

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

**DA DISSEMINAÇÃO SELETIVA À WEB SYNDICATION:  
uma proposta para a comunicação científica**

***FROM SELECTIVE DISSEMINATION TO WEB SYNDICATION:  
a proposal to the scientific communication***

Robson Lopes de Almeida (PPGCI/UnB, rlalmeida@unb.br)

**Resumo:** Este trabalho apresenta os resultados parciais de um estudo preliminar sobre a aplicação de um serviço de notificação automática de artigos de periódicos científicos eletrônicos. Comenta o conceito de Disseminação Seletiva de Informação (DSIs) a partir da perspectiva de um sistema que visa aumentar o grau de relevância na recuperação de informação para o usuário e explora as potencialidades de uma tecnologia emergente conhecida como *Web Syndication* para distribuição e agregação de conteúdos digitais. Discute as vantagens, formas de aplicação e eventuais barreiras para implementação deste recurso. Como piloto, foram selecionados 10 periódicos eletrônicos da área de Ciência da Informação (CI) agregados em um único repositório temático por meio de *feeds* RSS. Os resultados do modelo proposto demonstram que a tecnologia empregada incrementa os sistemas de disseminação seletiva. A economia de tempo com o recebimento de conteúdo selecionado confere uma grande vantagem à utilização desse tipo de serviço, capaz de reunir, em um único ambiente, conteúdos variados produzidos por diferentes fontes, sem a necessidade de o usuário ter que acessar cada sítio individualmente.

**Palavras-chave:** agregação de conteúdo. Disseminação Seletiva de Informação. periódico eletrônico. RSS. *Web Syndication*.

**Abstract:** *This paper presents the partial results of a preliminary study about the usage of one service of automatic notification of electronic scientific journals. The concept of Selective Dissemination of Information (SDI) is discussed from the perspective of a system that aims at increasing the level of relevance in the information recovery for the user. The potentialities of a recent technology, known as Web Syndication, for the distribution and gathering of digital contents are also explored. The advantages, ways of usage and circumstantial obstacles for its implementing are discussed as well. For testing, 10 electronic journals related to Information Science were chosen. They were put together into just one thematic area by RSS feeds. The results of the proposed model show that the technology which was used increases the systems of selective dissemination of information. Time saving with the receiving of the selected content provides a relevant advantage to the usage of this kind of service, which is able to gather, in just one environment, several contents produced by different sources. As a consequence, the user has no need to access each homepage individually.*

**Keywords:** *content aggregation. Selective Dissemination of Information. electronic journals. RSS. Web Syndication.*

## 1. Introdução

Um novo capítulo da história das comunicações começou a ser escrito a partir do início dos anos 90, com o advento da *World Wide Web* – a parte gráfica e multimídia da Internet. Em pouco tempo, as páginas da Web tomaram conta da rede provocando uma verdadeira “explosão informacional”, com impacto ainda maior do que o mesmo fenômeno observado na década de 40, após o surgimento de periódicos especializados após a Segunda Guerra Mundial. Esse movimento acabou impulsionando o desenvolvimento da própria Ciência da Informação, que teve sua origem a partir da revolução científica e tecnológica decorrente do pós-Guerra. (SARACEVIC, 1996, p. 42).

Observa-se que o volume de informação disponibilizado na Web nos últimos anos vem se tornando um motivo a mais de preocupação para a nossa sociedade. E a razão não está somente na dificuldade em localizar determinado conteúdo que atenda às reais 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 mais limites nos suportes digitais. Essa situação implica no aparecimento de vários problemas, como a falta de organização e estruturação das informações, dificultando a sua recuperação.

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 difusão de novas idéias. Por outro lado, a velocidade com que a informação digital circula na rede tem exigido maior atenção em relação à qualidade do conteúdo a ser publicado e mais discernimento quanto ao seu uso. A comunicação científica também foi afetada por essas transformações. Em particular, destacamos o aumento do número dos periódicos científicos eletrônicos disponibilizados em todas as áreas do conhecimento.

Profissionais das mais diversas áreas do conhecimento, comprometidos com a pesquisa e desenvolvimento de tecnologias voltadas para a recuperação da informação têm buscado soluções que possam lidar com essa situação. Historicamente, porém, já nos anos 50 surgiam os primeiros mecanismos capazes de selecionar informação relevante para um determinado usuário, considerando seu perfil de interesse. Esse conceito até hoje é conhecido pelo nome de “Disseminação Seletiva de Informação” (DSI), criado por Hans Peter Luhn com a finalidade de aperfeiçoar os serviços de alerta oferecidos pelas bibliotecas, centros de documentação e sistemas especializados de informação documental.

