sistemas de busca. Define *surfing* como a perseguição não sistemática de *links* e *browsing* como o folhear sistemático de páginas Web, auxiliado por coleções ou revisões. Usa o termo *searching* para o processo de busca com a ajuda de bases de dados construídas para este fim. Numa elucidativa comparação com o tradicional mundo dos livros, define: *surfing* equivale a folhear o livro ao acaso; *browsing* a procurar no sumário por determinado capítulo e *searching* a procurar fatos e palavras isoladas utilizando o índice do final do livro.

Uma das conclusões do estudo de Koch (1996), compartilhada por outros, como Pinkerton (1994), Shaw (1995) e Feldman (1997), é a identificação de uma concentração maior dos sistemas de busca, de um modo geral, em atender o usuário médio, razão pela qual as opções de busca avançada sempre se apresentam numa segunda tela, numa interface não muito amistosa e pouco trabalhada em termos de recursos. Segundo Pinkerton (1994) o número médio de palavras utilizadas numa questão de busca é de 1,5.

Se por um lado as limitações de hardware ou de software inibem a performance dos sistemas de busca, por outro, os desenvolvedores destes sistemas de busca ignoram o progresso realizado pelas tecnologias de rede e insistem na criação de bases de dados centralizadas, ou seja, multiplica-se o número de índices que possuem mínimo ou mesmo nenhum relacionamento entre si. Cooperação é uma das lições que a Biblioteconomia cedo aprendeu como indispensável ao fornecimento de bons serviços e produtos aos usuários.

Existem atualmente mais de 1.800 sistemas de busca na Web, a maioria deles altamente especializada e direcionada a assuntos especiais (Feldman, 1997), como o FindLaw¹⁶, especializado em informações sobre legislação ou o HealthAtoZ¹⁷, o qual indexa sites que versem sobre medicina ou saúde em geral. De resto, tornar-se-á impossível a manutenção dessas bases de dados com o crescimento vigoroso do número de páginas Web e do uso desses serviços de busca. Isso leva Koch (1996) a acreditar que, numa perspectiva internacional a médio/longo prazo, deverão surgir propostas para arquiteturas de buscas distribuídas.

Num excelente estudo publicado recentemente, Leighton & Srivastava (1997) apresentam uma alternativa viável e bem justificada para contornar o problema da diversidade de performance dos sistemas de busca como fator inibidor de estudos comparativos, conforme citado acima; fazendo, por esta razão, as vezes de síntese das pesquisas anteriores. Para comparar a eficácia dos cinco mais populares sistemas de busca utilizaram questões não estruturadas compostas de termos simples. Partem de dois princípios para justificar esta opção: o primeiro está no baixo índice de utilização de operadores booleanos pelos usuários médios, conforme já fora apontado por

¹⁶ <http://www.findlaw.com>

Pinkerton (1994), Shaw (1995), Solock (1996) e Feldman (1997); o segundo reside no fato de questões não estruturadas forçarem o sistema de busca a realizar a maior parte do trabalho, ordenando os resultados mais por algum algoritmo próprio do que pelos limites estabelecidos por qualquer tipo de combinação dos operadores pelos usuários dos sistemas.

Submeteram quinze questões aos sistemas de busca os quais foram avaliados por variados juizes. A avaliação utilizada, denominada *precisão dos vinte primeiros*, julga os serviços com base na porcentagem de resultados dentre os vinte primeiros retornados considerados relevantes ou úteis.

No intuito de contornar o problema do viés subjetivo inerente ao julgamento de valor (relevância), desenvolveram uma fórmula matemática a partir da divisão dos vinte primeiros *links* retornados em três grupos: os primeiros três, os próximos sete e os últimos dez. Estes grupos foram escolhidos porque os três primeiros normalmente constituem a primeira tela do usuário, os próximos sete são visualizados rolando-se a tela e os últimos dez formam a segunda página de resultados. O primeiro grupo é multiplicado por 20, o próximo grupo por 17 e o grupo final por 10. Este valores ponderados são então somados para gerar o numerador da fração. Por exemplo, se para um dado teste, o serviço retornou cinco bons *links*, seu score será $(3 \times 20) + (2 \times 17) = 60 + 34 = 94$, caso eles estivessem entre os cinco primeiros *links*, mas somente $(5 \times 10) = 50$ se todos eles estivessem nas posições de ordem entre o décimo primeiro e o vigésimo.

O denominador foi calculado pelo número de *links* retornados pelo serviço de busca (limitado ao máximo de vinte). Se o serviço retornou 20 ou mais *links*, então a soma de todos os pesos para o conjunto de vinte *links* ($(3 \times 20) + (7 \times 17) + (10 \times 10) = 279$) foi utilizada, sendo este o máximo denominador. O denominador foi ajustado caso houvesse menos que vinte *links* retornados, com fim de se dar algum benefício para a precisão verdadeira. O denominador para esta fração é gerado pela soma de todos os pesos para 20 (279), subtraindo-se 10 para cada número de *links* que não alcance os 20. Por exemplo, se um serviço retorna 15 *links*, o denominador é $279 - (5 \times 10) = 229$, enquanto se ele retorna um link, o denominador é $279 - (19 \times 10) = 89$.

¹⁷ <http://www.heathatoz.com>

O numerador é dividido pelo denominador para calcular a medida final. A fórmula completa é apresentada abaixo:

$$(Links\ 1/3 \times 20) + (Links\ 4/10 \times 17) + (Links\ 11/20 \times 10)$$

$$279 - [(20 - \text{Número de links recuperados, ou } 20) \times 10]$$

As conclusões apresentadas por Leighton & Srivastava (1997), muito embora não ofereçam grandes revelações, até mesmo confirmando certos fatos já conhecidos, como a supremacia do Alta Vista, o fazem com fundamentação científica, o que alça este estudo à categoria de representante dos demais.

Um sistema de recuperação uniforme, completo e funcional para uma base de dados como a que se encontra hoje (desorganizada) na Internet, simultaneamente disponível para cerca de cento e cinquenta milhões de pessoas no mundo todo nunca existirá, vaticina Koch (1996).

A avaliação final do estudo realizado por Feldman (1997) afirma-se, por ora, com força de lei: todo sistema de busca fornece bons resultados de busca algumas vezes, do mesmo modo como surpreendem o usuário com péssimos resultados, mas, nenhum sistema de busca fornecerá bons resultados de busca o tempo todo.

MODELOS DE RECUPERAÇÃO DA INFORMAÇÃO: LÓGICA BOOLEANA X LÓGICA NEBULOSA

Muitos autores tem discutido a necessidade de adoção de modelos lógicos alternativos para a recuperação da informação. A terminologia para essas lógicas varia tanto quanto sua função ou aplicação específicas. Tratar-se-á aqui da lógica nebulosa de forma genérica.

Apesar de a literatura brasileira em Ciência da Informação não tratar do assunto em profundidade, já de longa data a literatura estrangeira especializada vem cobrindo este tópico, a exemplo de Radecki (1979), cujo artigo foi apresentado em versão preliminar na VI International Conference on Mechanised Information Storage and Retrieval Systems, em 1977. Apresentando de forma muito técnica, que não prescinde de conhecimentos aprofundados de Matemática, as bases para a aplicação da lógica nebulosa em recuperação da informação, Radecki (1979) destaca que o método matemático adequado deve levar em conta que os sistemas de informação precisam estar aptos a fornecer respostas a uma variedade de estimativas e de conceitos de pertinência e relevância .

Com o enorme crescimento das bases de dados disponíveis *on-line*, bem como o incremento de *sites* institucionais ou pessoais e a conseqüente necessidade de aperfeiçoamento das técnicas para acessar a informação, está havendo um enorme interesse na área de recuperação não só pelos cientistas da informação mas por todos os profissionais ligados à informática de uma maneira geral.

Por muitos anos a pesquisa dos sistemas de recuperação foi realizada por comunidades pequenas sem impacto na indústria da informação, a exemplo do Projeto Cranfield, o qual a partir da proposição das mesmas questões a diferentes usuários de uma mesma coleção através da utilização de variados mecanismos de indexação, chega a conclusões surpreendentes, tal como a supremacia das linguagens que utilizam termos isolados (únicos) (Araújo, 1994; Lesk, 1997).

Muitas aplicações eram dirigidas às bases de dados bibliográficas e serviços de informação como o Dialog™, o qual baseava sua abordagem na lógica booleana (dentro da mencionada teoria da caça ou adequação/coincidência de palavras). Pouca atenção foi dada aos modelos de recuperação que hoje preocupam todos os grandes serviços disponíveis na rede como Infoseek e Lycos, os quais exibem os resultados da busca ponderando os documentos pelo número de ocorrência na coincidência de palavras.

Algoritmos para recuperação da informação têm sido estudados já há mais de trinta anos. Estabeleceu-se, portanto, um corpo teórico respeitável. O aumento vertiginoso da informação possibilitado pela Rede, entretanto, colocou algumas dificuldades para a adoção dessas técnicas tradicionais. Já são demais conhecidos os enganos (*false drops*) em recuperação nos catálogos e sistemas de busca da Internet. Todos os grandes sistemas tradicionais (inclusive esses catálogos e sistemas de busca) utilizam de forma mais ou menos elaborada a lógica booleana e seus operadores básicos. Esta lógica apesar de simples e de larga aceitação cria limitações evidentes para a recuperação da informação.

A lógica booleana baseia-se na premissa de que o que é recuperado é relevante e o que não é recuperado não é relevante. Isto pode ser verdadeiro para dados factuais, como dados cadastrais ou demográficos por exemplo. Em relação à informação menos factual como a própria literatura (seja de ciência e tecnologia, seja a literária propriamente dita), a premissa do recuperado ser igual ao relevante não procede. Relevância não é uma propriedade dicotômica, não ocorre, por exemplo, em proposições de pertinência ou não-pertinência, mas em graus de maior ou menor pertinência. Imprecisão, ambigüidade e determinismo são conseqüências naturais da lógica booleana. Por isto, uma abordagem mais probabilística têm sido desenvolvida através de uma outra lógica envolvendo até mesmo o uso da lógica nebulosa, a qual possui mais proximidade com a linguagem e o raciocínio humanos (Araújo, 1994).

A lógica nebulosa difere dos sistemas convencionais de lógica em que seu objetivo é prover um modelo para aproximação mais do que um raciocínio preciso. Na lógica nebulosa o que importa não é necessariamente o cálculo do valor absoluto (*pointwise*) das proposições; ao contrário, uma proposição nebulosa induz uma noção relativa, e "verdade" é um predicado nebuloso da mesma forma que, por exemplo, "alto" (Dubois, 1980©).

