

**VIII ENANCIB – Encontro Nacional de Pesquisa em Ciência da Informação  
28 a 31 de outubro de 2007 • Salvador • Bahia • Brasil**

GT 2 – Organização e Representação do Conhecimento  
Comunicação oral

**WEB 2.0, BIBLIOTECA 2.0 E CIÊNCIA DA INFORMAÇÃO (I):  
Um protótipo para disseminação seletiva de informação na  
Web utilizando *mashups* e *feeds* RSS**

***WEB 2.0, LIBRARY 2.0 AND INFORMATION SCIENCE (I):  
A prototype for the selective dissemination of information  
on the Web utilizing mashups and RSS feeds***

Luiz Fernando de Barros Campos (PPGCI – UFMG, lfbcampos@gmail.com)

**Resumo:** Este artigo objetiva descrever o que se denomina Web 2.0 e Biblioteca 2.0, estabelecendo ligações com a ciência da informação. Desenvolveu-se um protótipo de um sistema que filtra e dissemina metadados de artigos publicados em periódicos da Web, procurando (a) demonstrar empiricamente os preceitos da Web 2.0 e Biblioteca 2.0 caracterizados no artigo e (b) enfatizar como processos tradicionalmente conhecidos e estudados na ciência da informação ajustam-se perfeitamente ao ambiente informacionalmente carregado da Web atual. Para isso, foram automaticamente gerados os *feeds* RSS da revista Ciência da Informação do IBICT e utilizaram-se aplicações Web adequadas ao processamento e integração de conteúdo oriundo de mais de uma fonte em uma só interface (*mashups*). Concluiu-se que a Web 2.0 e a Biblioteca 2.0, pelo menos no momento atual, antes que teorias consolidadas, são abordagens predominantemente empíricas, que freqüentemente retomam e enfatizam princípios, conceitos e práticas em formação. Não obstante, as tendências sociais, econômicas e tecnológicas ressaltadas tencionam o campo da ciência da informação de modo duplo. Por um lado, há oportunidades com o surgimento de profícuos temas de estudo, criação de aplicativos interativos e leves, e a possibilidade e conveniência da aplicação de teoria e práticas da área, caso exemplar da disseminação seletiva de informação. Por outro lado, há desafios impostos aos profissionais da informação que podem demandar profundas alterações estruturais em sua atuação e formação.

**Palavras-chave:** Web 2.0. Biblioteca 2.0. disseminação seletiva de informação. *feeds* RSS. *mashups*.

**Abstract:** This article intends to describe what is denominated Web 2.0 and Library 2.0, establishing connections with information science. It was developed a system that filters and disseminates metadata of articles that were published in periodicals on the Web in order to (a) empirically demonstrate the traits of Web 2.0 and Library 2.0 that were described in the text and (b) emphasize how traditional and well-know processes of information science are adequate to the information overloaded Web environment. For this, feeds RSS of the electronic periodical Ciência da Informação from IBICT were generated, and Web applications fitted to collect, process and integrate information obtained from many sources into one interface (*mashups*) were utilized. It was concluded that Web 2.0 and Library 2.0 are not, at least at present time, consolidated theory, but mainly empirical approaches, which frequently retake and emphasize principles, concepts and practices in formation. Nevertheless, the social, economic and technological trends highlighted press the information science field in a two-fold way. By one side, there are opportunities when one takes into account the appearance of productive research themes, the creation of interactive and light-weighted applications, and the opportunity and convenience of applying theory and practices of the field, such as the selective dissemination of information. By the other side, there are threats that can demand transformations on the education of information professionals and the way they perform their duties.

**Keywords:** Web 2.0. Library 2.0. selective dissemination of information. RSS feeds. *mashups*.

## 1 Introdução

Recentemente, tem havido uma discussão intensa sobre a Web 2.0 que se reflete, em proporção, na Biblioteca 2.0. Máquinas de busca na Web, admitindo-se seus resultados como uma medida relativa e parcial de popularidade ou atualidade, retornam mais de 676.000.000 de páginas em língua inglesa para o primeiro termo e quase 2.300.000 para o segundo<sup>1</sup>. A bibliografia sobre o assunto começa a se avolumar.

Termos como esses já nascem controversos. Uma crítica comum refere-se à sua imprecisão (BERNERS-LEE, 2006). Outra, ao fato de realmente não constituírem inovações, mas retomarem conceitos e práticas e vendê-los como novidade (CRAWFORD, 2006). Na literatura que versa sobre esses temas, constantemente se nota um certo espírito de “carta de intenções”, animado de uma atitude revolucionária, afeita a inovações. Outras vezes, percebe-se uma descrição objetiva e precisa de tendências emergentes em ambientes de redes, do ponto de vista social, tecnológico e dos modelos de negócio.

É sob este último ponto de vista que a Web 2.0 e Biblioteca 2.0 são analisadas neste trabalho. Claramente, com será discutido, não se trata de teorias consolidadas. Nem se trata, necessariamente, de novidades. Contudo, os temas abordados retratam e sedimentam perspectivas, visões do mundo (“quase teorias”) e, principalmente, práticas sociais e tecnológicas que necessariamente repercutem intensamente no campo da ciência da informação, ao mesmo tempo que, em certo grau, são influenciadas por ele.

Assim sendo, este artigo, o primeiro de uma série prevista de três, procura descrever o que se denomina Web 2.0 e Biblioteca 2.0, estabelecendo ligações com a ciência da informação. Como, um dos argumentos centrais deste artigo, prevalecem as práticas e as abordagens empíricas na Web 2.0 e Biblioteca 2.0, descreve-se um protótipo implementado de um sistema que filtra e dissemina metadados de artigos publicados em periódicos da Web. Para isso, foram automaticamente gerados os *feeds* RSS da revista Ciência da Informação do Instituto Brasileiro de Informação em Ciência e tecnologia (IBICT) e utilizaram-se aplicações Web adequadas ao processamento e integração de conteúdo oriundo de mais de uma fonte em uma só interface (*mashups*). Em especial, dois pontos motivaram a construção do protótipo: (a) a intenção de demonstrar empiricamente os preceitos da Web 2.0 e Biblioteca 2.0 caracterizados neste artigo e (b) enfatizar como processos tradicionalmente conhecidos e estudados na ciência da informação, como a disseminação seletiva da informação, ajustam-se perfeitamente ao ambiente informacionalmente carregado da Web atual, onde se publica abundantemente material de qualidade desigual.

Na próxima seção, caracteriza-se a Web 2.0 mantendo-se um foco constante em tópicos comuns aos estudos da ciência da informação, e em suas relações com o contexto tipificado. A Biblioteca 2.0 é descrita na seção seguinte, assim como suas ligações com a Web 2.0. Nas duas seções posteriores, discorre-se sobre a disseminação seletiva da informação e os *feeds* RSS, bases da criação do protótipo, detalhado na próxima seção. Nos comentários finais, procura-se especificar o alcance e limites dos conceitos de Web 2.0 e Biblioteca 2.0, assim como algumas de suas implicações para os profissionais da informação, as bibliotecas, e os estudos da ciência da informação, traçando-se caminhos para futuros trabalhos.

## 2 Web 2.0

A Web 2.0 é um termo cunhado, por volta de 2003, pela empresa de mídia O'Reilly (uma editora de livros e revistas, e promotora de conferências e serviços *on-line*), fundada por Tim O'Reilly. Musser *et al.* (2007, p. 5) definem a Web 2.0 como um conjunto de “tendências econômicas, sociais e tecnológicas que coletivamente fundam a próxima geração da Internet – uma mídia mais madura e distintiva, caracterizada pela participação dos usuários, abertura, e efeitos de rede.” As características da Web 2.0 usualmente enfatizadas na literatura são a seguir expostas e discutidas, com base em Musser *et al.* (2007) e O'Reilly (2005), e em autores

que abordam temas correlatos. Durante a exposição, procurou-se destacar a conexão dos temas tratados com conceitos caros à ciência da informação, como informação, taxonomias, busca de informação, interfaces, mídias ricas, foco no usuário e sua participação ativa nos sistemas e serviços de informação.