Assim, se faz cada vez mais necessária a utilização de serviços de intermediação que atuem como filtros, permitindo a coleta e a distribuição seletiva de informação aos usuários de acordo com seu interesse. É dentro dessa perspectiva e, considerando, o caos informacional da Web, que pretendemos estudar o processo de divulgação da comunicação científica a partir da aplicação de uma tecnologia emergente conhecida como *Web Syndication*, popularizada pelo formato RSS, o qual permite oferecer aos usuários a possibilidade de receber notificações automáticas sobre a atualização de conteúdos disponibilizados na plataforma Web.

A contribuição do presente trabalho é demonstrar as potencialidades da referida tecnologia na agregação de funções para os periódicos eletrônicos, especialmente os que já dispõem de recursos que facilitem a implementação de *feeds* RSS<sup>1</sup>.

Adicionalmente, comenta as vantagens para os editores e leitores/usuários, além de descrever o progresso e resultados da experiência na produção dos primeiros *feeds* criados pelo autor a partir de uma amostra de nove periódicos eletrônicos da área de Ciência da Informação baseados no Sistema de Editoração Eletrônica de Revistas (SEER/OJS).

## 2. Disseminação Seletiva de Informação (DSI)

Luhn (1961, p. 132) define o processo de Disseminação Seletiva de Informação como sendo “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 serviço de referência, o objetivo de um DSI é reunir a literatura mundial corrente e anunciá-la seletivamente para uma grande comunidade de usuários (HOUSMAN, 1973). Para Longo (1978, p. 102-103), o processo de disseminação seletiva envolve uma série de fatores que contribui para que o usuário gaste menos tempo com o exame e a seleção da literatura corrente. Tais processos envolvem: 1) coleta da informação produzida; 2) indexação dessa informação; 3) divulgação da informação aos usuários; 4) acessibilidade da informação. Essas etapas dependem da eficiência das pessoas e serviços que estão envolvidos na execução dos mesmos (LONGO, 1978, p. 102).

Dessa maneira, o DSI atuaria como a automação de uma função clássica da biblioteca tradicional, consistindo em informar o usuário sobre as novas aquisições, compatíveis com seus hábitos e interesses de consulta (HOUSMAN, 1973). Antes, 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.

Sousa e Brigheti (1981, p. 28), 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.

Entre as vantagens desse tipo de serviço, Longo (1978, p. 103) 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. Apesar das vantagens, esse tipo de serviço também apresenta sérios problemas, como a dificuldade na padronização da produção de base de dados, a falta de reatualização por parte dos usuários e o elevado custo de operacionalização do serviço (LONGO, 1978, p.110-111).

Em alguns países os serviços de DSI tiveram início ainda na década de 60 e expandiram-se nos anos posteriores, tornando serviços correntes de disseminação nos grandes sistemas de recuperação de informações referenciais. No Brasil, a literatura registra algumas iniciativas que se destacaram, entre elas a do Sistema Orientado para Notificação Automática de Referências (Sonar), criado pela Comissão Nacional de Energia Nuclear/Centro de Informações Nucleares (CIN/CNEN), em 1972, e o serviço automatizado da Embrapa – Empresa Brasileira de Pesquisa Agropecuária, no final da década de 70.

Com as mudanças introduzidas pela tecnologia, os serviços de disseminação seletiva foram gradativamente incorporados aos formatos eletrônicos. Após o surgimento das bases de dados em CD-ROM, a partir da década de 80, as pesquisas bibliográficas ganharam agilidade e aumentaram as chances de se obter resultados satisfatórios. Atualmente os serviços de DSI continuam tendo um alto índice de aceitação entre os pesquisadores apesar das facilidades oferecidas pelas diversas opções de ferramentas de busca disponíveis gratuitamente na Internet.

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 contexto da Web, porém, existem algumas barreiras para a operacionalização de um serviço de disseminação seletiva nos moldes tradicionais, a começar pela dificuldade na construção do perfil do usuário de modo adequado, além da dificuldade em selecionar fontes de informação relevantes, haja vista o aumento da produção bibliográfica na Internet, além do tempo cada vez mais exíguo para aguardarmos pelo resultado de uma pesquisa mais sofisticada realizada por terceiros.

Para tentar amenizar esse problema, na segunda metade da década de 90 tiveram início algumas experiências com o envio de informação sob demanda para o usuário da Internet, a exemplo do serviço Crayon<sup>2</sup>. Outra iniciativa que obteve relativo sucesso na época foi o recurso “Active Channel”, desenvolvido pela Microsoft para o seu navegador Internet Explorer, em sua versão 4.0.

Esses dois projetos tinham em comum a missão de reunir, em um só local, conteúdos variados e dispersos na Web por meio de uma tecnologia batizada como *push*, já que a idéia era “empurrar” informações personalizadas para os usuários em vez de esperar que estes visitassem os sítios para “puxar” (*pull*) os conteúdos de que necessitavam. Empresas especializadas, como a norte-americana PointCast, bem que tentaram popularizar a tecnologia, mas os programas utilizados na época eram complicados demais para os usuários que ainda estavam se adaptando ao ambiente Web, recém criado.

