SILVIA LAURENTIZ

Tags e metatags? De Ted Nelson a Tim Berners-Lee

RESUMO

Há muito tempo o homem tenta resolver um grande quebra-cabeça: como armazenar, acessar e distribuir a quantidade de informação que produz. Quais as mudanças geradas por Theodor Nelson, desde a criação do termo hipertexto, e por Timothy Berners-Lee, quando propõe uma ampliação da *web* à Web Semântica, em nossas vidas? O que são dados, metadados ou metainformação? O que e quanto nos afetam estes sistemas de indexação, buscas, classificações e dados estruturados? Qual o pensamento por detrás dessas tecnologias, e como nos afetam? Como se comportam os artistas diante de tais mecanismos? Há uma visão crítica dos artistas? Essas são algumas das perguntas que suscitaram este artigo.

PALAVRAS-CHAVE

Dados; Linguagem; Arte; Web; Agentes.

TAGS E METATAGS? DE TED NELSON A TIM BERNERS-LEE

Introdução

Talvez o maior desafio para a Web Semântica (Web 3.0) seja resolver um problema de ontologia, que à parte ser uma propriedade geral filosófica, para o contexto computacional é aquela área que definirá um vocabulário comum entre homens e máquinas para que compartilhem informação. Definir ontologias é tarefa complicada, pois prevê um conjunto de métodos e técnicas automáticas ou semi-automáticas para aquisição de conhecimento utilizando textos, dados estruturados e semiestruturados, esquemas relacionais e outras bases do conhecimento.[1]

Enquanto na Web 2.0, ao acionarmos um mecanismo de busca ("search engines"), obtemos uma lista de respostas para aquilo que procuramos, a Web 3.0 dará um passo além, e o sistema será capaz de processar as informações, filtrando a lista de respostas encontradas a partir dos interesses solicitados, à procura da informação mais relevante ao problema lançado. Na verdade, estaríamos falando de um "mecanismo de decisão"[2], considerando que este teria autonomia para algumas resoluções durante o processo. Assim, relevância, autonomia, decisão são palavras-chave para os novos procedimentos da *web*. Entretanto, muitas mudanças já haviam sido previstas por ninguém menos que Theodor Nelson, aquele que cunhou os termos Hipertexto e Hipermídia, nas décadas de 50–60, e depois, em 80, questionou esses mesmos sistemas, principalmente quanto ao modo de armazenar e distribuir a informação. É o que pretendemos demonstrar neste artigo: o desenvolvimento dessas ideias nasceu naqueles primórdios e nem todas foram implantadas até agora.

Ted Nelson: um precursor

Theodor Nelson, ou simplesmente Ted Nelson, como é conhecido, propôs, em 1981, um sistema de armazenamento e publicação eletrônico universal[3] que, considerando a literatura um sistema contínuo de documentos interconectados, deveria seguir alguns critérios para ser realmente inovador:

1. Aquilo que o usuário vê (na tela ou outro dispositivo/interface de visualização) deveria ter a mesma estrutura intrínseca do material distribuído, e não, como em

1 GÓMEZ, 2003.

2 "Decision engine", utilizando um termo da Microsoft para apresentar seu novo mecanismo de busca – Bing, que foi criado para fazer concorrência à hegemonia do Google.

3 'Proposal for a Universal Eletronic Publishing System and Archive', de seu livro *Literary Machines*, 1981. In: WARDRIP-FRUIN, Noah; MONFORT, Nick (Ed.). *The New Media Reader*. Massachusetts: MIT Press, 2003. p. 441–461.

muitos processadores de texto, algo que se amalgama neste material combinando a ele uma série de convenções obstrutivas sobre a qual estará armazenado. Essa preocupação de Nelson se justifica, pois todos os códigos, formatações, protocolos e padrões de procedimentos de um processador de texto estarão orientando/editando aquilo que o usuário lê/vê na tela e, ao mesmo tempo, desenhos, traços, rabiscos, sons, imagens, ficarão descolados do conteúdo e em bancos de dados distintos, o que dificultará o compartilhamento entre eles. Logo, o problema é criar uma forma de representação e um sistema de armazenamento tão geral que, automaticamente, permitam trabalhar com quaisquer formas de estrutura de dados que um usuário pudesse vir a solicitar; 2. Um sistema de armazenagem de texto (e outras formas de representação estão aqui subentendidas) deveria manter aquela informação que parecesse estar velha e ineficiente a princípio, mas que poderia ser muito útil no futuro (ou seja, todas as versões de um documento deveriam ser mantidas de alguma forma); 3. Uma armazenagem de texto deveria guardar cada mudança e fragmento ocorridos, assimilando cada mudança às suas atualizações e mantendo as mudanças anteriores; integrando-as, juntas, através de um método de indexação que permitisse a qualquer instante ser reconstruído; 4. O gerenciamento do sistema deveria manter, automaticamente, o caminho das mudanças e suas partes, de forma que, quando se pedisse por uma parte de uma versão anterior, ela pudesse surgir na tela. O usuário poderia, então, referir-se não meramente à versão presente do documento: ele poderia ainda voltar no tempo para qualquer outra versão anterior; 5. Este sistema partiria do princípio de que a leitura se daria na tela, e não no papel. Isso já significa que não seria necessário ter todo o documento sempre à mostra, mas apenas parte dele estaria visível na tela. Para Nelson, o erro dos programas convencionais tem sido assumir que o documento todo tem de estar ali, produzido e pronto; 6. Neste sistema proposto, poder-se-ia estocar o mesmo material em diferentes versões no tempo, como, por exemplo, os sucessivos rascunhos de um romance. Enquanto o usuário de um sistema de processador de palavras tradicional usa a barra de rolagem num documento individual, neste, proposto por Nelson, o usuário poderia usar a rolagem para acessar diferentes tempos, tanto quanto espaços, observando as mudanças em uma passagem enquanto o sistema ordenaria suas modificações sucessivamente, sugerindo um estado em processo. Além disso, poder-se-iam ter versões alternativas ordenadas e eficientemente comparadas lado a lado. Nelson chamou a isso modelo de armazenagem prismático; 7. Num primeiro estágio de armazenagem, o sistema acumularia os dados no computador assegurando os registros de partes de um todo, e não apenas grandes blocos, e os agruparia instantaneamente em versões; permitiria a criação de links de qualquer tipo e mostraria quais partes se repetem e quais são suas diferenças, entre versões relacionadas. Nelson chamou a isso hyperfile; 8. Assumindo que estamos estocando materiais desta maneira em uma estrutura