### 2.1 A Web é concebida como uma plataforma

Considerar a Web como uma plataforma remete diretamente a seu alcance e pervasividade. Além disso, a idéia contrapõe-se ao sentido de uniformização imposta pela adoção de tecnologias específicas. Assim, trata-se de uma plataforma ubíqua e interoperável. Procura-se a independência de implementações específicas. Com um navegador, acessa-se informação abstraindo-se de sua localização ou de que tipo de servidor a hospeda. Empresas podem fornecer serviços livremente, bastando atender a alguns padrões. Muitos protocolos de compartilhamento de dados na Internet têm como base a descentralização geográfica, com objetos digitais residentes simultaneamente em muitas máquinas, como é o caso de protocolos de transmissão ponto-a-ponto. A concepção de plataforma envolve alcance geográfico, diversidade de tecnologias, padronização de protocolos e da comunicação, e geração de flexibilidade.

### 2.2 Aproveitando a inteligência coletiva

A “inteligência coletiva” ou “inteligência das multidões” (SUROWIECKI, 2005) é cada vez mais aproveitada na Web. A expressão refere-se à construção de uma arquitetura de participação, na qual a colaboração do usuário agrega valor. O valor pode ser criado de modo direto e explícito como ocorre nas comunidades virtuais. Nelas, o conteúdo é fornecido pelos usuários, que descarregam arquivos em áreas compartilhadas, comentam, filtram, recomendam, fazem buscas, conversam *on-line*, criam ponteiros para conteúdos de terceiros, interagem. Exemplos bastante conhecidos dessas comunidades são o Youtube ou Orkut, respectivamente um sítio de compartilhamento de vídeos e uma rede social. Ao mesmo tempo, tornou-se cada vez mais fácil publicar na Web. Surgiram verdadeiras comunidades em torno de notas publicadas freqüentemente na Web, no estilo de um diário, os *blogs*. Corriqueiramente, essas comunidades especializam-se e aprimoram ou complementam atividades do mundo não virtual, como ocorre habitualmente, por exemplo, em relação à cobertura da imprensa de assuntos políticos (HEWITT, 2007). Outras vezes verdadeiras obras coletivas, de qualidade inesperada, surgem, como é o caso da Wikipedia, uma enciclopédia *on-line*. Sítios colaborativos como esse são chamados *wikis*, também denominação das próprias intervenções individuais de criação de páginas ou artigos.

Nesse contexto virtual, a fronteira entre usuários, intermediários e organizadores da informação dilui-se. Um exemplo é o surgimento das *folksonomies*, neologismo criado em contraposição a taxonomias. Uma *folksonomia* é uma taxonomia gerada pelo próprio usuário para categorizar e recuperar conteúdo publicado na Web por meio da criação de rótulos. Elas constituem um caso ilustrativo das implicações ambíguas da participação ativa do usuário. Claramente, apresentam várias vantagens como o baixo custo de geração das entradas, vocabulário mais compreensível pelo usuário, capacidade para responder rapidamente às mudanças e adequação a mídias ricas, como vídeo e áudio. No entanto, apresentam também desvantagens, como a intensificação de problemas de polissemia, sinonímia e flexões, além da criação de rótulos idiossincráticos. Sítios como del.icio.us e Flickr foram pioneiros em sua utilização. Esses sítios oferecem, respectivamente, serviços de *bookmarking* social (que consistem basicamente em armazenamento e compartilhamento de preferências) e publicação/compartilhamento de fotos.

No entanto, o valor da arquitetura de participação também é construído de modo indireto e implícito, sem necessariamente envolver uma comunidade virtual. Um exemplo significativo é a utilização do algoritmo *Pagerank* em mecanismos de busca, que gerou uma melho-

ria expressiva na precisão das páginas recuperadas nas posições do topo da lista (o que é importante na Web, onde usuários dificilmente examinam além dos 10 primeiros documentos recuperados). Para possibilitar esse avanço, o Pagerank utiliza a estrutura de apontadores das páginas Web, criado pelos usuários em atividades cotidianas e naturais na rede (algoritmos clássicos, como o modelo vetorial, baseiam-se apenas no conteúdo textual dos documentos). Páginas para as quais apontam muitos *links* recebem muitos “votos” e são consideradas importantes ou populares (o que, na prática funciona como um indicador de sua relevância).

A participação dos usuários é intensificada pelos efeitos de rede. Segundo Musser *et al.* (2007, p. 16), leis empíricas instituem que o valor de utilidade de aparelhos em rede – como telefones ou computadores – não cresce linearmente, mas proporcionalmente ao quadrado de número de nodos na rede (Lei de Metcalfe). A lei de Reed postula que o valor de aplicações que facilitam a formação de sub-grupos na rede, como os sítios de relacionamento Orkut ou Myspace, cresce em ritmo ainda mais acelerado, à taxa exponencial de  $2^n$ . Tirando partido disso, empresas na Web 2.0 procuram potencializar a cooperação dos usuários fornecendo contextos significativos para a criação de conteúdo (ferramentas adequadas, personalização, agregação, filtros de recomendação, etc), confiando no usuário, fornecendo-lhes vantagens iniciais (quem, fora das corporações, paga por uma conta de *e-mail* hoje?) e facilitando a emergência de funcionalidades ao, dentro de limites, permitir usos não previstos de suas aplicações.

### 2.3 Dados são um fator de vantagem competitiva

Quase sempre, as aplicações na Web 2.0 são suportadas por um conjunto de dados. O mecanismo de busca da Google demanda um (gigantesco) índice invertido que é ininterruptamente atualizado por rastreadores. O sítio Youtube é baseado na grande coleção de vídeos descarregados por seus próprios usuários. A livraria virtual Amazon depende essencialmente do banco de dados de seus produtos, constantemente enriquecido pelos seus próprios clientes.

Dados geram valor. Mas de onde vêm os dados? Quem os possui? As estratégias para sua geração são variadas. Eventualmente os dados exigem alto custo e especialização na sua criação e são comprados. Esse é o caso dos mapas utilizados em muitos sítios, como Google Maps, adquiridos usualmente da Navteq. Algumas empresas concentram-se em prover infraestrutura de dados para seus clientes – é o caso, na Web, da Limelight Networks, uma empresa que presta serviços para aplicações que demandam banda larga e alta capacidade de armazenamento, como o sítio de relacionamento Myspace. Outras empresas na Web focam em prover acesso a dados anteriormente difíceis de obter na rede, como avaliações imobiliárias para vendas ou aluguel em regiões geográficas precisas, caso do sítio Zillow.

Essas estratégias, de certo modo, são mais convencionais, anteriores ao conceito de Internet. Certas classes de dados interessam mais no contexto da Web 2.0, justamente aqueles gerados pelos próprios usuários, integrantes da arquitetura de participação descrita no tópico anterior. Exemplos são os metadados criados e enriquecidos por eles, como as avaliações de produtos em sítios comerciais como Submarino ou Amazon, que implementam um sistema virtual de recomendação boca-a-boca, ou as indicações compartilhadas de notícias no sítio Digg. Esses dados têm valor agregado, demandam tempo para sua construção e, como comentado anteriormente, são muitas vezes criados de modo indireto em operações usuais na Web. Esse é o caso dos perfis e estatísticas de uso das aplicações que constituem informações contextuais passíveis de serem utilizadas para, entre muitos usos, o aprimoramento de mecanismos de busca. Isso ocorre com os dados de consultas efetuadas anteriormente ou a relação de sítios visitados mais freqüentemente: efetivamente, máquinas de busca, como a Google, já usam, quando autorizadas, esses dados, introduzindo comercialmente os primeiros rudimentos de *contexto* na recuperação de informação automática na Web.

## 2.4 Inovação na montagem

Quando adotam modelos de negócios vindos do passado, as empresas procuram resguardar seus direitos intelectuais na produção de *softwares*. No entanto, no contexto da Web 2.0, houve um deslocamento para o uso de padrões, protocolos, programas e sistemas operacionais abertos, cuja estrutura interna é conhecida e o código fonte, quando é o caso, e sob algumas restrições, é disponível gratuitamente e passível de alteração por qualquer um. Tecnologias largamente utilizadas na Internet e de reconhecida qualidade, como sistemas operacionais, servidores Web e linguagens de programação (Linux, Apache, PHP, por exemplo), são produtos de uma inteligência coletiva construídos com base em uma estrutura de cooperação.