Na maior parte dos campos de aplicação da lógica booleana está a teoria dos circuitos eletrônicos. Esses circuitos são modelados pelas expressões booleanas, as quais podem envolver somente negação, disjunção e conjunção. Aqui está o entrave que as aplicações da inteligência artificial vêm tentando resolver há algumas décadas, esta forma de modelação dicotômica (0-1; sim-não) traz implicações semânticas óbvias, não

resolvem questões de significado, ficando somente no plano sintático sua abordagem eficiente.

A adoção de uma lógica não-clássica propõe-se *per se*, naturalmente, isto é, o próprio processo de busca já é naturalmente nebuloso, pois, ao propor uma busca, o usuário parte do que Belkin (1980) chama de "estado anômalo do conhecimento", ou seja, o pesquisador busca informações e constrói o objeto de pesquisa dialeticamente. Tal processo tem como característica o alto grau de incerteza natural que o usuário tem sobre o assunto pesquisado; parte de um desejo informacional já absolutamente nebuloso, na maior parte dos casos, para construir de fato suas interrogações durante o processo de busca. O processo de recuperação da informação é qualificado, portanto, por sua própria natureza, como um processo nebuloso. Losee (1997) o define como a arte de localizar e obter documentos baseado nas necessidades informacionais expressas a um sistema numa linguagem de busca.

A lógica nebulosa enriquece a busca com a inclusão de níveis de tolerância para proposições imprecisas. Por meio do uso de aproximações apropriadas, um sistema nebuloso pode imitar o trabalho em ato de um bibliotecário de referência, o qual intuitivamente adiciona valor ao processo de busca (Girill & Luk, 1996). Alguns exemplos corriqueiros da adoção deste tipo de lógica já se encontram presentes em sistemas comerciais de largo alcance, como o processador de texto Word™ da Microsoft™, o qual permite corrigir o texto digitado por aproximação. Isto é extremamente relevante se se leva em conta que oitenta por cento dos erros de sintaxe do usuário envolvem caracteres colocados fora de ordem, conforme Girill & Luk (1996).

O uso dos operadores lógicos booleanos, até há pouco tempo de uso restrito dos profissionais da informação, integram, atualmente, as habilidades do usuário comum (Losee, 1997; Levacov, 1997). Esta popularidade advém, em sua maior porção, ao sucesso alcançado por esta lógica nos sistemas comerciais de recuperação da informação, mais do que por sua eficácia. Claro está que o usuário médio não faz uso dos recursos mais avançados da lógica booleana, conforme demonstraram alguns estudos já citados (Pinkerton, 1994; Shaw, 1995; Solock, 1996; Feldman, 1997), ficando ao nível dos operadores básicos E, OU e NÃO.

A proposta básica da lógica nebulosa é dada pela suposição de que os termos propostos numa busca não são totalmente independentes. Isto se dá pela identificação dos relacionamentos que a lógica booleana ignora. Por exemplo, o termo televisão tem mais probabilidade de co-ocorrer com termos como rádio ou jornal do que com termos como esclerose ou hanseníase.

O grau de dependência entre dois termos é capturado em parte pelo produto médio da frequência binária do termo para os dois termos (em todos os documentos) (Losee, 1997). Ou seja, o sistema pondera o número de vezes que os termos aparecem juntos num mesmo documento e, de acordo com o grau de incidência, passa a associá-los como correlacionais, com grande possibilidade de co-ocorrerem.

Os operadores booleanos mais comumente utilizados (AND e OR) são analisados por Losee (1997). O operadores booleano AND apresenta o seguinte *ranking*:

Lógica Booleana (operador AND)	
Publicações Eletrônicas	Direito Autoral
1	1

Lógica Nebulosa (operador AND)	
Publicações Eletrônicas	Direito Autoral
1	1
1	0
0	1

(Tabelas adaptadas de Losee, 1997)

Ou seja o resultado da busca na lógica booleana fica restrito à primeira linha da tabela. Somente os documentos que contenham os termos *publicação eletrônica* juntamente com *direito autoral* são recuperados. Para o sistema baseado em lógica booleana os documentos que contêm somente um dos termos e os documentos que não os contêm a nenhum, são tratados da mesma forma, como irrelevantes.

Assim ocorre também com o operador OR. O sistema nebuloso assume que o operador AND é o principal neste caso. Isto é , o documento que contêm ambos os termos pesquisados é o mais importante, deve, portanto, ocupar a primeira posição no *ranking*. Não se pode tratar os documentos que contêm ambos os termos da mesma forma.

Lógica Nebulosa (operador OR)	
Publicações Eletrônicas	Direito Autoral
1	1
1	0
0	1

Lógica Booleana (operador OR)	
Publicações Eletrônicas	Direito Autoral
1	0
0	1
1	1

(Tabelas adaptadas de Losee, 1997)

Alguns aspectos do comportamento de busca são universais, outros referem-se a situações específicas. Em ambiente de rede, as bibliotecas precisam desenvolver mecanismos de busca que reflitam a crescente diversidade de necessidade informacional de uma população crescentemente heterogênea. As bibliotecas servem tanto aos interesses locais — ou seja, funcionam como repositórios do conhecimento gerado localmente e, portanto, como ponto de partida para determinados tipos de busca — quanto aos interesses universais, considerando-se que é, no vasto ciberespaço informacional, somente mais um componente (*link*) de um emaranhado de fontes (Bishop & Star, 1996).

INTERFACES E AMBIENTES DE RECUPERAÇÃO DA INFORMAÇÃO

"Para lembrar-se da importância da percepção sensorial, do gesto e da resposta ao gesto, pense na última vez que você apertou um botão de elevador e a luz não acendeu (provavelmente porque estava queimada). A frustração é enorme: será que ele me ouviu?" (Negroponte, 1995).

Esta citação pode ser a síntese de um longo capítulo inteiramente dedicado à discutir a interface homem/máquina no livro de Negroponte. Este tem sido um dos maiores desafios para a área de informática como um todo.

O tempo máximo de espera por resposta a qualquer interação deste tipo deve ser de no máximo cinco segundos. Não é por outra razão que são criados ícones que se movem para indicar o cursor do mouse e as janelas têm sempre algum movimento. Diante da tela do Netscape Navigator™, por exemplo, em horários de pico na Internet, ter-se-ia a impressão de abandono ou de falha não fossem os cometas que se movem no canto superior direito da tela enquanto alguma informação é processada. A indicação de movimento na tela é a garantia para o usuário de que a máquina o entendeu e está processando uma resposta. Esta questão da resposta, do *feedback* — diriam os teóricos

da comunicação — é fundamental aliás para todo e qualquer tipo de interação (diga-se de passagem, interação significa ação que se exerce mutuamente). Basta lembrar das conversas ao telefone onde, não sendo possível ou não sendo proveitoso, exibir gestos de anuência ou de negação, depende-se dos "hums" produzidos como sinais de que se ouve e se entende o que é dito.

A pesquisa em recuperação da informação na Web também se depara com um grande desafio em questões de interface: a permanência ou validade de instrumentos num ambiente em constante desenvolvimento. A própria contradição do momento atual apresenta-se como desafio a ser vencido, pois se o crescimento tecnológico não pode ser aprioristicamente encarado como exponencial, ou seja, não se pode assumir que a tecnologia terá para os próximos dez ou quinze anos o mesmo avanço verificado para o mesmo período anterior, por outro lado, não cabe reproduzir as tecnologias de recuperação da informação pensadas para a informação bibliográfica impressa, porquanto existe um desenvolvimento real já verificado e, ainda que não se lhe arrisque afirmá-lo como exponencial, promissor.

Existe uma série de desafios técnicos para os quais são requeridas soluções urgentes, tais como: comunicação em alta velocidade; agentes inteligentes para indexação, descoberta, busca e acesso a coleções de fontes heterogêneas de informação (Desai, 1997). A solução de tais desafios visa a garantir a manutenção do privilégio das bibliotecas na ação de selecionar informações, posto que sua função de agente de entrega da informação está sendo amplamente questionada. Numa belíssima metáfora Lesk (1995) compara: se se pensa a informação no futuro como um oceano, não mais será função das bibliotecas fornecer água; mas sim o fornecimento de barcos.

Interface e ambiente são hoje problemas cruciais para as bibliotecas virtuais e suas estratégias de recuperação da informação. De dez problemas que possam ser elencados como restritores à eficácia de uma biblioteca virtual, pode-se afirmar sem grande margem de erro, pelo menos setenta por cento deles referem-se a problemas desta ordem.

Evidente que este paroxismo não atinge diretamente o usuário médio em nível técnico, pois a biblioteca virtual o desobriga de conhecer em profundidade procedimentos de interface, posto que o endereça diretamente aos documentos, não

importando se estão armazenados em endereços de FTP (File Transfer Protocol), Gopher, ou em bases referenciais acessíveis via catálogos de biblioteca (ambiente Telnet). No entanto, o problema é perceptível para os desenvolvedores de bibliotecas virtuais. Quer estas optem por desenvolver seus próprios mecanismos de busca (inflacionando o já tumultuado tráfego nas infovias) ou pela utilização dos sistemas já existentes, chocar-se-ão inevitavelmente com os protocolos de exclusão ou de limitação estabelecidos por alguns servidores para controlar a ação dos robôs. Somadas a operação de descoberta e posteriormente de verificação do *link*, um *site* muito concorrido pode receber, num intervalo de tempo indefinido, a visita de aproximadamente dois mil robôs, o que representa utilização excessiva da largura de banda disponível.

Largura de banda é um dos detalhes técnicos da configuração ambiental da Internet que mais se fazem sentir no uso corriqueiro dos recursos disponíveis, é em função dela que o protocolo de transferência de arquivos ou de pacotes informacionais controla o volume de tráfego na rede. Imagine-se, por exemplo, o acesso realizado a uma universidade australiana cujo *backbone* tenha largura de banda de 2 Mbps¹⁸. Durante a realização da transferência de um arquivo de 40 Kb (tamanho médio de um artigo de periódico) via FTP da referida universidade para o computador da PUCCAMP, por exemplo, o arquivo deixa a máquina original em um único pacote de 327.680 bps em direção aos Estados Unidos (1 segundo de transferência), seguindo na mesma velocidade até São Paulo¹⁹, de São Paulo a Campinas afunila-se a largura de banda (68 Kbps) obrigando o protocolo que gerencia o arquivo a enviar pacotes de 68000 bps, consumindo 4,81 segundos, somando um tempo total de 6,81 segundos de transferência. Este exemplo, entenda-se bem, é calcado numa proposição impossível de que o usuário tenha todas estas linhas disponíveis somente para ele.

