

CAVEAT WEB SURFER! RESPONSABILIDADE SOCIAL E RECURSOS DA INTERNET*

Thomas J. FROELICH*
tfroehli@slis.kent.edu

INTRODUÇÃO

Eu não sou *Luddite*, contrário às novas tecnologias ou às mudanças tecnológicas, mas tenho muitas questões informacionais, éticas e legais para suscitar em relação ao endosso irrefletido e maciço do admirável mundo novo da Infra-estrutura Global (GI) da Informação e em relação a essa ideologia escondida de acesso em larga escala à Internet e ao outro lado da superdevoção à tecnologia. Dada a grande variedade de provedores de serviços na Internet, fornecedores de programas do tipo Web Browsers, produtores de computador e varejistas, seria fácil para as pessoas serem levadas a acreditar que as promessas da era da informação pela Internet se cumpriram: o acesso generalizado gratuito ou barato à informação ao alcance de todos. Um pesquisador de informação iria simplesmente ligar o seu computador, montar o seu *web browser* como Netscape, apontar e clicar em direção ao caminho da informação nirvana. Tais visões utópicas podem, na verdade, sinalizar pesadelos. Os pesadelos

(*) Tradução concedida à Transinformação por T. Hogan Jr. from Information Today Inc. em 02/10/97.

(**) Professor Titular da School of Libray & Information Science da Kent State University, Ohio, USA.

incluem: oferta excessiva de dados, informação enganosa, falta de informação ou desinformação, sobrecarga de informação, inadequada falta de precisão e revocação, ausência de autoridade cognitiva, falta de acesso público ou uso educacional gratuito, responsabilidade legal dos provedores com respeito à violação dos direitos autorais na Internet, o desperdício de recursos, etc. *Caveat emptor* é um princípio comercial de que, sem uma garantia, o comprador assume para si mesmo o risco sobre a qualidade do produto. Na Internet muitos dos recursos não são comerciais, mas será que isso significa a ausência de responsabilidade dos navegadores da Internet com relação à natureza e qualidade das informações por eles obtidas? E no caso dos recursos comerciais, qual é a responsabilidade do produtor, se é que há alguma? Deveríamos declarar, assim como no mercado comercial: *Caveat Web Surfer!* Enquanto for verdade que não podemos e não devemos controlar os processos do pensamento de uma pessoa, os profissionais da informação, que seguem conscientemente os ditames da responsabilidade social, deveriam se esforçar em desenvolver instituições e estruturas que facilitassem a consciência crítica em relação à informação que os usuários finais podem obter enquanto navegam pela supervia da informação. Mas antes de sugerir possíveis soluções, a natureza dos problemas deve primeiramente ser compreendida.

2. BUSCA DE INFORMAÇÃO NÃO-MEDIADA: QUAL É O PROBLEMA?

2.1 Oferta Excessiva de Dados, Excesso e Falta de informação

Com o crescimento explosivo de padrões da Internet, está cada vez mais difícil achar informação qualitativa. Há oferta excessiva de informações, sobrecarga de informações, crescente quantidade de publicidade ligada a interesses comerciais, etc. Uma forma de se começar a enxergar o problema é contrastá-lo com alguns dos mais

úteis aspectos das bibliotecas públicas e acadêmicas. Cada biblioteca tem uma política de desenvolvimento de coleções: isso significa que a biblioteca desenvolve um plano específico que indica os tipos de materiais que ela irá e não irá colecionar. Existem certas diretrizes para essas políticas de desenvolvimento de coleções. Para se afirmar em termos do jargão corrente, é preciso estruturar estratégias de metadados a fim de facilitar as necessidades dos usuários. Dentre os objetivos estabelecidos em uma biblioteca, a coleção deveria ser completa, equilibrada e representativa. Ser completa, quer dizer: deveria obter tantos materiais quanto pudesse pagar. Deveria ser equilibrada: nenhum cargo, ponto de vista ou ideologia deveria ser expressa com peso indevido. Deveria ser representativa: colecionar uma diversidade de materiais representando a variedade de pontos de vista sobre qualquer questão ou temática, ou pelo menos, nas principais delas. É óbvio, dado que a Internet não possui uma autoridade central, que não existem políticas para coleções. Qualquer pessoa com vontade, com um pouco de experiência (ou dinheiro) e com acesso tecnológico pode montar a sua própria *Homepage* com qualquer coisa que queira anunciar, promover ou tornar disponível. Localmente, podem existir Políticas sobre o Uso do Computador (CUPs-Computer Use Policies) que restrinjam jogos, obscenidades, chamadas de paixão, etc., mas tais políticas não refletem ou não controlam o que está a disposição, mas somente o que está legitimado ou legalmente acessível em um dado *site*.

Se é que se pode falar em uma coleção de materiais na Internet, esta é na verdade uma anti-coleção: é uma miscelânea de itens, surgindo numa diversidade de formas, com pouca ou nenhuma autoridade ou controle, com pouca organização global, mecanismos de busca bem pobres, e é daí portanto, que pousa a maior dificuldade em achar itens de informação de valor e utilidade. O paradoxo é que com a proliferação de materiais, há uma extraordinária carência de informação na Internet. Para a maior parte, não se encontra disponível material com direitos autorais (a não ser que você tenha conta com um vendedor, que cobrando poderá acessá-la). Em geral, os problemas de desigualdade no acesso à informação são mais agravados do que

atenuados: os ricos têm acesso às informações e podem pagar os custos de informações cobradas. Os pobres não podem ter esse acesso e têm que superar vários níveis de dificuldades: pagar as tecnologias básicas (PCs, networks, etc.), assim como os direitos autorais das informações.