Ao mesmo tempo, os sítios evoluíram de uma estrutura estática, que expunha um conjunto determinado de informações, para um conteúdo dinâmico, apresentado de acordo com demandas específicas dos usuários. Sítios dinâmicos são usualmente suportados por um banco de dados acessado por algum protocolo ou linguagem de programação (caso típico da CGI, *common gateway interface*). Atualmente esses sítios foram além, caracterizando-se não mais como reservatórios de informação, mesmo que dinâmicos, mas como “plataformas que disponibilizam dados e serviços para facilitar novos e inteiros ecossistemas” (MUSSER et al., 2007).

Um exemplo ilustra bem o processo. No início de 2005, um engenheiro de software, Rademacher, procurava um apartamento utilizando um serviço na Internet de fornecimento de informações imobiliárias, mas queria ver onde os imóveis localizavam-se em um mapa de sua região. O programador, utilizando engenharia reversa, criou uma aplicação que plotava os dados dos imóveis em mapas fornecidos pelo Google Maps. Como a empresa Google reagiu? Ela viu nisso uma oportunidade e forneceu uma interface de programação, um conjunto de rotinas com funções e pontos de entrada conhecidos (uma *application programming interface* - API) para uso gratuito na rede, tornando possível que os interessados aproveitassem mais facilmente as funcionalidades oferecidas pelo Google Maps em seus sítios, para suas finalidades específicas. Atualmente esta é uma das API mais utilizadas para a produção de serviços ou aplicações combinados na Web 2.0.

A prática de fornecimento de API generalizou-se nos últimos tempos. Mais que isso, as API estão sendo integradas ao modelo de negócios da Web 2.0. A livraria virtual *Amazon* fornece várias delas. A API S3 viabiliza meios para a utilização do serviço de armazenamento *on-line* da *Amazon*, que é pago por uso efetivo, gerando escalabilidade. A API de serviços de comércio eletrônico permite que outras empresas utilizem em seus sítios recursos de busca no catálogo de produtos da *Amazon*, oferecendo algumas funcionalidades de um carrinho de compras. Essas empresas são afiliadas ao programa de vendas da *Amazon* e recebem um percentual que varia de 4 a 8,5% das vendas originadas em seus sítios. Além das API, são oferecidos componentes de mais fácil uso, os *widgets* ou *gadgets*, que são reutilizáveis, pequenos e permitem integração de conteúdo originado de várias fontes sem esforço de programação. O *Google AdSense* disponibiliza *gadgets* para que anúncios de seus clientes sejam expostos em páginas de terceiros, que, por meio de uma sistemática automatizada, recebem por anúncio visto ou clicado. A utilização de API e *widgets* gera parceiros de negócios ou clientes de modo automático, escalar, barato, desburocratizado e de modo transparente.

A aplicação de Rademacher, comentada anteriormente, que combinou serviços de diferentes fontes, foi um dos primeiros casos de *mashups* na Internet. Um *mashup* é uma aplicação ou sítio que integra conteúdo de mais de uma fonte em uma só interface. Utilizam-se API, *gadgets* ou *feeds* RSS, dentre outros recursos, para se construir *mashups*, como será demonstrado posteriormente. Esse é um exemplo de tecnologias, e diferentes concepções de seus usos, que permitem que inovações sejam geradas com pouco esforço de programação e baixo custo, ao se criar *montagens* com base em componentes padronizados e reutilizáveis.

## 2.5 Experiências ricas do usuário

Os avanços das tecnologias empregadas na Web 2.0 têm propiciado experiências mais ricas aos usuários em vários níveis. As vantagens de aplicações que são executadas localmente, no *desktop* – como a interatividade rica, alto envolvimento do usuário e desempenho superior – são combinadas com os pontos fortes das redes – independência de plataforma, acesso ubíquo e colaboração. Há uma tendência clara do deslocamento de aplicações do *desktop* para as executadas em navegadores (*browsers*). Aplicativos de escritórios, como editores de textos ou planilhas, que rodam *on-line* e armazenam os dados dos usuários já são oferecidos na Web sem nenhum custo. Tecnologias como *Ajax* permitem que haja troca de dados entre o servidor e o navegador cliente sem necessidade de atualizar a página, possibilitando que interfaces gráficas avançadas, com alto nível de interação com o usuário, sejam implementadas em navegadores. É o caso da simples e eficiente interface do Google Mail, por exemplo.

Interfaces na Web 2.0 procuram otimizar a simplicidade, usabilidade, personalização e “achabilidade” (*findability*<sup>2</sup>). Frequentemente os sítios fornecem mecanismo para pesquisar informações em sua estrutura, o que é facilitado por *widgets* do tipo “busque neste *site*” disponibilizados na Web por mecanismos de busca. Deixou-se o paradigma “metáfora da página”. Isso quer dizer que o conjunto de informações atual em uma página Web não corresponde mais necessariamente ao endereço da página (*uniform resource locator* – URL). Ao, por exemplo, localizar-se algum imóvel no Google Maps, aquele determinado mapa não tem o endereço específico da página. Assim, os sítios preocupam-se em prover mecanismos a seus usuários que garantam a *endereçabilidade* (como o recurso “Meus Mapas” do Google Maps). Ao mesmo tempo, obtém-se, desse modo, personalização. Os usuários podem especificar a disposição gráfica dos elementos em uma página e, frequentemente, o próprio conteúdo, criando assim páginas personalizadas. Ao mesmo tempo, as aplicações aprendem dinamicamente com o comportamento do usuário, antecipando necessidades, lembrando as preferências e adaptando-se quando apropriado – a maioria das páginas da *Amazon* são geradas no momento de sua carga e adaptadas ao perfil do usuário, indicando, por exemplo, produtos que provavelmente mais lhe despertarão o interesse.

Ainda, considerando a banda mais larga disponível e a maior e mais barata capacidade de armazenamento, o conteúdo cada vez mais se centra nas mídias ricas, como gráficos sofisticados, figuras elaboradas, sons, música e vídeos, ampliando o público e as aplicações possíveis na Web 2.0.

## 2.6 Percorrendo a cauda longa

A cauda longa é um nome coloquial para uma distribuição estatística característica da lei das potências e encontrada em muitas leis empíricas bibliométricas como a de Zipf. No contexto da Web 2.0, o conceito de cauda longa foi explorado pelo editor da *Wired Magazine* Chris Anderson (ANDERSON, 2004, 2006). Ele argumentou que os relativamente poucos produtos mais vendidos no mercado estão na cabeça da curva (vide Figura 1), amealhando boa parte da renda. No entanto, os numerosos produtos localizados na cauda da curva podem coletivamente corresponder a uma larga fatia do mercado e gerar renda que rivaliza com a dos mais vendidos. Isso implica uma produção diversificada, a distribuição por meio de agregadores que democratizem a produção, a constituição de mercados de nichos e a atuação de formadores de preferência, o que envolve a recuperação de informação sobre produtos nos nichos e a atuação da inteligência coletiva por meio, entre outros, de filtros de recomendação. Em suma, o que viabiliza e potencializa a extensão da cauda é informação, ou, como quer Anderson, aquilo que faz uma oferta concentrada encontrar uma demanda geograficamente pulverizada.

Anderson (2004) baseou-se, entre outros trabalhos, em Brynjolfsson; Hu e Smith (2003), que calcularam que o excedente do consumidor gerado devido à variedade de produtos vendidos pela livraria virtual *Amazon* é de sete a dez vezes maior que o excedente causado

pelo aumento da competição e diminuição de preços. Por causa da capacidade de catalogar, recomendar e disponibilizar um grande número de livros, bem superior ao das livrarias físicas, a *Amazon* comercializa, conforme os cálculos do trabalho, aproximadamente 40% dos livros na cauda da curva (definidos como aqueles títulos depois dos 100.000 mais vendidos). Posteriormente, Brynjolfsson; Hu e Simester (2007) obtiveram mais evidências empíricas de que as vendas pela Internet são menos concentradas que as vendas pelos canais usuais - a diferença apurada foi significativa mesmo quando controlando por diferenças entre consumidores e foi mais forte para indivíduos com mais experiência no uso da Internet.