Uma conexão telnet pode consumir ainda mais largura de banda, pois em se tratando de acesso remoto em tempo real, exige que a conexão permaneça aberta enquanto durar a sessão.

¹⁸ 1 Kbit=1000 bits, menor, portanto, que 1 Kb que equivale a 1024 bytes, sendo 1 bit a unidade mínima de informação e 1 byte o equivalente a 8 bits, o volume de bits necessários para representar uma caracter

¹⁹ Veja informação sobre o *backbone* da RNP em <http://www.rnp.br/1.3.bone.html>

Se a quantidade e qualidade das infovias é um problema que, de certo ângulo, foge ao controle das bibliotecas virtuais, o mesmo não se pode dizer dos procedimentos de interface e de desenvolvimento geral do *site*. Facilitar a navegação do usuário torna-se fator de economia do tempo gasto na Rede e, conseqüentemente, de redução de tráfego e uso da largura de banda.

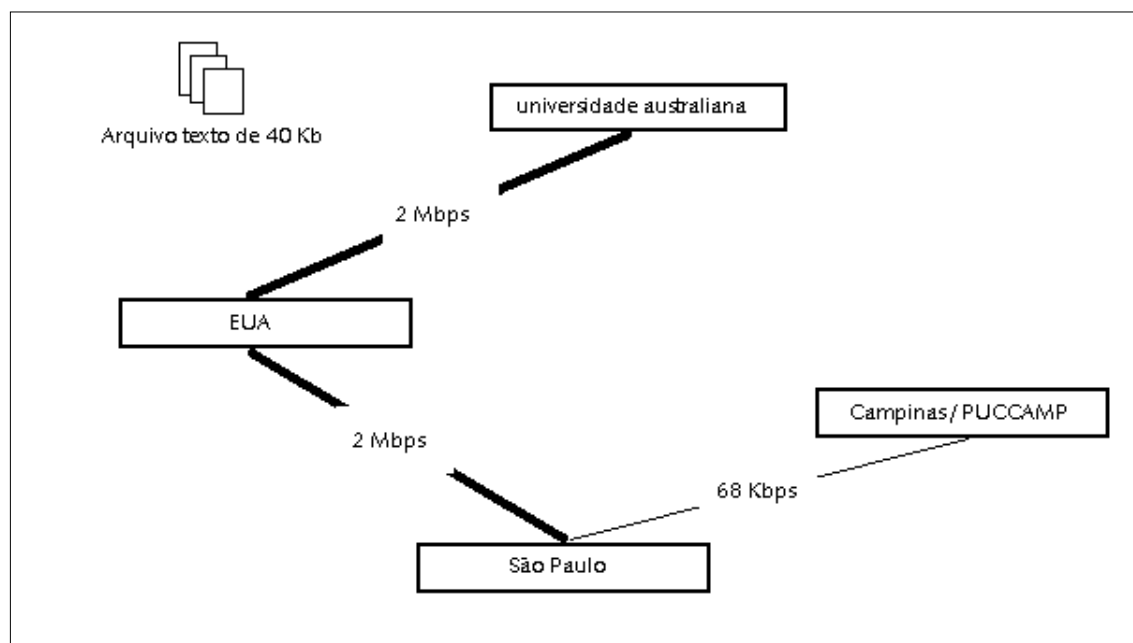


Figura 1 : Esquema de transmissão de dados via FTP

Devido à ação de uma diversidade de ferramentas de busca, cada qual com sua especificidade, que já começa a ser diluída, e ao constante incremento de páginas Web — uma página a cada quatro segundos, o que permite à Internet dobrar de tamanho a cada cinquenta dias — a falta de padronização quanto a forma em tais serviços tem obrigado os estudiosos da área a buscar soluções alternativas, que ponham alguma ordem no caos aparente. Um exemplo disto é o desenvolvimento de cursos voltados para a criação e manutenção de *sites* implantado recentemente pela University of British Columbia (Smith, 1994).

Para além dos conteúdos de caráter técnico, como linguagem de programação, estrutura de hipertexto e protocolos de transmissão de dados, e dos conteúdos de caráter administrativo, como as questões de direito autoral e netiqueta (normas de etiqueta na

rede), entre outras disciplinas, tais cursos voltam-se principalmente para o estabelecimento de um conjunto mínimo de padrões de interface.

Na linguagem comum confunde-se *homepage* com *site*. Essa confusão terminológica, longe de ser gratuita ou incoerente, baseia-se no fato de ser a *homepage* a própria interface, ou o primeiro e, portanto, o mais importante, contato do usuário com o conjunto de páginas Web denominado *site*. A *homepage* está para a biblioteca virtual assim como o balcão de referência — no que este tem de receptividade e responsabilidade pela imagem da biblioteca — está para a biblioteca tradicional.

A interface predominante no ambiente Internet é adaptada do já bem sucedido ambiente Windows™, expresso pelo acrônimo WIMPs (*windows, icon, mouse, pop-up menu*, ou seja, janelas, ícones, mouse e menus suspensos), também conhecida como GUIs (*graphical user interface*).

Apesar de estar fortemente vinculada, devido à sua popularidade, aos produtos da Microsoft™, foram a Xerox™ com o sistema Star™ e posteriormente a Apple™ com o aperfeiçoamento do Macintosh™, os pioneiros (Rowley, 1994).

A maior vantagem das GUIs é que os softwares desenvolvidos para esses ambientes incorporam as características do ambiente como um todo, fazendo com que o usuário esteja sempre em interação com uma plataforma já conhecida (WIMPs), o que diminui consideravelmente o tempo gasto na acomodação ao novo sistema. Isso torna-se patente quando compara-se a metáfora da janela com os tediosos sistemas operacionais de comandos, como o DOS™, por exemplo, ou o acesso WWW em comparação com o Telnet.

O maior problema em questões de interface, não obstante, está na ausência de um único índice completo dos recursos disponíveis na Internet, construção, aliás, considerada impraticável, conforme apontado na discussão sobre sistemas de busca. Existem ferramentas denominadas metaíndices que se propõem a fazê-lo — reproduzindo sonhos que remontam à Alexandria — como por exemplo o Metacrawler ou o Webtaxi, contudo apresentam visíveis limitações, como, por exemplo, a não utilização de todos os recursos de busca avançada, específicos de cada sistema de busca. A condução de uma busca intensiva na Internet por um tópico específico ainda esbarra

na proposição da mesma questão a várias ferramentas de busca, devido às características peculiares que cada ferramenta exige na formulação da estratégia de busca e de seu escopo de atuação.

Se ninguém discute que, numa ampla visão, a função da biblioteca virtual reproduz aquela já atribuída à biblioteca tradicional, com relação às tarefas básicas, isto não é tão pacífico assim. Os serviços básicos, essenciais mesmo, de identificar as necessidades dos usuários, relacionar as necessidades às fontes informacionais, prover e facilitar o acesso do usuário, são agora dependentes da configuração tecnológica atual, ou seja, não se pode mais pensar tais tarefas — porquanto ainda é preciso pensá-las — sem o conluio das aplicações tipificadas pela Web (Ridley, 1994).

Para Pereira(1995) está-se diante de princípios que ocupam já há muito a prática biblioteconômica: seleção/*aquisição* de fontes informacionais em função da comunidade usuária; seleção dos melhores lugares para o fornecimento de materiais atualizados e o agrupamento ordenado desse material de forma a orientar o usuário. "Um conhecimento tão velho quanto a profissão", arremata.

Se a assertiva da autora causa estranheza num primeiro momento, no seguinte já não se pode roubar-lhe a razão; claro que mais do que nunca a profissão revitaliza-se, pois, a despeito de todos os novos materiais informacionais e de toda gama de serviços marcados pelo auto-atendimento, está-se diante de um processo básico (sintetizado ao extremo) de organização e disponibilização de informação, campo de domínio do bibliotecário desde sempre, embora não exclusivo.

Ocorre que o ambiente de recuperação da informação deixa gradualmente de estar restrito às coleções das bibliotecas — seu acervo ou as bases de dados que assina — para ocupar um ambiente virtual, possível de realizar-se conforme a vontade do usuário, a qualquer tempo e em qualquer máquina conectada a um modem.

Novo ambiente exige, por certo, novos paradigmas ou revisão de antigos, como a que propõe Marchiori (1996) em função da orientação do desenvolvimento de coleções para as necessidades do usuário. Claro que os usuários influem na formação e desenvolvimento de coleções, mas o fazem principalmente enquanto indivíduos comprometidos com a organização maior que comporta a biblioteca. Assim, a natureza

da coleção "é influenciada por uma série de fatores, tais como as demandas dos usuários, as decisões administrativas, as ações governamentais, a disponibilidade de recursos de informação e mais claramente, as questões econômicas" (Marchiori, 1996).

Ambiente de recuperação da informação é antes de mais nada um problema tecnológico. Definida a biblioteca virtual como uma coleção de *links* com valor agregado, os critérios que mais pesam na inclusão/manutenção de um determinado link são, conforme Koopman & Hay (1994).

- a fonte, confiabilidade e completeza da informação contida;
- frequência de atualização e comprometimento do servidor em mantê-la;
- valor relativo da informação para o usuário potencial da biblioteca virtual e possível relacionamento da fonte eletrônica com uma fonte impressa presente na biblioteca;
- a funcionalidade do site, isto é, a facilidade de uso e de conexão.

O problema da confiabilidade da fonte e do comprometimento do servidor com sua atualização e manutenção pode ser encarado por dois ângulos. Um do ponto de vista tecnológico, que envolve a necessidade de constantes mudanças de endereço (URL), fator característico de um ambiente em formação; outro do ponto de vista social, que pode ser resolvido com o simples comprometimento de estabelecer um *link* para o novo endereço na página desativada por algum período.

Além do mais, a mudança é algo previsto, o que obriga também as bibliotecas virtuais a verificarem constantemente seus *links*, seja por meio do uso de agentes inteligentes, seja manualmente; poucos *sites* são deliberadamente excluídos, a maioria muda tão somente de provedor ou passa a ocupar uma outra ramificação da árvore do mesmo que já ocupava antes.

