

**DISSEMINAÇÃO SELETIVA DE INFORMAÇÃO NO
CONTEXTO DOS SERVIÇOS DE BIBLIOTECAS DIGITAIS**

ALMEIDA, R. L.

Departamento de Ciência da Informação e Documentação
Universidade de Brasília

DISSEMINAÇÃO SELETIVA DE INFORMAÇÃO NO CONTEXTO DOS SERVIÇOS DE BIBLIOTECAS DIGITAIS

ROBSON L. de ALMEIDA*
(rlalmeida@unb.br)

Resumo: Esse trabalho apresenta os resultados parciais de um estudo exploratório sobre serviços de notificação de conteúdo a ser integrado em projetos de bibliotecas digitais. Comenta o conceito de Disseminação Seletiva de Informação a partir da perspectiva de um sistema de filtragem, que visa aumentar o grau de relevância na recuperação de informação para o usuário. O trabalho propõe um modelo de uma mini-biblioteca digital temática na área de Ciência da Informação baseado na tecnologia de *Web Syndication* e discute as vantagens, formas de aplicação e eventuais barreiras para implementação deste recurso em serviços que compõem a infra-estrutura de uma biblioteca digital e/ou repositório institucional.

Palavras-chave: agregação de conteúdo; biblioteca digital; Disseminação Seletiva de Informação; RSS; *Web syndication*

SELECTIVE DISSEMINATION OF INFORMATION IN THE DIGITAL LIBRARIES CONTEXT

Abstract: This paper presents the partial results of an exploratory study about content syndication services to be integrated in projects of digital libraries. It comments on the concept of Selective Dissemination of Information from a perspective of a filtering system, which aims at increasing the degree of relevance when recovering information to the user. This work proposes a model of a micro thematic digital library in the area of Information Science, based on Web Syndication technology. It also discusses the advantages, forms of application and possible barriers to the implementation of this resource in services that make up the framework of a digital library and/or an institutional digital repository infrastructure.

Key-words: content aggregation; digital library; Selective Dissemination of Information; RSS; Web syndication

* mestrando do Programa de Pós-Graduação em Ciência da Informação da Universidade de Brasília (CID/ UnB)

1. INTRODUÇÃO

O volume de informação produzido nos últimos anos vêm se tornando um motivo a mais de preocupação para a nossa sociedade. A razão, segundo Filho e Delgado (2003), não está somente na dificuldade em localizar determinado conteúdo que atenda às necessidades do usuário mas, principalmente, no efeito de sua sobrecarga. O rápido desenvolvimento das tecnologias de informação e comunicação (TICs) têm colaborado para a disseminação desordenada de novos conteúdos, os quais não encontram limites nos suportes digitais. Parece ser consenso a necessidade de haver serviços de intermediação que atuem como filtros, permitindo a coleta e a distribuição seletiva de informação aos usuários de acordo com seu interesse.

É bem verdade que a simplicidade das atuais ferramentas de publicação na rede tem beneficiado tanto quem produz quanto quem consome informação, oferecendo mecanismos ágeis e de baixo custo para comunicação de novas idéias. Por outro lado, a rápida disseminação da informação digital tem exigido maior atenção em relação à qualidade do conteúdo a ser publicado e discernimento quanto ao seu uso. Essa preocupação é compartilhada por Feitosa (2006), que atribui ao excesso de informação os maiores problemas de acesso e de recuperação de qualquer tipo de conteúdo.

Profissionais das mais diversas áreas do conhecimento, comprometidos com a pesquisa e desenvolvimento de tecnologias voltadas para a área de recuperação da informação têm buscado soluções que possam lidar com essa situação (SARACEVIC, 1996). As valiosas contribuições de precursores na pesquisa de sistemas de recuperação e de disseminação de modo seletivo considerando o perfil do usuário até hoje produzem reflexos no desenvolvimento de sistemas de informação.

As bibliotecas digitais, assim como os repositórios institucionais, configuram-se como um desses sistemas modernos de recuperação devido às suas características de estruturação de dados e organização da informação, proporcionados por mecanismos eficientes e interoperáveis entre si. É dentro dessa perspectiva e, considerando, o caos informacional da Web, que pretendemos revisitar o conceito de Disseminação Seletiva de Informação (DSI) a partir de uma tecnologia emergente conhecida como *Web Syndication*, popularizada pelo formato RSS (*Really Simple Syndication*).

O propósito desse trabalho é propor um serviço de personalização de conteúdo digital a ser integrado em projeto de biblioteca digital temática no âmbito da área de Ciência da Informação. Sua contribuição é demonstrar as potencialidades de como este recurso poderá vir a ser utilizado por serviços que compõem a infra-estrutura de uma biblioteca digital.

2. REFERENCIAL TEÓRICO

A partir do advento da Word Wide Web (WWW), no início dos anos 90, o fluxo de informação digital passou a ser disseminado muito rapidamente e a custo baixo, porém de modo desordenado. De acordo com estudo realizado por Lyman e Varian (2003), anualmente o mundo todo produz cerca de 1,5 bilhões de gigabytes de informação registrada em diferentes suportes (impresso, magnético, ótico e imagético). Esse número representa uma média de 250 Mb por habitante do planeta. Outro indicador revelado pela pesquisa indica uma taxa de crescimento de 30% no volume de informação nova produzida no mundo.

O aumento do fluxo de informação que trafega na Internet conjugado com a inabilidade de ser convertida em efetivo conhecimento direcionado ao usuário tem trazido aos núcleos produtores de informação sérios desafios, os quais Araújo (1995, p. 70) diagnostica com precisão: "... seleção que não seleciona; indexação que isola e mutila; organização de arquivos que tem problemas quanto à sua própria integridade física e que se ampliam e repercutem no armazenamento; imprecisão e indeterminismo da análise e negociação de

questões; limitações e dicotomização da estratégia de busca/recuperação; incoerência e perplexidade na disseminação/acesso ao documento...”.