### 3. Web Syndication

A tecnologia de *Web Syndication* (sindicação na Web) pode ser entendida como a atividade de coleta e replicação de conteúdos dinâmicos em ambientes digitais. Teve início no final dos anos 90 como uma solução para minimizar o problema da sobrecarga da informação na Web. Atualmente vem se tornando popular devido à 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” ou “*feeds*”, 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 utilizada<sup>3</sup>. 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 atualizados com certa periodicidade.

Para aqueles já familiarizados com o código HTML (*HyperText Markup Language*), a estrutura de um documento RSS (*feed*) parecerá familiar. Porém, diferentemente da linguagem 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 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 ajudar a 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.

Tem-se, então, a possibilidade de reutilização automatizada da informação, podendo compartilhá-la mais facilmente com os usuários e/ou 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. Para o autor, considerado “pai” da Web e um dos precursores da Web Semântica, o princípio deste novo paradigma está localizado na valoração dos conteúdos e na capacidade de novas aplicações em interpretar e processar informações, além de compartilhar resultados com outros programas.

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 especificação 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 interessante da especificação 1.0 é a possibilidade de extensão com outros vocabulários RDF/XML, podendo incorporar outros modelos de metadados consagrados, como o *Dublin Core*, desenvolvido por bibliotecários e cientistas da informação para descrição de documentos. Este módulo 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. Esse conjunto de metadados fornece informação tanto sobre os *feeds* (no metadado *channel*) como sobre os itens individualmente, conforme exemplificado na Figura 1.

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.

```

<?xml version="1.0"? encoding ="UTF-8">
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns="http://purl.org/rss/1.0"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:dcterms="http://purl.org/dc/terms">

  <channel rdf:about="http://www.exemplo.com.br/news.rss">
    <title>Exemplo de um canal RSS</title>
    <link>http://www.exemplo.com.br/</link>
    <description>Essa é uma demonstração de um canal RSS </description>
    <dc:publisher>CID/UnB</dc:publisher>
    <dc:creator>Robson Lopes de Almeida</dc:creator>
    <dc:rights> CID/UnB. Todos os direitos reservados</dc:rights>
    <dc:date>2007-10-01T08:32:44Z</dc:date>

    <!-- Cada item representa um artigo diferente -->
    <items>
      <rdf:Seq>
        <rdf:li resource="http://www.exemplo.com.br/2007/10/01/">
        <rdf:li resource=" http://www.exemplo.com.br/2007/10/02/">
      </rdf:Seq>
    </items>
  </channel>

  <item rdf:about="http://www.exemplo.com.br/2007/10/01/">
    <title>Título do artigo 1</title>
    <link>http://www.exemplo.com.br/2007/10/01/</link>
    <description>Resumo do artigo 1</description>
    <dc:publisher>CID/UnB</dc:publisher>
    <dc:creator>
      <rdf:Seq>
        <rdf:li>Nome do autor 1</rdf:li>
        <rdf:li>Nome do autor 2</rdf:li>
      </rdf:Seq>
    </dc:creator>
    <dc:date>2007-11-01</dc:date>
  </item>

  <item rdf:about="http://www.exemplo.com.br/2007/10/02/">
    <title>Título do artigo 2</title>
    <link>http://www.exemplo.com.br/2007/10/02/</link>
    <description>Resumo do artigo 2</description>
    <dc:date>2007-11-02</dc:date>
  </item>

</rdf:RDF>

```

**Figura 1 – Exemplo de *feed* RSS (versão 1.0) utilizando módulo *Dublin Core***

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 dispo-

níveis 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 arquivados e apagados posteriormente, sem nenhuma complicação. Existe, ainda, a possibilidade de selecioná-los por assunto e/ou por data.