2.2 Suposições Problemáticas dos Usuários Finais, Acesso Problemático à Informação e Procedimentos dos Tomadores de Decisão

Têm havido e sempre haverá para todos nós, não importa qual o nosso nível de experiência, falta de senso crítico em diferentes graus e capacidades. Tais capacidades críticas deveriam se desenvolver conforme nos tornamos maduros. Ainda assim, dados certos contextos - principalmente o desenvolvimento de compensação imediata na vida americana, tal crescimento e tal tomada de consciência parecem estar frustados. Usuários finais tendem a fazer muitas suposições falsas a respeito da natureza, qualidade e compreensibilidade da informação disponível em computadores em geral e na Internet em particular. Do mesmo modo que alguns leitores de jornais caem como presas numa visão "factóide" da informação (i.e., se alguma coisa está impressa, tem que ser verdadeira; originalmente o termo "factóide" significava algo fictício apresentado como fato, sempre como a conseqüência de uma repetição contínua), usuários de computador parecem cair como presas na visão "tecnofactóide" de recursos de computador: se algo aparece em algum aparelho tecnológico, tem que ser verdade. Por qualquer que seja a razão, muitos usuários tendem a ser menos críticos com relação às informações que recebem em computadores, e a repetição e facilidade com que os dados vagueiam pela Internet tornam os buscadores de informação presas fáceis do investimento de credibilidade em recursos que podem não merecê-la. Além disso, os usuários finais sempre atribuem aos recursos do computador e da Internet uma compreensão, precisão e confiança que eles não possuem. Eles imputam uma autoridade cognitiva, a qual na realidade eles podem não ser capazes de averiguar. Supondo que eles não estejam explicitamente conscien-

tes da autoridade de um trabalho de referência em uma biblioteca, o contexto real desse ambiente adequa-se às expectativas. Não se pode dizer o mesmo dos recursos da Internet. Os usuários tendem também a não entender a natureza dos instrumentos de busca ou mecanismos de busca disponíveis na Internet, suas qualidades e, mais importante, suas deficiências e dificuldades que precisam ser dominadas para que se possa usar efetivamente esses mecanismos. Os usuários possuem pouco ou nenhum entendimento sobre como se obter maior precisão ou maior revocação ou de que tais demandas são possíveis nas buscas de informação.

Em adição ou em relação a essas suposições problemáticas dos usuários, existem também algumas questões relacionadas à psicologia e ao comportamento dos pesquisadores de informação. Há uma literatura significativa sobre como as pessoas usam mal e abusam do acesso à informação, ao processamento e às fontes, ações estas tendem a aumentar, dada a complexidade e a diversidade de informações disponíveis na rede. Paul C. Nutt (1) em *Making Tough Decisions: Tactics for Improving Managerial Decision Making* trata de muitas dessas dificuldades, que deverão agravar-se no campo da disponibilidade de informação na Internet. Enquanto se reconhece que o foco de seu trabalho esteja nas decisões mais difíceis, alguns dos pontos principais se generalizam em tomadas de decisões ou busca de informações na Internet. Tomadores de decisão e pesquisadores de informação reconhecem a informação seletivamente e, tendem a dar muita importância às informações, que estão imediatamente prontas. Eles cometem enganos ao identificar e pesar as informações pois (1) é difícil estimar precisamente a frequência dos acontecimentos, (2) esses pesquisadores dão muita importância a acontecimentos com os quais estão mais familiarizados, (3) enfatizam informação consistente com experiências vividas por eles, (4) são atraídos pelas informações mais significativas e afastam-se das informações pouco expressivas e, (5) há efeitos de ordem: a informação que é inicialmente reconhecida tem mais valor do que a reconhecida posteriormente (p. 69). É importante perguntar, dada a natureza do hipertexto de alguns recursos da Internet, se tais efeitos de ordem

não são mais agravados. Além do mais, o pesquisador de informação naturalmente tenta simplificar e reduzir esforços ao processar a informação, mas poderá cair em muitos erros ao fazer tal coisa: (1) a disponibilidade heurística, pela qual o tomador de decisão coloca um relacionamento causal entre duas coisas, pois este é marcante em situações onde muitos relacionamentos plausíveis entre coisas são possíveis; (2) a heurística representativa na qual os tomadores de decisão estabelecem uma decisão *a priori* para guiar uma escolha em curso, por exemplo, na opção do que poderia constituir os recursos da Internet a serem consultados; (3) heurística estereotipada, na qual o tomador de decisão através da supergeneralização estereotipa o comportamento, baseado, em respostas artificiais para determinadas situações: por ex. "Money Talks"; (4) uma relutância em ajustar avaliações preliminares baseadas em novas informações; informação significativa é provavelmente mais fácil de ser lembrada do que informação sutil, etc. O aumento de informação disponível na Internet parece que impulsiona os pesquisadores de informação, especialmente como tomadores de decisão, a recorrer mais prontamente a essas tendências e a esses processos errôneos para lidar não somente com o fôlego da informação disponível, mas também com as técnicas para alcançá-las.

2.2 Acesso Direto: Busca Não-Mediada

Assim como os vendedores, os produtores e os acionistas da supervia da informação gostam sempre de falar que há enormes vantagens no acesso direto à informação. Dada a lei do menor esforço, as pessoas preferem usar recursos próprios disponíveis em computadores pessoais ou de trabalho. Mesmo com todas essas tecnologias, existem aspectos indesejáveis nestas aparentes boas notícias. A ausência de quaisquer intermediários da informação, como bibliotecários ou especialistas em informação significa que qualquer habilidade interfere no acesso dos usuários finais, no uso ou no entendimento da informação torna-se não-existente ou reduzido. Não há meios de

se alertar os usuários quanto à qualidade da informação ou a sua fonte, ou se fizeram a busca correta ou usaram os mecanismos certos de busca ou se obtiveram resultados apropriados para a informação que procuravam.