2.1 BIBLIOTECAS DIGITAIS

Cunha (1999) considera que o conceito de biblioteca digital é o resultado de um processo evolutivo, desde a implementação dos catálogos em linha pelas bibliotecas, nos anos 70, até o uso regular de banco de dados e CD-ROM para recuperação de referências bibliográficas e textos completos de artigos de periódicos e outras fontes. “A partir de 1994, por exemplo, com o fenomenal crescimento da Internet, as possibilidades de acessar e recuperar informações aumentaram de forma nunca antes imaginada.” (CUNHA, 1999, p. 258). Em outro trabalho, o mesmo autor sintetiza o conceito, afirmando que “[...] bibliotecas digitais são simplesmente um conjunto de mecanismos eletrônicos que facilitam a localização da demanda informacional, interligando recursos e usuários” (CUNHA, 2000, p.77).

No entendimento de Camargo e Vidotti (2006), o termo biblioteca digital possui várias conotações dependendo dos diferentes pontos de vista dos autores. O importante é que este tipo de serviço implica no surgimento de novas funções em relação ao armazenamento, organização e recuperação de informação. Desse modo, uma biblioteca digital deve permitir acesso remoto e simultâneo, disponibilizando serviços e produtos, possibilitando recuperar documentos completos e bibliográficos, possuindo diversos tipos de registros (música, imagem) e utilizando sistemas inteligentes que ajudam na recuperação da informação.

2.2 RECUPERAÇÃO E FILTRAGEM DA INFORMAÇÃO

O termo “recuperação da informação” foi criado em 1951 pelo cientista americano Calvin Mooers e é definido como sendo “o nome do processo ou método em que um possível usuário pode converter a sua necessidade de informação numa lista real de citações de documentos armazenados que contenham informações úteis a ele” (MOOERS apud FEITOSA, 2006).

Segundo a concepção de Mooers, dentre os problemas tratados por esse campo destacam-se a representação e organização da informação, a especificação da busca por informação e o da criação de mecanismos para a recuperação. Particularmente, os Sistemas de Recuperação de Informação (SRIs) deveriam responder: a) como descrever intelectualmente a informação? b) como especificar intelectualmente a busca? c) que sistemas, técnicas ou máquinas devem ser empregados? (BAX et al, 2004).

O processo de refinamento nas buscas sobre os sistemas de recuperação é denominado “filtragem da informação”, termo cunhado em 1982 por Peter Denning. Frequentemente, o conceito é utilizado para descrever uma variedade de atividades envolvendo entrega da informação para pessoas que precisem dela. Assim, os sistemas de filtragem baseiam-se em descrições individuais ou grupos de informação preferenciais, geralmente chamados de perfil que, por sua vez, representa um interesse individual de longo prazo, podendo expressar tanto o que as pessoas desejam encontrar como também o que elas não querem (BELKIN e CROFT, 1992).

Apesar dos sistemas convencionais de recuperação e de filtragem possuírem relacionamentos estreitos quanto ao seu objetivo de selecionar informação relevante de acordo com a necessidade do usuário, Belkin e Croft (1992) identificaram diferenças primárias entre ambos. Primeiro, as preferências do usuário (perfil) em sistemas de filtragem de informação representam interesses de longo prazo, enquanto as expressões de busca nos sistemas de recuperação tendem a representar interesses de curto prazo, os quais podem ser satisfeitos imediatamente a partir do resultado da busca. Em segundo lugar, enquanto a recuperação de informação normalmente se preocupa com usuários individuais do sistema, a filtragem está focada com nos usuários que utilizam o sistema regularmente. Por fim, enquanto a filtragem

envolve o processo de “remoção” de informação irrelevante, a recuperação cobre o resultado de todos os mecanismos de busca de um fluxo.

Vale ressaltar que o conceito de filtragem de informação não é novo nem tampouco é limitado a documentos eletrônicos. De fato, esse processo de filtragem ocorre diariamente em nossas vidas. Por exemplo, quando preferimos a leitura de determinado tipo de publicação em detrimento de outra semelhante, mas que contenham informações redundantes ou irrelevantes aos nossos interesses (FOLTZ e DUMAIS, 1992).

Ainda em relação à atividade de localização de informação, Saracevic et al (1997) mostram que existem diversas abordagens relacionadas ao problema da especificação da busca por informação. Essas abordagens podem estar centradas no sistema, no usuário ou em ambos, como ocorre com os sistemas de disseminação seletiva, uma das primeiras formas de filtragem de informação eletrônica.

2.3 DISSEMINAÇÃO SELETIVA DE INFORMAÇÃO (DSI)

O processo de Disseminação Seletiva de Informação ou *Selective Dissemination of Information* (SDI), foi criado na década de 50 por Hans Peter Luhn, da IBM Corporation, com a finalidade de aperfeiçoar serviços de alerta oferecidos por bibliotecas, centros de documentação e sistemas especializados de informações documentais. Luhn (1961) define DSI como “aquele serviço dentro de uma organização que se refere à canalização de novos itens de informação, vindos de quaisquer fontes, para aqueles pontos onde a probabilidade de utilização, em conexão com o interesse corrente [do usuário], seja alta”.

Como um serviço de referência, o objetivo de um DSI é reunir a literatura mundial corrente e anunciá-la seletivamente para uma grande comunidade de usuários. (LONGO, 1978). Para Housman (1973), o DSI atuaria como a automação de uma função clássica da biblioteca tradicional, que consiste em informar ao usuário sobre as novas aquisições, compatíveis com seu interesse de consulta. Antes, segundo explica Longo (1978, p. 102), “existiam serviços manuais de alertas por meio do manuseio de títulos de periódicos e da compilação de resumos para as referências que fossem consideradas relevantes, distribuindo-as então pelos usuários”.

Como um serviço tipicamente voltado para a filtragem de informação, um dos aspectos mais importantes em relação à configuração de um DSI é a construção do “perfil de interesse” a partir de uma base de conhecimento específico sobre a necessidade do usuário, característica que o distingue dos demais serviços de alerta. Ainda em relação ao perfil, deve-se mencionar que o mesmo pode ser prontamente modificado de acordo com a mudança da necessidade de informação do usuário.

