



Maior banco de dados do mundo já existe na Internet

A Internet sempre teve vocação para ser um gigantesco arquivo sobre os mais variados assuntos. Mas o segundo semestre de 2001 viu surgir serviços que poderiam ser classificados como “meta-Internet”: a grande rede arquivando a si própria e falando sobre si mesma.

No dia 11 de dezembro, o Google incorporou aos seus 3 bilhões de páginas os arquivos dos últimos 20 anos da Usenet. São 700 milhões de mensagens, desde 1981, da mais importante rede de grupos de discussão, formando um registro histórico precioso. Neste banco de informações encontram-se pérolas como o anúncio de Tim Berners-Lee's , a respeito de sua “criação” – a World Wide Web -, ou a primeira mensagem de Linus Torvalds sobre o seu “projeto de estimação”, o Linux. (Leia o [artigo](#)).

Um outro projeto, lançado publicamente em 24 de outubro, é ainda mais excepcional, apesar de não ter sido tão divulgado, pelo menos no Brasil. Trata-se da [Wayback Machine](#) (WBM), a máquina do tempo da Web. Em vez de se ler mensagens, teorias ou discussões históricas sobre a rede, pode-se *ver* sua evolução, de 1996 até hoje.

A WBM forma a maior parte do projeto intitulado Internet Archive, uma espécie de biblioteca da Internet. Assim como numa biblioteca física encontram-se livros, jornais, revistas e outras publicações com datas passadas, no Internet Archive encontram-se sites e páginas classificadas por datas, muitos dos quais já nem existem mais, além de outras formas de conteúdo digital.

Fundada em 1996 por Brewster Kahle, presidente do serviço Alexa [www.alexa.com](#) , esta biblioteca virtual espalha-se por vários servidores, formando um acervo de mais de [100 terabytes](#) (100 trilhões de bytes) e mais de 10 bilhões de páginas. A inspiração para o projeto vem da Biblioteca de Alexandria, que existiu na Idade Antiga e, diz-se, dispunha de um exemplar de cada livro publicado naquela época.

Navegar pelo acervo da WBM é excitante. Tente procurar um site que você conhecia há algum tempo e que desapareceu, ou cujo layout mudou (talvez o seu próprio site), e sinta a emoção de encontrá-lo novamente, perpetuado no tempo. Este é o grande trunfo do Internet Archive, e principalmente da Wayback Machine. Enquanto os documentos de papel podem ser preservados por séculos, os dados digitais são voláteis (não é por outro motivo que os backups são tão importantes). E os sites são os que mais mudam.

Você quer ver como era o Uol em 1996, ano em que surgiu o provedor? Confira uma de suas *home pages* , em [23 de dezembro de 1996](#). E o Terra, na época em que ainda era Zaz? Veja um exemplo de [12 de dezembro de 1998](#).

Até InfoGuerra, lançado no final de 2000, quando ainda era um ilustre desconhecido, pode ser visto com o [layout anterior](#), produzido com as limitações gráficas de seu editor (clique em “entrar”).

Os exemplos acima podem ser enquadrados na categoria de curiosidade, mas a WBM também possui registros de páginas que hoje estão perdidas nos descaminhos da Web. Você lembra do antigo provedor gratuito [Super11.net](#), desaparecido no segundo semestre de 2000? Era fã de suas colunas [Analfabytes](#),



de Honório Pacheco, ou [Mundo Charlab](#), de Sérgio Charlab? Pois você pode acessá-las novamente, clicando nos links em destaque.

Em 1999, o centro de processamento de dados do jornal Folha de Londrina, um dos principais do Paraná, sofreu um incêndio, que destruiu boa parte de seu arquivo eletrônico, incluindo as páginas de seu site. Provavelmente, nem os servidores do jornal possuem mais estas [páginas](#), mas a WBM, sim. As notícias mais antigas registradas pela ferramenta, em [16 de janeiro de 1998](#), ainda podem ser lidas na Web.