Os últimos autores citados argumentam que as tecnologias de informação têm o dom de diminuir o custo de armazenamento e distribuição, tornando viável o aproveitamento de nichos de produtos. A melhoria da logística é extremamente significativa, viabilizada pelos sistemas corporativos integrados, armazéns de dados centralizados, intercâmbio de dados e serviços providos pela Internet. Mas há outros fatores, especialmente do lado da demanda. As comunidades virtuais constituídas na Web tratam de assuntos, produtos e serviços extremamente especializados. Os usuários interagem de forma ativa, informações sobre os produtos e serviços são obtidas de fontes pessoais e mídias ricas, e mecanismos de busca, sistemas de recomendação e ferramentas de amostragem possibilitam aos usuários localizar e obter informações sobre produtos incomuns, raros ou fora de sua área geográfica.

### **2.7 Novos paradigmas na produção de aplicações**

Concomitante com a intensificação das características da Web 2.0 anteriormente descritas, ocorreram profundas alterações no modo de produção das aplicações.

Como a Internet gradativamente passa a abranger equipamentos que não computadores, como telefones celulares, eletrodomésticos, sistemas de vigilância, etc, os programas são projetados *para além do nível de um dispositivo único*. Isso presume que as aplicações identifiquem o dispositivo no qual rodem (de fato, uma dimensão do contexto) e dados e informações sejam fornecidos em formatos adequados à interface utilizada e com o nível de detalhamento ou precisão propícios ao uso. Desse modo, por exemplo, surgem páginas Web especialmente construídas para visualização em telefones móveis.

Outro aspecto relevante é que o ciclo de produção de aplicações foi radicalmente alterado. Fala-se agora em *perpétuo beta*, em analogia à fase de testes dos programas. As aplicações não são mais artefatos, mas serviços contínuos (denominadas *software as a service* – SaaS). As modificações passam a ser incrementais e os usuários testam-nas em tempo real. Os serviços são estruturados para mostrar como os usuários utilizam de fato as aplicações. Assim, se pouco utilizada, uma alteração pode não permanecer, ou então, na situação contrária, cria-se uma funcionalidade ao constatar um conjunto de operações reiteradas pelos usuários. Perde-se um bocado o sentido de falar de “versão da aplicação”.

Por fim, procura-se desenvolver *modos leves de programação e escalabilidade efetiva de custos*. As idéias de controle e acoplamento enfraquecem-se. Os serviços na Web tendem a ser simples, renunciando ao controle estrito do que ocorre na outra ponta. Os sistemas passam a ser, sempre que possível, frouxamente acoplados. A simplicidade torna-se uma meta e só há controle quando estritamente necessário. Almeja-se a reutilização de componentes e a reutilização – daí o uso intenso das API e *gadgets*. Essas abordagens também geram escalabilidade. Os componentes são oferecidos no sistema de auto-atendimento: são os parceiros que decidem incorporá-los em seus sítios e atuar como intermediários no negócio de uma outra empresa, por exemplo. A modularidade desempenha um papel preponderante. O ponto básico é que, na Web 2.0, o desenvolvimento das aplicações está intrinsecamente ligado aos modelos de negócio abertos e flexíveis.

## **3 Biblioteca 2.0**

Com cada vez maior frequência, surge, no campo da ciência da informação e biblioteconomia, uma literatura que procura apontar as repercussões das mudanças atuais na Web

para as bibliotecas. A *Library 2.0* tem sido constantemente tema de artigos e livros (MANESS, 2006<sup>3</sup>; CASEY; SAVASTINUK, 2006, 2007, COURTNEY, 2007).

Nascido e debatido nas comunidades virtuais de blogueiros na Internet, o termo Biblioteca 2.0 tem sido utilizado de várias maneiras, como acontece com Web 2.0. Crawford (2006) examinou minuciosamente várias definições do termo, concluindo que o sentido gravita em torno de dois núcleos. O primeiro deles abrange um espectro de metodologias de aplicações e programas (*software* social, modularidade, interatividade, API, e outras) assim como um conjunto de conceitos sobre os serviços prestados pela biblioteca, boa parte deles antigos. Com discussão e habilidade, algumas dessas novidades e teses poderiam tornar as bibliotecas mais interessantes e relevantes. Mas o segundo conceito, “Biblioteca 2.0”, seria uma confrontação, uma visão negativa sobre as bibliotecas existentes, uma imputação de um papel não adequado às bibliotecas de fornecedoras de qualquer tipo de informação, uma ênfase em ferramentas e modelos de negócios que não atenderiam todas as variadas comunidades de usuários, e uma atitude de não valorização de um trabalho já existente.

A exposição de Crawford (2006) reflete as discussões acaloradas e visões extremadas que acompanharam o surgimento da expressão. Maness (2006, p.3), traz uma proposta concreta e metodologicamente assentada. Sua definição de Biblioteca 2.0 é “*a aplicação de tecnologias interativas, colaborativas e multimídia baseadas em Web a serviços de bibliotecas e coleções baseadas em Web*”. Duas características da sua definição sobressaem-se: (a) trata-se de uma abordagem empírica, de aplicações; e (b) é limitada para serviços baseados na Web, e não para os serviços e bibliotecas em geral, sendo que desse modo “*evita confusões potenciais e permite suficientemente que o termo seja pesquisado, depois teorizado, e o torna mais útil no discurso profissional*”.

Maness (2006) entendeu que uma teoria para Biblioteca tem quatro elementos essenciais. A teoria é centrada no usuário, que participa dinamicamente no consumo e criação do conteúdo. Ela valoriza a experiência multimídia – um aspecto particularmente enfatizado por esse autor. Também é socialmente rica, abrangendo comunicação síncrona, como programas de mensagens instantâneas, e assíncrona, como *blogs* e *wikis*, entre os próprios usuários, e entre usuários e bibliotecários. E, por fim, é comunitariamente inovadora, o que pressupõe que não apenas a biblioteca mude com os usuários, mas que se permita que os usuários mudem a biblioteca, constantemente alterando seus serviços e as formas de comunicação. Opina-se que esses elementos, embora oportunos e adequados, sejam mais uma declaração de (bons) valores, do que exatamente fundamentos de uma teoria

Para Casey e Savastinuk (2006, 2007), a base da Biblioteca 2.0 é a mudança com foco no usuário. Trata-se de um modelo que encoraja os usuários a participarem da criação dos serviços físicos ou virtuais que desejem, com base em uma avaliação constante e consistente dos serviços. Também tenta servir melhor os usuários existentes com serviços direcionados a suas necessidades e incorporar novos usuários, explorando a diversificação na cauda da curva, nos moldes da argumentação de Andersom (2006). Tecnologia não é um requisito essencial, mas, principalmente aquela relacionada à Web 2.0, desempenha um papel significativo para que a biblioteca mantenha-se atualizada com as necessidades dos usuários, criando novos serviços interativos ou ensejando formas originais de intercâmbio.

Desse modo, a ênfase está no usuário, e a tecnologia é um instrumento considerado essencial para viabilizar essa perspectiva. Esse argumento aparece na literatura que versa sobre Biblioteca 2.0 sob diversas roupagens. Assim, por exemplo, a obra organizada por Courtney (2007) descreve o uso de várias tecnologias (*folksonomies*, *tagging*, *podcasting*, bibliotecas de realidade virtual, etc) e discute seu potencial no ambiente da biblioteca de hoje. Em relação ao tema, John Blyberg publicou dois curtos artigos em seu *blog* (BLYBERG, 2006, 2007) que repercutiram positivamente na comunidade que discute a Biblioteca 2.0 (como em CASEY; SAVASTINUK, 2006 ou CRAWFORD, 2006). O primeiro deles propugnou que os fornece-



dores de sistemas integrados para automação e gestão de bibliotecas (*integrated library systems – ILS*) incorporem novas tecnologias mais rapidamente e, principalmente, construam os sistemas de modo a gerar maior autonomia para as bibliotecas no acesso a seus próprios dados. O segundo é uma defesa veemente, mas equilibrada, da necessidade de uma Biblioteca 2.0, ressaltando aspectos habitualmente menos enfatizados nas discussões recentes sobre a Biblioteca 2.0, como a imprescindível agilidade na adoção de novas tecnologias e de mudanças internas nas bibliotecas, decorrentes de uma reavaliação de sua missão e de uma reorganização em sua estrutura administrativa e cultural.