Entre as vantagens desse tipo de serviço, Longo (1978) destaca a redução considerável do tempo gasto pelos usuários durante o exame e seleção de literatura corrente, o maior uso da coleção das unidades de informação pelos usuários, além de incentivar uma mudança no comportamento do profissional da informação que passa a ter uma postura mais ativa em relação ao usuário.

Sousa e Brigheti (1981), por sua vez, apontam para a responsabilidade das bibliotecas de prover sua comunidade de serviços de recuperação e localização da informação compatíveis com a necessidade de sua clientela. Disseminar seletivamente a informação é uma das formas de disponibilizá-la ao usuário, uma vez que se trata de uma extensão dos serviços convencionais de referência.

Atualmente os serviços de DSI continuam tendo um alto índice de aceitação entre os pesquisadores apesar das facilidades oferecidas pelos bancos de dados online e das diversas opções de ferramentas de busca. Isso porque o grau de relevância da informação recuperada por meio de um DSI é bastante superior às encontradas nas pesquisas bibliográficas realizadas

pelo próprio usuário, sem contar com o auxílio de sistemas de filtragem ou a intermediação de um profissional habilitado.

No âmbito dos serviços oferecidos pelas bibliotecas digitais, parece que estes são orientados a um usuário genérico. Dificilmente são levadas em consideração as preferências de seus usuários quanto ao acesso à informação. Na visão de Speroni et al (2006), o aumento do número de bibliotecas digitais e de suas coleções implica no crescimento da quantidade de usuários e seu nível de exigência. Dessa forma, “as novas gerações de bibliotecas digitais devem oferecer serviços mais avançados que atendam às expectativas dos usuários e consequentemente, otimizem o processo de acesso às informações que sejam mais relevantes” (SPERONI et al, 2006, pg. 12-13).

3. WEB SYNDICATION

A tecnologia de *syndication* (sindicação), que pode ser entendida aqui como a atividade de coleta de conteúdos replicando-os em ambientes digitais, surgiu no final dos anos 90 como uma solução para minimizar o problema da sobrecarga da informação na Web. Atualmente vem se tornando popular devido a difusão do padrão RSS, adotado pelos principais serviços de informação que possuam atualização freqüente, como os sítios de notícias em tempo real, por exemplo. Por meio de seus “canais”, como também são chamados os arquivos RSS, o usuário tem a possibilidade de receber notificações automáticas de novos conteúdos, oriundos de quaisquer fontes de informação baseadas na Internet.

O nome “RSS” é um termo abrangente que representa, pelo menos, duas especificações diferentes (porém, paralelas). RSS pode significar “*Rich Site Summary*”, “*RDF Site Summary*” ou “*Really Simple Syndication*”, dependendo da versão utilizada¹. Independente de como são chamados ou do número de sua versão, basicamente trata-se de um dialeto ou vocabulário da família XML (*eXtensible Markup Language*), voltado especialmente para a captura automática e distribuição de conteúdos de sítios Web. Normalmente é usado para distribuição de conteúdos atualizados com certa freqüência.

Para aqueles já familiarizados com o código HTML (*HyperText Markup Language*), a estrutura de um documento RSS (*feed*) parecerá familiar, como pode ser visto na Figura 1. Porém, diferentemente do HTML, que se limita a fornecer um formato universal para representar a informação sem acrescentar referência quanto à estrutura e significado (semântica) dos dados contidos, o padrão RSS, como linguagem baseada nas regras de XML, é capaz de representar informação sobre recursos Web.

No entendimento de Nottingham (2005), as especificações RSS são essencialmente linguagens de representação de metadados essenciais que permitem a distribuição de resumos de sítios e listas de *links* em tempo real com o propósito de fornecer aos usuários informações rápidas para ajudarem se decidirem se querem ou não ter acesso ao conteúdo completo. Em sua forma mais simples, o conjunto de metadados utilizado por qualquer versão de RSS consiste no título, seguido de um *link* (localização da página) e uma breve descrição (resumo) do conteúdo a ser disponibilizado.

O uso de *tags* (rótulos) específicos, como <title>, <link> e <description> permitem tratar cada unidade de informação como um objeto distinto, permitindo estruturar um determinado conteúdo para que esse seja interpretado e tratado por *scripts* ou programas especiais de modo que esses dados se transformem em objetos qualificados como atributos.

¹ Para uma discussão mais aprofundada sobre o surgimento do formato RSS, leia: <http://en.wikipedia.org/wiki/Rss>.

```

<?xml version="1.0" encoding="UTF-8"?>
<rss version="2.0">
  <channel>
    <title>Revista Exemplo</title>
    <description>Descrição resumida do periódico</description>
    <link>http://www.cid.unb.br/</link>
    <language>pt-br</language>
    <lastBuildDate>Wed, 17 Jan 2007 14:02:11 -0200</lastBuildDate>
    <managingEditor>rlalmeida@unb.br</managingEditor>
    <pubDate>Wed, 17 Jan 2007 15:47:50 -0200</pubDate>
  <item>
    <title>Título do artigo 1</title>
    <link>http://www.cid.unb.br/artigo_1</link>
    <description>
      Resumo do artigo 2
    </description>
  </item>
  <item>
    <title>Título do artigo 2</title>
    <link>http://www.cid.unb.br/artigo_2</link>
    <description>
      Resumo do artigo 2
    </description>
  </item>
</channel>
</rss>

```

Figura 1 – Exemplo de um *feed* RSS (versão 2.0)

Tem-se, então, a possibilidade de reutilização automatizada da informação, podendo compartilhá-la mais facilmente com os usuários e com outros sistemas de informação (interoperabilidade), organizá-las em bancos de dados e realizar pesquisas automáticas. Como uma aplicação XML, a idéia geral é que tudo aquilo que seja passível de ser descrito por meio de <tags>, pode ser integrado via RSS.

Tais características se identificam com os princípios da Web Semântica, que, segundo Bernes-Lee et al (2001) reúne um conjunto de modelos conceituais que visam propor soluções para a melhor compreensão e gerenciamento dos conteúdos armazenados na Web atual, independente da forma em que estes se apresentem, seja texto, som, imagem e gráficos.