Os pioneiros da Web

A Wayback Machine elaborou sua própria coleção de sites memoráveis, a qual batizou de “[Pioneiros da Web](#)“. Nesta seção encontram-se, em seus primórdios, alguns dos sites que colaboraram para tornar a Web o que é hoje.

Por exemplo, o site da Amazon.com em [outubro de 1996](#). Fundado em meados de 1995, nesta época o site era bem diferente do atual. A página de abertura era plana, longa, com fundo branco e links azuis básicos, com poucas fotos e apenas livros em exposição.

Hoje, a Amazon.com é um verdadeiro shopping center virtual, que vende não só livros, mas também CDs, DVDs, computadores, telefones celulares, aparelhos eletrônicos, eletrodomésticos, brinquedos e até carros.

Lá também se vê o Yahoo, em [dezembro de 1996](#). Criado em 1994, em um trailer na Universidade de Stanford por dois de seus alunos, o Yahoo decididamente influenciou os rumos da Web. Entre os sites comerciais, foi um dos que menos mudou o visual ao longo dos anos, apesar de ter agregado vários novos serviços.

A revista eletrônica Feed, uma das primeiras iniciativas de jornalismo exclusivamente online, foi fundada em maio de 1995 e desapareceu em junho do ano passado, em meio ao esgotamento de recursos que assolou empresas pontocom. Uma parte da influência que exerceu, no entanto, está guardada nos servidores da WBM, em mais de 160 registros que o mecanismo fez do site. O primeiro é de [23 de dezembro de 1996](#).

Em 1991, a Web deu seus primeiros passos (em 2001 comemoraram-se 10 anos de existência da WWW e 30 do e-mail). Neste ano, um grupo de acadêmicos do laboratório de computação da Universidade de Cambridge, criou a primeira webcam do mundo. O laboratório, chamado de Trojan Room, tinha apenas uma cafeteira elétrica, cujo café era avidamente disputado pelos pesquisadores, que costumavam passar as noites em claro desenvolvendo seus estudos.

Alguns desses pesquisadores, residindo em outras partes do prédio da universidade, tinham de percorrer vários lances de escadas até chegar à cafeteria, muitas vezes para encontrá-la vazia. Foi então que eles tiveram a idéia de acoplar um captador de imagens em frente à máquina. O sistema foi ligado a um



servidor, que transmitia para todos os interessados as imagens da cafeteira atualizadas três vezes por minuto, em um circuito interno. Estava lançado o que foi chamado de *XCoffe*.

O projeto logo se tornou o principal assunto da universidade e virou tema de reportagens no ano seguinte. O captador de imagens sofreu uma pane, mas o sistema foi ressuscitado por dois pesquisadores do laboratório – Daniel Gordon e Martyn Johnson. As imagens passaram a ser transmitidas também na Web e eram vistas por centenas de milhares de curiosos. O sistema sobreviveu até 22 de agosto do ano passado, quando foi definitivamente desativado. [Dezenas de imagens](#) da cafeteira, tomadas em vários anos, ainda podem ser vistas na WBM.

Além destes e de outros sites que marcaram o desenvolvimento da Web, o Internet Archive também traz cerca de 5 mil páginas históricas relacionadas à [Arpanet](#), a rede militar surgida em 1969 e que deu origem à Internet. Há ainda coleções especiais, como a dos sites relacionados com os atentados de 11 de setembro, acervo de filmes históricos, e outras.

As implicações legais

Até aqui, foram apresentados apenas os aspectos históricos ou até mesmo lúdicos do sistema, e que dificilmente criarão polêmica. Mas o projeto tem implicações bem mais profundas. As semelhanças entre a WBM e uma biblioteca física terminam no momento em que é feita a coleta do material a ser arquivado.

Numa biblioteca convencional, as obras são compradas ou doadas, o que garante a preservação dos direitos autorais. Os servidores da WBM, ao contrário, simplesmente vasculham a Internet e copiam suas páginas, com tudo que elas contêm – textos originais, links, figuras, fotos e concepção visual. E boa parte desse material é protegida por leis de copyright.

