

O VOLUME DE INFORMAÇÕES NA INTERNET E SUA DESORGANIZAÇÃO: REFLEXÕES E PERSPECTIVAS

Claudio Henrique Schons

Resumo

Este artigo analisa os fatores que contribuíram para a evolução da Internet e seu volume de informações. Posteriormente, são avaliadas as tecnologias que visam apoiar a organização de informações na Internet. Por fim, são apresentadas algumas tecnologias que atuam como agentes inteligentes visando tornar a Internet um ambiente inteligente, com capacidade de aprender e entender as necessidades dos usuários em um contexto adequado.

Palavras-Chave

Organização da Informação; Internet; Sistemas de Informação; Recuperação da Informação

1 INTRODUÇÃO

Atualmente vivemos em um mundo competitivo em que os crescentes avanços da tecnologia e da comunicação vêm exercendo uma forte influência sobre a sociedade. Dentre os meios de comunicação desenvolvidos, destaca-se como o de maior amplitude e importância global, a Internet.

Com a Internet rompem-se paradigmas e intensificam-se relações sociais, fa-

zendo surgir um novo cenário onde a informação é o elemento chave. Desde seu surgimento, sua função é permitir o acesso e a partilha de informações. No entanto, no decorrer do tempo, por apenas se preocupar em absorver e suportar todo volume informacional criado, tornou-se um grande espaço emaranhado de nós interconectados desprovidos de qualquer organização.

A partir desse momento, as atenções voltaram-se para o desenvolvimento de

tecnologias visando a busca e recuperação eficiente das informações, visto que as dificuldades de encontrar informações relevantes em meio a um espaço não estruturado são grandes.

Nesse contexto, busca-se com este artigo apontar a partir de um recorte temporal os fatores que apoiaram a evolução da Internet e a explosão informacional, bem como avaliar as tecnologias presentes e futuras no processo de organização da informação. Espera-se que este artigo sirva de base, contribuindo para o desenvolvimento de estudos futuros numa dimensão mais aprofundada acerca do tema exposto.

O artigo está estruturado da seguinte maneira: primeiramente busca-se explorar os fatores que sustentaram o crescimento da Internet. Para tal, faz-se um rastreamento em sua anatomia a fim de evidenciar suas características. A seguir são analisadas quais tecnologias provêm apoio à organização informacional da Internet, mensurando suas funções de operacionalização. Por fim, avalia-se a utilização de tecnologias inteligentes, seus impactos e perspectivas diante do volume de informações existentes.

2 FATORES QUE FAVORECERAM O CRESCIMENTO DA INTERNET

A Internet pode ser definida como uma vasta e onipresente rede global. Não

é à toa que é chamada de “a rede das redes”, interconectando vários computadores em todo o mundo, alcançando níveis de abrangência e utilização jamais imaginados por seus idealizadores.

Desde sua criação (meados da década de 1970), tinha como propósito permitir a comunicação, a troca de informações e o compartilhamento de computadores e outros recursos. Seu público-alvo consistia de cientistas, pesquisadores e militares, que a utilizavam para fins educacionais e militares.

O processo de transição da Internet destacou-se a partir do final da década de 1980 quando os EUA liberaram a rede para uso comercial. No entanto, somente no início da década de 1990, com o surgimento das primeiras empresas provedoras de acesso comercial, e principalmente com o surgimento da *Web*, houve de fato a explosão popular da Internet. A partir dessa época até os dias atuais, a Internet passou a integrar-se progressivamente aos diversos segmentos da sociedade, contribuindo para o desenvolvimento de um novo tecido social, caracterizado por mudanças de hábitos e comportamentos.

A expansão da Internet, desencadeada pelo aumento do número de usuários conectados, tornou-se um fenômeno amparado pelo desenvolvimento

tecnológico¹ levando em conta dois pressupostos: a) novas tecnologias que possibilitaram a interconexão de *hosts* a partir de equipamentos de telecomunicação e computação cada vez mais rápidos; b) softwares de comunicação fáceis de serem utilizados que permitiram o acesso e partilha de informações na rede, bem como facilitaram a interação entre os internautas. Esse contexto favorável contribuiu para que o uso da Internet evoluísse continuamente a ponto de tornar-se um hábito em suas vidas.

Conseqüentemente, novos documentos foram produzidos e distribuídos no meio social através da *web*, criando assim uma cultura humana de produção.

