Universidade do Minho
Instituto de Ciências Sociais

Centro de Estudos de Comunicação e Sociedade
www.cecs.uminho.pt

O código oculto das notícias electrónicas •

Alberto Sá
Assistente

albertosa@ics.uminho.pt

Universidade do Minho
Centro de Estudos de Comunicação e Sociedade
Campus de Gualtar
4710-057 Braga
Portugal

• SÁ, A. (2005) "O código oculto das notícias electrónicas", *Actas do 4º SOPCOM - Congresso da Associação Portuguesa de Ciências da Comunicação*, Aveiro: Universidade de Aveiro, 20 e 21 de Outubro.

O código oculto das notícias electrónicas

Alberto Sá
Depº Ciências da Comunicação, Universidade do Minho

Resumo

Esta comunicação pretende analisar o fenómeno RSS na perspectiva do que está oculto. A *Web* está a modificar-se, e muita dessa evolução ocorre despercebida aos olhos do utilizador. O recurso aos metadados, apesar de não assentar em tecnologia inovadora (as suas raízes remontam aos anos 60), permite disponibilizar um filão informativo na penumbra do código, ao alcance do computador para extracção automática dos dados, sem necessidade de intervenção humana. Será sempre mais fácil relacionar conhecimento, evitando a info-poluição e aproximando-se da ideia original da *Semantic Web* de Tim Berners-Lee para a emergência de um meio universal de partilha de informação, atribuindo significado ao conteúdo dos documentos na *Web*, de um modo perceptível pela máquina, mas fazendo deste um meio colaborativo entre seres humanos.

Introdução

Historicamente, a noção de que os documentos em formato digital deveriam obedecer à separação entre conteúdo e forma data de finais da década de 60[1]. A ideia subjacente era a da troca e manipulação de ficheiros em formato normalizado e aberto, para edição, formatação e extracção de informação[2].

Tal como hoje em dia, a crescente proliferação de documentos, formatos e suportes de informação, com necessidades de publicação também distintas, motivavam a busca por uma solução integradora que combinassem a partilha universal de dados com o registo sobre a própria apresentação e estrutura dos mesmos. Durante os anos 80, das primeiras tentativas em normalizar a metodologia de estruturação e representação de documentos resultaria a criação da metalinguagem SGML (*Standardized Generalizad Markup Language*) [3], lançada publicamente em 1986 já sob a forma de norma ISO 8879, e desde então aplicada em todas as linguagens de anotação[4].

[1] Foram pioneiros os trabalhos de William Tunnicliffe, em 1967, na elaboração da linguagem *GenCode* da associação norte-americana *Graphic Communications Associations* (GCA) – actualmente "IDEAlliance" [www.idealliance.org.], e de Stanley Rice na elaboração de parâmetros de catalogação.
[2] Em 1969, Charles Goldfarb desenvolvia um projecto de investigação na IBM que conduziria à invenção da linguagem *Generalized Markup Language* (GML), que propunha resolver os problemas associados ao tratamento de documentos de diferentes plataformas.
[3] Cf. SGML Users' Group, 1990.
[4] Uma linguagem de anotação é um sistema complexo de descrição de informação, normalmente documentos, representados por texto descritivo fechado entre símbolos "<" e ">".

Já no final dessa década, o investigador do CERN, Tim Berners-Lee, adoptaria um conjunto de anotações extraído do SGML para anotar e formatar visualmente as páginas na sua aplicação de hipertexto. Da maturação do projecto nasceria o primeiro *browser* de divulgação pública da *WWW*, o *Mosaic*, que integrava a linguagem HTML (*HiperText Markup Language*), decisiva para a definição do panorama actual da *Internet*.

No entanto, o HTML limita-se a fornecer "apenas" um formato universal para representar a informação, sem acrescentar qualquer referência quanto à estrutura e significado dos dados contidos, pelo que se tornou comum aludir a *Internet* como uma incomensurável amálgama de dados não estruturados. O contributo para a ordenação do caos parece advir do XML (*eXtensible Markup Language*), definida como uma metalinguagem (dados sobre os dados) de anotação de documentos para a *Web*, derivado do SGML, mas sem a complexidade da linguagem original, e que descreve (codifica) a estrutura e conteúdo da informação legível por máquinas, num formato universal.