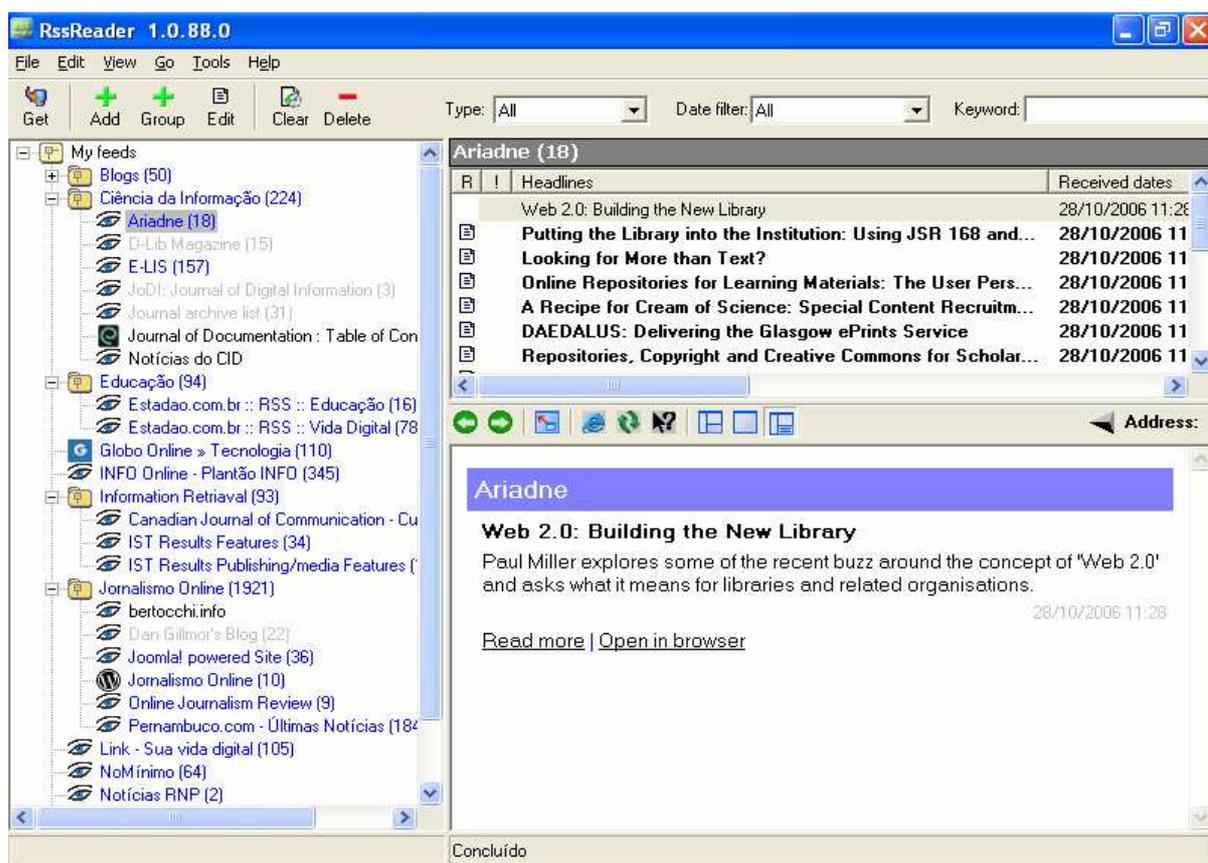


Figura 2 – Tela do RSS Reader, exemplo de aplicativo agregador de conteúdo

É importante salientar 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. Segundo especialistas, 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. RSS nos periódicos científicos eletrônicos

A comunicação científica, como parte integrante do desenvolvimento da ciência, se beneficia das vantagens oferecidas pelas tecnologias de informação, permitindo mais rapidez no desenvolvimento das pesquisas e colocando em contato pesquisadores localizados em diferentes partes do mundo (OLIVEIRA e NORONHA, 2005, p.1).

Além disso, as ferramentas utilizadas para o cumprimento de sua missão – troca de informações entre membros de uma mesma comunidade científica – acompanharam essa evolução e rapidamente adaptaram-se aos meios eletrônicos, produzindo novos formatos e funcionalidades, do suporte ao processo de publicação e comunicação científica.

No que diz respeito aos periódicos científicos, há um amplo reconhecimento da importância deste veículo como um dos principais canais da comunicação científica. Consequentemente, os periódicos também foram atingidos pelo impacto das tecnologias de informação e comunicação. Independentemente de suas vantagens ou desvantagens, o fato é que a crescente utilização da Internet pelos pesquisadores tem contribuído de maneira significativa para o desenvolvimento do processo de comunicação científica eletrônica promovida pelas diversas áreas do conhecimento.

Além dos modelos híbridos, em que as publicações eletrônicas co-existem com o seu equivalente impresso, atualmente existem várias revistas científicas puramente eletrônicas com reconhecida qualidade e prestígio, inclusive no âmbito nacional, tal como a DataGramZero (<http://www.dgz.org.br>), para ficarmos apenas na área da Ciência da Informação.

As publicações eletrônicas e seu envolvimento com a comunidade científica, contudo, não se configura como algo essencialmente novo. Em 1982, bem antes do aparecimento da Internet, um ambicioso projeto chamado BLEND (*Birmingham and Loughborough Electronic Network Development*) experimentou as possibilidades de comunicação por intermédio de revistas criadas, editadas e publicadas por computador. Assim, surgiu a revista *Computer Human Factors*, acessada via terminal remoto conectado a um computador central por meio de uma rede local. Outra iniciativa, intitulada “Quartet”, rendeu como fruto o sistema HyperBIT, um modelo para a publicação dos primeiros periódicos eletrônicos utilizando hipertextos.