No entanto, faz-se necessário comentar que devido a características de sua própria estrutura funcional, a produção e acesso de informações na Internet foram favorecidos. Dentre tais características, compreende-se como mais importantes nesse processo:

- Desregulamentada: a Internet não possui dono e, portanto, não é propriedade de ninguém. Não há um manual de regras e normas para utilizá-la. No âmbito legal, existe apenas uma organização internacional dos usuários da Internet chamada

*Internet Society*², que controla e promove a troca global de informações. No Brasil há uma organização similar chamada CGI³ (Comitê Gestor da Internet), criado em 1995, que infere decisões quanto a seu conteúdo;

- Descentralizada e aberta: a Internet se caracteriza por ser uma rede que todos podem acessar a qualquer momento e lugar. A rede não pára, ela interage com tudo e todos a todo momento. Desse modo, a Internet é um sistema ativo, em constante troca com o meio e altamente mutável. Vaz (2004) compreende que, com o surgimento da Internet, o conceito de rede foi significativamente alterado e passou a denotar um sistema aberto capaz de romper fronteiras, permitindo qualquer indivíduo participar. Para o autor, dentre as características da Internet, pode-se dizer que é infinita, sem um centro comum e multipolarizada;

- Não-hierárquica e interativa: pelo seu alto grau de interatividade, a Internet promove o remodelamento na estrutura do fluxo de informação, possibilitando o desdobramento hierárquico entre emissores e receptores, pois ambos atuam como nós na rede e possuem funções ativas (ambos produzem e distribuem informação). Nessa perspectiva, Lévy (1999) reflete acerca da contribuição entre os internautas e atribui o termo “inteligência coletiva”, porque todos

¹ O crescimento da *web* foi favorecido pelo avanço de equipamentos tais como: satélites, roteadores, servidores, *switches*, *backbones* entre outros.

² <http://www.isoc.org>

³ <http://www.cgi.br>

podem contribuir para a concretização de uma tecnodemocracia por intermédio de suas percepções e inteligências. Lévy (2003, p.195), ainda complementa afirmando que a Internet permite a “reciprocidade na comunicação e a partilha de um contexto”. Em outras palavras, possibilita a comunicação horizontal de todos para todos.

Percebe-se então que a inexistência de regras e padrões rígidos, bem como a dinamicidade, a informalização e a horizontalização da rede favoreceram a participação social no acesso e produção de informações.

Por meio do apoio e desenvolvimento tecnológico, as características da Internet foram fortalecidas, garantindo sua rápida expansão e tornando-a um espaço extremamente complexo, dotado de infinitas ramificações, conduzindo informação em diferentes formatos e lugares.

O fluxo abrangente de informação distribuído por uma gama de veículos de comunicação como jornais, periódicos, revistas, textos, *sites* de organizações, dentre outros, dinamizou ainda mais a troca de idéias e o compartilhamento de informações. A amplitude desse fluxo possibilitou que em diversos contextos, como entretenimento, cultura, ciência, educação, negócios, fossem estabelecidas comunicações não apenas de “um para um” e de “um pa-

ra muitos”, mas também de “muitos para muitos”.

Esse cenário interativo, descentralizado, aberto e favorável ao crescimento da rede, posteriormente desencadeou um grave problema, que é exposto a seguir.

3 A INTERNET E SEU VOLUME DE INFORMAÇÕES

Com o passar do tempo, tornou-se perceptível o crescimento exponencial da Internet, pela expansão de sua massa de informações, juntamente com o número de usuários que dela participam no processo de “inteligência coletiva”.

A facilidade de acesso e publicação de documentos na rede desencadeou novas possibilidades quanto à construção de um grande acervo informacional.

Tal fenômeno a princípio é positivo, mas no decorrer do tempo esse processo originou um grande problema na rede pelo fato de conduzir e fomentar o processo de explosão informacional gerando grandes conflitos, uma vez que o excesso de informações parece ser pior que a sua falta. Um exemplo pode ser colocado da seguinte maneira: um usuário, ao efetuar uma pesquisa na Internet, tende a ficar mais frustrado pelo excesso de “respostas” que pela falta delas. Isto porque o excesso de

informações não representa uma solução, mas um problema: a desinformação.

Bawden (2001) entende que diante da sobrecarga informacional, a informação disponível e potencialmente útil acaba desencadeando um obstáculo (ou atraso) para o indivíduo que dela necessita, em vez de uma ajuda. Para Terra e Bax (2003, p.5), “o excesso de informação está associado à perda de controle sobre a informação e à inabilidade em usar efetivamente a informação”.

Acerca do caos informacional, Johnson (2003) compreende que a *Web* é um espaço tremendamente desorganizado, um sistema em que a desordem cresce com o volume total. Isso se dá pela própria capacidade da rede de ser ilimitada, ou seja, não há limites para estocagem de informações. Diante desse cenário, percebe-se então um excesso de informações na rede, que excede a capacidade humana de percorrê-lo e explorá-lo, ocasionando assim uma grande tarefa temporal para o indivíduo acessar e processar a informação que deseja.

Lévy (1999, p.16) abrange o dilúvio informacional na Internet:

Fluída, virtual, ao mesmo tempo reunida e dispersa, essa biblioteca de Babel não pode ser queimada. [...] As águas deste dilúvio não apagarão os signos gravados: são inundações de signos.