em processo contínuo, a criação de vínculos para o material tornar-se-ia muita mais ágil. Poderiam ser criados links para comentários, marcadores de páginas, notas de rodapé, notas nas margens, saltos hipertextuais e outros inumeráveis usos. Mas, da maneira que estamos trabalhando em sistemas convencionais, esses vínculos são muito duros: eles se mantêm sempre fixos, num mesmo lugar. Na estrutura visionária de Nelson, qualquer link poderia estar firmemente associado a pedaços de dados em qualquer parte da estrutura, e este acompanhando-os para qualquer lugar que o enviassem e/ou o atualizassem. Qualquer tipo de link poderia ser criado dessa forma, e estes migrariam quando mudanças ocorressem a seus vínculos; 9. Qualquer documento poderia ser citado, mas nenhuma cópia do documento seria feita. Um símbolo de vínculo-citação (ou algo equivalente) seria colocado no percurso do link armazenado; desta forma, nenhuma cópia seria feita, e isso significa que nada afetaria a propriedade da autoria. E, desde que o material citado apenas residisse em seu lugar de origem, e não no documento que o citou, outros documentos que o citassem poderiam ser automaticamente atualizados quando seus autores os mudassem. Nelson cria então o conceito de *transclusão*, que é a inclusão de uma parte de um documento em outro documento por referência; 10. A integridade de cada documento seria mantida por esta separação: documentos derivados seriam permanentemente definidos nos termos de seus originais e de suas mudanças. Um documento poderia consistir meramente de mudanças de outros documentos. Assim, um artigo publicado em uma revista que fora modificado poderia ser pensado como sendo dois documentos: o original e outro, com as mudanças ocorridas nele; 11. Questões de versões alternativas e autoria também receberiam um toque inovador: um autor de um documento poderia criar arranjos alternativos do mesmo material, tudo num mesmo documento. Um outro usuário, entretanto, estaria livre para criar sua própria versão alternativa de um documento que não é de sua autoria. Mas isso aconteceria em um documento à parte, em uma outra janela do mesmo material. Documentos poderiam apontar relações entre outros documentos, que seriam chamados de documentos intercomparados. Poderia também ser criado documento composto e assim por diante… 12. Poderíamos ter um sistema de publicação eletrônico que alimentaria sua tela de computador com exatamente aquilo que você procurasse, tão logo fosse procurado, com *royalties* divididos entre o autor do documento na exata proporção de quanto seu material estaria sendo transmitido ou usado. Materiais privados seriam disponíveis apenas para seus próprios autores ou projetos; material publicado seria disponível para todos, rendendo os direitos autorais para seus autores. Documentos privados poderiam lincar e abrir uma janela para documentos públicos. Nelson cria ainda o termo *Transcopyright,* ou seja, a pré-permissão para republicação virtual; 13. Cada leitor deveria ser capaz de saber quais documentos estão conectados com o documento que está lendo. Entretanto,

isso poderia ser muito extenso. Por exemplo, imagine quantos vínculos já foram criados para *Alice no País das Maravilhas*? Deve haver milhões de links atualmente sobre esse tema. Por isso, uma espécie de filtragem deveria ser necessária. E, se houvesse filtros, deveria haver também distribuição em diretórios e categorias que orientassem os conteúdos filtrados. Não há nada errado com categorizações, diz Nelson; entretanto, elas possuem natureza passageira: sistemas de categoria têm uma meia-vida, e categorizações "começam a parecerem estúpidas" depois de alguns anos. Para resolver esse problema, deveríamos manter diretórios de categorizações fora do nível do sistema!; 14. Há ainda uma importante distinção entre "back-end" e "front-end" entre as propostas seminais de Nelson. Funções "front end" são aquelas particulares para um computador individual, enquanto que funções "back end" são as que se dão no nível amplo da rede; o local onde uma operação ocorre é crucial. Por exemplo, um usuário pode ajustar seu "front end" para filtrar certos tipos de conteúdo e este será um exercício de preferência. Mas, quando este conteúdo é removido do nível da rede, ele pode ser mais propriamente chamado de censura (*censorship*). Ou seja, uma coisa é filtrar informações que não interessam, outra coisa é *ela ser retirada da rede por uma censura*.

Essas foram ideias que inquietavam aquele que criou, nos anos 50, o termo hipertexto, e que até hoje não foram implementadas totalmente, mas que já nos deixam antever o que seria a grande inovação na *web*: a Web Semântica.

Tim Berners-Lee e a Web Semântica

Também no início dos anos 1980, Tim (Thimoty) Berners-Lee começou a desenvolver um sistema ligado em rede para publicação eletrônica de trabalhos científicos. Esse sistema, chamado *Enquire*, fora projetado para armazenar, recuperar, e vincular documentos pela rede de computador do CERN[4], na Suíça. Não chegou a ser finalizado, mas – influenciado por experiências com hipertexto, publicação digital e rede aberta – Berners-Lee expandiu seus conceitos e passou a explorar uma maneira de como um sistema de hipertexto poderia funcionar com a Internet. Por iniciativa própria, em 1990 Berners-Lee completou seu primeiro browser e software de servidor de *web*. Em 1991, começou a distribuir este software, então chamado de World Wide Web.

No cenário atual da *web*, desde seu surgimento, podemos notar que muitas das propostas de Nelson, lançadas nos anos 80, estariam hoje sendo anunciadas como as promovedoras de um novo estado da *web*, a que Berners-Lee, em 2001, denominou de Web Semântica e que ainda estaria por vir. Berners-Lee (2001) explica as inovações ocorridas na *web* a partir do seguinte exemplo:

4 Organização Europeia para Investigação Nuclear, mais conhecida pelo acrônimo CERN, do francês Conseil Européen pour la Recherche Nucléaire (Conselho Europeu para Pesquisa Nuclear), é um centro de estudos sobre física de partículas.