Uma década mais tarde, em 1992, o projeto TULIP (*The University Licensing Program*), da editora científica Elsevier, lançou 42 publicações científicas em rede para cerca de 15 instituições acadêmicas que participaram do projeto, incluindo as universidades de Harvard, Carnegie, Cornell e Princeton. O objetivo do Tulip era examinar os aspectos econômicos, legais e técnicos que envolvem os periódicos eletrônicos. Outros projetos voltados para a pesquisa de publicações eletrônicas se seguiram sempre influenciados pelos desafios, oportunidades e limitações das tecnologias vigentes (SILVA et al, 1996, p. 455).

A incorporação do formato RSS nos periódicos eletrônicos teve início desde a sua criação. Afinal, foi originalmente concebido para ser utilizado no portal “*My Netscape Network*” como padrão para construção de sistemas de publicação de *headlines* (manchetes) em páginas da Internet, servindo com um resumo do conteúdo de uma publicação com os respectivos *links* para as fontes originais de informação. Atualmente o RSS é amplamente empregado em jornais e revistas, começando a dar os primeiros passos no contexto das publicações científicas periódicas em diversas áreas.

Uma aplicação imediata da tecnologia nas publicações científicas consiste no serviço de alerta ou notificação para os novos artigos que são disponibilizados por meio de sumários. Nesse caso, temos uma nova abordagem para o conceito de “disseminação seletiva”, uma vez que cabe ao usuário a ação de selecionar os *web feeds* (RSS ou Atom) definidos pelos editores científicos e, em alguns casos, os termos que definem esses canais.

A escolha da tecnologia de sindicância ou, ainda, da especificação RSS dependerá do tipo de aplicação desejado pelo editor. No caso da utilização em publicações científicas, em

que os *feeds* deverão conter informações suficientes para que estes possam citar ou produzir uma citação para um determinado artigo, o ideal é adicionar um conjunto de metadados além dos principais elementos obrigatórios em um documento RSS na descrição de recursos.

Por esse motivo, especialistas sugerem a adoção da especificação RSS 1.0 por estar em conformidade com a especificação RDF, considerada ideal para a inclusão suplementar de metainformação e, por esse motivo, bastante útil para a integração em sistemas de Web Semântica. (NOTTINGAM, 2005). Desse modo, além do *Dublin Core*, capaz de introduzir novos elementos de metadados dentro de um *feed* RSS, existe ainda o vocabulário PRISM (*Publisher Requirements for Industry Standard Metadata*)<sup>4</sup>, que permite a extensão dos elementos do *Dublin Core* no que se refere à informação bibliográfica dos artigos, tais como: *issn*, *volume*, *number*, *startingPage*, etc. (HAMMOND et al, 2004).

O fato dos editores científicos disponibilizarem *feeds* RSS e, conseqüentemente, procederem à sindicância dos seus metadados traz vantagens significativas tanto para os autores dos artigos científicos como para quem os publica, no que se refere à divulgação e disseminação dos conteúdos científicos. O RSS, na sua essência, permite abranger uma área significativamente extensa da Web, permitindo uma maior visibilidade dos trabalhos científicos desenvolvidos pelos pesquisadores.

## 5. Metodologia

O presente estudo possui características de uma pesquisa exploratória por se tratar de uma aproximação inicial para a investigação de um fenômeno, no caso o potencial de integração da tecnologia RSS em sistemas de notificação e agregação de conteúdos de periódicos eletrônicos.

Para isso, procuramos desenvolver um protótipo de um mini-repositório temático na área de Ciência da Informação, construído a partir da agregação de *feeds* RSS de uma amostra de 10 periódicos nacionais, todos eles baseados no 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. Em março de 2007, mais de 900 periódicos utilizavam a plataforma OJS<sup>5</sup>, incluindo os principais títulos nacionais. Outro ponto forte que contribui para a escolha dos periódicos suportados pelo OJS é que o sistema já dispõe de um módulo RSS desde sua versão 1.X capaz de gerar *feeds* automaticamente a partir da publicação dos artigos.

Os periódicos selecionados foram: *Arquivística.net*, *Ciência da Informação*, *Pesquisa Brasileira em Ciência da Informação e Biblioteconomia* (Instituto Brasileiro de Informação em Ciência e Tecnologia), *Em Questão* (Universidade Federal do Rio Grande do Sul), *Informação e Sociedade: Estudos* (Universidade Federal de Pernambuco), *Informação e Informação* (Universidade Estadual de Londrina), *Perspectivas em Ciência da Informação* (Universidade Federal de Minas Gerais), *Revista ACB* (Associação Catarinense de Bibliotecários), *Revista Digital de Biblioteconomia e Ciência da Informação* (Universidade Estadual de Campinas) e *Transinformação* (Pontifícia Universidade Católica de Campinas).