A funcionalidade do site, sim, é uma questão que se pode afirmar, sem grande margem de erro, puramente tecnológica. Koopman & Hay (1994) apontam esse quesito como o maior responsável pela desistência do usuário. Problemas como *links* defeituosos, mal endereçados, ou *sites* tão populares que tornam sua conexão um

fenômeno raro, ou ainda *sites* que requerem softwares que a máquina usuária não possui (como as que exigem acesso telnet, por exemplo), ou o carregamento, já na *homepage*, de imagens *pesadas*, formam um exército de obstáculos à boa performance do site. Inserir um *link* telnet num hipertexto é tão fácil quanto criar um *link* referencial, mas viabilizar este tipo de acesso implica em que o usuário localize, transfira e configure o cliente telnet antes que o URL telnet possa funcionar em seu navegador.

Aliás, voltando a discutir a *homepage* enquanto interface do site, sua criação envolve mais habilidades artísticas do que requisitos tecnológicos propriamente, pois da mesma forma que o usuário não se disponibiliza a exercitar sua paciência observando a lenta formação de uma imagem de alta qualidade em sua tela, também não a emprega na leitura de extensas descrições. Estudos sobre o *design* da interface adequada ao usuário são por esta razão muito importantes; tanto para entender como os usuários usam um sistema quanto para descobrir quais aspectos adicionais são necessários ou requeridos no sistema. Tais estudos requerem equipes multi-disciplinares de profissionais da informação, cientistas da computação, cientistas sociais, especialistas em *design* e, claro, usuários finais (Larson, 1996).

A instabilidade deste ambiente informacional é uma das grandes vilãs da não aceitação, por grande parte da comunidade científica, das fontes eletrônicas como meio de atribuição de prestígio acadêmico. A avaliação positiva da informação eletrônica, diz Graham (1995), pela comunidade científica, depende de sua confiança em tal informação que é disponibilizada, com autenticidade e manutenção de integridade.

Como se vê, a discussão sobre o macroambiente da biblioteca virtual não poderia passar ao largo de seu aspecto social. Isto se dá principalmente porque são os periódicos os primeiros textos científicos (enquadrados neste padrão a priori) a iniciar a gradual transferência para o ambiente virtual e uma das principais funções do periódico, como se sabe, é a de instituição social, como meio de atribuição de prestígio e reconhecimento acadêmico.

O aspecto social da biblioteca virtual é discutido por Bishop & Star (1996), o fenômeno da biblioteca virtual, afirmam, altera algumas das atividades sociais básicas. Trata-se de um novo modo de produzir trabalho ou de se divertir, como aconteceu com o desenvolvimento do automóvel e do telefone e as maneiras como eles mudaram os

ritmos, a velocidade e o acesso de múltiplos setores na vida moderna. Definem a informática social das bibliotecas virtuais como o estudo das influências sociais, processos, práticas e efeitos relacionados ao modo como o conhecimento é estruturado e comunicado em bibliotecas virtuais.

O vasto conceito de informática social está calcado na interação entre sociedade e informática, no modo como tecnologias informáticas e configuração social influenciam-se mútua e dialeticamente.

A aludida não aceitação pela comunidade científica dos periódicos eletrônicos, ainda que estes guardem muitas semelhanças com os tradicionais, com seus fascículos, números e corpo editorial, é plena de contradições. Os cientistas aceitam o meio eletrônico como leitores, mas não como autores e a razão para isto não é de ordem econômica, com certeza, vez que economiza-se pelo menos trinta por cento com a publicação eletrônica e o lucro do pesquisador, sabe-se, é o prestígio acadêmico. O argumento básico está na suposta abertura que a Internet, considerada de forma genérica, possibilita: qualquer pessoa pode publicar e até se auto-publicar na Rede; nos periódicos tradicionais, somente os bons autores o conseguem. Esta visão é claramente obliterada pelo preconceito. Qualquer pessoa pode publicar suas idéias igualmente usando o meio impresso, mormente após o uso dos editores de texto, cada vez mais sofisticados; não o fazem, portanto, não por razões de ordem técnica, mas de uma outra ordem que envolve a distribuição. O mesmo fato ocorre na Rede. Claro que é reconhecidamente mais fácil auto-publicar-se neste ambiente, basta criar um site num servidor que os oferece gratuitamente, mas a difusão de idéias esbarra na confiabilidade da fonte e do autor. O erro primário está em comparar a seriedade de um periódico eletrônico como o Digital Libraries²⁰, por exemplo, com a página de um estudante de informática que, no intuito maior de aumentar seu círculo de amizade, disponibiliza algumas bobagens. Aliás, a associação da explosão documental com documentos de má qualidade não é exclusiva do ambiente eletrônico; o critério de avaliação, tal qual no ambiente impresso, é a maturidade do pesquisador em selecioná-los.

Levy & Marshall (1994) expressam por meio da equação *papel=fixo+permanente* e *digital=fluido+transiente* o modo como o senso comum alinha e distingue os

²⁰ <http://www.dlib.org>

documentos. Ou seja, acredita-se que a tecnologia do papel produz documentos que não se alteram (como se não houvessem as edições revisadas ou atualizadas) e que expressam opiniões de verdades absolutas (como se não fossem derrubadas inúmeras teorias todos os dias). Neste mesmo raciocínio acredita-se que as tecnologias digitais produzem apenas documentos fluidos (que se expandem como um gás, sem forma) e transientes (efêmeros, marcados por uma vida útil muito curta). O erro desta proposição toda, tanto em relação aos documentos impressos quanto em relação aos documentos digitais, está em que todos os documentos, a despeito da mídia, afirmam os autores, são fixos e fluidos ao mesmo tempo.

As novas tecnologias e as novas bibliotecas precisam trabalhar com essas contradições; integrar papel e formas digitais como mídias absolutamente válidas, e já validadas, e resolver os problemas de recuperação da informação em coleções híbridas.

A questão maior, entretanto, está na esterilidade de toda essa discussão: é inevitável que os periódicos científicos migrem para a Rede. Com certeza não se trata de uma atitude impensada o fato de estar o Elsevier, cujo volume de vendas de periódicos fica em torno de seiscentos milhões de dólares anuais, transferindo toda sua base para o meio eletrônico.

Em função do ambiente, portanto, é que os serviços e também os produtos são redesenhados. Nas bibliotecas virtuais não há mais a separação física entre os documentos e os registros que os descrevem; requer-se, por esta razão, mais uma webliografia (uma bibliografia de documentos eletrônicos disponíveis em páginas Web) do que o que o estreito conceito de catálogo local propõe como representação do que a biblioteca *possui*.

Não se discute que a representação ainda é necessária, até mesmo pelo simples fato de não contemplarem, muitos documentos (ou simulacros destes), todos os elementos necessários a sua localização e recuperação. Dentre as razões apontadas por Lagoze (1997) para reforçar a ainda necessária representação das fontes de informação em rede estão:

- tamanho da representação é normalmente menor do que o objeto;
- preocupações com as questões de direito autoral e

- a representação pode apresentar informações que não estão, necessariamente, expressas no objeto.

Da mesma maneira como muitos usuários demonstraram pouco conhecimento dos mecanismos de busca utilizados pelas bibliotecas tradicionais e pouca inclinação para despender esforços no sentido de conhecê-los profundamente (Bishop & Star, 1996), o usuário de fontes eletrônicas de informação, ainda maravilhado com o vasto material disponível, parece cultivar o hábito (ou o vício) de digitar algumas palavras e ver o que resulta delas quando está inquirindo algum sistema de busca. Por esta razão acredita-se que se o ambiente tecnológico é uma influência positiva, ainda que precária muitas vezes, e a questão proposta às bibliotecas virtuais não será solucionada sem um grande empenho da atividade profissional do bibliotecário ou profissional da informação (Levy & Marshall, 1994; Woodward, 1996), pois referem-se a problemas de comunicação entre sistema e usuário, de interface, em última instância.

Uma crítica comum envolvendo a questão da ambientação em bibliotecas virtuais são as falhas (*bugs*) encontradas num dado conjunto de *links*. Em maior ou menor grau o mesmo problema já vem acontecendo com as bibliotecas tradicionais desde há muito. Muitas vezes, e a literatura especializada está recheada de exemplos, o usuário deixa a biblioteca sem encontrar o que procura, seja porque não entendeu o funcionamento do catálogo (um problema de interface, como já foi dito), seja por falhas de estrutura no sistema mais abrangente, como, por exemplo, documentos perdidos, documentos fora de lugar ou simplesmente não encontrados (Bishop & Star, 1996). Não basta o fato de lidar com uma coleção física de documentos para garantir homogeneidade no tratamento e no atendimento informacional, nem tampouco oferecer a possibilidade (infinita, em tese) de saltos hipertextuais para garantir ao usuário a posse/acesso ao documento pretendido.

Espera-se que o bibliotecário utilize sua experiência em localizar documentos valiosos para sua comunidade na Rede, ainda que (ou por esta mesma razão) os documentos em bibliotecas virtuais ou dispersos pela Internet não sejam classificados e catalogados como a Biblioteconomia entendeu estes termos até aqui (Woodward, 1996).

A tabela abaixo, adaptada de Nürnberg et al. (1995) compara os materiais, suas representações e os processos que os envolvem na biblioteca virtual com os que se encontram nas bibliotecas tradicionais.

	Dado	Metadado	Processo
Traduções de elementos da biblioteca física	Livro Jornal Filme	Índice estático (vida mais longa, portanto, menos atualizado) Classificações Arranjo espacial	Aquisição de dados Sugestão de fontes Ajuda na localização de fontes
Novos elementos na biblioteca digital	Hipernovela Visualização científica Programas de computador	Índice dinâmico (vida mais curta, em compensação, mais atualizado constantemente) Estrutura personalizada (<i>bookmarks</i>) Anotações (modificações virtuais no documento)	Pesquisa em texto completo Apresentação personalizada (controle de acesso e layout da tela) Recuperação por agentes (programas que buscam dados de forma <i>autônoma</i> e relatam os resultados para o usuário)

Comparação dos elementos da biblioteca virtual com os da biblioteca tradicional
(adaptado de Nürnberg et al., 1995)

Na biblioteca virtual estão presentes, como se vê, elementos físicos traduzidos (escaneados, digitalizados), elementos conceituais da biblioteca tradicional adaptados ao ambiente digital (como, por exemplo, os esquemas de classificação ou de catalogação) e elementos digitais completamente novos, sem paralelo na biblioteca tradicional (Nürnberg et al., 1995). Mesmo que se repita o modelo da biblioteca tradicional, deve-se lidar com uma diferença fundamental: tanto os metadados nas bibliotecas virtuais quanto os processos são, ou podem ser, personalizados, isto é, construídos pelo usuário, e estão sujeitos às constantes mudanças das fontes informacionais, fato antes impensável para o modelo estático tradicional.