Esse quadro vem sendo acirrado pelo crescimento constante da quantidade de informação armazenada e disponibilizada via Internet. A fim de quantificar o cenário atual acerca da revolução informacional, são apresentados a seguir alguns dados.

Conforme pesquisa realizada por Lyman e Varian (2003), pesquisadores seniores da Universidade da Califórnia, a *Web* contém cerca de 170 *terabytes* de informação. Em volume, isto é 17 vezes maior do que as coleções impressas da Biblioteca do Congresso Americano. Segundo o serviço de medição *online* Netcraft⁴ existem na Internet 80 milhões de páginas publicadas.

A pesquisa⁵ também apresentou que a Internet dobrou o número de páginas em 3 anos, de 40 milhões em 2003 para 80 em 2006. O ponto crucial desta nova fase é como utilizar esta mega coleção de informações.

A explosão informacional trouxe à tona graves problemas quanto a falta de organização e estruturação de informações na Internet, dificultando a recuperação de informações relevantes aos usuários.

Nessa perspectiva, Marcondes e Sayão (2002) compreendem que a grande quantidade de informações produzidas e

⁴ <http://news.netcraft.com>

⁵ Disponível em:

http://idgnow.uol.com.br/internet/2006/04/10/idgnoticia.2006-04-10.1965396374/IDGNoticia_view. Acesso: maio 2006.

disponibilizadas por uma gama de atividades sociais, tornou cada vez mais difícil a identificação, acesso e utilização da informação.

Logo, diante do excesso contingencial de informações, a Internet passou a sofrer impactos e perder o controle da ordem, tornando sua estrutura funcional desordenada, anárquica, inconsistente, caótica, desenrolando o caos informacional de hoje.

A partir de então, as atenções (sobretudo dos profissionais da informação) voltam-se mais para a organização das informações que se dispõem, e não propriamente para como obter novas informações. O papel das tecnologias nesse contexto é analisado na próxima seção.

4 TECNOLOGIAS NO APOIO A ORGANIZAÇÃO INFORMACIONAL NA INTERNET

Devido à explosão informacional, a quantidade de informações cresceu de tal maneira que a estrutura da Internet apenas foi suportando e absorvendo todo volume informacional, não se preocupando com a representação padronizada do mesmo. Para Johnson (2003), as conseqüências negativas do fenômeno referente ao caos informacional na Internet ocorreram em razão de sua desorganização enquanto crescia. Logo, com o passar do tempo chegou-se a um estado de saturação, e a sociedade começou a ter dificuldades para a recu-

peração eficiente de informações e documentos.

Um estudo realizado por Bergman (2001) referente ao problema da recuperação de informações e documentos na Internet, expôs o seguinte quadro:

- O número de informações na Internet cresce em forma exponencial. Existem bilhões de páginas onde se publica tudo sobre tudo, de forma caótica;
- As informações/documentos disponibilizados não o são de forma ordenada, mas de maneira caótica;
- A informação disponibilizada na Internet se refere a uma vasta infinidade de temas, portanto altamente abrangente, dificultando a recuperação da informação; Não é à toa que se busca criar categorias ou vocábulos específicos para cada tema;
- Utilizam-se diferentes idiomas na disponibilização da informação, agravando o problema da categorização usada na recuperação;
- Os mecanismos de busca no processo de indexação das páginas apenas extraem palavras do texto HTML da página e armazenam estas palavras em bases de dados de forma isolada, sem compreender seus temas;
- Cada página HTML é reconhecida pelos programas-robôs dos mecanismos de busca durante a indexação de maneira

isolada e estática, desconsiderando que diversas páginas estão inter-relacionadas e que constituem um determinado *site*.

Segundo palavras de Johnson (2003, p.89) em referência à situação do caos informacional, há um grande paradoxo na Internet pois “quanto mais informação flui em seus reservatórios, maior se torna a dificuldade de achar um simples dado no seu oceano”.

De fato, com a explosão informacional, a necessidade crescente da recuperação da informação passou, cada vez, a vir à tona, necessitando que tecnologias mais eficientes surgissem no sentido de conter e organizar a avalanche de informações.

A pesquisa e desenvolvimento acerca de novas tecnologias tornaram-se essenciais para o bem da Internet, visto que esta possui conteúdo disperso em um ambiente não estruturado. Nesse contexto, o desenvolvimento de novas tecnologias visa atuar em duas vertentes: a) aperfeiçoar os recursos já existentes, de modo que novos padrões possam aprimorar o processo de representar e armazenar informações e documentos já disponíveis; b) prover novas metodologias e mecanismos para o registro, seleção, busca e representação da informação, de modo que o fluxo nessas atividades seja eficiente.

Em ambos os casos, estudos têm sido implementados no sentido de fazer frente à explosão informacional adotando como requisitos básicos a qualidade e a relevância da informação, ou seja, a informação certa no momento preciso.