Há no mínimo quatro áreas onde o acesso direto ou a busca não-mediada de informação propõe problemas: (a) dificuldade em achar informação útil e de saber o que não está disponível; (b) necessidade de alta revocação; (c) necessidade de alta precisão; (d) e as necessidades de avaliação da informação: julgar sua credibilidade.

Há muita dificuldade em se encontrar informação útil na Internet, e com certeza, os milhares de catálogos de biblioteca aparecem como um bom recurso, porém é necessário decidir qual é o mais apropriado para se usar (HYTELNET pode ser parcialmente útil em relação a esta questão) deve-se usar, então, qualquer mecanismo de busca que esteja ligado a aquela interface. Há uma grande variedade de tipos de catálogos e de formas de encontrá-los. Mesmo quando se acha uma referência a uma fonte, deve-se determinar como obtê-la localmente, rapidamente, convenientemente e de forma barata. O padrão Z.39.50 foi estabelecido para oferecer uma interface comum e muitos provedores de informação, incluindo bibliotecas, procuraram implementar tal interface para seus produtos. Esse padrão representa ao mesmo tempo boa notícia e má notícia. A boa notícia é que haverá uma interface comum com usuários que, se dispenderem tempo e estudo suficientes, poderão se tornar familiarizados com ela e que lhes permitirá encontrar bons recursos. Porém, para serem verdadeiramente efetivos e eficientes, os usuários devem ser capazes de entender os vários campos de um registro bibliográfico para que possam compreender o sistema de classificação do catálogo (por exemplo, classificação Decimal de Dewey ou a classificação da Biblioteca do Congresso) para compensar os vários erros de dados em tais catálogos. Por exemplo, Ohiolink contém o mesmo registro em duas diferentes formas: W. Forest Horton não escreveu dois livros, *Information Resources Management - Concepts and Cases* e *Information Resources Management: Concepts and Cases*, mas o software, ao criar o catálogo coletivo não identificou a natureza comum

de ambas as entradas. A má notícia a respeito de tais padrões como o Z39.50 é que eles são também "enfraquecedores": a maior parte da eficiência de muitas interfaces dos catálogos se perde quando eles são feitos para se adequar a um padrão comum, pois o padrão é freqüentemente ajustado para as menores capacidades do conjunto de softwares de busca ou das estruturas das bases de dados existentes e índices. Por exemplo, a habilidade de se limitar facilmente uma busca a um determinado campo ou à uma certa extensão não pode ser facilmente acomodada por tais padrões. Tais padrões facilitam os novos usuários até certo ponto, mas tendem a frustrar os usuários experientes e as buscas que visam alta precisão.

As ferramentas de busca da Internet, embora úteis para certos tipos de tarefas, são geralmente desalentadoras para quem busca alta precisão ou revocação adequada. Gophers e Veronica, de certa forma em declínio, provêem acesso ao *gopherspace*, a aquilo que conhecem a respeito ou a que estão conectados. Além disso, a busca está restrita aos nomes de arquivos para a maior parte (raramente bons indicadores do conteúdo intelectual de um arquivo) às vezes a palavras-chave, e raramente, ao texto todo. Não há, portanto, possibilidade de controlar ou classificar a natureza do resultado.

Arquivos são úteis também, principalmente para localizar software mas, não documentos. Novamente a busca é, em primeiro lugar restrita aos nomes de arquivos e diretórios - pobres indicadores da natureza do software (a não ser que se saiba o que se procura) ou da complexidade intelectual do documento. De novo, não há controle sobre o resultado, a não ser que se domine as excentricidades do arquivo de software, e que não haja classificação do resultado. O resultado dessa busca de um arquivo é uma listagem dos nomes de arquivos e locais deles (muitos dos quais não permitem *login* anônimo e que se tornam, por isso, inúteis aos pesquisadores de software), certamente útil se se souber o valor de um arquivo que se quizesse fazer um FTP. Contudo, o conteúdo intelectual de um arquivo ou a eficácia de um software (às vezes até mesmo a versão de um *software*) é quase impossível de se determinar.

Com a tecnologia WAIS dois aperfeiçoamentos ocorreram. Geralmente, essa tecnologia utiliza indexação de todo o texto, às vezes com palavras-chave dos autores, e ela faz a contagem de frequência de termos dos termos em documentos. Devido a esses cálculos de frequência, o resultado é passível de classificação. Mas existem problemas associados a isso, problemas que surgem em todos os mecanismos atuais de busca. Ninguém nunca sabe, com exceção de casos específicos ou muito simples, como a classificação ocorre, tanto em termos do que foi usado para criar a indexação em um recurso específico ou conjunto de recursos (e.g., texto completo, palavra-chave ou uma combinação) e nem quanto ao peso atribuído aos termos em questionamentos mais complexos. Por exemplo, se o usuário está procurando por três termos em um documento, podem ser classificadas em altas frequências de dois termos e, as baixas, de um termo melhor do que as frequências de todos os três termos? O paradoxo é que tais algoritmos sempre fazem parte dos "negócios secretos" de um determinado provedor de recursos, e ainda assim tal sigilo certamente impede que os usuários façam uma escolha apropriada acerca da adequação de um determinado mecanismo de busca para um determinado caso específico. Além do mais, será que o mesmo padrão se aplica a todas as buscas? Se é este o caso, isto é bem problemático, já que os diferentes tipos de busca poderão requerer diferentes pesos para os termos. No esforço de transformar a busca mais fácil para os pesquisadores de informação, tais tecnologias fomentam tendências com as quais os usuários, sem fazer uma avaliação concordam.