Esses periódicos foram objetos de análise durante os meses de janeiro e fevereiro de 2007, período em que procuramos investigar as principais características dessas publicações quanto aos formatos de divulgação de seu conteúdo, como serviços de alerta, mas, principalmente, se estes já disponibilizavam *feeds* para os seus leitores por meio da ativação do módulo RSS.

Uma vez que os periódicos escolhidos não dispunham de *feeds*, os mesmos foram criados manualmente com o auxílio de uma ferramenta de autoria chamada *FeedForAll*<sup>6</sup> (Figura

3). Os arquivos RSS tiveram como base a última edição de cada publicação, atualizados até o mês de maio de 2007. Posteriormente, os arquivos foram transferidos para um serviço de hospedagem com o propósito de formar uma base para que os *feeds* pudessem ser facilmente distribuídos ou sindicados.

Para a confecção do protótipo do mini-repositório temático – que funciona como um agregador de conteúdo – contamos com o auxílio do serviço Netvibes<sup>7</sup>, 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 jargão técnico como *Marshup*, termo é definido pela Wikipédia como “uma aplicação que combina conteúdo de mais de uma fonte em uma experiência integrada”.

O objetivo da criação desta aplicação foi demonstrar a simplicidade da utilização de sistemas agregadores de conteúdo, partindo do processo da criação de *feeds* RSS para os periódicos científicos em Ciência da Informação.