O sistema de entretenimento de sua casa estava tocando uma música dos Beatles, quando o telefone tocou. Quando Pete atendeu, seu telefone diminuiu o som enviando uma mensagem para todos os outros dispositivos locais que tivessem controle de volume na casa. Sua irmã Lucy estava ligando do consultório do médico: – "Mamãe precisa ver um especialista e precisará de uma série de sessões de terapia. Quinzenal ou algo assim. Eu farei meu agente marcar as consultas". Pete imediatamente concordou e disse que poderia ajudar a levá-la nas sessões. Do consultório do médico, Lucy instruiu seu agente da Web Semântica através de seu palm top. O agente prontamente recuperou do agente do médico as informações sobre o tratamento prescrito para a sua mãe; procurou em várias listas de provedores; verificou aqueles que atendiam pelo seu plano de saúde dentro de um raio de 20 milhas de sua casa; e, entre aqueles que tivessem recebido uma avaliação excelente ou muito boa em serviços de avaliação confiáveis. O agente então começou a tentar combinar entre tempos de consultas disponíveis (fornecidos pelos agentes de provedores individuais através de seus Sites da Web) e horários disponíveis da agenda de Pete e Lucy. Em alguns minutos o agente apresentou a eles um plano. Pete não gostou, porque o Hospital da Universidade era muito distante da casa de sua mãe, e ele teria que atravessar a cidade, e, além disso, na hora do rush. Ele deixa seu próprio agente fazer novamente a procura com preferências mais rígidas sobre local e horário. O Agente de Lucy, tendo completa confiança no agente do Pete no contexto da presente tarefa, automaticamente ajudou-o através de certificados de acesso e atalhos que ele já tinha percorrido. Quase imediatamente, o novo plano era apresentado: clínicas mais próximas, e horários mais cedo, foram apresentados, mas com duas notas de advertência: Primeira, Pete teria que replanejar alguns de seus compromissos menos importantes. Ele verificou que isto não seria um problema. O outro problema era algo sobre a companhia de seguro não incluir este provedor entre os terapeutas cadastrados: neste sentido o agente assegurou que, apesar de não constar entre os assegurados, poderia atendê-la pelo plano sim, e, em seguida, apareceu um link para maiores detalhes. Lucy registrou seu consentimento no mesmo momento que Pete solicitou: "poupe-me dos detalhes"; e estava tudo acertado. (Claro, Pete não podia resistir aos detalhes, e mais tarde ele acessa seu agente novamente e pede para explicar como achou aquele provedor, embora não estivesse na lista adequada). [tradução livre].

5

Metadado é o termo utilizado para designar "dado sobre dado". Para exemplificar, vejamos o que acontece com um documento de título "Minha página": a representação HTML desta informação (ou dado) seria <title>Minha página</title>. Neste caso, o conjunto de tags <title> e </title> tem a força de metadados, já que descreve que "Minha página" é o título da página.

6

RDF é um sistema para auxílio ao desenvolvimento de metadados (veja nota anterior). RDF é composto por três tipos de objetos: recursos, propriedades e triplas. Um recurso é o que será descrito por uma expressão RDF. Todo recurso é identificado por um URI (Uniform Resource Identifier, incluindo o Uniform Resource Locator – URL). Uma propriedade é qualquer característica utilizada para descrever um recurso. Uma tripla é formada por um recurso, uma propriedade e um valor para a propriedade daquele recurso.

7

O World Wide Web Consortium (W3C) começou a trabalhar em meados da década de 1990 em uma linguagem de marcação que combinasse a flexibilidade da SGML (Standard Generalized Markup Language) com a simplicidade da HTML. O princípio do projeto era criar uma linguagem que pudesse ser lida por software e integrar-se com as demais linguagens. Esta é a função de XML.

A Web Semântica pretende ampliar os princípios da *web*. Essa extensão permitirá que dados sejam eficazmente compartilhados entre comunidades e automaticamente processados, seja por ferramentas, seja manualmente pelo homem.

Algumas características da Web Semântica: 1. Permitir aos dados emergirem na forma de dados reais, assim, um programa não tem que se privar de sua formatação, imagens, anúncios de uma página da *web* e o sistema, sozinho, acharia onde os dados e o conteúdo estão; 2. Permitir que pessoas escrevam seus arquivos que explicariam para uma máquina a relação entre conjuntos diferentes de dados; 3. Permitir às máquinas seguirem vínculos e, consequentemente, integrar dados de muitas fontes diferentes automaticamente.

É quase imediato relacionarmos tais características com algumas das propostas de Nelson já apresentadas. Mas, para que estas metas fossem atingidas, muita tecnologia teria que ser implementada desde aqueles primórdios. Como exemplo, podemos citar que alguns conceitos importantes que participam da Web Semântica patrocinariam tais características, como: Metadados[5]; RDF Resource Description Framework[6]; XML[7]; CSS (Cascading Style Sheets)[8]; SGC – Sistema de Gerenciamento de Conteúdo[9]; Mashup[10], OWL[11].

Além disso, cada conceito em si já é complexo, a filosofia do XML, por exemplo, seria incorporada por vários outros princípios importantes: a) separação do conteúdo da formatação; b) simplicidade e legibilidade, tanto para humanos quanto para computadores; c) possibilidade de criação de ***tags*** sem limitação; d) criação de arquivos para validação de estrutura (chamados DTDs); e) interligação de bancos de dados distintos; f) orientação a partir da estrutura da informação, e não da sua aparência; etc.

Em 2004, o termo Web 2.0[12], cunhado em uma conferência da O'Reilly Media, referiu-se a uma assim chamada "segunda geração" de aplicações *web*, caracterizadas por um grau maior de interação e colaboração entre usuários. O termo passou a ser constantemente utilizado pelo mercado, através do rápido crescimento do número de blogs, comunidades virtuais, Wikis e outras aplicações. Em *What is the Web 2.0* (http://www.oreilly.de/artikel/web20.html), Tim O'Reilly define o conceito a partir de um conjunto de princípios e práticas: 1. Web como plataforma de serviços: oferta de serviços, e não pacotes de software; 2. Arquitetura focada em participação; 3. Escalabilidade; 4. Mistura de fontes de dados e de transformação de dados; 5. Software utilizável em vários tipos de dispositivos; 6. Aplicações que atuam como potencializadores da inteligência coletiva.