#### 4 Disseminação seletiva da informação

Tendo em vista a proliferação de informações na Web, suas características e tendências atuais delineadas anteriormente e as repercussões nas bibliotecas rascunhadas na última seção, apresenta-se aqui o conceito de disseminação seletiva da informação (*selective dissemination of information – SDI*), vista como uma abordagem a ser enfatizada pela ciência da informação no contexto da Web 2.0.

O conceito de SDI foi empregado prematuramente por Luhn (1958, p. 315) ao reportar o desenvolvimento de um sistema de inteligência de negócios na IBM Corporation que pretendia criar perfis de unidades ou indivíduos na organização que determinassem o recebimento de metadados gerados automaticamente sobre documentos que provavelmente lhes despertariam o interesse. Assim, o objetivo do sistema era “*prover informação adequada para fornecer suporte às atividades executadas por indivíduos, grupos,..., ou unidades maiores*”, os chamados pontos de ação, criando meios para “*disseminar informação para cada um dos pontos de ação de acordo com suas requisições e desejos.*”

Tratava-se de uma automatização de uma função clássica da biblioteca, informar os usuários de aquisições que poderiam lhe interessar. Há diferenças significativas em relação à *busca retrospectiva* (HOUSMAN, 1973), na qual a base de dados é formada primeiro, indexada, e então disponibilizada para responder a consultas dos usuários. No tipo de *busca prospectiva* da SDI, a consulta do usuário vem em primeiro plano e há um foco em material novo, atual. Se, em sistemas retrospectivos, poucas consultas são confrontadas com inúmeros documentos, ocorre o contrário em um sistema de SDI, com relativamente poucos documentos sendo cotejados com mais consultas.

Habitualmente, um processo de SDI engloba as seguintes fases, não necessariamente nessa ordem: (a) definição de perfis de usuários; (b) acesso periódico às fontes de informação buscando as referências para recursos novos ou alterados e, quando for o caso, armazenando as indexações efetuadas em um repositório de dados; (c) casamento dos descritores armazenados com os perfis; (d) disseminação dos recursos ou das referências para os usuários selecionados, eventualmente em formato discriminado em seu perfil.

O perfil do usuário pode ser comparado com uma consulta que permanece no sistema e continua a produzir resultados. Assim, a literatura sobre SDI tem como uma de suas constantes preocupações a formação desse perfil (SOUTO, 2006) e sua alteração no decorrer do tempo (HOUSMAN, 1973), fatores críticos para o sucesso da disseminação. Segundo Bax *et al.* (2004), sistemas de SDI podem ser baseados em conteúdo, quando a determinação da adequação do recurso para determinado usuário funda-se na análise de conteúdo do recurso, ou baseados em colaboração, caso em que a adequação é avaliada com base na comparação entre o perfil do usuário e perfis de usuários com interesses em comum ou pertencentes a comunidades padrão para os quais o recurso foi considerado relevante. Esse é o princípio de filtragem colaborativa, vastamente empregada na Web 2.0, como foi visto.

Essa breve revisão bibliográfica provocou surpresa pela atualidade do tema e das preocupações de seus primeiros autores. Luhn (1958) concebeu a SDI como um sistema automatizado, preocupado com a taxa de produção e utilização da informação cuja aceleração já se vislumbrava na época. Housman (1973), em uma revisão na *Annual Review of Information*

*Science and Technology* no começo da década de 70, destacou tópicos como o atendimento a indivíduos com base em um perfil do grupo ao qual pertence (o que converge com a idéia de servir comunidades virtuais proposta por MANESS, 2006), o problema das numerosas citações não-relevantes e a atualização dinâmica dos perfis com base no contexto. Considerando o que a distingue, a Web 2.0, embora apresente problemas semelhantes aos descritos, potencializa a resolução de vários deles.

### 5 Feeds Rss

No contexto da Web 2.0, como efetuar dinamicamente a disseminação de informação? Os *feeds* RSS oferecem um meio simples e flexível de fazê-lo. O termo *syndication* refere-se originariamente à distribuição do trabalho de um colunista ou cartunista para muitos jornais, assim como à venda de, por exemplo, séries televisivas para estações locais. *Web syndication* é o processo de disponibilizar conteúdo publicado em um sítio para utilização por outros sítios. RSS, em sua última forma, é um acrônimo para *Really Simple Syndication*. *Feeds* RSS pertencem a uma classe de “alimentadores” utilizados para atualização constante de conteúdo na Web, como ocorre em *blogs* ou sítios de notícias.

A utilização de *feeds* RSS possibilita que uma pessoa mantenha-se atualizada com seus sítios preferidos de notícias, *blogs*, jornais, periódicos, ou outros, de maneira automática, sem ter que realizar uma checagem manual. Em geral, o processo funciona assim: os sítios disponibilizam os *feeds* RSS, que podem ser “assinados” por programas leitores de *feeds* ou “agregadores”. Agregadores podem ser programas funcionando em *desktops* (como o *FeedReader*) ou serviços fornecidos na Web (como o serviço *Bloglines*). Eles têm como funcionalidade básica convergir as informações publicadas por vários sítios em um só ponto, verificando regularmente a existência de novo conteúdo. Do mesmo modo, sítios podem agregar e apresentar conteúdo dinâmico proveniente de outros sítios, referenciando adequadamente seus *feeds* RSS.

Basicamente, *feeds* RSS são documentos XML (eXtensible Markup Language) utilizados para representar uma unidade ou objeto de informação, como uma notícia ou um artigo em um periódico. Assim são fornecidas informações rápidas para ajudar o usuário a decidir se querem acessar o conteúdo completo. Rótulos (ou *tags*) habitualmente utilizados são <title>, <description> e <link>, para delimitar o título, descrição e localização na Web do objeto informacional. Assim, se por exemplo, um sítio de determinado periódico publica um feed RSS, ele simplesmente indica a localização de um documento XML que descreve os artigos do periódico (normalmente a edição atual ou as mais recentes), discriminando, usualmente, seu título, resumo, localização (*link*) e data de publicação. Esse documento XML pode ser acessado por programas como os agregadores, referenciado por outros sítios, ou mesmo ser lido por humanos; o fundamental é que há possibilidade de tratamento automático desses dados.

### 6 Um protótipo para disseminar e filtrar metadados de artigos em periódicos científicos na web

Nesta seção, é descrito um protótipo que visa agregar metadados de artigos científicos oriundos de diversos periódicos da área da ciência da informação que versem sobre o tema “Web”. Considera-se esse um caso simplificado de disseminação seletiva de informação. O perfil do usuário é definido, em parte, pela própria escolha das fontes a serem consultadas (os periódicos de interesse) e, principalmente, pela definição do filtro, que sugere o tema de interesse para o usuário no momento (no caso, Web). A disseminação ocorre pelos recebimentos dos *feeds* de modo centralizado.

O protótipo procura também demonstrar os princípios da Web 2.0/Biblioteca 2.0 discutidos. Assim consideraram-se como pressupostos para sua implementação: (a) uso de ferramentas e protocolos abertos e de livre acesso, disponíveis gratuitamente, como as API, *widgets*, XML etc; (b) interoperabilidade entre diversas aplicações através da Web, concebendo-a

como uma plataforma; (c) acesso à informação livremente disponível na Web; (d) redução do esforço de programação ao máximo (simplicidade); e (e) inovação com base em reutilização de componentes e montagem, evitando retrabalho.