Com o advento de mecanismos de busca mais sofisticados, tais como o *InfoSeek* e o *Lycos*, questionamentos mais complexos e controlados podem ser construídos, mas estes também trazem outra série de problemas, e não resolvem a questão de como o usuário pode vir a compreender a forma pela qual suas indagações pessoais foram propostas e classificadas em uma base de dados de recursos. Antes de mais nada, os mecanismos de busca são confusos para os usuários finais, pois variam consideravelmente entre como devem ser usados e suas reais capacidades. Alguns mecanismos permitem a

utilização dos operadores Boolean (*InfoSeek*), outros têm operadores posicionais (próximos ou seguidos por - *Open Text*), alguns fazem truncagem automática (*Lycos*) e outros não fazem nada, e o uso deles, de mecanismo para mecanismo, não é consistente. Se você utiliza duas palavras adjacentes em um sistema, ele é tratado como se as palavras fossem O Red (a não ser que as aspas sejam usadas) e em outro é tratado como termos adjacentes (na mesma ordem). O *Open Text* é indiferente à situação, o *InfoSeek* não o é.

Talvez o mais problemático seja o que esses mecanismos usam com base de busca. Por exemplo, o *InfoSeek* utiliza o URL, as referências de URL e o texto completo (os primeiros 50 K do texto). *Lycos* usa URL e as referências de URL, título, cabeçalhos, vinte linhas de texto, 100 palavras de peso. *Open Text* usa URL, referências de URL, texto completo (incluindo palavras não significativas). Há vários problemas relacionados à utilização de tais termos para indexação: URLs, como os títulos de muitos livros e artigos são sempre indicadores pobres do conteúdo intelectual do artigo - nas ciências exatas eles podem ser os melhores indicadores, mas em ciências sociais ou artes e humanidades, eles não o são. E se se usar arbitrariamente as primeiras 1000 palavras do texto, um outro tipo de problema será introduzido: para alguns artigos e literaturas, as primeiras 1000 palavras podem ser as mais significativas (por exemplo, o parágrafo principal na base de dados de jornais); para outros, pode ser que as últimas 1000 palavras (por exemplo conclusão de artigos científicos) . Por isso, quando um padrão arbitrário é aplicado pelo instrumento de busca, pode causar mais problemas do que soluções. Mesmo se um texto completo for incluído para indexação, o pesquisador de informação ou usuário final é vítima da linguagem utilizada pelo autor, a qual pode ou não se adequar à escolha do pesquisador quanto aos termos ou quanto àqueles da literatura sobre determinado assunto. Por isso, tais mecanismos de busca não resolvem os problemas associados à busca do texto completo, principalmente na ausência de um vocabulário controlado. Mesmo o resultado alcançado por esses mecanismos de busca é problemático. O *InfoSeek* faz a classificação de relevância pelo local de frequência da palavra. Obviamente, é

impróprio comparar os resultados de tais buscas, porque elas possuem bases diferentes; e em qualquer caso, geralmente são incapazes de produzir qualquer precisão. Além do fato de esses mecanismos de busca serem defeituosos, há algum indício de que muitos deles não sejam tão objetivos quanto aparentam ser. Há um artigo de David Corn no *Washington Post* (2) de 7 de julho de 1996 na seção *Outlook*, página 65 chamado: "*Anatomy of a Netscan: Why Your Internet Search May not Be as Honest as You Think*". Este artigo relata que a Iron Mountain Global Information Systems adquiriu as palavras "eleições", "Republicano", "Democrático", e outras diversas palavras no *site* Yahoo. O que isso significa é que onde quer que você busque uma dessas palavras, sua lista de recuperação irá incluir um banner anunciando: "Os dez melhores *sites* políticos". Os "dez mais" não são aqueles tão julgados por *experts* ou *cybersurfers*, mas sim por aqueles mais dispostos a pagar por eles.

Dada uma informação específica, o pesquisador pode estar consciente do que achou, considerando que o resultado é valioso. Mas sempre o problema mais importante está no que o pesquisador de informação não encontrou e que poderia ou deveria ter encontrado, seja através de um mecanismo de busca pobre, ou do uso pobre de um mecanismo de busca, ou de um catálogo pobre, ou por qualquer meio que seja. Como pode o pesquisador de informação saber o que não alcançou, mesmo que estivesse disponível nos serviços gratuitos na Internet, e como pode avaliar o que seria apropriado achar em sua busca, mas que não se encontra disponível (por exemplo: informação com direitos autorais)? Admitindo-se que ninguém consegue atacar essa questão satisfatoriamente, pelo menos os bibliotecários e profissionais da informação estão conscientes da variedade de fontes que podem ser empregadas na busca de informação em geral ou em uma área de assunto específico e podem avaliar ou realizar avaliação cruzada das fontes com possíveis erros. Eles também estão conscientes dos pontos fortes e fracos dos vários instrumentos de busca. Em uma situação de busca não-mediata, ignorância é contentamento. Se o pesquisada de informação, não conseguiu encontrar informação suficiente ou adequada, não ficará sabendo pois terá pouca ou nenhuma maneira de conferir suas

inadequações como pesquisador ou de avaliar as inadequações dos instrumentos de busca empregados. Ele não tem noção de quais informações deveriam estar ausentes, mas poderiam estar presentes.