É nesse contexto que os *mashups* se inserem, por exemplo, e são considerados um dos tipos de aplicação da chamada Web 2.0. Podemos citar, entre eles, Flickr (um site da *web* de hospedagem e partilha de imagens fotográficas e, eventualmente, de outros tipos de documentos gráficos, como desenhos e ilustrações); Organizr (uma ferramenta para organização de fotos, grupos, coleções e suporte à localização no

8 É uma linguagem de estilo utilizada para definir a apresentação de documentos (por exemplo, fonte, cor, espaçamento, cor de fundo…) escritos em uma linguagem de marcação, como HTML ou XML. Seu principal benefício é prover a separação entre o formato e o conteúdo de um documento. Em vez de colocar a formatação dentro do documento, o desenvolvedor cria um link (ligação) para uma página que contém os estilos. Quando se quiser modificar o estilo, basta alterar esta página e todos os documentos que a seguem se modificarão automaticamente.

9 Um Sistema de Gerenciamento de Conteúdo – SGC (em inglês Content Management Systems – CMS) é um gerenciador de websites, portais e intranets que integra ferramentas necessárias para criar, gerenciar (editar e inserir) conteúdo em tempo real, sem a necessidade de programação de código, cujo objetivo é estruturar e facilitar a criação, a administração, a distribuição, a publicação e a disponibilidade da informação. O grande diferencial de um CMS é permitir que o conteúdo de um website possa ser modificado de forma rápida e segura de qualquer computador conectado à Internet.

10 Um *mashup* é um website ou uma aplicação *web* que usa conteúdo de mais de uma fonte para criar um novo serviço completo. O conteúdo usado em *mashups* é tipicamente código de terceiros por meio de uma interface pública ou de uma API. Mais adiante, no corpo principal do texto, haverá exemplos de *mashup*.

mapa do Yahoo! Maps – chamado de Geotagging, ou Georreferência –, semelhante ao Google Maps); Google Maps (um serviço de pesquisa e visualização de mapas e imagens de satélite da Terra gratuito na *web* fornecido pela empresa Google); YouTube (um site na Internet que permite que seus usuários carreguem, assistam e compartilhem vídeos em formato digital); Del.icio.us (O del.icio.us é o nome de um site que foi desenvolvido por Joshua Schachter e entrou no ar no final de 2003. Ele oferece um serviço on-line que permite que você adicione e pesquise bookmarks sobre qualquer assunto. É uma ferramenta para arquivar e catalogar seus sites preferidos para que você possa acessá-los de qualquer lugar); Yahoo Pipes (possibilita a integração de múltiplas fontes e bancos de dados partindo das preferências do usuário em que cada um monta seu próprio espaço *mashup*).

E, finalmente: WIKI

WikiWeb (fundada pelos princípios da Web 2.0), ou simplesmente Wiki, permite que os documentos sejam editados coletivamente com uma linguagem muito simples e eficaz, utilizando apenas um navegador *web*. O maior exemplo aqui é a *Wikipedia.org*, a enciclopédia livre e colaborativa da *web*. Uma das características definitivas da tecnologia Wiki é a facilidade com que as páginas são criadas e alteradas – geralmente, não existe qualquer revisão antes de as modificações serem aceitas, e a maioria dos Wikis são abertos a todo o público, ou pelo menos a todas as pessoas que têm acesso ao servidor Wiki.

Além disso, podemos trabalhar na Wiki com sistemas de gerenciador de conteúdo (CGC), folhas de estilo (CSS), *templates*, acrescentar *mashups*, trabalhar com metadados (RDF), etc. Outra vantagem é que a Wiki trabalha com versões. É possível guardar e acessar as versões anteriores, e o sistema registra as modificações ocorridas na versão atual. Dessa forma, caso seja necessário voltar à última versão, isso é facilmente arranjado. Outro problema sanado com esse artifício é que podemos detectar rapidamente algum caso de vandalismo que estiver acontecendo no sistema. No momento em que constatarmos algo irregular nas modificações ocorridas, rapidamente voltamos à versão anterior. Isso sem que tenhamos que repassar página a página o conteúdo do site, pois o sistema mostra as modificações ocorridas (e não a página modificada inteira). Basta acompanharmos os acessos e as modificações, para mantermos o sistema aceitável e estável.

Primeiras considerações

O interessante é que as ideias de Nelson e Berners-Lee ainda não foram totalmente colocadas em uso. Algumas mudanças de percursos agora estão sendo reavaliadas,

11 OWL é uma linguagem de marcação semântica para publicação e compartilhamento de ontologias, capaz de descrever classes e relacionamentos entre elas.

12 Esta ainda não atingiria a Web Semântica, que viria a ser denominada Web 3.0, embora tenha sido apresentada, enquanto conceito, anteriormente, em 2001.

muitas questões sendo retomadas e, com certeza, ainda estamos caminhando para grandes transformações neste assunto. Problemas como autoria, censura e relevância de conteúdo estão sugerindo novas formas de ação e envolvimento coletivo. Estilos, padronizações e design gráfico estão sendo remodelados pela hipermídia adaptativa, que oferece ao usuário a escolha e a definição de seus próprios interesses e preferências. E "agentes inteligentes", programas especializados, passam a mediar ações e a definir tarefas a partir de novas formas de categorizações e filtragem de informações.