Neste sentido, um documento possui dois tipos de informação, os dados em si e as anotações sobre os mesmos (Ramalho & Henriques, 2002: pp. 25 e ss.).

O XML foi concebido para que os documentos massivamente estruturados pudessem ser usados na *Web*. Esta linguagem fornece uma maneira de descrever o conteúdo e a estrutura lógica da informação (recorrendo à anotação) mas sem alguma preocupação quanto ao formato ou aparência visual de como esta se deve apresentar, por exemplo, num *browser*, separando, deste modo, o conteúdo da forma[5].

É, assim, o herdeiro natural dos ideais lançados na década de 60. Não sendo, portanto, propriamente inovador no actual panorama informático, o XML tem vindo progressivamente a ser usado para estruturar a informação, tanto no domínio lato da *Web* como no âmbito empresarial. A sua vantagem advém da capacidade em transferir qualquer tipo de dados, mantendo-os estruturalmente coesos e inteligíveis, por meio da sua estrutura declaradamente *markup* (<tag>value</tag>), ou seja, por uma linguagem de anotação.

[5] Para formatar graficamente um documento XML ou uma página HTML, foi criado um conjunto de especificações de estilo que darão aspecto visual ao conteúdo do documento: a norma ISO/IEC 1079:1996, conhecida por DSSL (*Document Style Semantics and Specification Language*), as regras de associação de folha de estilo CSS (*Cascading Style Sheet*), e uma evolução destas normas existentes, o XSL (*eXtensible Stylesheet Language*) (Ramalho & Henriques, 2002, pp. 169 e ss).

O XML não tem propriedade comercial, antes foi desenvolvido sob a égide do W3C (*World Wide Web Consortium*) de acordo com as experiências anteriores, como o SGML ou o HTML. É um tipo de linguagem neutra e independente com respeito às plataformas de *software* e de *hardware*, o que, de certa forma, explica a invasora transversalidade desta norma em todos os domínios de aplicação informáticos[6].

Ao isolar o conteúdo da forma, a mesma informação pode servir modelos de utilização diferentes em obediência a variadas plataformas de transmissão, sem, no entanto, perder a essência dos dados. Na prática, o aproveitamento recorrente da informação permite adequação às diferentes aplicações, em respeito pelas necessidades dos variados públicos, e através de suportes de difusão distintos, seja num *website*, num *PDA* ou num telemóvel.

Ao contrário do HTML, o XML classifica por meio de <tags> a estrutura dos dados, permitindo às aplicações informáticas maior eficiência em tarefas de procura, processamento e transferência de informação, requisitos por excelência da *Era da Informação*.

A aptidão para encontrar nos dados exactamente a informação necessitada através da leitura dos <tags> que a definem abriu excelentes possibilidades para a partilha dessa informação com outros programas. Neste contexto, surgiu o RSS.

O que é o RSS?

O RSS é um derivado da metalinguagem XML, específico para o fornecimento de conteúdos na *Web*, daí estar assinalado com um identificador colorido que tanto alude a uma como a outra sigla, sobre um rectângulo cor-de-laranja: XML ou RSS.

São várias as interpretações quanto ao significado deste acrónimo. Seja “Rich Site Summary”, “RDF Site Summary” ou “Really Simple Syndication”, as variações revelam a falta de consenso pela tutela de mais um fenómeno de crescimento rápido na *Web*, tendo

[6] Com efeito, a anotação de documentos tem sido recorrente nas mais variadas comunidades *Web*: na publicação electrónica (*DocBook*, na edição de livros técnicos e o *TEI*, utilizado em documentos nas áreas das Ciências Humanas), na representação musical em formato textual descritivo (*MusicML*), na representação do genoma em Biologia, na representação da informação em Sistemas de Informação Geográficos, na representação de estruturas moleculares na Química (*CML – Chemical Markup Language*), na Matemática com a edição e formatação em *MathML*, entre outras. Para uma panorâmica dos domínios de intervenção do XML, cf. Ramalho & Simões, 2004, e, igualmente, Ramalho, Simões & Lopes, 2005.

levado já ao despertar da cobiça por parte dos gigantes informáticos Microsoft e Google[7].