Enquanto as pesquisas sobre o tema seguem em plena ebulição, há quem afirme que a tecnologia RSS já represente a prática da Web Semântica mais difundida fora do círculo acadêmico (DOWNES, 2005). De acordo com o pesquisador, RSS é o que surgiu, de fato, para descrição de conteúdos em tempo real, sendo utilizado por milhões de sítios na Web para descrever não somente recursos, mas até mesmo pessoas, lugares e objetos. Por ser baseado no padrão XML, o RSS também permite a troca de informações entre diversas plataformas e possibilita a descrição de dados em arquivos-texto.

No caso da versão RSS 1.0, baseada no padrão RDF, o entendimento da tecnologia como representação da informação torna-se ainda mais clara já que arquivos RDF são recomendados pelo W3C desde 1999 para criação de um modelo simples de dados, com uma semântica formal, sendo a linguagem ideal para se representar conteúdos informacionais na Internet.

Outra característica do formato RSS que contribui para o nosso entendimento deste recurso como forma de representação da informação é a possibilidade de incorporar outros modelos de metadados consagrados, como o Dublin Core, desenvolvido por cientistas da informação para descrição de documentos. Esse modelo inclui elementos úteis tais como dc:date, para associação de datas entre itens, dc:subject, destinado à categorização de

informações e dc:rights, para direitos de propriedade ou intelectual sobre determinado item, entre outros.

A popularidade da tecnologia RSS está relacionada com a agilidade com que este formato confere à leitura de novos conteúdos por dispensar o acesso ao sítio no qual se encontra a informação publicada originalmente. Outra razão para a sua rápida adoção pode ser explicada pelo fato de que é possível implantar e utilizar este recurso sem nenhum custo, tanto para leitura quanto para a disponibilização de conteúdos por se tratar de um documento que contém apenas texto ASCII, facilmente interpretado pela maioria das aplicações de computador.

A maneira mais prática que o usuário comum tem para se beneficiar dessa tecnologia é dispor de um *software* “agregador de conteúdo”, um tipo de aplicativo capaz de ler e interpretar os *feeds* RSS. Tais programas geralmente são gratuitos, de fácil instalação e, na maior parte, se assemelham a um leitor de correio eletrônico. Do mesmo modo, existem serviços disponíveis na Web que realizam a tarefa de “agregação” em tempo real. Basta um simples cadastro e fazer um *login* na página toda vez que o usuário desejar receber as atualizações sobre os assuntos escolhidos.

Por meio dos leitores de RSS é possível fazer uma espécie de “assinatura” dos conteúdos de diversas fontes de informação e examinar rapidamente as novidades de maneira condensada, de acordo com a periodicidade desejada. Quando um determinado tema despertar interesse, basta clicar no título para exibir a página com o texto integral.

Na Figura 2, que mostra uma tela típica de um desses programas, é possível ver na coluna da esquerda todos os canais escolhidos e adicionados, os quais podem ser lidos na coluna da direita. Na parte superior dessa coluna encontram-se listados os tópicos (*headlines*), enquanto na parte inferior, observa-se parte ou todo conteúdo selecionado. O ato de “clicar” duas vezes no título do tópico fará com que, na janela inferior, seja exibido o texto integral, apontando para a informação originalmente publicada na Web. Os *feeds* RSS podem ser armazenados ou apagados e existe, ainda, a possibilidade de selecioná-los por assunto e/ou por data.