Dentre as tecnologias emergentes que buscam colaborar na organização eficiente da informação no contexto da Web, destacam-se:

Metadados e XML

Os metadados (dados sobre dados) são amplamente utilizados, permitindo a descrição de documentos e informações, possibilitando que seres humanos, robôs ou agentes de *software* possam avaliar sua relevância e associá-los a documentos para que possam ser recuperados e manipulados com eficiência. A linguagem XML (*eXtensible Markup Language*) é um padrão estabelecido que visa apoiar a codificação dos metadados de um documento eletrônico, facilitando que pessoas e programas o compreendam.

Inicialmente, para permitir que os robôs ou agentes de *software* dos mecanismos de busca façam a indexação com maior qualidade, passou-se a utilizar metadados no texto das páginas via *tags* especiais da linguagem HTML (*tags* META que indicavam o conteúdo e palavras-chave da página). Como o uso dessas *tags* para co-

dificação de metadados possuía algumas limitações, passou-se a discutir o uso de metadados juntamente com a linguagem XML.

Acerca dessa associação, Marcondes (2007) diz que ela possibilita a explicitação da descrição do conteúdo e, sobretudo, a estruturação dos documentos eletrônicos sob a forma de marcações, tornando os documentos mais organizados. O XML também pode ser utilizado em conjunto, por exemplo, com folhas de estilos CSS (*cascating style sheet*) para referenciar como um documento deve ser exibido, podendo também fazer referência a parâmetros de outros arquivos (XML Schema ou DTD – *Document Type Definition*), especificando regras de validade para o documento e garantindo precisão na sua formatação.

Os metadados podem ser utilizados conforme padrão Dublin Core (proposto pela *Dublin Core Metadata Initiative* – DCMI)⁶, que provê melhores possibilidades para a descrição e recuperação de informações. Seu propósito é permitir, de modo simples e auto-explicativo, que os próprios criadores dos documentos possam descrevê-los ao publicá-los. Segundo Marcondes (2007), o padrão DCMI compreende o uso de 15 elementos descritivos, simples e auto-explicativos, o chamado *Dublin Core Metadata Element Set* (DCMES). Os meta-

dados disponíveis no DCMI envolvem as seguintes possibilidades:

- *Subject* (assunto);
- *Title* (título);
- *Creator* (autor, criador do documento);
- *Publisher* (publicador, quem disponibiliza o documento na Internet);
- *Contributor* (demais colaboradores, por exemplo, editores, tradutores);
- *Description* (descrição, sumário);
- *Date* (data de publicação);
- *Type* (tipo de recurso: *homepage*, romance, poesia, *software*, dicionário);
- *Format* (formato do arquivo que contém o documento eletrônico: texto, PDF, HTML ou outro);
- *Identifier* (geralmente o URL de um documento eletrônico);
- *Relation* (relacionamentos com outros documentos, por exemplo, versões);
- *Source* (fonte ou origem);
- *Language* (idioma do documento);
- *Coverage* (cobertura - espacial ou temporal, sobre que lugar, ou sobre que época o texto se refere);
- *Rights* (texto livre especificando qualquer restrição referente a direitos autorais).

⁶ <http://dublincore.org/>

Um ponto interessante a ser mencionado é que o uso de metadados, seguindo o padrão Dublin Core e codificação XML, pode ser exemplificado pelo protocolo *Open Archives Initiative Protocol for Metadata Harvesting* (OAI-PMH)⁷. Este protocolo é bastante utilizado na comunidade científica onde os próprios autores efetuam a publicação *online* de seus arquivos de acesso aberto, os denominados “open archives”.

Ferramentas de busca

Cabe às ferramentas de busca a tarefa de analisar e indexar os documentos existentes na *Web* e, posteriormente, armazená-los em uma base de dados de modo que a cada pesquisa solicitada pelo usuário, suas necessidades sejam atendidas.

Em meio ao enorme volume informacional espalhado na *Web*, as ferramentas de busca representam um meio para localizar informações. Johnson (2003) considera que tais ferramentas são verdadeiros antídotos criados pelo homem para o natural caos na *Web*. No entanto, apesar de minimizarem o estado caótico atual, tais ferramentas ainda não proporcionam uma solução definitiva e perfeita. Marcondes e Sayão (2002) compreendem que o uso de ferramentas de busca representa soluções parciais para o problema da explosão informacional.

Salienta-se, neste sentido, a necessidade de aprimoramento operacional dos mecanismos de busca por meio de melhor qualidade de indexação, além da adoção de novas ferramentas e funcionalidades no processo de busca.

Tal abordagem vai ao encontro do pensamento de Gesteira (2006, p.5)

A forma de recuperar a informação, nos *sites* de busca, tende a se aprimorar a cada dia, mas o quadro que se verifica no momento é da inexistência de mecanismos capazes de filtrar a informação de forma a proporcionar a seleção restrita dos temas de interesse.

De fato, os mecanismos de busca ainda necessitam de aperfeiçoamentos; no entanto, se analisarmos a história de seu desenvolvimento, percebemos grandes avanços tecnológicos.