O RSS, embora não sendo inovador nem tão pouco recém-nascido (é um derivado do XML, nascido no âmbito do *Netscape* em 1999), apenas recentemente ganhou popularidade pela simples razão de permitir, de modo gratuito e simples, a recepção de conteúdos (como as notícias *online* ou os últimos comentários num blogue) sem necessidade de aceder aos sites originários. A maioria dos conteúdos em RSS são relativos a cabeçalhos de notícias, a informações de última hora ou ao mais recente comentário num blogue, o que faz desta uma aplicação recorrente no universo dos sites noticiosos *online* e da blogosfera. Ter um leitor de RSS é como compor um jornal apenas com matérias de interesse ao utilizador, apresentado instantaneamente no ambiente de trabalho do computador, de modo automático e centralizado, a degustar quando mais apetecer. A sensação de desconforto provocada pela avalanche informativa que emerge do turbilhão da Internet pode ter, no RSS, uma panaceia para abreviar a procura de conteúdos e proporcionar mais tempo para os apreciar.

No entanto, a sua aplicabilidade não se esgota nestes domínios pois tudo aquilo que seja passível de descrever por meio de anotações (<tags>) pode ser integrado via RSS. A crescente popularidade do RSS é o reflexo do progressivo alargamento nos âmbitos de aplicabilidade desta tecnologia, assegurando indiscutíveis vantagens em ambientes como os dos tópicos de discussão em fóruns, os da publicitação de eventos com notificação de alterações de programa e de calendarização, os das ofertas promocionais de cariz comercial, os da publicação editorial, os da notificação sobre novas versões de *software*, os da previsão meteorológica, os da interacção gestora e financeira entre a sede empresarial e filiais, os das alterações recentes a uma página *Wiki*, entre outros tantos exemplos distintos. Não esquecer, todavia, uma das mais recentes facetas de aplicação, o *Podcasting*, que consiste na difusão de ficheiros áudio e mesmo vídeo por RSS[8].

Ao acto de tornar os conteúdos disponíveis em simultâneo para uma plêiade de inte-

[7] A Microsoft decidiu incorporar as extensões RSS 2.0 no sistema operativo *Vista*, em 2006 (a Apple já o incorpora no *browser Safari*). O Google anunciou o lançamento de uma nova ferramenta, "Google Reader", que fornece o suporte para acolher *feeds* RSS.

[8] O *Podcasting* é tido como um modelo comercial de elevado interesse para os agentes musicais, sendo utilizado no domínio dos *talk-shows*, em tutoriais, em debates, etc. Aos *Podcasts* estão geralmente associados metadados que descrevem o ficheiro de som ou vídeo, o que permite ao utilizador, contrariamente à rádio e à TV, receber apenas o género musical subscrito, desfrutando quando mais desejar. A transferência desses ficheiros para um leitor portátil *iPod* pode ser automaticamente programada.

ressados denomina-se de *syndication*. Quando aplicado ao contexto da *Web*, significa tornar disponível ao público algumas secções de um *website*, sem necessidade prévia de licenciamento[9]. Na prática, representa uma fonte de informação que está acessível para consulta, partilha e mesmo republicação em outros *sites*. Este é o cerne de todo o processo, perceptível pela leitura da definição mais consensual do acrónimo RSS: *Really Simple Syndication*.

Por *really simple* podemos entender a simplicidade de processos, uma vez que a intervenção humana é reduzida, dado tratar-se de uma comunicação entre máquinas. Ao utilizador basta instalar uma aplicação (*software*) que funciona como congregador dos conteúdos de múltiplas e diferenciadas fontes num único espaço visual – o *news aggregator*. Mas também pode ser entendido como uma *simple syndication* pois a apropriação dos conteúdos é simplesmente gratuita e sem licenciamento, autorizando a partilha e a republicação dos conteúdos, sem direitos proprietários.