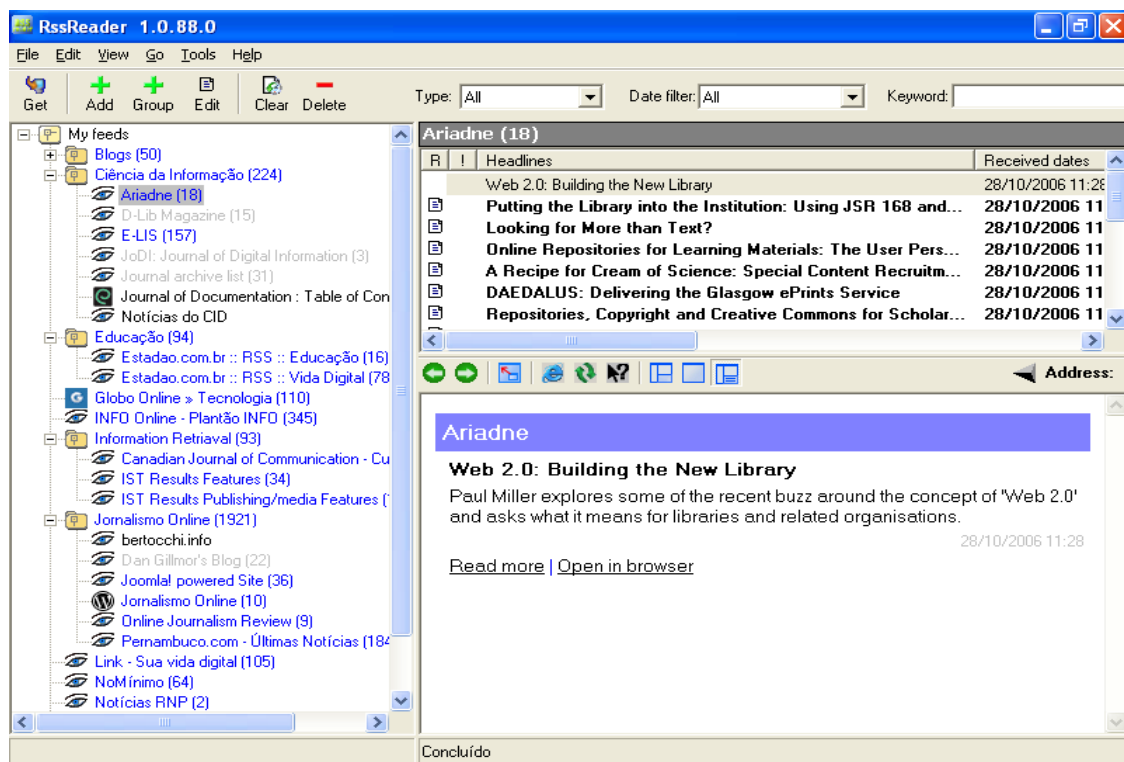


Figura 2 – tela do RSS Reader, exemplo de aplicativo agregador de conteúdo

É importante frisar que a tecnologia de *Web Syndication* não deve ser entendida como sinônimo de RSS, já que existe pelo menos um outro formato igualmente importante e também voltado exclusivamente para distribuição de conteúdos na Web. Trata-se do Atom, criado com a finalidade de ser 100% neutro, aberto e facilmente implementado por qualquer desenvolvedor. Tal como o RSS, o Atom é baseado no padrão XML, mas seu desenvolvimento é tido como mais sofisticado. De acordo com definição da Wikipédia, trata-se de uma proposta de unificação do RSS 1.0 com o RSS 2.0 e deverá ser seu substituto natural, já que conta com o apoio de grandes corporações.

4. METODOLOGIA

O presente estudo possui características de uma pesquisa exploratória, ou seja, trata-se de uma primeira aproximação para a investigação de um fenômeno, no caso a possibilidade de integração da tecnologia *Web Syndication* em projetos de bibliotecas digitais.

Para demonstrar o potencial de sua utilização, procuramos desenvolver um protótipo de uma mini-biblioteca temática na área de Ciência da Informação, construída a partir da agregação de *feeds* RSS já disponibilizados por alguns periódicos eletrônicos estrangeiros, blogs e repositórios digitais. Os conteúdos dos periódicos nacionais da área foram adicionados posteriormente, já que não havia *feeds* disponibilizados pelas publicações da área. Como método de trabalho, procuramos criar manualmente esses arquivos para poder integrar alguns periódicos nacionais ao nosso projeto.

Para a confecção do protótipo da mini-biblioteca, contamos com o auxílio do serviço Netvibes (<http://www.netvibes.com>), que oferece uma página personalizada a qualquer usuário cadastrado, capaz de gerenciar módulos criados a partir de *feeds* RSS/Atom. Trata-se de uma aplicação Web 2.0 disponível gratuitamente na Internet e conhecida no meio técnico como Marshup, termo é definido pela Wikipédia como uma aplicação que combina conteúdo de mais de uma fonte em uma experiência integrada.

Os *feeds* dos periódicos nacionais foram construídos para 10 publicações eletrônicas da área da Ciência da Informação que utilizam o Sistema Eletrônico de Editoração de Revista (SEER), uma ferramenta voltada para administração do processo editorial de periódicos eletrônicos adaptada do *Open Journal System* (OJS) pelo Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT/MCT). A preferência pelas revistas baseadas no OJS se justifica pelo fato deste ser um sistema já consolidado para publicação e gerenciamento de periódicos revisados por pares. Os periódicos selecionados foram: Arquivística.net, Ciência da Informação, Em Questão, Informação e Sociedade, Informação e Informação, Perspectivas em Ciência da Informação, Pesquisa Brasileira em Ciência da Informação e Biblioteconomia, Revista ACB, Revista Digital de Biblioteconomia e Ciência da Informação e Transinformação.

Para a construção dos canais RSS de cada periódico, contamos com uma ferramenta de autoria chamada FeedForAll (<http://www.feedforall.com>). Posteriormente, os arquivos foram transferidos para um serviço de hospedagem com o propósito de formar uma base para que os *feeds* pudessem ser facilmente sindicados.

5. RESULTADOS

Uma vez instalada, a ferramenta de autoria facilitou a tarefa de construção dos canais sem a necessidade de escrever códigos próprios do formato RSS. Primeiramente, foi necessário preencher as informações básicas do canal: *<title>* (título do canal), *<link>* (URL da página correspondente ao canal) e *<description>* (uma breve descrição sobre o conteúdo do canal ou *feed*). Em seguida, alimentamos o conteúdo dos itens propriamente ditos, que correspondem à entrada de metadados dos artigos. As informações obrigatórias são as mesmas para a criação de um *Web feed*: *<title>* (do artigo), *<link>* (localização da página do

artigo) e <description> (resumo do artigo). A Figura 3 ilustra o preenchimento dos campos requeridos da aba Itens.