Algo ainda a destacar, a ideia de que a máquina possa vir a compreender a linguagem humana, subentendida pela Web Semântica: significa dizer que a *web* será capaz de criar inferências, deduzir, gerar raciocínio lógico e abstrato, relacionar e processar informação, interagir de forma cooperativa no conhecimento. O que, em última instância, significa também dizer que a máquina será capaz de interpretar signos, reconhecer e gerar novos significados – pois ***reconhece*** o significado dos documentos e ***infere*** novos conhecimentos a partir deste reconhecimento – e, ainda mais, representar associações "abstratas" entre coisas. Isso significa que uma questão primordial será definir padrões e relações entre dados, para que se possa ter um entendimento comum passível de ser compartilhado entre todos ou, pelo menos, entre um mesmo domínio. E, para se definirem padrões e relações entre dados, é necessário, em primeiro lugar, uma conceitualização, ou seja, uma "visão abstrata" daquilo que queremos representar a partir de um propósito. E ontologia é a responsável para se especificar essa conceitualização. Mas a ontologia precisa de mecanismos para que se torne eficiente. Como nos explica Júnio César de Lima, a representação de conceitos acontece

> *através de uma taxonomia e um conjunto de regras de inferência. A taxonomia define as classes de objetos e as relações que se estabelecem entre elas. A camada lógica é responsável por definir mecanismos para se fazer inferência sobre os dados. Uma ontologia define um conjunto comum de termos que são usados para descrever e representar um domínio, como medicina, biblioteca etc. Estas definições de conceitos básicos de domínios e seus relacionamentos são processáveis por computadores e são expressas em linguagens baseadas em lógica.*[13]

Mas é mais complicado do que isso: 1. A palavra "perna" refere-se a uma "perna humana" ou a uma "perna de cadeira"?; 2. Como explicar para uma máquina que "carro" pode ser o mesmo que "veículo" em determinado contexto?; 3. Como saber os diferentes contextos (pessoal, profissional, social, cultural) de que uma "pessoa" possa estar participando? Como evitar, por outro lado, a reutilização de contextos

13 LIMA, 2005.

já utilizados anteriormente, sobrecarregando o sistema?; 4. Como diferenciar para uma máquina o que é uma "rosa-flor", de uma "rosa-dos ventos", de uma "Rosa-nome de uma pessoa", e ainda dizer que tudo pode se modificar a partir do contexto no qual se insere? Como interpretar, reconhecer e criar associações através de dados não estruturados?[14]

Retomando as palavras de Nelson, categorizações "têm uma meia-vida, e 'começam a parecerem estúpidas' depois de alguns anos" e, somado ao fato de que o conhecimento na *web* não é estático: há mudanças de domínios, adaptações a diferentes tarefas, ou mesmo mudanças na conceitualização, já podemos perceber que há um grande desafio ainda a se resolver. E essa deve ser uma ação de todos, entre diferentes áreas do conhecimento, e nossa proposta agora é demonstrar a participação crítica de alguns artistas brasileiros nesse contexto, a partir de alguns trabalhos.

Tags, mecanismos de buscas, serviços

Antes, vamos "olhar" mais de perto algumas destas novas características da *web* – com a intenção de nos aproximarmos um pouco mais dos trabalhos artísticos que serão apresentados –, a partir de um exemplo cotidiano: quando estamos procurando por algo na Internet, utilizamos um mecanismo de busca que realiza uma pesquisa entre milhões de páginas existentes (dentre aquelas já indexadas). O resultado, geralmente apresentado por ordem de relevância, constitui-se de uma lista de URLs, cujo número varia de poucas unidades até milhares. Na maioria dos casos, apenas os 10 resultados mais relevantes são exibidos na primeira página do resultado. É aí que começa a ditadura dos TOP-10.

Martha Gabriel[15], pesquisadora e artista, apresenta um estudo detalhado sobre esse assunto, além de nos oferecer sugestões de como podemos otimizar os sistemas de busca na *web*. Para dar conta da análise de trabalhos de arte na *web*, inicialmente apresentaremos algumas das ideias lançadas em seu livro *Marketing de otimização de buscas na web*[16]. Uma pessoa ou organização que possuam uma página na *web* e queiram ser encontradas por usuários interessados em suas informações procuram, evidentemente, estar na lista TOP-10. Isso porque a maioria dos usuários que procuram por uma informação não passa da primeira página (ou no máximo vai até a segunda). Assim, estar listado na décima primeira posição do Google, por exemplo, pode significar deixar de receber a visita de muitos usuários, já que, a cada página, são lançados 10 endereços por vez. A primeira coisa que podemos perguntar: quais são os fatores de relevância entre os mecanismos de busca e por que mecanismos de busca diferentes podem apresentar resultados divergentes na primeira página? A princípio, dois fatores afetam o posicionamento de um site em uma busca na *web* – sua relevância (ranking de sua página) e as palavras-chave relacionadas a ele.

14 As imagens de vídeo são dados não estruturados. Um grande empenho nesta área tem sido conseguir fazer com que se identifiquem as faces de pessoas em imagens capturadas por câmeras de vigilância em locais públicos, por exemplo. Tal reconhecimento seria um feito da computação e inteligência artificial, num esforço de estruturar os dados capturados das imagens de vídeo, que originalmente são de natureza complexa e não estruturada.

15 GABRIEL, 2008.

16 Ibidem.

Relevância é um índice de 0 a 10 que mede a "importância" de uma página. Os fatores que medem a "importância" da página são todos em função do relacionamento que a página mantém com outras páginas na *web* e do fluxo que a página atrai para si. Assim, quem e quantos apontam para a sua página e para quem e quantos você aponta determinam esse relacionamento. Portanto, a quantidade de links que sua página possui significa que se relaciona com outras e, por isso, seus conteúdos são considerados mais relevantes do que os daquelas páginas soltas ou isoladas.

Links apontando para a sua página demonstram a sua popularidade, este é o pensamento que está por detrás do sistema, explica-nos Gabriel[17]. Quanto mais links apontam para uma página, mais créditos ela possui. Se estes links forem de websites com alta relevância, isso também aumentará a relevância da sua página. Se uma página está conectada à sua e esta página aponta para poucos links, você deve ser mais relevante do que se sua página é apenas uma entre centenas de links na página que aponta você.

Só que há também outros fatores aqui envolvidos, que não são tão ideais assim – ou são, isso é uma questão de ponto de vista. Por exemplo, alguns "buscadores" vendem palavras-chave, de modo que, quando algum usuário busca aquele assunto relacionado àquela palavra, receberá o resultado em uma área de destaque, a dos "links patrocinados", aparecendo em uma coluna separada dos demais resultados. Além disso, podemos criar algo na sua página como se fosse uma espécie de "isca" para atrair links para ela. Palavras-chave posicionadas no título ou no início de um documento resultam em um maior valor de ranking. Além destes, é levado em consideração também o número de vezes que as palavras-chave ocorrem no corpo do documento. E, partindo da hipótese de que uma página mais acessada que as demais é provavelmente mais relevante, esta também estará com maior contagem no ranking das TOP-10.