Hoeschl (2006) apresenta uma classificação temporal baseada em gerações quanto aos mecanismos de busca na *Web*. A primeira geração, para fazer frente ao grande volume de conteúdo na Internet, deu-se com os diretórios ou catálogos (Yahoo e similares), e a descoberta, avaliação, descrição e inclusão dos recursos eram feitas por profissionais de informação. Na segunda geração, esse processo foi automatizado com os robôs digitais (Altavista). A terceira geração veio com os metabuscadores, juntando em um único resultado as informações de vários motores de busca (MetaMiner). Na quarta geração os resultados são ainda mais refinados (All the

⁷ <http://www.openarchives.org/OAI/openarchivesprotocol.htm>

Web). A quinta geração corresponde à geração atual, tendo como grande exemplo o Google, utilizando várias tecnologias, dentre as principais, a PageRank, comentada posteriormente. Na figura 1, pode-se observar a utilização de alguns recursos do Google. A letra "A" indica a função "em cache" por meio da aplicação de metadados de preservação, onde uma cópia da página é armazenada no momento que foi indexada. Essa função é interessante quando o usuário precisa acessar determinado conteúdo e o site não se encontra mais *online*. No entanto, esse recurso é disponibilizado apenas para as páginas mais visitadas.

Outra funcionalidade interessante, representada na figura pela letra "B", é a de "páginas semelhantes", que utiliza o resultado da busca para oferecer ao usuário outras páginas da *Web* com resultado semelhante ao solicitado na busca. Uma outra utilização de metadados é o resultado apresentado na letra "C" da figura, em que, por meio das *tags* é apresentado o conteúdo do site:

```
<meta name="Description" content="UFSC -
Universidade Federal de Santa Catarina.
Campus Florianópolis - Santa Catarina –
Brasil www.ufsc.br - +55 (48) 3331-
9000" />
```

[Efetuar login](#)

Web [Imagens](#) [Grupos](#) [Notícias](#) [mais >](#)

ufsc [Pesquisa avançada](#)
[Preferências](#)

Pesquisar: a web páginas em português páginas do Brasil

Web Resultados 1 - 10 de aproximadamente 3.090.000 para ufsc (0,07 segundos)

UFSC - Universidade Federal de Santa Catarina - Brasil
UFSC - Universidade Federal de Santa Catarina. Campus Florianópolis - Santa Catarina -
 Brasil www.ufsc.br - +55 (48) 3331-9000.
www.ufsc.br/ - 2k - [Em cache](#) - [Páginas Semelhantes](#) C

[Vestibular UFSC 2006](#) - www.vestibular2006.ufsc.br/index.php?s=inscricoes
[PRPG - Pró-Reitoria de Pós-graduação](#) - www.prpg.ufsc.br/arquivos/Programasdepos.htm B
[Agecom](#) - www.agecom.ufsc.br/index.php?id=4600&uf=ufsc
[UFSC EAD > Cursos](#) - www.ead.ufsc.br/portal/index.php?section=13
[Mais resultados de www.ufsc.br >](#)

[Comissão Permanente do Vestibular](#)
 Seminário de Acesso ao Ensino Superior, Ações Afirmativas · Concurso Público para
 Professores / UFSC · Certificação LIBRAS www.prolibras.ufsc.br.
www.ufsc.br/coperve/ - 5k - [Em cache](#) - [Páginas Semelhantes](#) A

Universidade Federal de Santa Catarina - Sistema de Bibliotecas ...
Universidade Federal de Santa Catarina - Sistema de Bibliotecas - Biblioteca Universitária
 - Filiada a CBBU.

Figura 1 – Ambiente de resultado após realização da busca

Web semântica

Frente a diversas tecnologias apontadas como possíveis soluções diante do caos informacional, a *web* semântica⁸ vem ganhando abrangência, estabelecendo no-

vos padrões e metodologias para o tratamento e recuperação da informação. Seu grande objetivo é possibilitar que a informação seja mais bem representada para, conseqüentemente, ser recuperada com eficiência.

⁸ <http://www.w3c.org/2001/sw/>

Sua filosofia consiste em dividir as informações em categorias para serem entendidas por seres humanos, mas que também possuam significado para os computadores através do uso de metadados, tornando-as desse modo processáveis por máquinas. O princípio utilizado é que pessoas e computadores possam trabalhar cooperativamente, salienta Berners-Lee, Hendler e Lassila (2001).

No entanto, para que tal interação ocorra, é fundamental que as tecnologias sejam padronizadas, ou seja, que se estabeleçam padrões para descrição dos dados, além de uma linguagem comum que permita a construção e codificação de significados compartilhados.