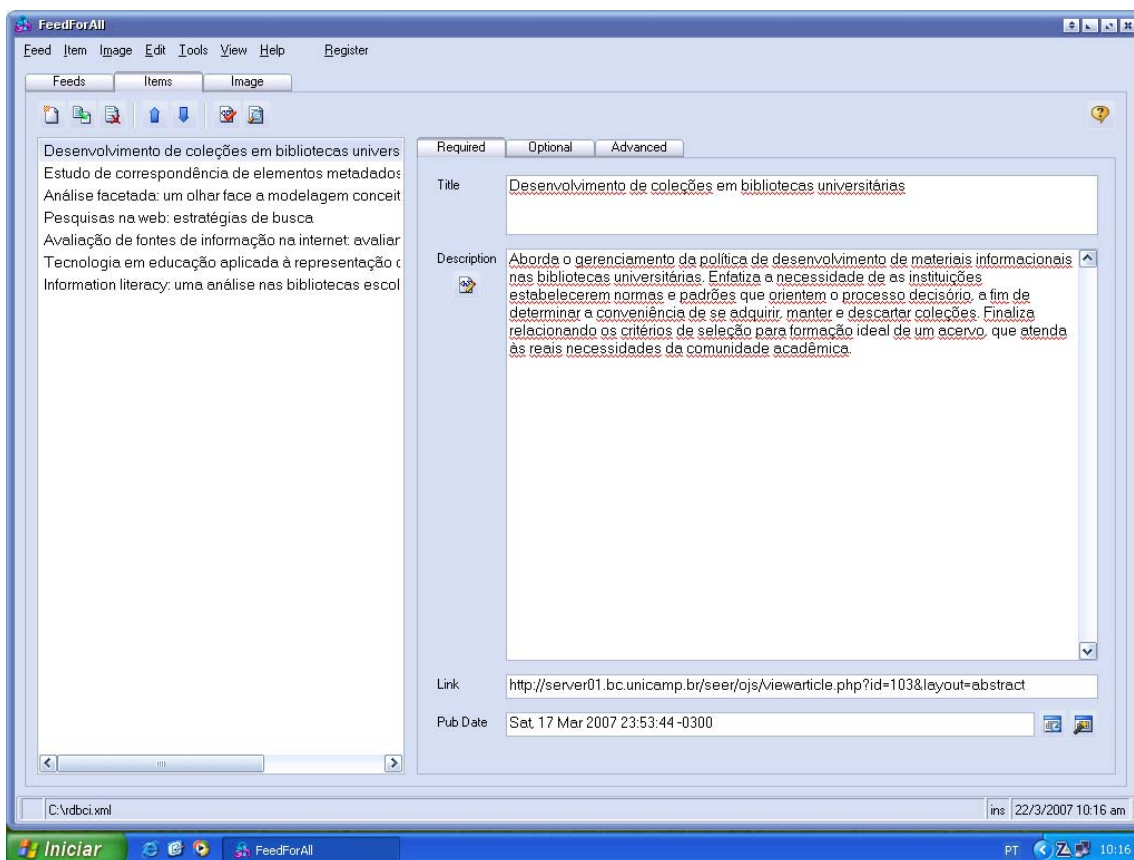


Figura 3 – Tela do FeedForAll, ferramenta para criação de feeds RSS

A mini-biblioteca digital em Ciência da Informação reúne referências de periódicos nacionais e internacionais, repositórios e blogs. Cada módulo criado no âmbito do serviço Netvibes representa o conteúdo equivalente de um *feed* RSS/Atom (Figura 4).

A função geral do sistema é realizar uma varredura periódica de todos os canais cadastrados, de acordo com a preferência do usuário. Assim, a mini-biblioteca identificará automaticamente as atualizações nos conteúdos dos sumários dos periódicos, *posts* dos blogs e depósitos mais recentes dos repositórios. Se este usuário desejar acessar o texto completo do artigo ou do *post* em questão, ele poderá remeter-se diretamente à página correspondente a fonte de informação.

O modelo apresentado demonstra uma das principais aplicações referentes à utilização de *Web feeds* atualmente. No caso, apresentamos a dinâmica de um mecanismo capaz de reunir, em única página Web, a indicação dos artigos (com respectivo resumo) dos periódicos nacionais e estrangeiros, além de blogs e repositórios em Ciência da Informação. O modelo poderia ser estendido para agregar, ainda, base de dados ou outras fontes relacionadas, desde que disponham de *feeds* associados.

Uma vez tendo todo esse conteúdo agregado, é possível fazer buscas simultaneamente em todos os tipos da coleção. Assim, se o usuário desejar fazer uma pesquisa pelo termo “ontologia” em todos os periódicos nacionais, por exemplo, o resultado da busca trará todos os artigos que contenham o termo “ontologia” em seu título ou no resumo, independente da publicação.