É possível, através do uso de uma série de ferramentas automáticas e semi-automáticas de análise, diagnóstico e aconselhamento, adequar um website com o objetivo de que este apareça em uma melhor posição para uma dada busca, conclui Gabriel[18]. Uma forma de otimização, continuando com as sugestões da autora, é criar uma página de entrada no site que possua palavras-chave atrativas, e mesmo falsas, para conseguir bom posicionamento nas buscas. Abarrotar de palavras-chave a página para tentar conseguir um posicionamento melhor também é usado. O uso de "Texto Invisível" também funciona, o que consiste em escrever palavras-chave na página com a mesma cor de fundo da página. As palavras escritas na mesma cor do fundo da página são "vistas" pelos mecanismos de busca, mas não pelos usuários no navegador (browser). E, ainda, podemos contratar empresas que oferecem o serviço de submeter a sua URL para todos os mecanismos de busca, em curtíssimo prazo.

17 Ibidem.

18 Ibidem.

Digital oracles

"*Verdadeiros oráculos digitais*" é como Martha Gabriel, agora a artista, encara os mecanismos de buscas. Em sua proposta poética *Digital oracles*[19], Gabriel cria um paralelo entre oráculos, que ajudam desde os tempos mais remotos o homem a escolher seus caminhos, e os mecanismos de buscas, que, sem que percebamos, nos auxiliam em nossas escolhas e influenciam nossas vidas, na sua forma digital, como o Google, o Yahoo!, o Altavista, etc. Trata-se também de um alerta, pois nos faz perceber que, muito provavelmente, a maioria das pessoas que acessam tais "oráculos" não tem consciência do controle que possuem sobre suas escolhas e ações. A influência que esses "buscadores" têm é tamanha, principalmente porque não nos damos conta desse poder deles sobre nós. Ou seja, "seus conselhos" são tão poderosos quanto aqueles dos oráculos antigos, mas sua força se dá subliminarmente – o que deflagra algo de maior potência, pois, quando sabemos de onde vem uma força, podemos tomar uma atitude sobre ela, enquanto que não saber de onde vem nos torna meras ferramentas, vítimas de suas orientações. Evidente que essa é uma interpretação paradoxal, pois, dada a complexidade da Web hoje, não conseguiríamos navegar sem estes "buscadores/auxiliadores". Portanto, há um caráter libertador também! Nesse contexto, a artista vem questionar esta "ditadura dos TOP-10", a relação entre "sobrevivência e existência" na rede, a real "consciência de quem procura um resultado" na *web*, e o "poder, controle e confiança" desse resultado.

Quando acessamos o site deste trabalho, algumas palavras vão surgindo na tela, sucessivamente e em diferentes locais, que orientam estas dúvidas, tais como: "utilidade", "poder", "ditadura dos TOP-10", "confiança", "privacidade", "manipulação". Aparentemente, essas palavras aparecem numa sequência aleatória, que se repete no tempo de duração de seu acesso, num reforço constante de que algo ali deve ser questionado. Principalmente pelo teor das palavras, que nos remetem à sugestão da dúvida.

Em seguida, há o apelo da participação, quando se é solicitado a: "Digite sua pergunta aqui"; e logo ao lado: "Escolha um oráculo". Entre os oráculos possíveis de escolha estão Google, AOL Search, Cadê?, eXcite, Ask, msn, altavista, HotBot, UOLBusca, Yahoo!, mamma, dogpile, Vivisimo. Assim que uma pergunta for feita e um buscador escolhido, o sistema enviará para aquele buscador solicitado e mostrará a primeira página da busca desejada. Quando se encerra a pesquisa, o sistema fecha a janela e apresenta as seguintes frases:

"Something to think about…"; "Você CONTROLA as suas ESCOLHAS?"; "Não estar entre os resultados significa NÃO EXISTIR?"; "Por que os resultados TOP 10 são os TOP 10?"; "Alguém consegue MANIPULAR os resultados?"; "Existe PRIVACIDADE durante a busca?".

19 GABRIEL, Martha. Disponível em: <http://www.digitaloracles.com.br/>.

YouTag

YouTag[20], de Lucas Bambozzi, mixa imagens recuperadas da rede por mecanismos de busca a partir de palavras digitadas pelo usuário, que recebe em troca um vídeo através de seu e-mail. A ideia de tags e indexadores está relacionada também à problemática dos mecanismos de busca já comentados. O artista, então, traz para a estética questões da remixabilidade, que por sua vez está sobrecarregada de referências tanto modernas quanto pós-modernas, como: apropriação, colagem, fotomontagem, sampleamento, as questões lançadas pelos DJs e pelos VJs e suas ferramentas de ação em tempo real e, agora, a ideia expandida de uma "remixagem coletiva". Mas é mais interessante que isso, pois faz de todos nós, que já postamos alguma imagem (ou vídeo) na rede, colaboradores. Em consequência, nos transforma em tags também…

Como acontece: o usuário pode participar escrevendo 3 palavras, um título ou uma frase. A partir deste momento, estas serão as palavras-chave utilizadas pelo mecanismo de busca, que selecionará vídeos a partir delas. Um processo de remixagem entre as imagens gera um novo vídeo, remix de 3 resultados oferecidos pelos buscadores, que é enviado para seu e-mail. Dessa forma, conforme explicado no site, o *youTag* cria um vídeo a partir de um título de sua preferência. Você digita o título e seu vídeo customizado será enviado para o seu e-mail em alguns minutos.