É possível afirmar que, admitindo a disponibilidade de recursos apropriados na Internet, um instrumento de busca iria provavelmente funcionar muito bem em questionamentos que teriam sucesso na busca do texto completo: por exemplo, se um pesquisador de informação quisesse saber sobre o "Heimlich maneuver", poderia achar alguns recursos adequados, se usasse a sintaxe correta para o instrumento de busca empregado, devido a natureza que é peculiar da busca, o característico dela é o desenvolvimento do contexto da natureza da frase. Porém em muitas buscas, principalmente naquelas que exigem alta precisão, tais mecanismos tendem a funcionar muito mal. Além disso, esses mecanismos de busca e catálogos também falham em buscar alta revocação. Enquanto muitas buscas na Internet produzem várias citações, a maioria delas são redundantes e excessivamente irrelevantes, e qualificar essas buscas como sendo de grande revocação significa cometer uma grave injustiça para com a idéia tradicional de "grande revocação", pois em uma recuperação tradicional on-line na qual a busca estava restrita à base de dados relacionados a assuntos, haveria, portanto, deslizos mas a maioria dos alvos permaneceriam no local de domínio de determinado assunto. Com os mecanismos de busca, pode-se ter uma cacofonia de deslizos-alvos que contém os termos, mas como não existe nenhum controle (como a base de dados relacionados a determinado assunto tal como a Disclosure, ou uma coleção de uma biblioteca), a proporção de deslizos é astronomicamente alta. As razões para tanto são várias: as habilidades precárias do pesquisador, o peso do algoritmo do mecanismo de busca; o mecanismo de busca empregado e sua base de dados (isto é agravado no Netscape pela suposta neutralidade na provisão de mecanismos de busca - enquanto for interessante alternar os mecanismos de busca propostos ao se clicar em "Net search", por outro lado, usuários mais inexperientes ou ingênuos podem vir a pensar que esses mecanismos de algum modo, sejam todos equivalentes). Além do mais, os mecanismos de busca geralmente tornam

difícil a intensificação da estratégia de busca para obter mais informações; por exemplo, empregar facilmente uma variedade de sinônimos ou termos relacionados em uma busca, qualificar posteriormente a busca, ou filtrar (retirar) informações inapropriadas (rejeitando certos termos ou dando um significado específico a um deles) ou reduzir as informações a um conjunto controlável (por meio de mecanismos de filtragem: os mecanismos de busca não permitem a restrição dos termos a vocabulário controlado ou a descritores principais). Por razões similares, tais mecanismos de busca geralmente falham nas buscas por alta precisão, um conjunto de citações que se encontra precisamente no alvo, sendo a única exceção o local de frases únicas ou distintas, como se nota em seguida: a falta de disponibilidade de um vocabulário controlado é provavelmente a razão mais significativa para o insucesso de tais buscas. Apenas quando o indexador sistematicamente designa termos ao documentos, usando consistentemente o mesmo termo para significar a mesma coisa, pode-se recuperar os documentos que estão exatamente relacionados aos tópicos de interesse. As tentativas dos fabricantes de software de desenvolver um esquema de indexação automática que alcance alto nível de precisão nunca atingiram seus objetivos por causa da natureza dinâmica pela qual os seres humanos criam e interpretam símbolos ou sinais.

O que é mais problemático a respeito das fontes da Internet é a ausência de autoridade. Qualquer pessoa com um computador e acesso à Internet pode montar uma *webpage* que faça reivindicações para ou sobre alguma coisa. Bem recentemente 60 Minutes, um programa da CBS, fez uma reportagem que demonstrava o quão fácil era montar uma página e sustentar que esta pertence a outra pessoa. Mesmo se alguém reclamar como sua a autoria da página pode-se dizer qualquer coisa que queira. Eles podem afirmar que a queda do vôo da TWA 800 se deve a ações do governo, da Máfia ou do Papa. Os mecanismos de busca da rede colecionam tais *sites* obedientemente, colocam os termos nas bases de dados, e os pesquisadores de informação desavisados (ou aqueles que querem satisfazer seus caprichos) podem achar dados que sustentem suas afirmações (ver

tendências em processar acesso e processar informação acima). Um dos principais indícios quanto à confiabilidade de uma fonte não está simplesmente no que esta diz, mas quem ou o quê a recomendou - sua autoridade cognitiva. Com a dificuldade de obter-se autoridades cognitivas sobre os recursos na Internet, há poucas ou nenhuma pista quanto à qualidade dos recursos oferecidos na Internet e a habilidade de localizá-los é ruim já que a natureza das classificações e indexações não permite a determinação da qualidade dos recursos de informação. Um dos principais problemas com os recursos de informação na *net* é a "descontextualização". Recursos convencionais de informação fornecem pistas a respeito da qualidade dos recursos, como os editores, conselhos editoriais, uma história das publicações qualitativas (indicadores gerais da autoridade cognitiva): tais indicadores criam bases para o pesquisador de informação fazendo-o acreditar que são verossímeis, os quais não são obviamente um mecanismo infalível (dada a ocorrência de fraudes científicas), mas geralmente bom.

Enquanto tais mecanismos podem acompanhar informações eletrônicas futuramente, no momento a maioria das fontes carecem de um contexto para se fazer julgamentos qualitativos e de devida relevância quanto a natureza da fonte. Para se avaliar recursos, são necessárias muitas habilidades, dependendo da informação que se necessita: alfabetização, alfabetização em Inglês, (a linguagem das interfaces, dos mecanismos de busca e da maioria de recursos), alfabetização em computação, alfabetização informacional (habilidade de entender registros de informação e de outros campos de informação, por exemplo, registros bibliográficos e suas esferas), alfabetização científica e/ou tecnológica (por exemplo, habilidade em ler e digerir os estudos e artigos científicos, como entender um projeto experimental adequado, medidas e técnicas estatísticas apropriadas) alfabetização cultural (a habilidade em compreender a sociedade e a cultura que produziram um recurso de tal forma que este seja entendido devidamente), alfabetização crítica ou avaliativa (habilidade que combina algumas das habilidades acima citadas, como também o entendimento da lógica, de enganos lógicos, de métodos

apropriados de linguagem e analogias). Infelizmente, a capacidade de compreender informação e avaliá-la são provavelmente as competências mais desejadas durante a busca na *Net*- não há um curso ou parte dele requerido nos níveis educacionais de graduação ou pós-graduação.