A primeira fase da construção do protótipo foi a definição dos periódicos eletrônicos a serem utilizados. A princípio, a condição seria que o periódico disponibilizasse os feeds RSS. Assim, selecionaram-se três periódicos internacionais de qualidade reputada (*Journal of Documentation*<sup>4</sup>, *Journal of the American Society for Information Science and Technology*<sup>5</sup> e *Information Development*<sup>6</sup>). Não se deparou com tal facilidade em relação aos periódicos nacionais da área da ciência da informação. Usualmente esses periódicos utilizam o Sistema Eletrônico de Editoração de Revista (SEER), uma adaptação da ferramenta de administração do processo editorial de periódicos eletrônicos *Open Journal System* (OJS) efetuada pelo IBICT. Embora o OJS tenha desde a versão 1.x um *plug-in* que permite a geração de feeds RSS dos artigos (ALMEIDA, 2007), não se encontrou nenhum periódico da área que os disponibilizasse. Almeida (2007), trabalhando com notificação automática de conteúdo em ciência da informação, selecionou dez edições de periódicos nacionais e, com o auxílio de um programa, criou manualmente seus feeds RSS. Dado o contexto do protótipo aqui desenvolvido, optou-se por uma solução automática.

Para isso, foi escolhido, pela antiguidade e reconhecida qualidade, um periódico nacional, a revista *Ciência da informação* do IBICT, e criaram-se os feeds RSS de sua coleção, utilizando-se um serviço disponibilizado na Web que possibilita a varredura automática de sítios e extração de informações para a criação de *mashups*, clipagens, feeds, e outros, o *Openkapow*<sup>7</sup>. Tendo em vista o foco em material recente nos processos de disseminação seletiva de informação, implementou-se um serviço que gera os feeds RSS das edições a partir de 2006 em tempo real<sup>8</sup>. Não obstante, considerando o caráter exploratório do protótipo, o mesmo serviço foi adaptado para criar os feeds das edições de 1995 até 2005 em um arquivo que foi armazenado na rede.<sup>9</sup> Mostra-se parte do código XML desse feed RSS na Figura 2.

Desse modo foram definidas as cinco fontes de informação do protótipo, os feeds dos três periódicos internacionais, o feed gerado em tempo real das edições recentes da revista *CI-IBICT* e o feed das edições antigas disponibilizado na rede. Nessa segunda fase do desenvolvimento do protótipo, restava filtrar os cinco feeds e agregar os resultados em um só aplicativo. Constatou-se que usualmente os aplicativos leitores e agregadores de feeds enfatizam a coleta e exibição dos feeds. No entanto, há aplicativos direcionados para a aquisição de dados de formatos e fontes diversas na Web e seu processamento, antes de remixá-los para gerar *mashups*. Tal é o caso do *Pipes*<sup>10</sup>, um aplicativo *on-line* baseado em uma interface visual fornecido pela Yahoo, empregado no protótipo. O serviço (*pipe*)<sup>11</sup> criado lê os feeds, seleciona aqueles referentes à Web (isto é, baseado no critério estipulado de existência do termo Web em seu título ou no resumo) e ordena-os com base na data de publicação. Pode-se visualizar a interface gráfica do aplicativo na Figura 3. O resultado, que abrange a relação dos artigos filtrados, seus resumos e links, pode ser visualizado no próprio *Pipes* ou lido por meio de outro programa agregador (vide Figura 4).

## 7 Comentários finais

Discutir temas em formação como Web 2.0 e Biblioteca 2.0 sempre é arriscado e envolve uma avaliação pessoal. No entanto, acredita-se ser razoável efetuar algumas considerações. Como se mostrou no decorrer deste texto, a Web 2.0 e a Biblioteca 2.0 têm sido caracterizadas tipicamente de modo empírico, com destaques para as boas práticas. Desse modo, entende-se que, no presente momento, a Biblioteca 2.0, assim como a Web 2.0, não é uma teoria; trata-se mais de pertinentes postulados empíricos, que muitas vezes retomam e enfatizam pontos sedimentados nos estudos de informação, como a interatividade, a adequação dos serviços aos usuários, a importância da avaliação dos serviços de informação, e outros.

Desse modo, embora o caráter inovador e revolucionário de termos com Web 2.0 e Biblioteca 2.0 seja freqüentemente evidenciado, em muitos casos eles representam a consolidação e valorização de conceitos, princípios e práticas já em formação. Nesse sentido, muitos dos postulados da Web 2.0 refletem conceitos oriundos da programação orientada a objetos, como a modularidade e reaproveitamento de componentes. Ao mesmo tempo, isso aponta uma interessante questão de estudo: as similaridades entre Web 2.0 e Web Semântica. Há semelhanças evidentes, pois a Web semântica pretende estabelecer uma extensão da Web com semântica compartilhável, criando padrões e interoperabilidade entre agentes de *software* e humanos, permitindo estruturação e utilização automática de informação existente na Web, e empregando recursos como XML, ontologias, e outros (BERNERS-LEE *et al.*, 2001). Por outro lado, certas diferenças parecem despontar, como a ênfase da Web 2.0 em aspectos como interatividade social, simplicidade e fraco acoplamento entre aplicações.

No entanto, o último ponto levanta a questão do desempenho e da segurança. Eventualmente, diversos aplicativos baseados em princípios da Web 2.0 não se prestam a emprego em ambientes corporativos. Tecnologias Web 2.0 serão aplicadas nas corporações em determinadas circunstâncias e com objetivos específicos. Mesmo Musser *et al.* (2007), uma obra bastante voltada para a Web 2.0 em corporações, destacaram situações em que exigências de segurança ou desempenho ensejam o uso de forte acoplamento e aplicativos “pesados”. Por outro lado, acredita-se que essas demandas sejam menos restritivas em ambientes de biblioteca no que diz respeito aos dados da coleção (mas não necessariamente em relação às transações dos usuários).

A esse respeito, em certa medida, a Biblioteca 2.0 é uma reação a pouca flexibilidade, interatividade e agilidade em incorporar as novas tecnologias nos sistemas integrados para a automatização e gestão de bibliotecas (BLYBERG, 2006). Na literatura, aparecem referências constantes a processos de personalização, indicação de recomendações e serviços de alerta (todos processos típicos da Web 2.0) em catálogos de acesso público *on-line* (OPAC) (CASEY; SAVASTINUK, 2006; MANESS, 2006). Talvez, esse seja um ponto central para a compreensão do conceito de Biblioteca 2.0: em certa medida, as novas tecnologias da Web, em convergência com a ampla aceitação da valorização do usuário em estudos da ciência da informação, colocam em tensão a prática das bibliotecas. Isso, ao mesmo tempo, é uma oportunidade e um desafio. A proliferação de *blogs* e *wikis*, publicações não editadas e revisadas pelos pares, e a conseqüente necessidade de inseri-los em coleções (pelo menos alguns, dada sua manifesta qualidade), vai exigir proficiência e experiência de bibliotecários, ou mesmo a reavaliação de conceitos como autoridade ou confiabilidade da fonte (MANESS, 2006)<sup>12</sup>. Seja um outro exemplo: a tentativa de sanar problemas advindos do uso das *folksonomias*, alguns deles citados anteriormente, pode exigir novas práticas de intervenção dos profissionais, como a combinação manual de *folksonomias* com vocabulários controlados ou o acompanhamento do aperfeiçoamento dinâmico e parcialmente automatizado de taxonomias. A participação mais ativa do usuário tem ocasionado certa diluição entre seu papel e o do profissional da informação (MANESS, 2006). Conseqüentemente, pressupõem-se qualificações diversificadas e transdisciplinares dos profissionais, o que coloca em pauta a questão da formação acadêmica e do treinamento.

Além disso, os movimentos atuais na Web têm profundo impacto em temas do domínio dos estudos de informação, com variadas implicações, mas são relativamente pouco explorados no campo da ciência da informação. A tradição transdisciplinar dos estudos do campo tem motivado perspectivas múltiplas, de ordem política, econômica, social e cognitiva, que, como este artigo procurou indicar, são apropriadas para estudar com originalidade os fenômenos de informação que se manifestam na rede.