A *web* semântica, segundo Souza e Alvarenga (2004), representa a evolução da *web* atual baseada em documentos hipertextuais escritos segundo a linguagem HTML, que só permitem a indexação automática por palavras chaves, extraídas do texto. Para Marcondes (2007), a infraestrutura da *web* semântica consiste em páginas utilizando XML, que, além do conteúdo, terão metadados utilizando vocabulários e relações muito poderosas (ontologias e RDF - *Resource Description Framework*) para expressar a semântica das novas páginas *Web*. Nesse ambiente, atuarão agentes inteligentes que irão realizar tarefas que envolvem conhecimento, raciocínio e dedução. Desse modo, espera-se a

concretização de vínculos cooperativos no intuito de que a vida dos usuários seja facilitada, suas necessidades sejam atendidas antecipadamente, uma vez que o sistema aprende com o tempo, os hábitos e necessidades de seus usuários.

5 O USO DA INTELIGÊNCIA NA WEB

A abordagem do uso de tecnologias inteligentes enquanto elemento potencial para incrementar o processo de revolução digital e informacional vem se constituindo objeto de importantes projetos de estudo. A idéia de tornar a *web* capaz de aprender (armazenar, recuperar e processar informações) de forma inteligente, similarmente a um grande cérebro global, vem ganhando consistência. Oliveira e Vidotti (2004) defendem que, para a formação de uma inteligência coletiva mais dinâmica, é fundamental que a própria rede descubra e aprenda a melhor organização para si mesma.

Considerando-se como parâmetro a mente humana, o conhecimento e significado decorrem de um processo de aprendizado em que, quanto maior o uso de determinados conceitos, mais fortemente eles se conectam. Para a *web* a análise é semelhante; com base nos caminhos mais percorridos pelos internautas, algumas conexões se tornam mais fortes, enquanto os

links pouco utilizados se tornam mais fracos.

Nesse contexto, Oliveira e Vidotti (2004, p.16) comentam que

Uma heurística simples pode propor possíveis candidatos para novos *links* (conexões): se o navegador navega de A para B e de B para C, é provável que não haja somente uma relação entre A e B mas também entre A e C e suas recíprocas. Desta maneira, novos *links* (conexões) potenciais seriam continuamente gerados mas somente aqueles que obtivessem determinada força seriam selecionados e tornados visíveis ao navegador.

Esse princípio de identificar ligações entre os *links* vem sendo utilizado pelo Google desde a implantação em seu sistema de buscas através do uso da tecnologia PageRank⁹ que possui uma proposta de ordenamento e priorização, e efetua o *ranking* dos resultados de uma busca levando como base *sites* mais referenciados por *link*. Seu uso visa ranquear as páginas em importância levando em consideração a estrutura de *links*. Primeiramente contabiliza os *links* de determinada página a outra, interpretando cada *link* da página A para a página B como um voto da página A para a página B. Posteriormente ele analisa o volume de votos que uma página recebe, e também a página que dá o voto, levando em consideração um critério de peso, em que as páginas mais importantes possuem

um maior peso na nota e também ajudam a tornar outras páginas importantes. A cada busca realizada essa nota é gravada. O Google também pesquisa e contabiliza quantas vezes a palavra pesquisada aparece nas páginas e examina o aspecto delas (conteúdo das páginas ligados a ela), tentando tornar o resultado o mais relevante possível.

Além disso, o Google também analisa a proximidade de cada termo pesquisado dentro da página, dando preferência a resultados em que os termos de pesquisa estejam mais próximos.

Sem dúvida, a tecnologia utilizada pelo Google representando a quinta geração das ferramentas de busca oferece um ambiente sofisticado para análise de informações. Baseado na classificação das ferramentas de busca proposta por Hoeschl (2006), a sexta e sétima geração ainda se encontram em período de desenvolvimento e testes.

A sexta geração compreende a junção de vários tipos de arquivos diferentes em um mesmo processo de busca (A9)¹⁰. Já a sétima geração de ferramentas de busca traz consigo uma tecnologia inovadora, baseada em ontobuscadores, mesclando tecnologias inteligentes com conhecimentos milenares e filosóficos, penetrando na essência dos conceitos e objetos.

⁹ Disponível em:
http://www.google.com.br/why_use.html. Acesso em: maio 2007.

¹⁰ <http://www.a9.com>

Um exemplo desta tecnologia é o OntoWeb¹¹, um buscador inteligente que é baseado em ontologias e técnicas de inteligência artificial, capaz de “pensar” enquanto seleciona as informações. Seu grande diferencial é que, conforme Hoeschl (2006), utiliza semânticas e estruturas valorativas para contextualizar as buscas e refinar resultados. Além disso, seu motor de busca efetua a hierarquização de conteúdos baseando-se em métricas de similaridade e engenharia do conhecimento. Segundo Hoeschl (2006), além das ontologias, o sistema utiliza diversas outras tecnologias como PCE (Pesquisa Contextual Estruturada), RC2D (Representação do Conhecimento Contextualizado Dinamicamente), técnicas de mineração de dados e raciocínio baseado em casos. Outro grande diferencial deste buscador é sua facilidade para comparar textos, pois enquanto outros buscadores como o Google aceitam até 256 caracteres, o OntoWeb permite entradas de até 7000 caracteres. Até o momento, a ferramenta possui uso apenas no campo do governo eletrônico, mas futuramente pretende-se utilizá-la como um buscador comum para todos os tipos de assunto.

