

Obra aberta - Bancos de dados na Web anulam o mito da privacidade

Omar Kaminski*

A recente onda de "evasão de privacidade" em programas como o "Na Real" da MTV e "Casa dos Artistas" no SBT, matéria de capa do semanário Istoé - "Intimidade exposta - atrações televisivas como a Casa dos Artistas, autobiografias pouco recatadas, a Internet como palco individual e câmeras em locais estratégicos derrubam o mito da privacidade" - se mostra extremamente oportuna para um estudo, principalmente diante das recentes medidas espãs como o Magic Lantern do FBI, que já está sendo planejado para devassar a vida dos cidadãos na Internet.

A experiência de se obter uma quantia razoável de informações sobre estranhos online não é novidade - a página pessoal é uma das tradições mais antigas no ciberespaço. Mas é um prato cheio para coleções de dados estouvados e ao mesmo tempo fascinantes, tornando possível a obtenção de uma gama variada de informações acerca da vida privada de diversas pessoas.

De modo estimulante, derivado do hipercomercialismo da Internet, esses bancos de dados não irão propiciar a descoberta de fatos tais como quem comprou um computador de último tipo nos últimos 5 minutos, ou quais websites são mais populares entre mulheres na faixa etária de 14-24 anos. Em vez disso, seremos brindados com estatísticas pelas quais apenas um Freud da era dos computadores poderia se interessar, como por exemplo, com o que passaram a sonhar as pessoas após 11/09.

Esses bancos de dados são interessantes para o internauta que deseja penetrar nos recônditos da vida alheia, mas também representam tesouros para pessoas que buscam a expansão do poder de análise dos dados e de obtenção de textos (o mining).

Utilizando ferramentas especializadas na busca de padrões textuais, os pesquisadores do futuro estarão aptos a utilizar esses bancos de dados pessoais como uma forma de mapear todos os movimentos e alterações, desde desejos humanos até padrões de migração em épocas de recessão econômica.

Um lugar ideal para buscas sobre sentimentos humanos na Internet é o banco de dados do Random Access Memory ou Memória de Acesso Randômico, um repositório de mais de 13.000 pensamentos exteriorizados por usuários anônimos em um período de cerca de dois anos. Os visitantes podem buscar as recordações por assunto, data ou apelido do confidente público.

Então, por exemplo, o arquivo de 1983 do banco de dados mantém informações sobre o videogame Atari, e o assunto "virgindade" apresenta dúzias de menções desde os anos

70 até o presente; algumas estúpidas, outras eloqüentes, e outras ainda muito dolorosas do ponto de vista humano. "Perdi minha virgindade com uma vagabunda de 17 anos" alguém orgulhosamente revela. Outra confissão pública, bastante breve, diz que "não foi bem um estupro, mas definitivamente não foi consensual... Eu tinha 15 anos."

Aprofundar-se no conteúdo desses arquivos é como mergulhar no inconsciente digital coletivo, um mundo de desejos e impressões irracionais.

O designer por detrás do Random Access Memory é Eric Liftin, gerente da Mesh Architectures, uma empresa nova-iorquina especializada em construções de tijolos e cimento, bem como no design de websites. Ele conta que sua fascinação por espaços construídos inspirou-o a criar esse banco de dados bastante incomum.

"Eu estou interessado na maneira como as pessoas povoam um website como se fosse um espaço físico, real, então eu criei um banco de dados vazio e disponibilizei-o para as pessoas viverem ali", disse Liftin. "Quando você disponibiliza um banco de dados tão aberto, como as pessoas irão reagir ao preenchê-lo?" Liftin quis que o banco de dados fosse simples e convidativo, mas também que servisse para o exercício das experiências de associação livre da memória em si.

Realizar uma busca no banco de dados é equivalente a examinar os pensamentos de outra pessoa - ainda assim esses pensamentos estarão intimamente ligados a uma consciência coletiva.

Quando o banco de dados foi desenvolvido, Liftin encorajou outras pessoas a criar projetos semelhantes. Um banco de dados RAM voltado aos ataques de 11/09, por exemplo, foi erigido em <http://www.randomaccessmemorial.org>. Uma memória exteriorizada nessa coletânea, como tantas outras, exaure-se em sua própria fragilidade. "Foram tantas vidas", alguém disse, simplesmente.

Uma coleção de dados íntimos muito maior e muito mais desorganizada pode ser consultada no repositório de textos "alt.sex.stories". O grupo de discussões da extinta USENET resultou do trabalho de uma organização sem fins lucrativos, que é responsável pela hospedagem de uma enorme coletânea de histórias eróticas disponibilizadas por aspirantes a escritores, fãs e amadores cheios de excitação. Muito embora o banco de dados em si estivesse no ar desde 1996, muitas das histórias remetem ao reinado das BBS (Bulletin Board Systems) nos anos 70 e 80 e mais recentemente a páginas pessoais de caráter referencial.