Para resumir, usuários finais da Internet são postos seriamente em perigo quando usam a Internet, sabendo ou não deste fato. Os instrumentos e técnicas para localizar ou verificar da informação na Internet, seja a busca por palavras-chave (via Veronica), sejam os diretórios de rede da base de dados ou mecanismos de busca, são completamente inadequados: (1) para acesso cognitivo (o acesso por assunto encontra-se ausente ou lastimavelmente deficiente), (2) para localizar uma boa informação (com poucas exceções, não há fontes direcionando os usuários finais aos locais específicos de bons recursos) (3) para determinar a qualidade dos recursos fornecidos (não há nenhuma ou poucas autoridades na Internet que avaliem a qualidade de recursos). Além do mais, o tempo requerido para localizar e realizar um *download* das informações sempre se contrabalançam com o benefício potencial. Enquanto que ao se usar um *web browser* como o Netscape, as imagens gráficas de várias *homepages* podem levar muitos minutos para serem visualizadas. Além disso, ao se usar um mecanismo de busca, pode-se obter dentro de alguns segundos um *gigabyte* de alvos dos termos de busca; pode-se levar horas para se encontrar e avaliar sua qualidade, veracidade ou utilidade. Sempre é mais rápido e confiável obter-se uma fonte de informação da biblioteca do que obtê-la da Internet.

Pesquisadores de informação estão sendo atualmente incentivados a recorrer aos vendedores comerciais na Internet, tais como Dialog. Os serviços de Informação Knight-Ridder parecem estar bem mais interessados em prover acesso direto de suas bases de dados aos usuários finais. Mas, muitos das questões que agravam a mineração da Informação na Internet, ocorrem no meio comercial: dados pobres ou contaminados, indexação pobre em muitos casos, dados incompletos ou imprecisos, etc. E em ambos os casos de alta qualidade ou de baixa qualidade das bases de dados, os usuários

finais são induzidos a pensar que uma busca é uma questão fácil e que as habilidades de restringir a recuperação para um tamanho apropriado e a um nível de qualidade são insignificantes como ao se fazer escolhas em um cardápio. Os interesses comerciais são mais dirigidos a fazer dinheiro do que em ajudar os pesquisadores a alcançar informação apropriada e confiável. E não é necessária uma reflexão extensiva para se perceber que se os pesquisadores de informação realizam uma busca mal feita, *DIALOG* e Knight Ridder ganham **mais** dinheiro. Claramente, os usuários finais necessitam de ajuda e precisam ser alertados quanto à ajuda que eles virão precisar. Mas de quem é esta responsabilidade?

3. POLÍCIA E VENDEDORES DE ESTRADA, MAS SEM PARQUES, GUIAS E CENTROS DE TURISMO?

Um profissional com um pouco de bom senso percebe logo que os benefícios potenciais da Internet são extremamente limitados, como uma fonte de recurso de informação de qualidade se se quer encontrá-la com alta precisão e revocação altamente *apropriada* (a não ser que *sites* específicos sejam conhecidos de antemão ou a não ser que alguém queira pagar pelos direitos autorais da informação através de vendedores comerciais; a revocação deve ser apropriada no sentido acima, no qual para muitos mecanismos de busca, revocação significa que a referência contém os termos usados na busca, mas a taxa de alvo alcançado, pode variar muito devido à natureza do mecanismo de busca, seus algoritmos a partir das categorias de resultados, das missões). A questão é: pertence a alguém a responsabilidade de garantir ou facilitar o acesso aos recursos ou informações de qualidade? Se sim, a quem pertence? E deveria alguma estrutura regulamentadora deste assunto ser posta em vigor?

Duas abordagens são inaceitáveis: (1) o controle da Internet per se através do estabelecimento de alguma agência regulamentadora ou (2) o controle dos usuários finais (exceto nos casos em

que eles fazem parte de uma organização que possa estabelecer políticas de usuário de computador (CUPS)). Ambas abordagens, apesar das tentativas mal empregadas dos políticos de regulamentar certos tipos de informação, são impraticáveis, impossíveis e podem ser inclusive inconstitucionais. Quais outras abordagens existem? Enquanto, ao terminar a busca, os usuários finais são definitivamente responsáveis pelo seu acesso e uso da informação, há muita coisa que pode ser feita para facilitar a tomada de decisão dos usuários finais e uso apropriado dos recursos da informação. Há duas abordagens que poderiam ser incentivadas: abordagens para garantir o interessado público e abordagens que poderiam ser empreendidas por profissionais da informação.

3.1 Responsabilidade Social e Interesse Público

O trajeto da Internet está tomando rapidamente a forma de *shopping-centers*: a presença de alguém e a sua utilização é apenas estimulada se este alguém quiser tornar-se um consumidor. Infelizmente, as questões de interesse público estão sendo suprimidas ou deixadas de lado. As abordagens de interesse público podem ocorrer em dois níveis: no micro e no macro. Em nível macro, a política pública precisa ser desenvolvida e implementada. Enquanto o foco desse trabalho não se dirige às questões de política *per se*, há dois tipos de questões que possuem considerável significado: (a) as questões de políticas explícitas (b) as questões de políticas tácitas.