Ao conteúdo difundido aplica-se o termo *feed*. Um *feed* é um documento conforme a especificação XML que representa os metadados de um canal e do seu conteúdo (item), que na versão RSS 2.0 deve ter no código a seguinte estrutura[10]:

<?xml version="1.0" encoding="ISO-8859-1"?>	Declaração que define a versão e codificação do documento, no caso, o conjunto de caracteres especiais latinos.
<rss version="2.0">	Declaração da versão RSS.
<channel> **<title>**Departamento de Ciências da Comunicação**</title>**	O título com o nome do canal, isto é, a denominação pelo qual é referido e que corresponde geralmente ao nome do site.
<description>Página do Departamento de Ciências da Comunicação da Univ. do Minho, disponibiliza informações sobre…….**</description>**	Descrição, um curto parágrafo contendo o propósito do canal.
<link> http://www.comunicacao.uminho.pt**</link>**	Endereço URL do site acima designado.
<docs>http://blogs.law.harvard.edu/tech/rss**</docs>**	Foi atribuída à *Escola de Direito da Universidade de Harvard* a missão da especificação da norma RSS 2.0, numa tentativa de obter consenso.
<language>pt**</language>**	A linguagem de redacção do canal
<lastBuildDate>Thu, 13 Oct 2005 12:20:38 +0100**</lastBuildDate>**	A data de criação electrónica do canal.
<managingEditor> albertosa@ics.uminho.pt**</managingEditor>**	Identificação do gestor do canal.
<pubDate>Thu, 13 Oct 2005 10:49:07 +0100**</pubDate>**	A data de publicação electrónica do canal.

[9] O termo é igualmente usado em outras áreas, como a televisão, a rádio e a imprensa.
[10] Cf. especificação oficial da norma RSS 2.0.1 (Outono de 2002) em Winer, 2005. Cf, igualmente, Hammersley, 2005a.

<image> **<url>**http://www.comunicacao.uminho.pt/dcc_icoxml.gif**</url>** **<width>**144**</width>** **<height>**60**</height></image>**	Especifica o logotipo identificativo do canal, consistindo na localização URL de uma imagem de formato GIF, JPEG ou PNG, acrescido das dimensões da mesma em *pixels*.

No seguimento, enumeram-se os itens que compõem o canal, alusivos aos conteúdos. Um canal pode conter inúmeros itens, cada qual com a seguinte especificação:

<item>	
<title> Lançamento da Revista Comunicação e Sociedade 7**</title>**	O título, podendo ser o cabeçalho de uma notíca;
<description>O presente número da revista Comunicação Sociedade é dedicado à Economia Política da Comunicação**</description>**	A sinopse do item.
<link> http://www.cecs.uminho.pt/publ/revistas/com_soc_7.htm**</link>**	O endereço URL apontando à versão completa da referência.
<author>Alberto Sa**</author>**	
<pubDate>Fri, 16 Sep 2005 09:43:12 +0100**</pubDate>**	Indica a data de publicação do respectivo item
</item> **</channel></rss>**	Fecha as declarações entretanto abertas.

Como referido, basta ao utilizador a instalação no computador de um pequeno programa responsável pela recolha dos *feeds*, e que recebe várias denominações: *RSS Readers*, *feed readers, feed aggregators* ou *news readers*[11]. A proliferação de nomenclaturas é reflexo da indefinição que ainda caracteriza o sistema, atendendo à diversidade de incidências do RSS. A *syndication* não contempla a mera notificação de notícias, tem uma abrangência mais vasta como processo de transmissão de conteúdos (*podcasting*, como exemplo já citado), pelo que tem sido recorrente a adopção de um termo mais envolvente e generalista como o de *aggregator*.