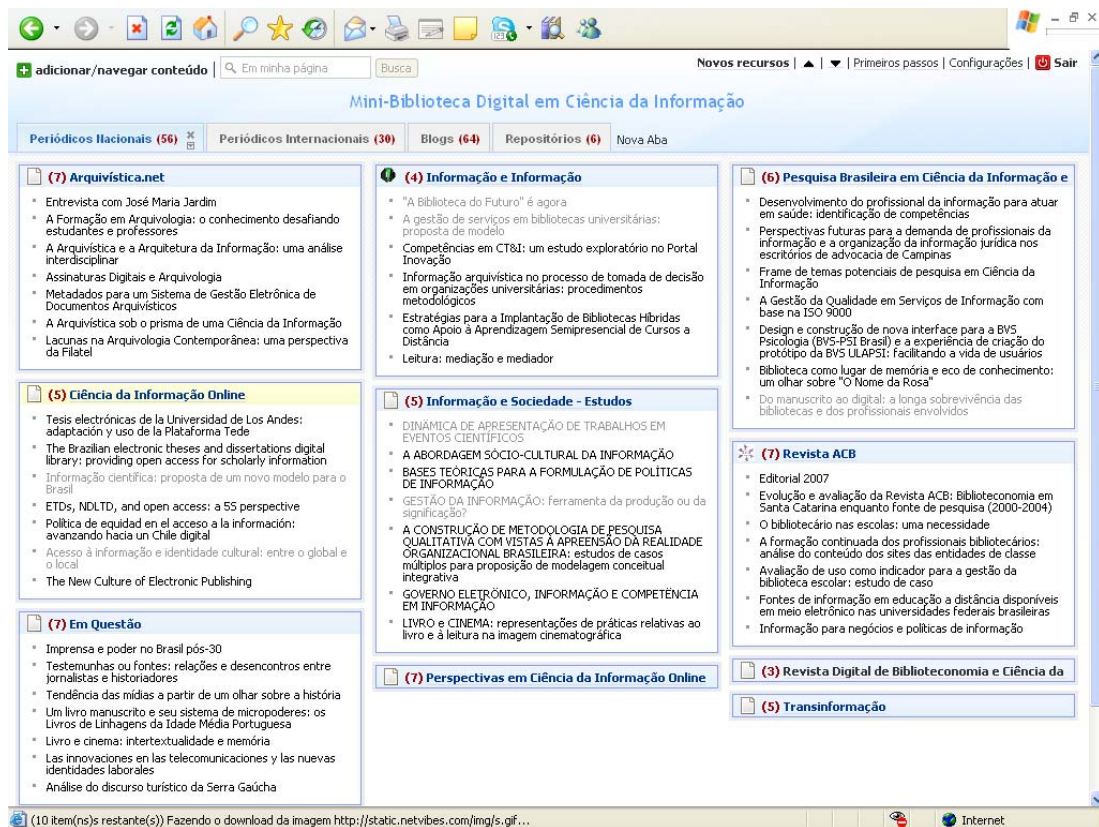


Figura 4 – Tela do protótipo da Mini-Biblioteca Digital em Ciência da Informação

6. CONSIDERAÇÕES SOBRE A WEB SYNDICATION

Atualmente, a maioria dos periódicos científicos que utilizam o sistema de publicação OJS já dispõe de um serviço de notificação que oferece ao usuário a possibilidade de se cadastrar para receber, por meio de correio eletrônico, um aviso com o sumário das novas edições a medida em que forem publicadas. Entretanto, se este usuário desejar acompanhar os trabalhos publicados sobre uma determinada área, terá que repetir o procedimento de cadastro em cada um dos periódicos, o que significa dizer que o mesmo receberá uma notificação diferente para cada atualização. Com a utilização de um serviço de *Web syndication*, o usuário não precisará, necessariamente, fazer um cadastro para se manter atualizado e ainda tem a vantagem de manter a sua caixa postal vazia.

Contudo, verificamos barreiras que podem impedir a ampla difusão deste tipo de recurso. O principal empecilho, na nossa opinião, é o próprio desconhecimento da tecnologia de sindicância, tanto por parte dos projetistas de serviço de informação quanto de seus usuários. No Brasil, de uma maneira geral, apenas os grandes jornais de circulação nacional e alguns outros sítios especializados disponibilizam *feeds* para seus usuários. Mesmo no ambiente acadêmico, com exceção dos cursos de informática, ainda é grande o desconhecimento a respeito de termos como *feeds*, *syndication* ou *aggregator*.

Outro problema ocorre com relação ao próprio procedimento para a inclusão de um *feed*: copiando a URL de um *link* (normalmente com extensão .xml ou .rss.) e colando de volta no aplicativo agregador de conteúdo. Embora seja um processo simples, ainda não é algo tão óbvio. Ao acessar um arquivo com extensão .rss, os programas navegadores (*browsers*) normalmente exibem o código XML do documento, algo incompreensível e que, às vezes, assusta os usuários principiantes.

Acreditamos que o usuário/leitor aos poucos aprenderá a lidar com um “poder” até hoje pouco imaginado: o de personalizar o conteúdo de informação que deseja consumir, bem como produzir novos conteúdos utilizando-se dos recursos oferecidos pelos formatos

RSS/Atom, sem a intervenção de intermediários. Desse modo, ao ler um artigo sobre “Web Semântica”, por exemplo, o usuário poderá, se desejar, assinar outros *feeds* sobre o mesmo tema e que estejam sindicados a partir de outros artigos, bases de dados, repositórios, etc. Tudo passará a ser interoperável graças a compatibilidade proporcionada pelas aplicações baseadas na linguagem XML e seus derivados. Na visão do usuário não haverá mais distinção entre o periódico A, B ou C de determinada área. Para ele é como se houvesse uma única fonte de informação, facilmente disponível a partir de sua “biblioteca digital pessoal”.

Nessa direção, nos próximos anos deverá haver uma explosão de novos Sistemas de Recuperação de Informação (SRIs) que contemplem *Web feeds*. Desse modo, uma vez tendo recuperado os resultados a partir de uma pesquisa em uma base de dados, o usuário tem a opção de assinar o *feed* relacionado àquela expressão de busca e, assim, manter-se atualizado em relação aquele determinado assunto. Um dos poucos serviços nesses moldes e que já se encontra em funcionamento é oferecido pela Agência Brasil (<http://www.agenciabrasil.gov.br>), que, além de disponibilizar *feeds* RSS para mais de 120 assuntos distintos, permite que o usuário crie novos canais a partir do resultado de sua pesquisa.

Uma iniciativa bem sucedida em termos de uso de *Web Syndication* em bibliotecas digitais é o IMesh Toolkit Project², do International Digital Libraries Programme (JISC/NSF), responsável pelo desenvolvimento de um módulo que permite o compartilhamento de conteúdos que utilizem o protocolo OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting), disponibilizando *feeds* RSS a partir de consultas realizadas nos repositórios das bibliotecas.

O próprio *software* Dspace (<http://www.dspace.org>), um dos mais utilizados para a construção de repositórios institucionais, já dispõe de um módulo para geração de RSS 1.0 e 2.0. Com uma simples alteração no arquivo de configuração <dspace.cfg>, é possível habilitar o programa de modo que o usuário do serviço receba uma notificação toda vez que um novo documento for depositado. Apesar do recurso já se encontrar disponível, não identificamos até o momento qualquer repositório que esteja adotando a prática de disponibilizar *Web feeds* para seus usuários.

7. CONCLUSÕES

No que se refere às aplicações relacionadas às bibliotecas digitais, nos parece conveniente a inclusão de *feeds* RSS tanto para efeito de divulgação sobre um novo depósito de material, por exemplo, como para manter o usuário atualizado a respeito de um determinado tema de seu interesse. Além disso, existe a possibilidade de integração com outros tipos de fontes de informação (bases de dados, diretórios, catálogos), devido à compatibilidade proporcionada pela linguagem XML.

Após nossa investigação sobre o potencial da utilização do RSS como recurso capaz de incrementar as funcionalidades de bibliotecas digitais, podemos resumir nossas conclusões com base nos seguintes tópicos:

- O conjunto dos recursos propostos pelos padrões da *Web syndication* configura-se como inovação tecnológica no campo dos novos serviços de referência para unidades de informação, bem como para o desenvolvimento de potencialidades de bibliotecas digitais e periódicos eletrônicos;
- Verifica-se o RSS como uma ferramenta de sucesso comprovado para a notificação (alerta) automatizado de conteúdos em ambientes Web;

² The IMesh Toolkit Project (<http://www.imesh.org/toolkit>)

- O comprometimento dos editores e desenvolvedores é fundamental para a disseminação de novos produtos que façam uso das tecnologias de sindicância de conteúdo;
- A tecnologia pode ser facilmente implementada nos Sistemas de Recuperação de Informação e de Disseminação Seletiva;
- Mais estudos são necessários para ampliar a discussão sobre o tema, buscando-se novos enfoques e aplicações;
- Trata-se de uma tecnologia baseada no paradigma do compartilhamento de informação. Portanto, contribui para a geração de novos conhecimentos.