Os blogs são formas de publicação onde qualquer pessoa pode dispor e emitir suas ideias, seja como diário pessoal, seja como informações jornalísticas, vídeo, foto, etc.[21] Temos também os serviços do Youtube, do Flickr, etc. As redes, portanto, disponibilizam o compartilhamento e a troca de arquivos de diversos formatos em qualquer lugar do mundo. E, através de buscadores, podemos interceptar, acessar, recombinar, otimizar todas estas informações. Isto, por si, já bastaria para percebermos fatores importantes. Mas não podemos esquecer os processos de filtragens ocorridos na informação. Há uma tendência no sistema, formulada pelo mecanismo de busca e seus critérios de relevância, e palavras-chave associadas, como já explicado anteriormente. Há também todas as questões sobre o modo como estão sendo elaboradas as páginas na tentativa de otimizar seus rankings. Então, o que a princípio estaria se apresentando como um padrão estético pelas imagens e pelos vídeos gerados, passa por um desvio ou interferência que o Marketing de Otimização de Buscas possa estar causando. É interessante que no trabalho há a seguinte explicação:

> *Em tempo de tags, metatags e indexadores de busca, o quê é o nome da 'coisa' e o quê é o nome possível da representação da 'coisa'? O que acontece nas vísceras dos search engines? Que tipo*

20 BAMBOZZI, Lucas, 2008. Disponível em: <http://www.youtag.org/>.

21 Podemos inserir aqui também o Twitter enquanto um miniblog, mas este será um assunto para um próximo artigo.

> *de sentido é produzido quando tudo passa a ser regido pelas formas de indexação atuais?*[22]

O que parece é que este trabalho, ao mesmo tempo em que utiliza seus recursos, critica o sistema. Aparentemente, há um processo interno que seleciona aleatoriamente as imagens recuperadas pelo mecanismo de busca, o que, num certo sentido, funcionaria como um equilibrador, uma retroalimentação de ajuste do sistema, que poderia suavizar a dinastia dos Otimizadores de Buscas. É instigante, pois alerta que há muito que pensar sobre os "indexadores" na Web, o "nome da coisa" é um processo muito mais complicado do que os entusiastas da Web Semântica parecem supor e, acima de tudo, a filtragem de informação está nos oferecendo padrões estéticos que estão sendo julgados por indicadores de preferências que não são, exclusivamente, estéticos!

Freakpedia

A *Freakpedia*[23] é uma experimentação artística na rede Internet de Fábio Fon e Edgar Franco. Faz uso da tecnologia WIKI, que possibilita que qualquer um crie, edite e insira conteúdos a qualquer tempo.

Como os próprios autores explicam, "uma enciclopédia digital e colaborativa na Internet – em que todos podem participar – propõe não só um novo meio de múltiplos autores como também uma nova maneira de encarar relevâncias".[24] Este trabalho é uma crítica declarada à liberdade proclamada pela *Wikipedia*, enciclopédia digital altamente reconhecida da *web* – e já comentada anteriormente. Contam os autores que, após constatarem a fragilidade nos processos de eliminação da informação da enciclopédia eletrônica, por falta de "relevância" do conteúdo, apontada pelos editores administradores, nasceu a ideia do trabalho *Freakpedia*:

> *tudo aquilo que eles acreditam não ter 'relevância' suficiente para permanecer no interior de sua magnífica publicação está sujeito a ser eliminado e estar banido permanentemente. Nesta prática não há critérios bem definidos e muitas vezes, essa eliminação está sujeita ao repertório pessoal e o desconhecimento dos assuntos tratados.*[25]

Após algumas eliminações registradas na *Wikipedia* – em qual estiveram envolvidas algumas das produções dos próprios autores –, Edgar Franco e Fábio Oliveira Nunes, artisticamente conhecido por Fábio Fon, tiveram a ideia de criar um espaço no qual a relevância não seja motivo de censura, excluindo-se apenas conteúdos que possuam restrições de direitos autorais:

22 BAMBOZZI, 2008.

23 Disponível em: <http://www.freakpedia.org/>.

24 NUNES, 2007.

25 Ibidem.

> *Na Freakpedia abre-se espaço para falar das pequenas coisas, daquilo que possui importância muito pessoal e que não possui qualquer pretensão grandiosa. Tudo que estiver ali presente é obrigatoriamente insignificante. Assim como na Wikipédia, qualquer internauta está livre para publicar verbetes e editar verbetes já existentes, basta que inscreva-se no site.*[26]

Projeto *Assina*

O Projeto *Assina: do texto ao contexto,* de Cícero Inácio da Silva, infelizmente não está mais on-line, mas é um importante questionador das redes. Esse projeto propõe investigar regimes de autenticidade e autenticação, questionando o nome próprio, a assinatura, o texto, a legitimidade, a autoria e o reconhecimento. O seu ponto de partida foi: "Tendo em vista que posso depositar na rede textos, imagens, vídeos, sons e tudo mais, sem ter 'crivo' algum que me autentique, pergunto: quem irá fazer o papel de cartório nas redes e nos novos meios de arquivamento?".[27]

A ideia é basicamente esta: textos são gerados eletronicamente e "assinados" pelos algoritmos que foram "batizados" com nomes de autores como Deleuze, Horkheimer, Platão, etc. É claro que isso causou muita confusão e vem conferir autenticidade ao próprio nome, às marcas e às referências autorais na Internet. A partir de uma experiência realizada na Internet que dissemina vários sites fictícios, entre revistas "científicas" eletrônicas, institutos de pesquisa e textos aleatórios, o autor nos pergunta ironicamente: "afinal, por que não posso batizar um algoritmo de Platão?".[28]

Além destes "textos assinados", há também a geração de 20 institutos de pesquisas falsos, que também indagam sobre a veracidade e a credibilidade das informações na rede Internet. Responda à pergunta: Você dá crédito a informações retiradas da Web? Uma pesquisa acadêmica pode estar fundamentada em informações retiradas da Wikipedia? Silva relata algumas consequências de seu projeto: "já localizei três citações dos textos do Instituto Gilles Deleuze e, em duas delas, os autores são pesquisadores brasileiros de mestrado que utilizam o texto em espanhol, convertendo-o para o português novamente, criando um efeito interessante."[29]

Continuando com as provocações, Silva ainda cria um mecanismo de controle de produtividade próprio, denominado ISSN (que explica ser abreviatura de "Interstellar Synchronism Setup Noise"), que, não por acaso, é homônimo de ISSN, abreviatura de "International Standard Serial Number", regulador da produção científica para publicação de artigos.