É claro que há uma grande diferença entre um indivíduo que se apropria sumariamente de um trabalho alheio e o expõe como se fosse seu, e uma organização sem fins lucrativos, como a WBM, que cataloga este mesmo trabalho para referência. Mesmo assim, o mecanismo pode enfrentar problemas legais.

“Detentores de direitos autorais eventualmente arrastarão o sr. (Brewster) Kahle aos tribunais”, prevê Lawrence Lessig, renomado professor de direito da Universidade Stanford, em um artigo publicado pelo jornal *The New York Times*. Lessig é um estudioso da influência das novas tecnologias sobre o comportamento da sociedade, mas também um entusiasta da WBM. Tanto que já afirmou que teria reescrito seu [livro](#), *The Future of Ideas: The Fate of the Commons in a Connected World*, lançado recentemente, se tivesse visto o projeto a tempo.



Em sua opinião, a WBM pode mudar as forças envolvidas no debate sobre as leis de copyright e o acesso à propriedade intelectual de livros, música e filmes. Na sociedade digital, tais forças dividiram-se agora entre detentores de direitos autorais e seus advogados, de um lado, e professores universitários e técnicos de informática, de outro. “Finalmente temos um exemplo claro e tangível do que está em jogo”, afirma Lessig no mesmo artigo. “Brewster está definindo o domínio público”.

O advogado paranaense Omar Kaminski, que já chegou a assistir a um curso ministrado em Harvard, tendo Lessig como um dos professores, também está atento ao impacto que o novo sistema pode trazer. “A situação concreta da violação de direitos autorais pela WBM ainda não foi analisada por nenhuma corte ou tribunal, mas essa análise não tardará, do contrário estaremos admitindo que, a curto ou médio prazo, todo conteúdo veiculado na Internet é de domínio público, pode ser apropriado por terceiros e livremente distribuído, o que certamente não irá agradar aos interesses monopolistas”.

Kaminski lembra ainda que a WBM pode beneficiar os detentores de copyright, ao invés de prejudicá-los. No caso de alguém se apropriar de conteúdo alheio, “a violação ficará evidente, documentada e até perpetuada, por mais que a página original tenha sido tirada do ar”. E que a WBM poderá vir a se tornar um ótimo meio de prova judicial.

Há também a questão da territorialidade, pois a WBM está distribuída por servidores que se encontram fora da competência brasileira. “Não podemos ignorar o aspecto territorial e de legislação local. Há certas burocracias que tornam o processo ainda mais tortuoso e demorado: a necessidade de expedição da chamada ‘carta rogatória’, que é uma carta de citação, de cumprimento de ordem, ao estrangeiro, e que precisa passar por consulados. E a observação de tratados e convenções internacionais vigentes. Mas é claro que no ciberespaço isso tudo toma uma outra conotação”.

Estas questões, no entanto, ainda estão no campo da teoria. Na prática, é muito fácil para o responsável por um site impedir que suas páginas sejam registradas por mecanismos de busca, incluindo a WBM. Basta acrescentar um arquivo de texto no servidor, chamado de “robots.txt”, o qual serve para dar instruções aos softwares de rastreamento, indicando quais páginas devem ficar fora da pesquisa. Além disso, a WBM acata os pedidos para eliminar de seu banco de dados as páginas que já foram registradas, se seus responsáveis assim o quiserem.

Mais delicado ainda é o aspecto da perpetuação de sites considerados criminosos, como os de pedofilia, softwares piratas e outros, mesmo que eles sejam retirados do ar pela polícia ou por seus responsáveis. Aqui vão alguns exemplos, relacionados ou não a atividades ilícitas:

1) Em dezembro último, o FBI desbaratou uma quadrilha internacional de piratas de software, considerada uma das maiores do mundo. Chamado de DrinkOrDie (Beba ou Morra), o grupo possuía um site, no qual apresentava seus “produtos”, e que foi retirado do ar. Uma busca na WBM, no entanto, mostra [mais de 50 registros](#) do site, desde 1997, com muitas páginas ainda funcionais.