Rey del Sexo, uma das duas pessoas que mantém esse site afirmou que o repositório teve início "no meu dormitório da universidade, quando eu era ainda um calouro" e a motivação foi o desejo de "ver crescer uma coleção erótica." Del Sexo deu início ao site simplesmente arquivando mensagens trocadas em um grupo de discussões moderado chamado "alt.sex.stories", mas gradativamente mais pessoas começaram a adicionar suas próprias coleções.

O que desperta a curiosidade ao se visitar o site do ASSTR é a sensação que o pesquisador cultural sente de ter obtido acesso a um banco de dados com informações sem precedentes, sobre assuntos que as pessoas normalmente ficariam muito embaraçadas em discutir de maneira franca e aberta no mundo "real". A dúvida é: como

poderia o pesquisador minerar as informações realmente úteis sobre o comportamento humano?

É neste ponto que as novas e sofisticadas técnicas de garimpagem de dados entram em cena.

Marti Hearst, uma professora de ciência da informação na Universidade de Berkeley é uma especialista em ferramentas informáticas, que possibilitam aos usuários pesquisar quantias gigantescas de textos em busca de padrões e estruturas específicas. Na maior parte das vezes ela utilizou essas ferramentas para pesquisar bancos de dados biomédicos, mas admite com jocosidade que há meios de se utilizar seus métodos de pesquisa igualmente em bancos de dados como o do "alt.sex.stories".

"Primeiro, você precisa converter este texto em um banco de dados," explica. "Nós transformamos cada arquivo em uma listagem de palavras, e para cada palavra iremos criar uma listagem de documentos onde cada uma destas palavras pudesse ser encontrada. Você pode seguir essa lógica utilizando frases também." Estas listas, por sua vez, serão inseridas nos vastos bancos de dados.

Para obter informações sobre uma frase sexual de conotação picante como "asiáticas fazendo sexo oral", por exemplo, o pesquisador poderia utilizar-se de ferramentas de busca de texto, executando uma análise estatística sobre todas as palavras e frases no banco de dados, para então descobrir quantos documentos contém essa frase, quando esses documentos foram disponibilizados, e possivelmente até, quais os tipos de palavras são comumente associadas a essa frase. As opções são surpreendentes.

"Diversos objetos que não são normalmente interpretados como se fossem verbos podem adquirir essa conotação, como acontece com algumas gírias da língua inglesa," disse Hearst. Por exemplo, após o caso Monica Lewinsky, poderemos esperar um aumento na utilização da palavra charuto como verbo (to cigar em uma tradução livre seria algo como "charutear") em coletâneas como a da ASSTR. Para as ferramentas de busca de dados de texto, verbos são uma grande fonte de problemas, quando o critério leva em conta a palavra apenas como substantivo e não como verbo.

Freqüentemente, a ferramenta deve ser "treinada" (inteligência artificial) para reconhecer substantivos que se transmutaram em verbos - pesquisadores precisam ajustar a "sintonia fina" da ferramenta toda vez que encontrarem um erro ou incongruência, o que significa "ensinar" a essa ferramenta como entender alterações particulares nos significados originais. Outro problema é o dos duplos sentidos, muito comuns na linguagem sexual.

Ferramentas de coleta de textos deverão entender e diferenciar o sentido da palavra "gostosa", por exemplo, que não se aplica popularmente apenas à degustação de alimentos.

Uma espécie muito diferente de bancos de dados online é aquela cujo conteúdo revela muito sobre a psique humana - ou pelo menos sobre a psique dos fanáticos por tecnologia (os chamados geeks) - como a coleção de centenas de milhares de mensagens disponibilizadas no site da comunidade Slashdot. Uma breve busca por "Quake III" revela centenas de mensagens e provavelmente mais informações sobre o sangrento

jogo do que qualquer pessoa poderia desejar, ou mesmo que qualquer apreciador ou fanático iria precisar.

E em São Francisco, EUA, o site da comunidade de Craig Newmark tem muito a acrescentar sobre a situação financeira de milhares de residentes que foram atingidos pela recessão econômica. Não se trata de informações sobre o psiquismo de pessoas ou sobre desejos ocultos, mas sim o de fornecer pistas de assuntos indiretamente ligados a como as pessoas estão reagindo a determinadas situações.