No fim o objetivo de toda e qualquer biblioteca, seja virtual ou tradicional, é recuperar informações. Só há, então, um motivo para organizar uma coleção de documentos (uma biblioteca); as limitações em torno de sua eficácia ficam por conta da tecnologia que é utilizada, ou de como essa tecnologia é aplicada.

Fato bastante comum na Biblioteconomia / Ciência da Informação (mas não somente nessas disciplinas) tem sido a adoção de modelos de outras ciências para a realização de algumas de suas descobertas e/ou reflexões. Assim, tem-se utilizado de conceitos da Física (termodinâmica, entropia, Teoria do Caos etc.); da Administração (sistemismo — amplamente abraçado pela Biblioteconomia, notadamente nos anos 80); da Psicologia (behaviorismo e a recente técnica do *sense making*) e da Sociologia (funcionalismo). Não deixa de causar certa estranheza, entretanto, a adoção de um conceito da Biologia, disciplina aparentemente sem conexão com todas as demais citadas, para explicar a mudança de paradigma que vem sendo discutida aqui, qual seja o conceito de *foraging*, expresso pelo controle da relação que existe entre a busca de sustento e a maximização da energia pelos animais.

A caça de sustento informacional refere-se às atividades relacionadas à procura, à avaliação e ao manuseio das fontes em ambientes informacionais (Pirulli & Card, 1995). Autores como estes, além de Sandstrom (1994); Cronin & Hert (1995) e Bishop & Star (1996), traçam analogias entre o modo como o animal se comporta quando busca por alimento e o comportamento de busca por informações demonstrado pelo pesquisador. O modelo tradicional de recuperação da informação é comparado ao modo como o lobo obtém alimento — perseguindo sua caça — enquanto o novo modelo que se desenha para a recuperação de fontes de informação na Rede aproxima-se mais da estratégia utilizada pela aranha, a qual arma sua teia e espera que sua presa venha até ela.

O objetivo da analogia proposta, conforme Pirulli & Card (1995), é estabelecer alguns pressupostos básicos para alicerçar os estudos de comportamento de usuários que versem sobre a relação custo/benefício da aquisição de informação no ambiente virtual, ou seja, aqueles que procuram identificar a correlação entre o valor (ou a relevância) da informação adquirida e as atividades dispensadas em tarefas relacionadas com a interação homem/máquina, ou indivíduo/computador.

Os pesquisadores em busca de sustento intelectual e originalidade procuram em diferentes locais e em diferentes períodos de tempo com intensidade variável. Esse locais incluem sua coleção pessoal, a biblioteca local e as remotas, os serviços de acesso

²¹ Tradução livre de Information Foraging Theory

a documentos, as possibilidades de interação informal e os recursos Web (Cronin & Hert, 1995). Estes últimos recursos, aliás, retomam todos os demais na perspectiva da interação virtual ou digital; a coleção pessoal, por exemplo, é enriquecida pelo próprio *bookmark* do pesquisador, neste sentido uma *quasi*-biblioteca virtual.

Em comparação com o ambiente criado pelas bases de dados, totalmente cliente-orientado, a WWW é muito mais livre; por esta razão a metáfora do *foraging* é apropriada ao contexto, pois, busca-se pelo sustento num terreno acidentado e imprevisível, não definido, da Rede.

O princípio que norteia os sistemas tradicionais de recuperação da informação é o da melhor coincidência de palavras (*matching*). Segundo Cronin & Hert (1995), *armazenamento* e *recuperação* são metáforas baseadas no esquema *contêiner*, o que equivale a *encaixar* ou *empacotar* as informações em compartimentos organizados de "e" e "ou". O desafio consiste em encaixar as caixinhas mentais com as caixinhas da máquinas.

Na Web normalmente não se trabalha com uma busca rigidamente específica; sempre se está aberto a novas possibilidades, novas interações. Disso resulta a aplicabilidade da retroalimentação de relevância e sua relação de proximidade com a estrutura hipertextual, permitindo a condução de buscas que orientam-se (ou desorientam-se?) pelo pensamento tangencial. A técnica da retroalimentação de relevância (*relevance feedback*) visa a possibilitar a seleção de novos documentos que guardem similaridades com os que já foram encontrados; a partir dos documentos marcados como relevantes cria-se, assim, uma nova questão de busca, a qual pode ser refinada até que se atinja a proposição adequada (Paijmans, 1997).

Grande parte da já discutida frustração dos usuários com os recursos Web está na pré-disposição destes em tentar repetir o modelo do *contêiner* no ambiente WWW. É óbvio que qualquer internauta deve lançar aqui um fatídico argumento contra o otimismo do texto e, principalmente, da lentidão do tráfego informacional na Internet, ainda mais em se tratando de um discurso orientado a uma profissão que, por excelência, intermedia a transferência de informação; não obstante, não se pode negar que tais problemas estão sendo/serão solucionados e não se pode esperar até que isto aconteça para que sejam apontadas alternativas.

Se há algo que o ambiente WWW tem de atraente e de motivador este é, com certeza, a possibilidade de descobrir novidades (novos nutrientes) (Cronin & Hert, 1995), com maior facilidade, ainda que com as limitações de hardware ou de telemática impostas, do que no ambiente impresso.

É preciso estudar os hábitos dos usuários e dos provedores de recursos eletrônicos de informação para determinar com alguma precisão quais são os locais onde são caçadas as informações. Um dos problemas levantados por Bishop & Star (1996) refere-se ao que se chama comumente de *propaganda enganosa* na linguagem comum; os robôs que rastreiam a Internet catalogando recursos são particularmente vulneráveis às palavras não-relevantes (em sua relação com o texto) acrescentadas por criadores de páginas Web para atrair usuários, utilizando, de modo análogo, a mesma estratégia com a qual a planta carnívora e sua bela aparência atrai os desavisados insetos.

Claro que a superficialidade com que é tratada aqui, onde o objetivo é somente apresentá-la, não reflete a amplitude da teoria em questão. De todo modo, o que estes estudiosos querem é mudar o paradigma da *melhor coincidência de palavras* para o da *caça de sustento informacional*. A dificuldade maior reside na viabilização de um conjunto de critérios pelo qual o usuário possa julgar a relevância de um determinado recurso em termos de um nutriente específico. Esta prática é considerada por Cronin & Hert (1995) como indispensável nas áreas da ciência da informação e da recuperação da informação, a medida que estas se afastam cada vez mais de populações claramente definidas de usuários e de conjuntos claramente definidos de documentos.

Se por um lado a relação custo/benefício pode ser de difícil mensuração, em razão das limitações da avaliação de recursos informacionais com a mesma precisão com que a Biologia cuida da relação calorias/nutrientes (Sandstrom, 1994), por outro, os custos operacionais já se fazem sentir, sendo representados, no ambiente de recuperação/acesso, pela operacionalidade dos sistemas de busca, razão pela qual é necessário também validá-los como ferramentas úteis.

RECUPERAÇÃO DA INFORMAÇÃO HIPERMÍDIA

O termo hipermídia é resultado de uma sutileza conceitual. A designação genérica hipertexto e sua natural vinculação com o elemento textual, indicada pelo próprio termo, não representam com precisão o tipo de processo predominante nos sistemas de informação modernos. Como a Web é um local onde circulam em igual proporção arquivos sonoros, imagens e textos a já conhecida designação *multimídia* ganha maior abrangência e cria-se a partir desta fusão (hipertexto + multimídia) o termo hipermídia, com a finalidade de representar hiperdocumentos que utilizem a tecnologia multimidiática (texto, som e imagem) com ligações, ou nós, entre si.

Na verdade a discussão sobre recuperação da informação hipermídia ainda é incipiente, sendo poucos os estudos que relacionam esta técnica com as funções/atribuições da biblioteca virtual. Ainda assim, a discussão sobre hipermídia em Ciência da Informação permanece fortemente atada às formas de tratamento/acesso/recuperação de imagens (estáticas — em maior número — ou em movimento) em detrimento de outras mídias possíveis, como o som ou as simulações científicas, amplamente utilizadas pelas ciências duras.

A riqueza da Internet, enquanto rede de comunicação planetária que é, está na multiplicidade; na possibilidade que esta tecnologia cria para que pessoas de variadas nacionalidades, de variadas linguagens, lançando mão de variadas mídias, possam interagir. Não há porque, então, limitar a atuação da biblioteca do futuro, como alguns chamam a biblioteca virtual, à informação textual. Nem há razão para a hegemonia da língua inglesa (americana) deste veículo, pontua Borgman (1997), a qual revela a assustadora e expressiva taxa de aproximadamente setenta por cento das fontes atualmente disponíveis na Rede estarem em língua inglesa.

O tratamento e a recuperação da informação hipermídia, bem como os desafios e as possibilidades que este tipo de material oferece em termos de arquivamento ou acesso em tempo real, evidenciam de modo acentuado as já amplamente discutidas (e desgastadas, por esta razão) diferenças entre a biblioteca tradicional e seu subsistema de recuperação e a biblioteca virtual. Muitos autores concordam que o maior problema

envolvido na acessibilidade a este tipo de informação está no desafio dado pelo desenvolvimento de técnicas adequadas ao seu processo de indexação como um todo.

Procura-se desenvolver um conjunto de técnicas que possibilite fornecer localização, acesso e recuperação que seja ao mesmo tempo comum a uma ampla gama de mídias (fotografias, mapas, gravuras, textos digitalizados, vídeos e arquivos sonoros), respeitadas suas diferenças de forma e função, e que não dependam da intermediação, e a conseqüente limitação, das descrições textuais.

Não se pode creditar eficácia, mas também não se deve negar a relativa eficiência — que, aliás, já não mais satisfaz os usuários mais exigentes ou mais especializados — da busca de imagens por meio de seus elementos textuais oferecidas por sistemas de busca como o Lycos ou por base de dados seletivas de recursos Web como o Yahoo!. Ainda que se critique suas (evidentes) limitações, exemplos notáveis vem surgindo, como por exemplo, o WebSeer²² e seu banco de dados de mais de setecentas mil imagens, permitindo localizá-las por assunto, tamanho, cor ou origem.