Controvérsia

O XML é demasiado genérico para ser empregue sem a definição da sintaxe exacta sobre a informação, pelo que o RSS corresponde a um esforço de normalização (ainda que não consensual) quanto à forma como os dados devem ser descritos. Pelo facto do XML não ter propriedade comercial, a estandardização é mais difícil de alcançar. O caso do RSS

[11] Existem no mercado várias soluções: as que se instalam directamente no computador do cliente (quer sob a forma de aplicação independente, quer por meio de um *plug-in* que se integra no editor de correio electrónico ou no *browser*) e a alternativa *web-based* que consiste no acesso a um site fornecedor deste serviço.

não é excepção, havendo distintas versões que reflectem as sucessivas modificações e rebaptismos verificadas ao longo dos últimos anos, a partir da base original enunciada pelo grupo *Netscape* (RSS 0.90), em Março de 1999: o *Rich Site Summary* corresponde à versão RSS 0.91, o *RDF Site Summary* corresponde às versões RSS 0.9 e 1.0, o *Really Simple Syndication* corresponde à versão RSS 2.0 (Pilgrim, 2002). Pelo facto de serem suportadas por grupos comerciais distintos, não é garantida a plena compatibilidade entre os formatos.

Paralelamente, surgiu a versão similar *Atom Syndication Format*, que acolheu a preferência do serviço blogue do Google, o que indicia esta contenda como um negócio de gigantes pela dominância de formatos *syndication*. Fundamentalmente, difere do RSS pelas nuances semânticas e no rigor na preservação dos metadados porque é inflexível quanto à constância dos atributos <title>, <last-updated> e <id> URI único, tanto no canal como no item (Hammersley, 2005b), o que em muito beneficia o arquivo automático de ficheiros.

Em 1998, havia já emergido o *NewsML* como indústria de estruturação de notícias multimedia, delineado em modelo aberto pela *Reuters* e doado à IPTC (*International Press Telecommunications Council*) para normalização. Esta tecnologia comercial integra os metadados na estruturação das notícias o que permite serem seleccionadas segundo critérios adequados ao meio transmissor, à plataforma de suporte e ao tipo de conteúdo (Levine, s.d.).

Em comum, todas estas iniciativas partilham, ora mais próximo, ora mais afastado, a base em XML, salientando-se que esta miscelânea não parece travar o êxito do RSS.

Benefícios

Independentemente da forma como são apelidados, é inegável que os *aggregators* vieram reduzir os factores tempo e esforço necessários para pôr em dia a leitura das últimas novidades, de entre os *sites* considerados referenciais para o utilizador, confinando ao ambiente de trabalho todo um espaço informativo ajustado ao interesse pessoal, e, nunca é demais frisar, gratuito. Esta aplicação é capaz de fixar um endereço URL que aponta para um ficheiro de tipo ".xml", de monitorizar a informação nele contido em intervalos de tempo pré-definidos e de recolher os conteúdos correspondentes no momento em que o fornecedor os actualizar. Em acréscimo, existe a possibilidade de se efectuar rotinas de pesquisa

sobre os *feeds* recebidos.

Trata-se de uma tecnologia "de puxar" em vez de "empurrar", uma vez que os conteúdos não são forçados aos consumidores. O RSS pressupõe a intencionalidade do utilizador porque em dado momento ocorreu uma tomada de iniciativa para "puxar"a informação a partir de um certo *website*, podendo, a qualquer momento, esse canal ser eliminado e a ligação findar. Este é um elemento importante, pois permite ao utilizador vigiar a sua "dieta" informativa, sem intermediários, e sem vínculos.

Os benefícios deste sistema fazem dele um dos maiores sucessos recentes da *Internet*.

Para o produtor de conteúdos, seja um jornal *online* ou uma empresa de comércio electrónico, o RSS significa uma maneira rápida e eficaz de levar a um público-alvo as últimas novidades, num *continuum* comunicativo automático e directamente orientado – os clientes escolheram intencionalmente receber *feeds* desse produtor, havendo implicitamente fidelização de público, o que promove relações de confiança e fortalece a identidade da marca. Deste pressuposto resulta uma maneira prática de promover um *site* sem gastos de publicidade ou de integrar dispendiosos canais de distribuição comercial, evitando a barreira anti-*spam*. Tem, no entanto, o óbice de desviar os olhares das mensagens publicitárias inseridas no *website*, restando apurar a reacção dos anunciantes face à quebra de audiência.