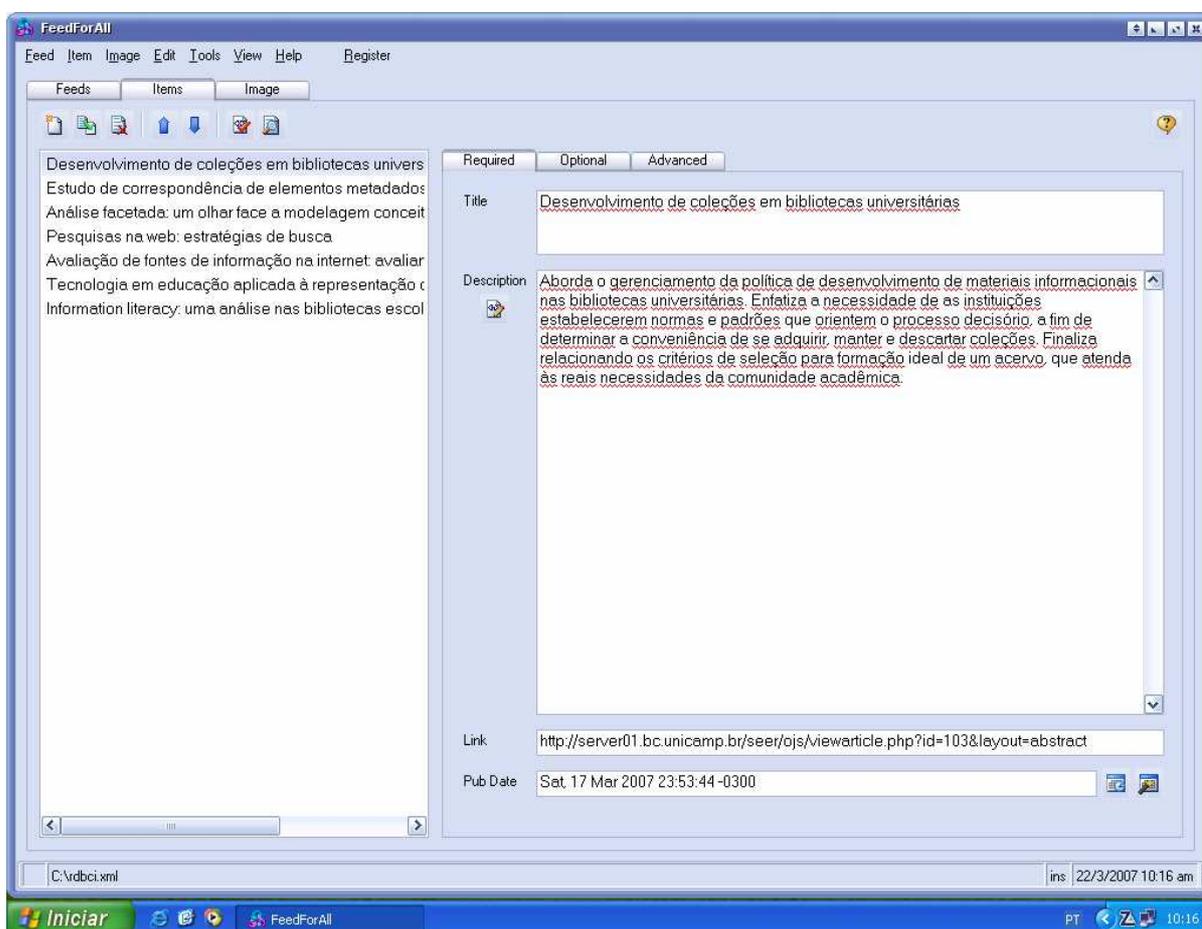


Figura 3 – Tela do FeedForAll, ferramenta de autoria para criação de *feeds* RSS

## 6. Resultados

O mini-repositório digital de periódicos em Ciência da Informação reúne referências das 10 revistas selecionadas. Cada módulo criado no âmbito do serviço Netvibes representa o conteúdo equivalente de um *feed* RSS correspondente a um periódico, como demonstrado na 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 aplicação identificará automaticamente as atualizações nos conteúdos dos sumários dos periódicos. Se este usuário desejar acessar o texto completo de um determinado artigo, ele poderá remeter-se diretamente à página correspondente à fonte de informação com um simples “clique” sob título do artigo.

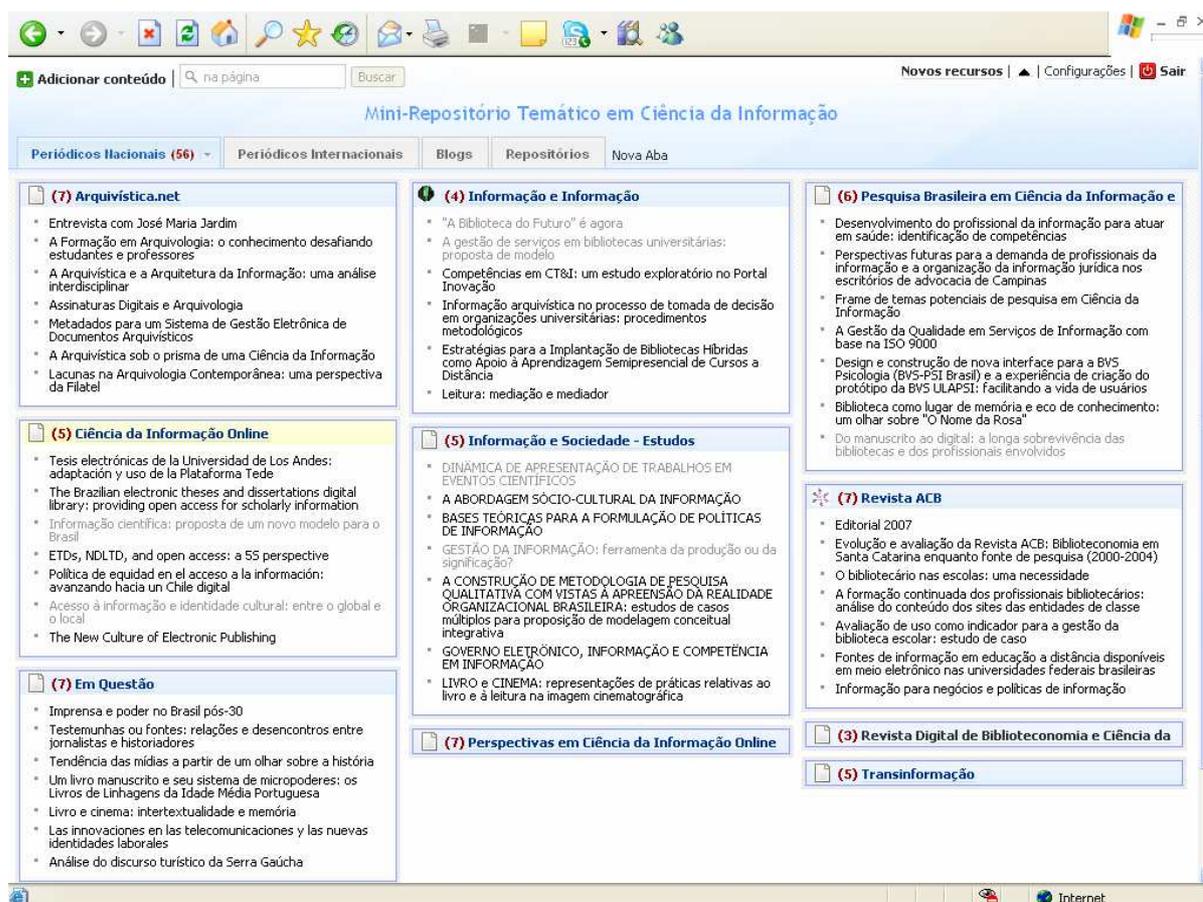


Figura 4 – Tela do protótipo do “agregador” de periódicos nacionais em Ciência da Informação

O modelo apresentado demonstra uma das principais aplicações referentes à utilização de *web feeds* atualmente. No caso, apresentamos um mecanismo capaz de reunir, em única página Web, a indicação dos artigos (com respectivo resumo) dos periódicos nacionais em Ciência da Informação. Se desejássemos, poderíamos incluir outros módulos cadastrando os *feeds* de periódicos estrangeiros, blogs e repositórios. 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 toda a coleção. Assim, se o usuário desejar fazer uma pesquisa pelo termo “ontologia” em todos os periódicos, por exemplo, o resultado da busca trará todos os artigos que contenham o termo “ontologia” em seu título ou no resumo, independente do periódico.