O desafio nesta área está na absoluta novidade destes elementos, sem paralelo e, em conseqüência, sem a tradição técnica e histórica empregada nas fontes tradicionais. A biblioteca enfrentou sem grandes percalços a fase da automação porque nesta fase não se inventou novas ferramentas de recuperação da informação, apenas melhorou-se as que já existiam; agora, sim, há que se inventá-las. É oportuno lembrar que qualquer discussão envolvendo organização e disponibilização de arquivos de imagens inclui, necessariamente, os textos digitalizados. A digitalização de textos impressos, apesar de não ser uma tarefa técnica ou operacionalmente difícil, resulta num objeto cujo conteúdo pode significar alguma coisa para um ser humano quando mostrada na tela, mas nada além de um conjunto de dígitos compondo uma forma gráfica para a unidade central de processamento do computador (Alexander, 1997). De modo inverso, informa este mesmo autor ao relatar a experiência da British Library, a criação de um índice destas imagens geradas a partir de digitalizações de textos só se realiza com a reconversão das mesmas, o que está sendo feito com o auxílio da tecnologia de reconhecimento ótico de caracteres (OCR).

²² <http://webseer.cs.uchicago.edu>

Esta é apenas uma das muitas alternativas possíveis propostas. Borgman (1997) entende que em projetos desta natureza há que se considerar as muitas opções para descrever sons: com palavras, com outros sons (produzindo um tom e buscando outro similar, por exemplo); ou para descrever imagens: com palavras, desenhando um objeto semelhante ou selecionando um exemplar em alguma árvore.

Recursos sonoros como entrevistas ou palestras, por exemplo, em tempo real ou armazenados, ou ainda recursos audiovisuais da mesma natureza, exigem igualmente novas técnicas de recuperação que lhes permitam ser efetivamente localizados. Por ora inexistem soluções gerais para tratamento e recuperação de tais registros.

Weibel & Miller (1997) consideram que os fatores intervenientes na acessibilidade da imagem impostos pela limitação do desenvolvimento da tecnologia telemática, em virtude das penalidades de tempo/largura de banda para a recuperação de imagens serem sempre altas, tornam-se mais prementes do que os que venham a viabilizar sua recuperação efetiva.

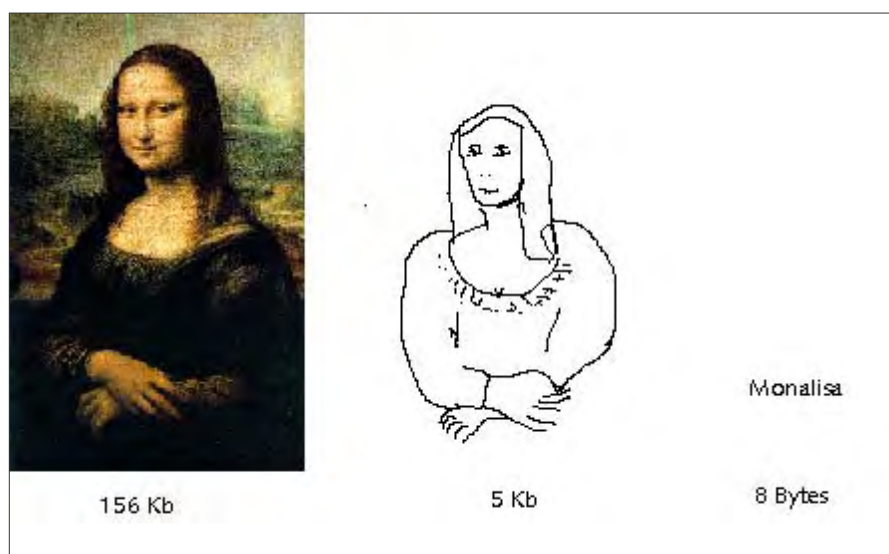


Figura 2: Comparação de tamanho entre fotografia digitalizada, gravura e texto

Óbvio que não se pode tirar-lhes a razão, como não se pode também deixar de imputar-lhes a contradição presente em seu enunciado. Dadas as reais limitações impostas à transferência de arquivos *pesados*, mais do que nunca é necessário desenvolver mecanismos que possam representá-las adequadamente e desenvolver-se *pari passu* com os avanços da telemática. A figura abaixo ilustra a diferença, em termos

de tamanho, de um texto, uma gravura e uma foto escaneada, tendo como elemento comum o mesmo referencial de informação.

Estes mesmos autores apostam ainda na descrição textual. Relatando o workshop sobre metadados para representação de imagens realizado pela Coalition for Networked Information (CNI) em conjunto com o Online Computer Library Center (OCLC), identificam como consensual a idéia de que as imagens não são tão diferentes do que chamam *objetos que se parecem com documentos (document-like-objects)* e que, por esta razão, a aplicação do padrão Dublin Core com algumas pequenas modificações presta-se satisfatoriamente para representá-las.

Por outro lado, destacam uma importante diferença: se com relação à informação de ordem textual os documentos e os elementos que os representam (metadados) não estão mais separados, o mesmo não se dá com a informação hipermídia.

A questão maior é que as diferenças entre o tipo de informação contido em uma determinada base e sua relação com a comunidade-alvo (no limite um relacionamento comum a todos os sistemas de informação) aumenta o grau de especificidade a ser exigido do padrão de representação. Por esta razão é que são rejeitados os modelos baseados na busca pela descrição textual. Um desenhista industrial pode, por exemplo, manifestar interesse em localizar uma dada imagem em função da composição de cores que apresenta; um ilustrador em função da figura mesma; um artista plástico em função da técnica adotada e assim indefinidamente.

Para representar uma imagem astronômica, por exemplo, o conjunto de metadados deve contemplar elementos como: o telescópio utilizado, a frequência de observação, a posição no céu e a descrição do objeto na imagem. Quem ilustra o fato é Plante (1997) do National Center for Supercomputing Applications da Universidade de Illinois, órgão que desenvolveu em conjunto com a NASA uma biblioteca digital de imagens astronômicas composta atualmente de cinco mil imagens em 13 Gb de dados, cuja missão é coletar imagens astronômicas em formato FITS (um formato padrão de imagens astronômicas) e disponibilizá-las à comunidade acadêmica ou ao público interessado via WWW.

A NCSA Astronomy Digital Image Library realiza, de uma certa maneira, o sonho projetado pelos projetos de ação cultural nas bibliotecas públicas, pois, além de possibilitar ao usuário a transferência de imagens ou de permitir buscas específicas em árvores de assunto, funciona como um local de armazenamento compartilhado das imagens astronômicas, dando ao usuário a vantagem de arquivar sua coleção de imagens (padronizadas) nesta base.

Qualquer pessoa que tenha o hábito de freqüentar bibliotecas virtuais percebe claramente que casos como o da citada biblioteca são antes exceções do que práticas comuns. A questão da informação hipermídia ainda não encontrou guarida nas bibliotecas virtuais. São raríssimas as que oferecem bancos de imagens ou de outro material que não seja o textual aos usuários. Acreditar que a biblioteca — não importando sua natureza — não precisa adquirir e/ou armazenar informação eletrônica implica em ignorar a obviedade de que alguma instituição precisa responsabilizar-se pela permanência e disponibilidade das informações geradas neste ambiente. Esta atribuição tem sido, ao longo da história, relacionada visceralmente às funções específicas da instituição biblioteca (Graham, 1995).

Com referência à migração de alguns documentos em mídia impressa para o formato hipermídia a coisa não é tão pacífica.

"A mudança de suporte implica tão-somente a preservação da informação e não a preservação física do documento na mídia original. O suporte digital resolve, portanto, apenas uma parte da questão e cria novos problemas: o da obsolescência das tecnologias de preservação, armazenamento e recuperação (hardware, software, sintaxe etc.) e, dada a facilidade de manipulação de dados pela mídia digital, o da autenticidade" (Levacov, 1997).

Se por um lado a deterioração física, inimiga universal da mídia impressa, já não assusta tanto, por outro, a obsolescência tecnológica dos dispositivos de leitura e armazenamento eletrônicos configura-se como o maior risco para a questão da preservação.

Uma biblioteca tradicional que disponibilize um banco de imagens ou de arquivos sonoros, ainda que tacitamente isto esteja dentre o que se possa enquadrar como pertencente aos serviços tradicionalmente por ela prestados, sempre foi considerada como especial. A própria coleção de material audiovisual que a compõe recebe

denominações específicas, tais como multimeios, material não-convencional, entre outras, termos que expressam sua especificidade em relação ao conceito de biblioteca tradicional e etimologicamente vinculado aos livros e material impresso em geral.

A biblioteca virtual, nesta acepção, é não-convencional por natureza, tem nisto uma característica intrínseca. A informação audiovisual (armazenada ou em tempo real) é elemento natural da composição do que se pode chamar, para preservar a analogia, seu acervo. Assim também a compreendem Bishop & Star (1996), os quais a definem como sistemas de informação e serviços que ofereçam documentos eletrônicos, incluindo arquivos textuais, som e vídeo digital, disponibilizados em repositórios dinâmicos ou de arquivos.

Cria-se uma demanda natural por estudos relacionados à recepção da imagem. Para além da característica puramente perceptiva, Jørgensen (1996) inclui outras que interferem no modo como o usuário irá representá-las e/ou acessá-las, como a cognição, afeição ou interpretação, como as que descrevem localização espacial, semântica ou características emocionais. No experimento que realizou, seis imagens foram projetadas por dois minutos cada e solicitado aos oitenta e dois participantes dispostos em dois grupos, que as descrevessem utilizando a linguagem natural. A um grupo foi solicitado que simplesmente descrevesse o que notavam na imagem; a um segundo solicitou-se descrever a imagem do modo como esperavam encontrá-la posteriormente numa coleção. Depois de quatro semanas solicitou-se do primeiro grupo que descrevesse o que se lembravam de cada imagem. Criou-se assim um conjunto de atributos, dispostos em três tipos principais: perceptivo (como cor), interpretativo (como estilo) e reativo (reações pessoais à imagem, como incerteza, confusão ou ligações com outros fatos). Não seria necessário mencionar a conclusão deste estudo, a não ser pelo fato de torná-la ainda mais óbvia e ululante: é necessário conhecer os hábitos do usuário.

Uma vez que a biblioteca virtual incorpora uma grande variedade de material ou de fontes digitais, é necessário conhecer como as pessoas recepcionam não somente textos ou imagens, mas também músicas e arquivos sonoros em geral, filmes de vídeo e outras mídias.

Os sistemas de mais fácil desenvolvimento são aqueles cujas aplicações e população de usuários estão claramente definidas. Estas são, contudo, condições raras

hoje; mais comum está se tornando o desenvolvimento de sistemas abertos que servem não somente a uma população local, mas também a populações remotas e até mesmo desconhecidas, cujas aplicações são impensáveis (Borgman, 1997).