No contexto da Web 2.0 e Biblioteca 2.0, as teorias, conceitos e práticas advindos da ciência da informação fundam iniciativas (como a dos algoritmos de ordenamento baseados

em análise de *links*, essencialmente uma antiga prática bibliométrica) e, ao mesmo tempo, encontram terreno fértil para proliferar. Isso acontece de modo bem claro, por exemplo, considerando os níveis intensos, complexos e desiguais de publicação na Web, com os processos de disseminação seletiva de informação. A construção do protótipo de filtragem e envio automáticos de informações sobre artigos de periódicos eletrônicos procurou demonstrar a pertinência da aplicabilidade dos conceitos da ciência da informação nos ambientes em rede e a possibilidade da instituição de novos modos de operacionalizá-los. Assim, o protótipo implementou uma função de distribuição seletiva de informação utilizando sistemas livremente disponíveis na Web, inovando com base em reaproveitamento de componentes e, demandando, desse modo, pouco esforço de programação. Ambientes em rede como os retratados motivam a utilização de modos ágeis e leves de desenvolvimento, resultando em produtos adaptados a contextos e situações específicos, além de mais interativos e direcionados a necessidades determinadas e transitórias dos usuários, o que pode mostrar-se útil no domínio das bibliotecas.

### Referências bibliográficas

- ALMEIDA, R. L. Notificação automática de conteúdo em ciência da informação: uma experiência na utilização de *feeds* RSS em periódicos eletrônicos In: Encontro Nacional de Ensino e pesquisa da Informação – CINFORM, VII, Salvador. *Anais...* (CDRom). 2007.
- ANDERSON, C. *A cauda longa – do mercado de massa para o mercado de nicho*. Rio de Janeiro: Elsevier/ Editora Campus, 2006.
- ANDERSON, C. The Long Tail. *Wired*, v. 12.10, 2004. Disponível em: <<http://www.wired.com/wired/archive/12.10/tail.html>>. Acesso em 02/08/2007.
- BERNERS-LEE, T. *et al.* The semantic web. *Scientific American*, maio, 2001.
- BERNERS-LEE, T. Podcast interview for IBM. 2006. <<http://www-128.ibm.com/developerworks/podcast/dwi/cm-int082206.txt>>. Acesso em 02/08/2007.
- BLYBERG, J. 11 reasons why Library 2.0 exists and matters. 2006. Disponível em: <<http://www.blyberg.net/2006/01/09/11-reasons-why-library-20-exists-and-matters/>>. Acesso em 02/08/2007.
- BLYBERG, J. ILS customer bill-of-rights. 2005. Disponível em: <<http://www.blyberg.net/2005/11/20/ils-customer-bill-of-rights/>>. Acesso em 02/08/2007.
- BRYNJOLFSSON, E., HU, Y. J.; SIMESTER, D. Goodbye Pareto Principle, Hello Long Tail: the Effect of Search Costs on the Concentration of Product Sales. 2007. Disponível em: <<http://ssrn.com/abstract=953587>>. Acesso em 02/08/2007.
- BRYNJOLFSSON, E., HU, Y. J.; SMITH, M. D. Consumer Surplus in the Digital Economy: Estimating the Value of Increased Product Variety at Online Booksellers. 2003. *MIT Sloan Working Paper*, n. 4305-03. Disponível em: <<http://ssrn.com/abstract=400940>>. Acesso em 02/08/2007.
- CASEY, M. E.; SAVASTINUK, L. C. Library 2.0 - Service for the next-generation library. *Library Journal*, 2006. Disponível em: <<http://www.libraryjournal.com/article/CA6365200.html>>. Acesso em 02/08/2007.
- CASEY, M. E.; SAVASTINUK, L. C. *Library 2.0: A Guide to Participatory Library Service*. Information Today Inc, 2007.
- COURTNEY, N. *Library 2.0 and Beyond: Innovative Technologies and Tomorrow's User*. Portsmouth (U.S): Libraries Unlimited Inc, 2007.
- CRAWFORD, W. Library 2.0 and “Library 2.0”. *Cites and Insights*, v. 6, n. 2, 2006. Disponível em: <<http://citesandinsights.info/civ6i2.pdf>>. Acesso em 02/08/2007.
- HEWITT, H. Blog – entenda a revolução que vai mudar seu mundo. Rio de Janeiro: Thomas Nelson Brasil, 2007.

- HOUSMAN, E.M. Selective dissemination of information. *Annual Review of Information Science and Technology*, p. 221-24, 1973.
- LUHN, H. P. A business intelligence system. *IBM Journal of Research and Development*, v. 2, n. 4, p. 314-319, 1958. Disponível em: <<http://www.research.ibm.com/journal/rd/024/ibmrd0204H.pdf>>. Acesso em 02/08/2007.
- MANESS, J. Library 2.0 Theory: Web 2.0 and Its Implications for Libraries. *Webology*, v. 3, n. 2, 2006. Disponível em: <<http://www.webology.ir/2006/v3n2/a25.html>>. Acesso em 02/08/2007.
- MILLER, P. Do libraries matter?: the rise of library 2.0 (A Talis White Paper), *Talis*, 2005a. Disponível em: <[http://www.talis.com/applications/downloads/white\\_papers/DoLibrariesMatter.pdf](http://www.talis.com/applications/downloads/white_papers/DoLibrariesMatter.pdf)>. Acesso em 02/08/2007.
- MILLER, P. Web 2.0: building the new library. *Ariadne*, n. 45, 2005b. Disponível em: <<http://www.ariadne.ac.uk/issue45/miller/#2>>. Acesso em 02/08/2007.
- MORVILLE, P. Ambient Findability. Sebastopol: O'Reilly Media, Inc., 2005.
- MUSSER, J.; O'REILLY, T. *et al.* *Web 2.0 – Principles and best practices*. Sebastopol: O'Reilly Media, Inc., 2007.
- O'REILLY, T. What is web 2.0? Design Patterns and Business Models for the Next Generation of Software. 2005. Disponível em: <<http://www.oreillynet.com/lpt/a/6228>>. Acesso em: 02/08/2007.
- SUROWIECKI, J. *The Wisdom of Crowds*. Anchor, 2005.

---

<sup>1</sup> Consulta efetuada em 10/10/2007 utilizando a máquina de busca da Google.

<sup>2</sup> Definida, por Morville (2005), com a facilidade com que um objeto é encontrado, descoberto ou localizado, ou a medida em que um sistema ou ambiente suporta a navegação ou recuperação.

<sup>3</sup> Há uma tradução em português do artigo: Maness, J.M. Teoria da Biblioteca 2.0: Web 2.0 e suas implicações para as bibliotecas. *Inf. & Soc.: Est.*, João Pessoa, v.17, n.1, p.44-55, jan./abr. 2007.

<sup>4</sup> Feed RSS disponível em [http://www.emeraldinsight.com/info/RSSFeeds/InsightFeeds/jd\\_rss.xml](http://www.emeraldinsight.com/info/RSSFeeds/InsightFeeds/jd_rss.xml).

<sup>5</sup> Feed RSS disponível em <http://www3.interscience.wiley.com/rss/journal/76501873>.

<sup>6</sup> Feed RSS disponível em <http://idv.sagepub.com/rss/recent.xml>.

<sup>7</sup> <http://openkapow.com/>

<sup>8</sup> Esse serviço (*robô*) está disponível em <http://service.openkapow.com/LuizCampos/GeraFeedsCI.rss>.

<sup>9</sup> Esse *feed* RSS pode ser acessado em <http://www.lfbcampos.xpg.com.br/antigasedrssCI.xml>.

<sup>10</sup> <http://pipes.yahoo.com/pipes/>

<sup>11</sup> Publicado em [http://pipes.yahoo.com/pipes/pipe.info?\\_id=3PNBNZga3BGrK9oNCB2yXQ](http://pipes.yahoo.com/pipes/pipe.info?_id=3PNBNZga3BGrK9oNCB2yXQ).

<sup>12</sup> Entre as referências que fundamentaram este trabalho, duas das consideradas mais relevantes foram curtos artigos publicados em um *blog*: Blyberg (2005, 2006).