As questões explícitas estão relacionadas com a política e/ou a lei. Elas envolvem privacidade, informação com direitos autorais e similares. Estão envolvidas aqui sérias questões, os esforços recentes da WIPO em Genebra em dezembro de 1996 e a administração Clinton no Ato de Proteção e Infra-estrutura dos direitos autorais da informação Nacional pesam contra o interesse público: a doutrina do uso justo e a doutrina da primeira venda, generosamente apreciadas na vida americana em público e em bibliotecas acadêmicas, estão sendo desagregadas por interesses comerciais. Espaços públicos e questões de interesse público são cada vez menos e menos repre-

sentadas em deliberações sobre a supervia da informação e nos regulamentos propostos para administrar aspectos da privacidade, dos direitos autorais e da segurança. Todos os americanos, incluindo os profissionais da informação, têm que perseguir metas de responsabilidade social protegendo seus interesses, através do contato com membros do Congresso e do Senado e órgãos estabelecadores de leis. Por exemplo, a American Library Association tem sido atuante em desafiar a censura não apropriada e o controle da supervia da informação, mas existem muitos outros grupos ativos para se juntar e apoiar: CPSR, *NetAction*, EFF e o *Digital Futures Coalition*.

Deveriam haver também políticas tácitas, a respeito da *netiquette*, na organização de uma página da rede, encaminhando mais para um código de ética. Esse código não seria posto em vigência por uma agência específica, mas seria estabelecido como uma política convencional para a criação de páginas de rede e provisão de informação.

Este código de ética seria originalmente para os provedores de informação (e outros criadores, mas neste momento meu foco são as páginas de rede como recursos de informação, como que opostos aos recursos comerciais da Internet e outros tipos.) Ele iria especificar os elementos que deveriam estar explicitamente evidentes nele ou acessíveis através de uma *home webpage*: autoria (se o autor quiser se manter anônimo, deve indicar o uso de pseudônimos ou anonimato), responsabilidade (se não for o autor, por exemplo, o editor ou mantenedor - a pessoa que é responsável pela página de rede e sua manutenção) data de criação, data da mais recente atualização; fontes (se usadas), fonte corporativa ou outros provedores de informação (se cabíveis), contexto, isto é, filosofia, objetivos, metodologia, política e orientação), dependendo de onde for aplicável, agência patrocinadora (se cabível), público interessado (se cabível), natureza do mecanismo de busca (se cabível) e indicação dos elos *aossites* que contenham informação com direitos autorais. Para compilações ou diretórios de assuntos ou *gateways*, a filosofia, a coleção, a filtragem,

a classificação ou os métodos de classificação devem ser tornados explícitos. Websites devem ser ecológicos (amigos do meio ambiente), através da minimização da utilização dos recursos de *downloading*: isto é, ao se carregar e imprimir informação de um site, alguns *downloads* causam a impressão de múltiplas *short pages*, causando, como conseqüência um enorme desperdício de espaço.

Enquanto longas *webpages* podem ser esteticamente agradáveis ou facilitarem a navegação, o *download* de determinada página é normalmente mais eficiente, porque o texto e/ou as imagens completam seqüencialmente um conjunto de páginas de impressão, mas que são altamente segmentadas, quando interligam páginas de rede desperdiçam muito papel, pois cada uma delas demanda sua própria impressão. (Enquanto se pode argumentar que múltiplas páginas de rede segmentadas podem, na verdade, impedir o carregamento desperdiçador, minha experiência é de que eles não o fazem). Além do mais, é essencial distinguir importantes questões de outras menos significativas. Por exemplo, há um crescimento de literatura na Internet avaliando os recursos de informação e sua qualidade (por exemplo [http:// www/cyber.com/guide1-html](http://www.cyber.com/guide1-html)), mas muitas delas não parecem ter um entendimento do que é de cunho importante: velocidade de *download*, primeiras impressões, utilização de gráficos ou multimídia são raramente questões de **qualidade** de recursos como **substância** da informação. A qualidade está relacionada a **conteúdo** - à autoridade, à confiabilidade, à verificabilidade, à precisão e à falta de. Enquanto alguns usuários ficar desinteressados em usar um certo *site* por causa de seus gráficos ruins e navegação pobre, isto não significa que o *site* não seja um recurso valioso. Além do mais, há uma questão ética notada acima sobre os algoritmos da classificação de mecanismos de busca, que é o método pela qual eles decidem que um documento é mais relevante do que outro; por exemplo, uma busca que envolva três termos, um documento que contenha dois termos de freqüência significativa e um termo de freqüência moderada é mais importante do que aquele que tenha uma freqüência moderada dos três termos. Tais algoritmos parecem ser propriedade ou seja, parte

dos "segredos comerciais" de um determinado provedor. Contudo tal classificação é tendenciosa, pois a natureza pode ser mais ou menos exagerada, e é obscura, dependendo da busca e ainda os podem não ser "objetivos".

O ponto é que a conscientização do interesse público e as políticas devem ser promovidas, de modo explícito e tácito. As políticas de ação explícitas tentariam preservar o "uso justo" e a "liberdade de informação", enquanto políticas tácitas iriam exigir metadata anexados a documentos que proveriam os usuários com alguns fundamentos para avaliação.

3.2 Responsabilidade Social e Profissionais da Informação

Enquanto questões de responsabilidade social devem ser preocupação de todos ao promover acesso universal, os profissionais da informação parecem ter uma responsabilidade especial em que pese serem eles o maior grupo de profissionais sensíveis aos problemas de acesso cognitivo e organização do conhecimento, os quais são necessários para facilitar a recuperação efetiva e eficiente na Internet. Há diversas formas de realizarem isso. Devem comandar e promover *sites* para propagar a tradicional norma de intermediários na filtragem, classificação, avaliação e referência da informação. Eles, talvez principalmente os das bibliotecas públicas, deveriam promover serviços de referência on-line, cujos navegadores de redes poderiam consultá-los no caso de não obter a informação desejada tais serviços poderiam direcionar os pesquisadores de informação aos recursos de rede específicos ou poderiam consultar coleção local para resolver as necessidades de informação. Eles poderiam usar sua experiência especializada em organização da informação e manter o alto nível da classificação dos materiais e para a provisão de múltiplos pontos de acesso aos recursos da informação. Eles poderiam insistir em índices qualitativos e diretórios. Eles poderiam ser os defensores das políticas de desenvolvimento de fortes coleções de *sites* que aparentam ser entradas de assuntos ou diretórios. As técnicas tradicionais de