O metadado textual é novamente combatido. Desta vez em razão de sua inadequabilidade para representar documentos em vídeo. Dada sua superabundância de informações, tanto de ordem visual (movimentos da câmera, mudanças de cena, cores) como de ordem auditiva (ruídos, silêncio, diálogos, músicas) é impossível vincular esta mídia ao tedioso e incompleto processo de descrição textual (Christel, 1995).

Por causa de seu tamanho, um vídeo pode levar um longo tempo para ser transferido de um local para outro. Recuperar 30 minutos completos de vídeo (algo imponderável para as condições atuais de configuração da Internet) quando somente um minuto é relevante é muito pior do que recuperar um livro inteiro quando somente um capítulo interessa. Isso em se tratando de uma transferência a partir de um repositório de arquivos; se a fonte de origem for um repositório dinâmico, este ato é não somente imponderável como impraticável pois, se para um texto o usuário/leitor estabelece sua própria taxa de leitura/processamento, o mesmo não é possível com relação aos documentos audiovisuais, cuja taxa de recepção é contínua, invariável em relação ao usuário.

A solução proposta por Christel (1995) é a partição do vídeo em pequenos cliques e sua representação alternada em parágrafos, como seqüências que dividem a mesma locação espacial.

A definição de critérios para a representação descritiva e temática da informação hipermídia com vistas a sua localização e acessibilidade, como se vê, não pode ser realizada sem o concurso da cuidadosa avaliação das especificidades das bibliotecas virtuais, da comunidade usuária (ainda que imaginária ou virtual), das características intrínsecas a cada mídia e do nível de progresso da telemática.

(IN)CONCLUSÕES

É praxe concluir um trabalho teórico revisitando e discutindo os pontos mais importantes do texto. Caso optasse por este expediente deparar-me-ia com a desatualização precoce do discurso, em função do rápido crescimento da área, e assim trôpego, esbarraria em muitas armadilhas, como a que o próprio título já apresenta. "Biblioteca tradicional X biblioteca virtual" enseja uma polarização que ao se "concluir" se despolariza no mesmo movimento. Não hesitei, contudo, em mantê-lo, pois, na verdade, ambos os modelos de recuperação da informação se complementam, quase se fundindo, e a contradição expressa pela oposição "versus" reflete a própria contradição histórica produzida por esta interpenetração que, como se fora um jogo, termina empatado, para o bem do público espectador.

Como se pôde perceber existem inúmeros pontos que aproximam e outros que distanciam as bibliotecas tradicionais das bibliotecas virtuais. Ainda não há consenso sobre quais elementos serão repetidos, quais serão objetos de transformações radicais para que se alojem no ambiente virtual de forma eficaz e quais serão simplesmente descartados. Há consenso, não obstante, na irreversibilidade da mudança paradigmática que envolve o trinômio organização, armazenamento e disseminação que regeu o

tratamento da informação até aqui. Como também há consenso sobre a requisitada experiência secular da Biblioteconomia para o tratamento da ampla gama de informações dispersa pela Internet, da qual não pode esquivar-se. A biblioteca, sabe-se, não é uma instituição que estabelece por si mesma sua missão — como, aliás, toda e qualquer instituição — à margem da sociedade; neste sentido é que a Biblioteconomia pode tornar-se mais próxima da sociedade do que o foi até agora.

A biblioteca tradicional não é o *alter ego* da biblioteca virtual ou vice-versa. Nem é a biblioteca virtual o modelo ideal para a instituição biblioteca, seria tolice supô-la, uma vez que a biblioteca é uma instituição prestadora de serviços valiosos pela configuração mesma que possui. Parece não restar dúvidas, entretanto, que doravante a biblioteca tradicional, sob pena de permitir que seja julgado à revelia sua tradição em prestar serviços informacionais, virtualizar-se-á em algum grau.

Biblioteca tradicional e biblioteca virtual estabelecem a partir de agora uma relação de mútua dependência, dialética. A criação ou manutenção de bibliotecas, da forma como foi dado ao mundo conhecê-las, depende do apelo inexorável dos novos serviços possibilitados pela biblioteca virtual, da mesma forma como esta, quiçá pagando o preço de carregar a carga etimológica do vocábulo biblioteca, não pode ser pensada sem o conjunto de tecnologias tipificados pelas bibliotecas tradicionais.

Muitas bibliotecas já dão mostras dessa fusão, seja disponibilizando seus catálogos em rede (passo inicial já dado por muitas instituições brasileiras), seja criando serviços especificamente eletrônicos como, por exemplo, o balcão de referência virtual e seus serviços de disseminação da informação mediados pelo correio eletrônico ou por listas de discussão.

O modelo de interação face-a-face da biblioteca tradicional e sua característica predominantemente oral funde-se paulatinamente com a criação de uma nova categoria de comunicação — a qual já representa uma fusão — possibilitada pelo ambiente eletrônico, qual seja a da escrita oralizada ou da oralidade escrita. Em outros termos, a biblioteca continua a interagir com o usuário por meio da escrita e da oralidade, mas o faz com a vantagem da interação escrita oralizada em tempo real e da quebra de barreiras geográficas, tornando-se, por esta razão, potencialmente acessível a um maior número de pessoas.

A potencialidade de atendimento a uma comunidade maior faz ecoar naturalmente a utopia da democracia informacional. Claro que ainda não se a alcança, mas, que não sirva de consolo, se a atualiza, tornando-a uma utopia pós-moderna. Assim como a biblioteca tradicional e o mundo escrito de forma geral excluiu — e ainda exclui — os analfabetos, a biblioteca virtual e o ambiente virtual exclui, necessariamente, os não alfabetizados em assuntos de informática.

Embora possa parecer irônica a proposição acima, na verdade, não se admite mais a biblioteca ausente da Rede, pois, ao fazê-lo, distancia-se do usuário (não o contrário) e arrisca-se a ser adjetivada com o antigo conceito de virtualidade, no que este tem de potencial e sem relação direta com o real, inacessível em última instância.

Com o desenvolvimento da dissertação tornaram-se evidentes alguns obstáculos à plena consolidação das bibliotecas virtuais:

- a questão do direito autoral ou das relações sociais de propriedade ainda impossibilita o perfeito fluxo de grande parte dos novos nutrientes informacionais;
- urge a definição de padrões de representação das fontes eletrônicas (metadados) e de metodologias para o desenvolvimento de bibliotecas virtuais capazes de lidar com a diversidade;
- ainda não estão firmes as políticas de compartilhamento de recursos, de custos, de arquitetura de sistemas distribuídos, tão necessárias ao ambiente de rede e
- mesmo que todas as questões acima estivessem resolvidas hoje, a biblioteca virtual ainda esbarraria no atual estágio de desenvolvimento da telemática e suas conseqüentes limitações por ora impostas.

Como se vê, existe uma montanha de desafios a ser escalada. Não basta, nem é adequado, entretanto, assumir o comportamento do neófito afoito para escalá-la. Para lançar-se à tarefa de forma eficaz é necessário equipar-se e aprimorar-se nas técnicas adequadas. Antes que me perca em metáforas, o que quero dizer é o seguinte, os desafios estão dados e é inócuo assumir qualquer postura de rejeição ou de

ressentimento, necessário, sim, é acomodar-se a nova ordem e permitir-se ser modificado por ela, para que possa também modificá-la.

A nova biblioteca que surge desse casamento entre tradição e modernidade (ou virtualidade) depara-se com um novo ambiente interacional, repleto de novos materiais informacionais clamando por novas técnicas de recuperação, e tem sua eficácia condicionada ao desenvolvimento da telemática. Tudo isto aguça o crescimento desta instituição e da profissão que se lhe relaciona, pois, em que pesem todas as contradições, incertezas e turbulências do quadro atual, há uma certeza alentadora: ainda é necessário agregar valor à informação.

REFERÊNCIAS

ALEXANDER, Michael Retrieving digital data with fuzzy matching **Management Information**, v. 4, n. 1/2, p. 34-37, jan./feb., 1997

ARAÚJO, Vânia Maria R. H. de **Sistemas de recuperação da informação: nova abordagem teórico-conceitual** Rio de Janeiro: UFRJ, 1994 (Tese de Doutorado)

BARKER, Phillip Electronic libraries: visions of the future **The Electronic Library**, v. 12, n. 4, p. 221, aug. 1994

BARRETO, Aldo de Albuquerque A questão da informação **São Paulo em Perspectiva**, v. 8, n. 4, p. 3-8, 1994

_____ A transferência de informação e as tecnologias intensivas: reposicionamentos **Informare**, v. 2, n. 2, p. 50-52, jul./dez. 1996

BELKIN, Nicholas J. Anomalous state of knowledge as a basis for information retrieval **Canadian Journal of Information Science**, v. 5, p. 133-140, 1980

- BISHOP, Ann Peterson & STAR, Susan Leigh Social informatics of digital library use and infrastructure **Annual Review of Information Science and Technology**, v. 31, p. 301-401, 1996
- BORGMAN, Christine L. Multi-media, multi-cultural and multi-lingual digital libraries: or how do we exchange data in 400 languages? [on-line] URL <http://www.dlib.org/dlib/june97/06borgman.html> 1997
- BUSH, Vannevar As we may think **The Atlantic Monthly**, v. 176, n. 1, p. 101-108, 1945. [on-line] URL <http://www.isg.sfu.ca/~duchier/misc/vbush/vbush-all-shtml>
- BUTTERFIELD, Kevin L. Cataloger's and the creation of metadata systems: a collaborative vision at University of Michigan [on-line] URL <http://www.oclc.org/oclc/man/colloq/butter.html> [Arquivo capturado em 1997]
- CADERNOS DE PROBLEMAS BRASILEIROS A democracia eletrônica São Paulo, n. 319, jan./fev. 1997 23 p. Encarte especial.
- CAMARGO, Emília da Conceição Navegar na Net: conceito impar em todo fragmentado **Trans-in-formação**, Campinas, v. 8, n. 2, p. 57-62, mai./ago. 1996
- CAPLAN, Priscilla "You call it corn, we call it syntax-independent metadata for document-like objects" **The Public Access Computer Systems Review**, v. 6, n. 4, 1995 [on-line] URL http://www.oclc.org:5406/oclc/research/conferences/metadata/dublin_core_report.html
- CHRISTEL, Michael G. Addressing the contents of video in a digital library [on-line] URL <http://www.cs.tufts.edu/~isabel/christel/christel.html> 1995
- COSTA, Icléia Thiesen Magalhães Memória institucional e representação: do mundo das formas (árvore) ao universo do pensamento (rizoma) **Informare**, v. 2, n. 2, p. 67-72, jul./dez., 1996
- COVE, J. F. & WALSH, B. C. Online text retrieval via browsing **Information Processing & Management**, v. 24, n. 1, p. 31-37, 1988