A adoção de um conjunto de tecnologias inteligentes que fazem uso de semânticas, ontologias, redes neurais e inteligência artificial, parece constituir a base de e-

lementos essenciais no futuro para que o processo de aprendizagem torne-se, de fato, realidade na grande rede, propiciando sua reestruturação.

Nessa perspectiva futura de construção da Internet como um espaço voltado para a aprendizagem, apontamentos têm sido realizados acerca do surgimento de uma nova era das redes, intitulada *Web 3.0*¹². Esta nova rede atuaria como um especialista, respondendo perguntas dos usuários a partir de sua própria análise. Desse modo, os usuários não precisariam efetuar longas pesquisas para emitir conclusões porque a própria rede levantaria as melhores soluções. Um exemplo prático pode ser apresentando da seguinte maneira: um usuário quer se hospedar em um hotel. Nesse caso a *Web 3.0* efetuará consultas em suas bases de dados levando em consideração, por exemplo, os depoimentos de terceiros para recomendar o melhor hotel conforme perfil do cliente.

A *Web 3.0* enquanto protótipo tem como premissa levar em conta o sentido de cada palavra do usuário, efetuando conexões entre elas para que o resultado seja

¹¹ <http://www.ontoweb.com.br>

¹² O avanço da Internet atual (*Web 2.0*) será representado com o surgimento da *Web 3.0*. A *Web 2.0* apresenta princípios de operacionalização através de novos ambientes funcionais possibilitando facilidades quanto ao uso dos *sites* e sobretudo permitindo que os diversos usuários da rede participem diretamente na criação e publicação de conteúdos. Dentre as principais iniciativas desse fenômeno estão os *wikis* tendo como principal exemplo a *Wikipédia*.

preciso conforme o contexto do usuário. Ela atuará com inteligência e intuição.

Acerca do surgimento de uma nova fase da *web*, sendo dotada da capacidade de “aprender”, “raciocinar” e “entender”, Johnson (2003, p.94) aponta que:

A Web pode nunca se tornar auto-consciente de maneira a se assemelhar à consciência humana, o que não quer dizer que não seja capaz de aprender. Uma rede de informação adaptável, capaz de reconhecer padrões complexos, poderia vir a ser uma das invenções mais importantes de toda a história da humanidade.

No entanto, apesar da fantástica evolução tecnológica que possibilita melhoras consideráveis na qualidade das buscas e ordenamento de documentos, tais tecnologias ainda carecem de estudos visando seus aperfeiçoamentos, visto que o usuário da informação digital quer sempre mais e melhores informações.

Quando esta etapa estiver mais avançada, a organização da Internet estará mais eficiente, ou seja, o conjunto de suas redes associativas representará sistemas complexos auto-organizáveis envoltos de elementos inteligentes, gerando uma melhor otimização na rede.

6 CONSIDERAÇÕES FINAIS

Devido a fatores intrínsecos à rede, potencializados pelo avanço tecnológico, gerou-se maior interatividade e dinamiza-

ção nos fluxos de informação na Internet, rompendo o limite próprio aos meios de comunicação de massa e promovendo a participação de todos com todos, em um contexto propício ao acesso e produção de informações.

Por consequência, desencadeou-se um estado caótico na Internet devido ao grande volume de informações, trazendo impactos significativos no processo de busca e recuperação da informação.

Diante da evolução anárquica e desestruturada dos conteúdos na rede, busca-se ordenar o crescimento informacional a partir da criação e aprimoramento de métodos e padrões que viabilizem a recuperação eficiente da informação.

Destaca-se o uso de tecnologias e linguagens como os metadados, a XML e, sobretudo, as ontologias, no processo da busca inteligente, de modo que as informações relevantes sejam resgatadas com precisão, obedecendo a controles terminológicos adequados em um contexto específico. O aperfeiçoamento de tecnologias inteligentes sustentando o processo de aprendizagem, possibilitará novas perspectivas para o futuro, como por exemplo, a *Web 3.0*.

Conclui-se, de maneira geral, que, se por um lado a tecnologia tem participado ativamente no combate ao problema da recuperação da informação de forma a mini-

mizar seus impactos, por outro resta um caminho a ser percorrido em relação à representação da informação objetivando sua interoperabilidade.