2) Pouco depois dos atentados de 11 de setembro, o governo britânico prendeu Sulayman Balal Zainulabidin, acusado de fornecer treinamento e instruções para confecção de armas de fogo, explosivos e material químico, biológico ou nuclear, os quais poderiam ser usados em ataques terroristas. Seu site,



Sakina Securities, foi fechado no mesmo dia de sua prisão, mas continua completamente **ativo** nos servidores da WBM.

3) Em dezembro de 2000, o grupo de crackers brasileiro Prime Suspectz invadiu o site www.jaru.ro.gov.br, que deveria servir à prefeitura da cidade de Jaru, em Rondônia. Inexplicavelmente, o site passou meses alterado, sem que ninguém se desse conta disso. A prova está em pelo menos dois registros da WBM, de [janeiro](#) e [fevereiro](#) de 2001.

4) Em agosto do ano passado, o Brasil foi apresentado, pelo Programa do Jô e pela revista Veja, a um “hacker” de carne e osso – Juliano Carneiro. Jô Soares chamou-o de “um dos maiores hackers do Brasil” e Veja [informou](#) (link para assinantes) que Carneiro “cobra 8. 000 reais pela proteção de uma empresa de pequeno porte”. Carneiro é responsável pelo site [Internet Segura](#).

Na época de sua aparição em público, a página que servia de apresentação dos serviços trazia uma informação enganosa para o público, além de absurda. “Somos afiliados à Truste.org e NCSA. Só trabalhamos com empresas que respeitam os direitos à publicidade”.

O site nunca foi “afiliado” à TRUSTe, principal organização de *certificação de privacidade* online dos Estados Unidos. Evidentemente, a TRUSTe também não tem nada a ver com o “direito à publicidade”, seja lá o que isto signifique. O site sofreu uma remodelação recentemente, mas a referida [página](#) ainda pode ser vista

5) Durante meses, a Microsoft manteve no ar alguns sites com nomes como Hmtest.com, Hotmailtest3.com, e outros. Eles eram idênticos ao do Hotmail e, aparentemente, serviam para testar o serviço. Os endereços nunca foram divulgados para o público. Até que, no final de 2000, um hacker australiano os descobriu e revelou o fato em seu site.

InfoGuerra enviou vários e-mails para a Microsoft, no Brasil e nos EUA, solicitando informações sobre os servidores, mas não obteve resposta. Publicou então uma [entrevista](#) com o hacker. Poucos dias depois, os sites foram tirados do ar e não mais retornaram, apesar de ainda estarem registrados em nome da companhia. Mas a WBM arquivou pelo menos um deles, o Hmtest.com. ([Confira aqui](#)).

Com tantas questões novas que podem surgir com o simples advento da Wayback Machine, parece haver uma necessidade de se criar critérios de seleção ou regulamentação das páginas a serem registradas. Mas propor leis para a Internet sempre foi uma tarefa polêmica. Omar Kaminski considera que a rede ainda está imatura para ser regulamentada, pois não se sabe qual o impacto negativo que isto trará para seu futuro.

“A necessidade ou não de regulamentação, incluindo protocolos, nomes de domínio e o fluxo em si é o âmago de toda a questão. Diz respeito a todas as medidas protetivas ou de caráter normativo que vêm surgindo antes e depois de 11 de setembro. A Internet é um meio de comunicação, sim, mas outro ponto interessante é o potencial em si da rede. Penso que a maioria ainda não se deu conta, e outra grande parte desconhece ainda esses recônditos dúbios para o Direito e para o futuro. E o mais grave é que, se nem os estudiosos sabem ao certo o impacto de uma iniciativa tecnológica do porte da WBM, quem dirá os legisladores?”.

Date Created



17/01/2002