Pelo acompanhamento do número de vagas no mercado de trabalho nos últimos dois anos, Newmark pôde gerar gráficos que descrevem a situação. De outubro de 1999 a outubro de 2000, por exemplo, anúncios buscando engenheiros de software cresciam no ritmo de 300 a 350 por semana. Após a uma queda dramática no início de 2001, essa listagem mantinha-se gerando 150 vagas por semana.

Numa tendência relacionada, e de acordo com as estatísticas da Craig's List, ocorreu um aumento estável nas vagas disponíveis para o aluguel. O número de apartamentos colocados à disposição atingiu as alturas no último ano: de 500 por semana em outubro de 2000 para mais de 5.000.

À medida que a Internet se desenvolve e mais usuários se prendem a ela, parece inevitável que o peculiar mundo público/privado do ciberespaço irá começar a rever padrões presentes em nossos mais profundos pensamentos, bem como em nossa vida econômica e social. A Net é, apesar de tudo, uma ferramenta humana. E quanto mais buscamos, mais desejamos descobrir sobre a razão da existência.

Na esfera jurídica, que nos interessa tanto quanto a esfera comportamental, o Professor da Universidade Federal de Santa Catarina e Doutor em Filosofia do Direito, Aires José Rover, autor do livro "Informática no Direito - Inteligência Artificial. Introdução aos Sistemas Especialistas Legais" (Editora Juruá, 2001), discorre com propriedade sobre a aprendizagem, raciocínio e representação do conhecimento, apresentando modelos lógicos e matemáticos que culminam em uma nova epistemologia para o Direito, por meio da formalização, modelagem e algoritmização das informações.

Entende o Dr. Aires que, para que haja uma diminuição nas complexidades tanto técnica como administrativas do Sistema Jurídico, deve-se abranger várias ações:

"1. empenho permanente dos juristas em implementar racionalidade ao sistema, restringindo ou diminuindo o seu caráter técnico nos níveis em que racionalmente é admissível pela sociedade, bem como, traduzível por sistemas de computação. O objetivo é, assim, democratizar e popularizar o conhecimento das normas jurídicas, rompendo com a perspectiva tecnocrática do conhecimento jurídico;

2. empenho permanente dos juristas, em conjunto com os engenheiros de software, em simplificar o mundo jurídico através de sistemas inteligentes;

3. empenho permanente dos juristas, em conjunto com os técnicos de comunicação e software, em desenvolver e melhorar uma tecnologia que permita o acesso ao conhecimento jurídico às grandes massas." (ob.cit., pág. 246)

A Lei nº 8.078/90, Código de Defesa do Consumidor, prevê importantes subsídios jurídicos acerca dos bancos de dados em seus arts. 43 e 44, bem como a Lei nº 9.610/98 dos Direitos Autorais.

Por fim, não podemos esquecer da Wayback Machine, que vem demonstrando potencial para se tornar a nova Biblioteca da Alexandria. Sobre este alvo, o jurista Amaro Moraes e Silva Neto aborda com sabedoria e oportunidade em uma alocução particular:

"(...) apesar de ter sofrido tantos e sórdidos ataques durante quase mil anos, essa fonte de saberes antigos continuou resistindo até que fosse, definitivamente, destruída pelos árabes em 646 d.C. - que já haviam destruído incontáveis livros de 'magia' no Islã e na Pérsia. A lógica a justificar a destruição era simples: o importante e básico está no Alcorão. Se lá se encontra, ¿por que outros livros? Se lá não se encontra, ¿não reflete a verdade do Profeta! Logo..."

O banco de dados da Wayback Machine é tão especial que sequer conseguimos visualizar, hoje, todas as possibilidades trazidas por esse megacompêndio.

Porém a mais importante, saliente e diríamos, inocente, é a de perpetuar os avanços e documentar as evoluções das páginas e sites da Web.

(Agradecimentos especiais à Annalee Newitz, escritora de São Francisco na Califórnia, onde a experiência comparada, base para este ensaio teve lugar; e ao SF Gate, veículo de cunho jornalístico que acolheu o estudo da pesquisadora de fenômenos cibernéticos e Editora de Cultura).

Publicado originalmente na Revista Consultor Jurídico de 29 de novembro de 2001.
(<http://cf6.uol.com.br/consultor/view.cfm?numero=7650&ad=c>)

*advogado e editor de Direito e Tecnologia da revista Consultor Jurídico

KAMINUSKI, Omar. **Obra aberta - Bancos de dados na Web anulam o mito da privacidade.** Disponível em <
<http://www.direitovirtual.com.br/artigos.php?details=1&id=48>>. Acesso em 23/07/06.