REFERENCIAS

- BAWDEN, David. Information overload. *Library and information briefings*, v. 92, p.1-15, Jan. 2001. Disponível em: <<http://litc.sbu.ac.uk/publications/libs/libs92.pdf>>. Acesso em: 15 jun. 2006.
- BERGMAN, Michael K. The deep web: surface hidden value. *Journal of Electronic Publishing*, v. 7, n. 1, Ago. 2001. Disponível em: <<http://www.press.umich.edu/jep/07-01/bergman.html>>. Acesso em: 12 maio 2006.
- BERNERS-LEE, Tim.; HENDLER, James.; LASSILA, Ora. The semantic web: a new form of web content that is meaningful to computers will unleash a revolution of new possibilities. *Magazine Content*, May, 2001. Disponível em: <<http://www.scientificamerican.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21&catID=2>>. Acesso em: 19 jan. 2007.
- GESTEIRA, Ivana Lins. A biblioteca e os novos modos de convivência social. *Revista Tecitura*, v. 1, n. 1, 2006. Disponível em: <<http://tecitura.jts.br/viewarticle.php?id=41&layout=html>>. Acesso em: 12 maio 2007.
- HOESCHL, Hugo. Ontoweb: a nova era das ferramentas de busca. *Estado de São Paulo*, São Paulo, 1 fev. 2006. Caderno ConJur. Disponível em:<<http://conjur.estadao.com.br/static/text/41493,1>>. Acesso em: 22 jan. 2007.
- JOHNSON, Steven. *Emergência: a dinâmica de rede em formigas, cérebros, cidades e softwares*. Rio de Janeiro: Jorge Zahar, 2003.
- LÉVY, Pierre. *Cibercultura*. São Paulo: Editora 34, 1999.
- _____. A revolução contemporânea em matéria de comunicação. In:
- MARTINS, Francisco M.; SILVA, Juremir M. da (Org). *Para navegar no século 21: tecnologias do imaginário e cibercultura*. Porto Alegre: Edipucrs, 2003.
- LYMAN, Peter; VARIAN, Hal.R. *How much information*. 2003. Disponível em: <<http://www2.sims.berkeley.edu/research/projects/how-much-info-2003/index.htm>>. Acesso em: 20 jul. 2006.
- MARCONDES, Carlos Henrique. *Metadados: descrição e recuperação de informações na Web*. Disponível em: <<http://www.professores.uff.br/marcondes/CapMarcondes-Metadados.doc>>. Acesso em: 02 maio 2007.
- MARCONDES, Carlos Henrique; SAYÃO, Luis Fernando. Documentos digitais e novas formas de cooperação entre sistemas de informação em C&t. *Ciência da Informação*, Brasília, v. 31, n. 3, p. 42-54, set./dez. 2002.
- OLIVEIRA, Walter Clayton; VIDOTTI, Silvana Aparecida Borsetti Gregorio. Auto-organização do ciberespaço: uma visão holística. *Textos de la CiberSociedad*, n. 4, 2004. Disponível em: <<http://www.cibersociedad.net/textos/articulo.php?art=35>>. Acesso em: 03 ago. 2006.
- SOUZA, Renato Rocha; ALVARENGA, Lídia. A web semântica e suas contribuições para a Ciência da Informação. *Ciência da Informação*, Brasília, v. 33, n. 1, p. 132-141, jan./abr. 2004.
- TERRA, José C.; BAX, Marcello P. Portais corporativos: instrumento de gestão da informação e de conhecimento. In: PAIM, Isis (Org.). *A gestão da informação e do conhecimento*. Belo Horizonte:UFMG, 2003. p.33-53

VAZ, Paulo. Mediação e tecnologia. In: MARTINS, Francisco Menezes; SILVA, Jurimir Machado da (Org.). *A genealogia do virtual: comunicação; cultura e tecnologias do imaginário*. Porto Alegre: Sulina, 2004.

Claudio Henrique Schons

Graduado em Administração (UFSC) e Ciência da Computação (UNISUL)
Mestrando em Ciência da Informação (UFSC)
claudioschons@cin.ufsc.br

Title

The Internet's information volume and its disorganization: reflections and perspectives

Abstract

This paper analyzes some factors that have contributed for the Internet's evolution and its volume of information. Lately, the technologies that aim at to support the organization of information in the Internet are evaluated. Finally, some technologies that act as intelligent agents aiming to turn the Internet into an intelligent environment, with the possibility to learn and to understand the necessities of the users in an adequate context are presented.

Keywords

Information Organization; Internet; Information Systems; Information Recovery.

Título

El volumen de la información en el Internet y su desorganización: reflexiones y perspectivas

Resumen

Este trabajo analiza los factores que contribuyeron para la evolución de la Internet y de su volumen de información. Más adelante, se evalúan las tecnologías que tienen como objetivo apoyar la organización de la información en la Internet. Algunas tecnologías actúan como agentes inteligentes y tienen como objetivo convertir la Internet en un ambiente inteligente, con capacidad de aprender y comprender las necesidades de los usuarios en un contexto adecuado.

Palabras Clave

Organización de la información en la Internet, Sistemas de Recuperación en la información

Recebido em: 10.04.2007

Aceito em: 17.06.2007