indexação e catalogação descritiva e de assunto deveriam prover a base para a análise de metadados de *web sites* e objetos. Felizmente, há uma crescente presença de tais *sites* na Internet: o Argus Clearinghouse for Subject Oriented Internet Resource Guides no <http://www.clearinghouse.net/>, o SOSIG (social Science Information Gateway) no <http://www.esrc.bris.ac.uk/>, o Internet Public Library no <http://ipl.sils.umich.edu/>, o Infomine no <http://ib-www.ucr.edu/>, para citar alguns. Eles poderiam também se incumbir de um papel adicional, o de consultores da informação, no qual eles poderiam ser capazes de ajudar os pesquisadores de informação a achar informação qualitativa, a prevení-los do que não está disponível na Internet ou daquilo que não tem custo, e ajudá-los a evitar uma interpretação errônea quanto ao contexto ou quanto a intenção de uma fonte de informação. Obviamente, isto significa que escolas de Biblioteconomia e Ciência da Informação precisam incluir em seus currículos dos cursos, naquilo que se refere aos instrumentos e recursos da Internet, seus méritos e deficiências, além de disciplinas em habilidades de pensamento crítico. Profissionais da informação devem se incumbir de promover políticas públicas mencionadas acima. Eles deveriam procurar prover espaços públicos na Internet, para biblioteca públicas, forums de informação gratuita ou outras formas gratuitas de armazenamento de informação, compartilhamento de informação e recuperação. Ele têm que se comprometer com a tarefa de educar o usuário quanto aos méritos dos recursos da *Internet*, assim como suas deficiências e seus problemas, e estimular o público a identificar suas limitações em atingir certos tipos de necessidade de informação, de tal maneira que possam buscar a assistência de profissionais. Eles precisam insistir, através da educação pública e profissional, na continuação das políticas de uso adequado da informação, reconhecendo, ao mesmo tempo, o direito dos produtores de informação de procurar compensação pelas informações disponibilizadas com fins comerciais. Eles precisam também promover políticas que protejam a privacidade, confiabilidade e segurança.

Profissionais da informação precisam ser crítico a respeito dos instrumentos da rede e dos *web sites*, indicando, em fóruns

públicos, na *Net* ou outro lugar, a origem das limitações e problemas com os mecanismos de busca, metaíndices e diretórios. Através do criticismo público eles podem encorajar ou forçar os provedores de recursos, tanto comerciais como não-comerciais, a desenvolver sites e instrumentos, por exemplo a remover dados contaminados, a estabelecer controles de autoridade para suas informações, a promover sistemas de classificação imparciais e adequados que favoreçam o usuário, tornando os mecanismos de busca tão poderosos e controláveis quanto possível (similares aos tipos de busca disponíveis em alguns sistemas comerciais, tal como a busca por termos em campos específicos).

Com essas atitudes, os profissionais da informação podem incumbir-se da responsabilidade social em quatro frentes: atuando como provedores de informação e consultores, como defensores das questões de interesse público, como educadores a respeito da utilização adequada e abusiva dos recursos da Internet e, como críticos dos *sites* ou instrumentos ineptos, inadequados, sem autoridade ou tendenciosos.

CONCLUSÃO

Para atingir os problemas de excesso de informação e qualidade de dados para os buscadores de informação, a Internet precisa de mais especialistas da informação, certamente não mais de software, particularmente de mecanismos de busca (exceto para aperfeiçoamentos). Os tecnólogos e vendedores parecem ter prazer em propagar o mito de que alguns programas de software irão resolver todos os problemas de acesso aos assuntos determinados, enquanto 3000 anos de Biblioteconomia tenham indicado o contrário. A classificação e os sistemas de classificação que os seres humanos criam, estão cheios de falhas como eles mesmos tem-se mostrado, no entanto ser mais precisos e úteis para as outras pessoas, especialmente quando eles procuram por alta precisão ou alta revocação. Uma das maiores ironias da sociedade da informação é que apesar da

informação ser ruim ou inadequada, as pessoas anseiam por pagá-la contanto que venha pelo computador ou *network*, mas não querem pagar por aquilo que faria da informação algo verdadeiramente útil: classificação humana, indexação, catalogação, etc. Elas irão imediatamente pagar por tecnologias de computação, mas não por tecnologias cognitivas. Portanto, bibliotecários e profissionais da informação estão sendo desafiados em duas frentes: **na decadência do acesso público e gratuito à informação e na falta de vontade de usuários e vários produtores de informação em pagar pela organização profissional do acesso à informação e pelo acesso a esta. E ainda conforme os interesses da responsabilidade social e sua sensibilidade a esses problemas, eles devem incumbir-se do desafio de que os usuários finais possam finalmente achar na Internet recursos que sejam apropriados às suas necessidade. Pode ser que os navegadores de redes encontrem mais e mais problemas em suas buscas, venham a entender e apreciar os procedimentos valiosos que os profissionais da informação desempenham.** (grifo nosso)

REFERÊNCIAS

- (1) NUTT, Paul C. *"Making Tough Decisions: Tactics for Improving Managerial Decision Making"*, Jossey-Bass Publishers, San Francisco, CA 1990.
- (2) CORN, David. "Anatomy of a Netscam: *Why Your Internet Search May Not Be as Honest as You Think*," Outlook Section, Washington Post, July 7, 1996, p. C5.