- CRONIN, Blaise & HERT, Carol A. Scholarly foraging and network discovery tools
Journal of Documentation, v. 51, n. 4, p. 388-403, dec. 1995
- DESAI, Bipin C. Supporting discovery in virtual libraries **Journal of The American Society for Information Science**, v. 48, n. 3, p. 190-204, 1997
- DUBOIS, Didier & PRADE, Henri **Fuzzy sets and systems: theory and applications**
San Diego: Academic Press, 1980 ©.
- EAGAN, Ann & BENDER, Laura Spiders and worms and crawlers, oh my: searching
on the world wide web [on-line] URL
<http://www.library.ucsb.edu/untangle/eagan.html> 1996
- EAGAN, Ann & BENDER, Laura Spiders and worms and crawlers, oh my: searching
on the World Wide Web. [on-line] URL : <http://www.library.ucsb.edu/untangle/eagan.html> 1996
- ELECTRONIC INFORMATION GROUP Research issues: a discussion of Internet
information retrieval [on-line] URL http://ftp.tamu.edu/global_info/research-issues.html 1995
- FELDMAN, Susan "Just the answers, please": choosing a web search service [on-line]
URL <http://www.infotoday.com/searcher/may/story3.htm> 1997
- FRANCO, Marcelo Araújo **Ensaio sobre as tecnologias digitais da inteligências**
Campinas: Papirus, 1997
- GIRILL, T. R. & LUK, Clement H. Fuzzy matching as a retrieval enabling technique
for digital libraries [on-line] URL: <http://www.asis.org/midyear-96/girillpaper.html> 1996
- GRAHAM, Peter S. The digital research library: tasks and commitments [on-line]
URL <http://csdl.tamu.edu/DL95/papers/graham/graham.html> 1995
- JÖRGENSEN, Corinne Indexing images: testing an image description template [on-
line] URL <http://www.asis.org/annual-96/ElectronicProceedings/jorgensen.html>
1996

- KELLY, Julia & ROBBINS, Kathryn Changing roles for reference librarians **Journal of Library Administration**, v. 22, n. 2/3, p. 111-121, 1996
- KOCH, Traugott Internet search services [on-line] URL <http://www.ub2.lu.se/tk/demos/DO9603-meng.html> 1996
- KOCHEN, M. **Principles of information retrieval** Los Angeles: Melville, 1974
- KOOPMAN, Ann & HAY, Sharon Swim at your own risk — no librarian on duty: large-scale application of Mosaic in an academic library [on-line] URL <http://www.ncsa.uiuc.edu/SDG/IT94/Proceedings/LibApps/hay/WWWPap.html> 1994
- LAGOZE, Carl From static to dynamic surrogates: resource discovery in the digital age [on-line] URL: <http://www.dlib.org/dlib/june97/06lagoze.html> 1997
- LALANDE, André **Vocabulaire technique et critique de la philosophie** Press Universitaire de France, 1951. p. 1106
- LANCASTER, F. W. **The measurement and evaluation of library services** Washington: Information Resource Press, 1977
- LARSON, Ray R. Search capabilities for users of digital libraries: tools and paradigms [on-line] URL: http://sherlock.berkeley.edu/ucla_dl_workshop/node1.html 1996
- LEIGHTON, H. Vernon & SRIVASTAVA, Jaideep Precision among world wide web search services (sistemas de busca): Alta Vista, Excite, Hotbot, Infoseek, Lycos [on-line] URL <http://www.winona.msus.edu/is-f/library-f/webind2/webind2.htm> 1997
- LESK, Michael Why digital libraries? [on-line] URL: http://www.ukoln.ac.uk/follett/lesk/why_digital_libraries.html 1995
- _____ The seven ages of information retrieval [on-line] URL : <http://community.bellcore.com/lesk/ages/ages.html> [Arquivo capturado em 1997]
- LEVACOV, Marília Bibliotecas virtuais: (r)evolução? **Ciência da Informação**, Brasília, v. 26, n. 2, p. 125-135, mai./ago. 1997

- LEVY, David M. & MARSHALL, Catherine C. **What color was George Washington's white horse?: a look at assumptions underlying digital libraries** [on-line] URL <http://www.cSDL.tamu.edu/DL94/paper/levy.html> 1994
- LEVY, David M. Cataloging in the digital order [on-line] URL <http://cSDL.tamu.edu/DL95/papers/levy/levy.html> 1995
- LEVY, Pierre **As tecnologias da inteligência: o futuro do pensamento na era da informática** São Paulo: 34, 1993
- _____. **O que é o virtual?** São Paulo: 34, 1996
- LOJKINE, Jean **A Revolução Informacional** São Paulo: Cortez, 1995.
- LOSEE, Robert M. Comparing boolean and probabilistic information retrieval systems across queries and disciplines **Journal of The American Society for Information Science**, v. 48, n. 2, p. 143-156, 1997
- MACHADO, Arlindo **Ensaio sobre a contemporaneidade** 1995 (em disquete)
- MARCHIORI, Patrícia Zeni Acessar ou possuir, eis a questão ... In: IX SEMINÁRIO NACIONAL DE BIBLIOTECAS UNIVERSITÁRIAS, 1996, Curitiba. **Anais ...** Curitiba: UFPR/PUCPR, 1996
- _____. "Ciberteca" ou biblioteca virtual: uma perspectiva de gerenciamento de recursos de informação **Ciência da Informação**, Brasília, v. 26, n. 2, p. 115-124, mai./ago. 1997
- MARX, Karl Processo de trabalho e processo de valorização In: _____. **O Capital: crítica da economia política** 3.ed. São Paulo: Nova Cultural, 1988. p. 142- 156 (Os Economistas)
- MORSE, P. M. Browsing and search theory In: RAWSKI, C. H. **Toward a theory of librarianship** New Jersey: The Scarecrow Press, 1973. p. 246-261
- MOSTAFA, Solange Puntel Biblioteconomia e história: uma abordagem dialética **Revista Brasileira de Biblioteconomia e Documentação**, v. 14, n. 1/2, p. 47-51, jan./jun. 1981

- NEGROPONTE, Nicholas **A vida digital** 2.ed. São Paulo: Companhia das Letras, 1995
- NOTESS, Greg R. On the net: Internet search techniques and strategies [on-line]
URL: <http://www.onlineinc.com/onlinemag/julOL97/net7.html> 1997
- NÜRNBERG, Peter J. et al. Digital libraries: issues and architectures [on-line] URL :
<http://csdl.tamu.edu/DL95/papers/nuernberg.html> 1995
- PAIJMANS, J. J. Models in information retrieval [on-line] URL:
<http://pi0959.kub.nl:2080/Paai/Onderw/Ir-models/ir-models.html> 1997
- PEREIRA, Maria de Nazaré Freitas Bibliotecas virtuais: realidade, possibilidade ou alvo de sonho? **Ciência da Informação**, Brasília, v. 24, n. 1, p. 101-109, 1995
- PINKERTON, Brian Finding what people want: experiences with the WebCrawler [on-line] URL
<http://www.nscs.uiuc.edu/SDG/IT94/Proceedings/Searching/pinkerton/WebCrawler.html> 1994
- PIROLI, Peter & CARD, Stuart Information foraging in information access environments [on-line] URL
<http://www.acm.org/sigchi/chi95/proceedings/papers/ppp-bdy.htm> 1995
- PLANTE, Raymond L. The NCSA Astronomy Digital Image Library: the challenges of the scientific data library [on-line] URL
<http://www.dlib.org/dlib/october97/adil/10plante.html> 1997
- RADECKI, Tadeuz Fuzzy set theoretical approach to document retrieval **Information Processing & Management**, v. 15, p. 247-259, 1979
- RIDLEY, Michael Innovation and implementation: adopting and managing World Wide Web services in academic libraries [on-line] URL
<http://www.nscs.uiuc.edu/SDG/IT94/Proceedings/LibApps/ridley/ridley.html> 1994
- RIJSBERGEN, C. J. Van **Information retrieval** 2.ed. London: Butterworths, 1979

- RODRIGUES, Adriano Duarte **Comunicação e cultura**: a experiência cultural na era da informática Lisboa: Presença, 1994.
- ROWLEY, Jennifer **Informática para bibliotecas** Brasília: Briquet de Lemos/Livros, 1994
- SANDSTROM, P.E. An optimal foraging approach to information seeking and use **Library Quaterly**, v. 64, n. 4, p. 414-449, 1994
- SAUNDERS, L. M. The virtual library today **Library Administration and Management**, v. 6, p. 66-70 1992
- SHAW, Russel Crawlers, spiders and worms [on-line] URL <http://www.webweek.com/95June/news/crawlers.html> 1995
- SMITH, Alastair G. Librarians and the web: a report on a study tour **LASIE**, v. 25, n. 1/2, p. 4-15, jul./oct. 1994 [on-line] URL <http://www.vuw.ac.nz/~agsmith/libnweb.html>
- SOLOCK, Jack Searching the Internet: part I: some basic considerations and automated search indexes [on-line] URL: <http://rs.internic.net/nic-support/nicnews/archive/september96/enduser.html> 1996.
- STALKER, John & MURFIN, Marjorie E. Frequently asked questions: an effective way to store and retrieve reference information? **Reference Services Review**, v. 24, n. 4, p. 31-40, winter, 1996
- STANLEY, Tracy The great search engine debate [on-line] URL <http://www.ariadne.ac.uk/issue2/engines> 1996
- UM BIBLIOTECÁRIO à sua disposição **Informática Exame**, p. 80-82, abr. 1996
- VILAN FILHO, Jayme Leiro Hipertexto: visão geral de uma nova tecnologia de informação **Ciência da Informação**, Brasília, v. 23, n. 3, p. 295-308, set./dez. 1994
- WEBSTER, Kathleen & PAUL, Kathryn Beyond surfing: tools and techniques for searching the web [on-line] URL <http://magi.com/~mmelick/it96jan.htm> 1996

- WEIBEL, Stuart & MILLER, Eric Image description on the Internet: a summary of the CNI/OCLC Image Metadata Workshop [on-line] URL <http://www.dlib.org/dlib/january97/oclc/01weibel.html> 1997
- WOODWARD, Jeannette Cataloging and classifying information resources on the Internet **Annual Review of Information Science and Technology**, v. 31, p. 189-220, 1996
- XU, Amanda Accessing information on the Internet: feasibility study of USMARC format and AACR₂ [on-line] URL <http://www.oclc.org/man/colloq/xu.html> [Arquivo capturado em 1997]