## 7. Discussão e considerações sobre a proposta

Atualmente, a maioria dos periódicos científicos que utilizam a ferramenta de publicação da plataforma 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 alguns entraves que podem impedir a ampla difusão deste tipo de recurso. O principal empecilho, na nossa opinião, é o próprio desconhecimento da tecnologia de syndicação, 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.

Apesar das barreiras identificadas, 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 de maneira rápida e 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 a um 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.

## 8. Conclusões

Após nossa investigação sobre o potencial da utilização do RSS como recurso capaz de incrementar as funcionalidades de periódicos eletrônicos, 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 periódicos eletrônicos, bibliotecas digitais, etc.;
- Verifica-se o RSS como uma ferramenta de sucesso comprovado para a notificação (alerta) automatizado de conteúdos em ambientes Web;
- 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 na adoção da tecnologia de *Web Syndication* em periódicos científicos eletrônicos são imensas, a começar pelos benefícios para os usuários/leitores, que passarão a dispor de uma poderosa ferramenta por meio da qual serão capazes de realizar buscas simultâneas em cima de várias fontes de informação, aumentando a relevância dos termos recuperados.

Já os editores, por sua vez, agregarão valor ao conteúdo de sua publicação e aumentarão sua visibilidade na medida em que a inclusão de um *feed* RSS possibilitará aos usuários manterem-se informados pelo próprio *software* agregador, sem precisar visitar a página de cada periódico. Isso pode parecer ruim à primeira vista, mas na verdade incrementa a visibilidade da publicação, 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 retornam ao mesmo sítio todos os dias. Ao prover um *feed* para a sua publicação, os editores estarão permanentemente em contato com seus usuários, aumentando as chances de que um determinado artigo desperte a sua atenção.

## 9. Referências

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

DOWNES, Stephen. Semantic networks and social networks. **The Learning Organization**; Vol. 12 Issue: 5. 2005. Disponível em: <<http://www.downes.ca/cgi-bin/page.cgi?post=31624>>. Acesso em: 03. jan. 2007.

HAMMOND, T. et al. The Role of RSS in Science Publishing Syndication and Annotation on the Web. **D-Lib Magazine** [on-line], 10 (12). 2004. Disponível em: <<http://www.dlib.org/dlib/december04/hammond/12hammond.html>>. Acesso em: 13. fev. 2007

HOUSMAN, Edward M.. Selective dissemination of information. **Annual Review of Information Science and Technology**. V. 8, Washington, American Society of Information Science, p. 221-241, 1973.

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.

OLIVEIRA, E. B. P. M. ; NORONHA, D. P. A Comunicação científica e o meio digital. **Informação & sociedade: estudos**, João Pessoa, v. 15, n. 1, p. 1-12, 2005.

SILVA, Luiz Antonio Gonçalves da et al. Buscando soluções para se publicar na Internet: a experiência do IBICT com a Ciência da Informação on-line. **Ciência da Informação**, Brasília, v.25, n. 3, p. 454-460, set./dez.1996.

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**. 2005. Disponível em: <<http://www.mnot.net/rss/tutorial/>>. Acesso em: 02 jun. 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.

---

<sup>1</sup> Um *feed* RSS é um documento baseado em XML, normalmente com extensão .xml, .rdf ou .rss, que contém um conjunto de metadados utilizados para representar um tipo de informação estruturada, como os artigos de periódicos, por exemplo.

<sup>2</sup> Crayon (<http://crayon.net>) é o acrônimo de *Create Ypur Own Newspaper*, um serviço de informação personalizada que oferece ao leitor a possibilidade de criar seu próprio jornal com *links* para centenas de fontes de notícias disponíveis na Internet.

<sup>3</sup> Para uma discussão mais aprofundada sobre o surgimento da tecnologia RSS, leia: <<http://en.wikipedia.org/wiki/Rss>>

<sup>4</sup> <<http://www.prismstandard.org>>

<sup>5</sup> Uma lista dos periódicos suportados pelo sistema OJS encontra-se disponível na página: <<http://pkp.sfu.ca/ojs-journals>>

<sup>6</sup> <<http://www.feedforall.com>>

<sup>7</sup> <<http://www.netvibes.com>>