26 NUNES, 2007.

27 SILVA apud BEIGUELMAN (http://p.php.uol.com.br/tropico/html/textos/2491,1.shl).

28 Idem.

29 Idem.

Considerações finais

A título de exemplo, e sem querer esgotar o assunto, através desses trabalhos podemos perceber uma visão crítica dos artistas, que aparece em suas propostas poéticas. Além da questão ontológica, em si um desafio, eles nos alertam que também devemos ficar atentos para outros fatores, como a relevância regida pelos domínios, a responsabilidade de cada um de nós com a informação, a distribuição e o conhecimento colaborativo, a importância dos dados vinculados e como todas as relações entre dados, em si mesmas, influenciam-nos. Ademais, o controle que exercem sobre nossas vidas acaba por demonstrar novos vínculos e "taguear" passa a ser um dos grandes verbos do futuro, que surge para definir a ***organização da informação***. Assim, se sempre estivemos nos perguntando quais as melhores formas de armazenar e distribuir a informação, hoje, nosso principal foco passa a ser como organizar tudo isso, para melhor armazenar e distribuir a informação. Pois o "nome da coisa", as "relações entre coisas", "rótulos ou etiquetas", palavras-chave que caracterizam um conteúdo, podem modificar todo o rumo de uma história. É instigante perceber como tudo se tornou mais complexo, como o pensamento acompanha a tecnologia de seu tempo e como uma 'tag' mal formulada pode significar informação perdida no meio de tantas geradas na Internet. Enquanto estar "tagueado" é promessa de não ficar a esmo e, se não encontrado, ser encontrável já é em si um potencial desejado. Além disso, processos autorais estão se modificando, sendo revistos. A proposta de Ted Nelson era de *royalties* divididos entre o autor do documento na exata proporção de quanto seu material estaria sendo transmitido ou usado, através de um processo de ***Transcopyright***, uma espécie de pré-permissão para republicação virtual. Essa pode não ser a melhor solução, mas, seja como for, não dá para perpetuar os processos tradicionais nesses novos formatos. E, enquanto a questão da autenticidade vem tentando se resolver, outro problema é reconhecer ou ser reconhecido como legítimo, pois o movimento natural contrário a tudo isso é justamente o descredenciamento da informação. E, finalmente, o quanto uma "censura disfarçada" (funções back end) e mesmo a não disponibilidade de "informação inalterada" passa a ser alvo de atenção. Tim Berners-Lee[30] pregava, em palestra apresentada na TED.com, a necessidade de que dados inalterados, puros, sejam livremente disponibilizados. Dizia ele: – RAW DATA NOW! – cobrando a participação e relembrando o compromisso e a responsabilidade de todos nós para com a informação.

30 BERNERS-LEE, 2009.

REFERÊNCIAS

LAUFER, R.; SCAVETTA, D. Texte, Hypertexte, Hypermédia. *Paris: PUF, 1995.*

BEIGUELMAN, Giselle. F for Fake 2.0. *Disponível em: <http://p.php.uol.com.br/tropico/html/textos/2491,1.shl>. Acesso em: 2009.*

BERNERS, T.B.; HENDLER; J., LASSILA, O. The Semantic Web. Scientific American, *maio 2001. Disponível em: <http://www.scientificamerican.com/2001/050lissueberners-lee.html>. Acesso em: 2007.*

_____. The semantic web: a new form of web content that is meaningful to computers will unleash a revolution of new possibilities. Scientific American. *Mai. 2001: 34–43.*

BERNERS-LEE, T. Talks Tim Berners-Lee e a próxima Web, fevereiro de 2009 (postado em março de 2009) Disponível em: <http://www.ted.com/talks/lang/por_br/tim_berners_lee_on_the_next_web.html>. Acesso em: 20 nov. 2009.

DZIEKANIAK, Gisele Vasconcelos; KIRINUS, Josiane Boeira. Web Semântica, Encontros Bibli, *n. 18, p. 20–39, julio-diciembre 2004.*

GABRIEL, Martha C. C. Marketing de Otimização de Buscas na Web. *São Paulo: Esfera, 2008.*

GÓMEZ-PÉREZ, A.; MANZANO-MACHO, D. A survey of ontology learning methods and techniques. *Projeto OntoWeb, Delieverable 1.5, 2003. Disponível em: <http://www.sti-innsbruck.at/fileadmin/documents/deliverables/Ontoweb/D1.5.pdf>. Acesso em: 20 nov 2009.*

LIMA, Júnio César de; CARVALHO, Cedric Luiz de. Ontologias – OWL *(Web Ontology Language), Technical Report – RT-INF_004-05 – Relatório Técnico Junho 2005. Disponível em: <http://ia.ucpel.tche.br/~lpalazzo/Aulas/IWS/m04/Recursos/OntoOWL.pdf>. Acesso em: 2009.*

NELSON, Theodor. Proposal for a Universal Eletronic Publishing System and Archive. In: WARDRIP-FRUIN, Noah; MONTFORT, Nick (Ed.). The New Media Reader. *Massachusetts: MIT Press, 2003. p. 441–461.*

NUNES, Fábio Oliveira; FRANCO, Edgar. Freakpedia: *A Ironia da Liberdade. Maio 2007. Disponível em: <http://www.fabiofon.com/freakpedia.html>. Acesso em: 2009.*

SILVIA LAURENTIZ

Docente do Departamento de Artes Visuais da Escola de Comunicações e Artes da Universidade de São Paulo. Doutora em Comunicação e Semiótica pela PUC-SP, mestre em Multimeios pela Unicamp (Campinas – SP), bacharel em Artes pela FAAP (SP). Artista Multimídia e pesquisadora em Arte e Novas Tecnologias, ela tem desenvolvido trabalhos em realidade virtual, multimedia e *web* arte, participando de exposições e conferências nacionais e internacionais. Lidera o Grupo de Pesquisa Realidades – da realidade tangível à realidade ontológica (ECA-USP) e tem lecionado no bacharelado de Multimídia e Intermídia e no Programa de Pós-Graduação em Artes Visuais, da ECA-USP, desde 2002.