De acordo com a nossa percepção as vantagens para o usuário de biblioteca digital que disponibilize *Web feeds* são imensas. Ele passará a dispor de uma poderosa ferramenta por meio da qual poderá manter-se atualizado com relação à diversas fontes de informação e, ao mesmo tempo, será capaz de realizar pesquisas simultâneas em cima de conteúdos relevantes, refinando a sua pesquisa e aumentando a relevância dos termos recuperados.

A própria biblioteca digital, por sua vez, agregará valor aos serviços oferecidos e aumentará sua visibilidade na medida em que a inclusão de um simples *feed* RSS ou Atom possibilitará aos usuários/leitores manterem-se informados pelo próprio programa agregador, sem precisar visitar constantemente a página da biblioteca. Isso pode parecer ruim num primeiro momento, mas na verdade incrementa a visibilidade dos serviços, tornando mais fácil a sua divulgação independentemente da ação do usuário. Já que existem muitas fontes de informação na Web, a maioria dos usuários não voltam ao mesmo sítio todos os dias. Ao prover um *feed* para um serviço de informação, os editores estarão permanentemente em contato com seus usuários, aumentando as chances de que um determinado artigo desperte a sua atenção.

Finalmente, apesar das barreiras identificadas neste trabalho, o estudo sobre as potencialidades das tecnologias de *Web syndication* aponta que estas são plenamente viáveis para integração de qualquer sistema de informação baseado na Web, como é o caso das bibliotecas digitais.

8. REFERÊNCIAS

ARAÚJO, V. M. R. H. de. **Sistemas de informação : nova abordagem teórico-conceitual**. Ciência da Informação, Brasília, v. 24, n. 1, p. 54-76, jan./abr. 1995.

BAX, Marcello P., ALVARENGA, Lidia, PARREIRAS, Fernando S., BRANDÃO, Wladimir C. **Sistema Automático de Disseminação Seletiva de Informação**. In: IFLA M&M, 2004, São Paulo, Anais. São Paulo: USP, 2004. Disponível em <http://www.netic.com.br/docs/publicacoes/pub0009.pdf>. Acesso em: 13 fev. 2007.

BELKIN, N. J; CROFT, W. B. Information Filtering and Information retrieval: two sides of the same coin?. **Communications of the ACM**, V. 35, n. 12, dez.1992.

BERNES-LEE, T.; HENDLER, J.; LASSILA, O. The semantic web: a new form of web content that is meaningful to computers will unleash a revolution of new possibilities. **Scientific American**. New York, mai. 2001. Disponível em: <http://www.sciam.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21>. Acesso em: 02 fev. 2007

CAMARGO, Liriane S. A. de ; VIDOTTI, Silvana Ap Borseti Gregorio. Arquitetura da Informação para Biblioteca Digital Personalizável. **Encontros Bibli** (UFSC), v. esp 3, p. 45, 2006.

CUNHA, Murilo Bastos. Desafios na construção de uma biblioteca digital. **Ciência da Informação**, Brasília, v.28, n.3, p.255-266, set./dez 1999.

_____. Construindo o futuro: a biblioteca universitária brasileira em 2010. **Ciência da Informação**, Brasília, v.29, n.1, p.71-89, jan/abr.2000.

DOWNES, Stephen. Semantic networks and social networks. **The Learning Organization**; Vol. 12 Issue: 5. 2005

FEITOSA, Ailton. **Organização da Informação na Web: das tags à web semântica**. Brasília: Thesaurus, 2006, p.28-30.

FILHO, Antônio Mendes da Silva; DELGADO, Maria Viviane Monteiro. A sobrecarga da informação na era da Internet, **Revista Espaço Acadêmico** [On Line]. n.23. Abril 2003. Disponível em: <<http://www.espacoacademico.com.br/023/23amsf.htm>>. Acesso em: 20 mai. 2007.

FOLTZ, P. W.; DUMAIS, S. T. Personalized information delivery: an analysis of information filtering methods. **Communications of the ACM**. V. 35, n. 12, dez.1992.

HOUSMAN, Edward M. (1973). Selective dissemination of information. **Annual Review of Information Science and Technology**. Washington, American Society of Information Science, p. 221-241.

LONGO, Rose Mary Juliano. Disseminação seletiva da informação (SDI): “estado da arte” e tendências futuras. **Revista de Biblioteconomia de Brasília**, 6 (2), p.101-120, jul/dez. 1978.

LUHN, H. P. (1961) Selective dissemination of new scientific information with the aid of electronic processing equipment. **American Documentation**. V.12, p.131-138.

LYMAN, Peter; VARIAN, Hal R. "**How Much Information**", 2003. Disponível em: <<http://www.sims.berkeley.edu/how-much-info-2003>>. Acesso em: 08 nov. 2006.

SPERONI, Rafael de Moura et al. Uma Biblioteca Digital Aberta com Serviços Personalizados. In: **II Workshop de Bibliotecas Digitais**, 2006, Florianópolis. Anais do II Workshop de Bibliotecas Digitais. Florianópolis:2006 : Editado por Renata de Matos Galante, Ronaldo dos Santos Mello, 2006. p. 11-20.

SOUZA, C.G.; BRIGHENTI, N.C. Disseminação seletiva da informação: um serviço de referência. **Boletim ABDF**, v.4, n.1, p-28-37, jan./mar. 1981.

NOTTINGAM, Mark. **RSS Tutorial for Content Publishers and Webmasters**. Disponível em: < <http://www.mnot.net/rss/tutorial/>>. Acesso em: 02 jan. 2007.

SARACEVIC, Tefko. Ciência da informação: origem, evolução e relações. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 1, n. 1, p .41-62, jan./jun. 1996.

SARACEVIC, T.; SPINK, A.; WU, M. **User and intermediaries in information retrieval: what are they talking about?**1997. In: Jameson, A.; Paris, C.; Tasso, C. User