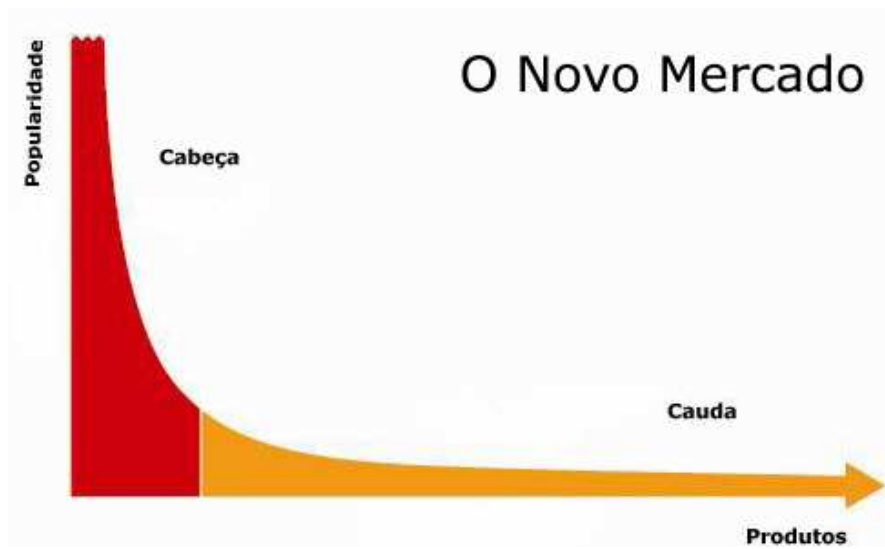


Figura 1 - A cauda longa do novo mercado. Fonte: Adaptado de Anderson (2004).

Código-fonte: <http://www.lfbcampos.xpg.com.br/antigasedrssl.xml> - Mozilla Firefox

```

<generator>Criado por Luiz Campos utilizando um serviço Openkapow - RoboSuite 6.2. </generator>
<image>
<title>Feeds extraídas da revista Ciência da Informação - Edições 2005 e anteriores</title>
<url>http://setup.openkapow.com/api/logo.gif</url>
<link>http://www.ibict.br/cionline/archive.php</link>
</image>
<item>
<title>Mathématique et statistique en science de l'information et en science de la communication</title>
<link>http://www.ibict.br/cionline/include/getdoc.php?id=1691&amp;article=913&amp;mode=pdf</link>
<description>L'application de la mathématique et de la statistique à l'étude des phénomènes informationnels a ent
<category>Resumo</category>
<dc:creator>Yves F. Le Coadic</dc:creator>
</item>

```

Figura 2 - Trecho do *feed* RSS (versão 2) gerado para as edições de 1995 a 2005 da revista CI-IBICT.

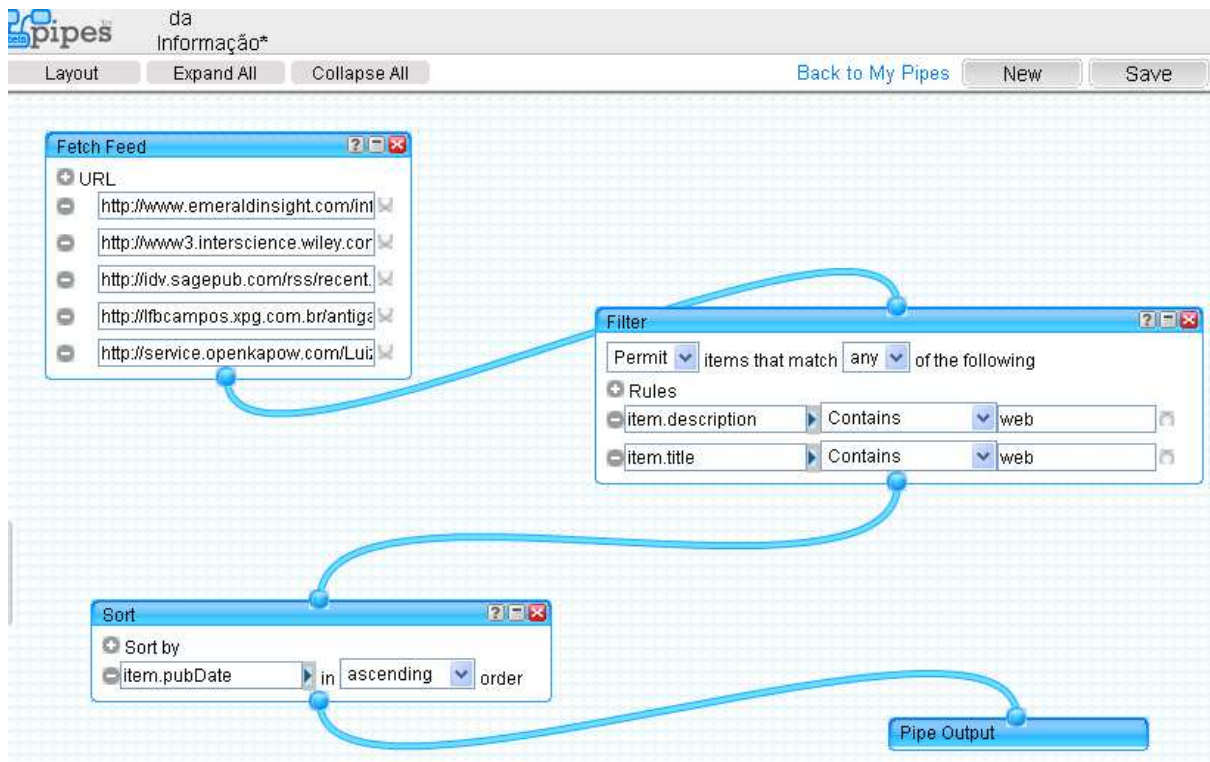


Figura 3 - Serviço (*pipe*) criado para filtragem e ordenamento dos *feeds* RSS agregados.

The screenshot shows the Feedreader 3.08 application window. The title bar reads 'Feedreader 3.08' and the menu bar includes 'Arquivos', 'Visualizar', 'Ferramentas', and 'Ajuda'. The main content area is titled 'Agregador\_Artigos\_Web' and features a 'Next unread' button. A table on the left lists filtered articles with columns for 'Titulo' and 'Data'. The 'Data' column for all items shows '05:37'. The selected article is 'Data cleansing for Web information retrieval using query independent features', dated '20:21 23/7/2007' and sourced from 'Agregador\_Artigos\_Web'. The preview on the right shows the beginning of the article's abstract.

Titulo	Data
Hoje	
Redes neurais e sua aplicação em ...	05:37
Uma proposta para gerenciamento...	05:37
Estratégias de produção e organiz...	05:37
Taxonomias web de clubes de fútb...	05:37
Buscando soluções para se publica...	05:37
Conversão de base de dados micro...	05:37
Un balance muy positivo para el nu...	05:37
Web sites governamentais, uma es...	05:37
Ferramentas de busca na Web	05:37
Introdução às linguagens de marcas	05:37
Avaliação do acesso a periódicos el...	05:37
Da bibliometria à webometria: uma...	05:37
Ferramentas alternativas para mo...	05:37

**Data cleansing for Web information retrieval using query independent features**  
 ☆ 20:21 23/7/2007, Agregador\_Artigos\_Web

Understanding what kinds of Web pages are the most useful for Web search engine users is a critical task in Web information retrieval (IR). Most previous works used hyperlink analysis algorithms to solve this problem. However, little research has been focused on query-independent Web data cleansing for Web IR. In this paper, we first provide analysis of the differences between retrieval target pages and ordinary ones based on more than 30 million Web pages obtained from both the Text Retrieval Conference (TREC) and a widely used Chinese search engine, SOGOU (www.sogou.com). We further propose a learning-based data cleansing algorithm for reducing Web pages that are unlikely to be useful for user requests. We found that there exists a large proportion of low-quality Web pages in both the English and the Chinese Web page corpus, and retrieval target pages can be identified using query-independent features and cleansing algorithms. The experimental results showed that our algorithm is effective in reducing a large portion of Web pages with a small loss in retrieval target pages. It makes it possible for Web IR tools to meet a large fraction of users' needs with only a small part of pages on the Web. These results may help Web search engines make better use of their limited storage and computation resources to improve search performance.

Figura 4 – Visualização parcial dos metadados dos artigos filtrados utilizando o agregador Feedreader.