Para o utilizador, o RSS permite receber no computador a informação de particular interesse de modo rápido e contínuo, com mínimo de esforço, dado ser desnecessário aceder à página principal do site. Porque a informação se encontra congregada na aplicação RSS, os utilizadores acedem a mais informação em menor período de tempo. Havendo o desejo pela versão integral da notícia, basta percorrer o *link* para o efeito incorporado.

O acesso directo e fidedigno à fonte da informação evita o "ruído" intermédio, permitindo evitar alguns dos males que afectam as *newsletters* via correio electrónico, como sejam, a recepção de informação não solicitada (o *spam*, com todos os constrangimentos advindos em relação ao espaço ocupado na caixa de mensagem com inutilidades, e a recepção de mensagens de teor indesejado e até agressivo), e a captação abusiva do endereço de correio electrónico pessoal para integrar bases de dados não-autorizadas. Dada ser uma tecnologia baseada em processos a pedido, como referido, os conteúdos não aparecem inadvertidamente na área do utilizador, evitando o incómodo da persistência, podendo os *feeds*

indesejados serem instantânea e definitivamente removidos. Estes são, no conjunto, fortes argumentos que prefiguram o RSS como alternativa viável ao correio electrónico enquanto canal privilegiado de distribuição de recursos.

Sabendo que a actualização dos sites varia consoante a sua natureza – se no caso dos agências noticiosas electrónicas pode ser minuto a minuto, nos blogues ou em sites de referência é indefinida –, o RSS liberta o utilizador da obstinência em percorrer os sites favoritos em busca de actualizações. Deste modo, a cadência das interacções passa a ser ritmada pela própria fonte, em tempo real e automático, alheia à predisposição do utilizador.

Inconvenientes

A proliferação de formatos em torno da distribuição de conteúdos pode ser um mal menor desde que as aplicações informáticas sejam transversalmente compatíveis.

Os *aggregators* reproduzem o conceito de jornal recorrendo ao escalonamento da notícia em cabeçalho e um breve parágrafo descritivo. Atendendo ao modo redutor como a informação se apresenta ao leitor, sugere uma certa superficialidade na abordagem aos conteúdos, havendo o inconveniente de, caso o cabeçalho ou a descrição não estejam bem elaborados, a mensagem poder passar despercebida.

Se numa primeira fase existe pró-actividade, o utilizador visitou um site e escolheu receber os respectivos *feeds*, no seguinte momento pode verificar-se o revés da iniciativa, ao confirmar-se a queda em catadupa de *feeds* no leitor de RSS. Neste sentido, verifica-se uma aproximação ao conceito de espectador televisivo, cuja passividade pode não ir mais além da alternância entre canais, sem procurar a fundo o desenvolvimento da notícia.

Os atributos do RSS demarcam-no do e-mail como veículo distribuidor de conteúdos e de informação, como anteriormente assinalado. Porém, algumas questões ainda se encontram por resolver, nomeadamente quanto à forma como serão compensados os meios publicitários que financiam em larga escala a manutenção dos sites envolvidos. Nesse sentido, cabe ainda resolver o modo de integrar os *feeds* na estatística de visitas do site fornecedor. A integração de publicidade nos *feeds*, a verificar-se, traduz-se num revés ao modelo, mas é inegável que a publicidade constitui um importante modelo de financiamento da Internet, pelo que ainda é prematuro apontar caminhos que permitam sair da encruzilhada.

Conclusão: O XML e a investigação jornalística

A euforia pelo fenómeno RSS pode ser interpretado como um sintoma de viragem. A crescente adopção do XML em amplos domínios de aplicação faz desta a linguagem de próxima geração capaz de elevar a funcionalidade da *Internet*. O facto do XML ser pensado precisamente para o contexto da *Internet* (neutra dependência a *hardware* e *software*) e de se centrar apenas nos dados recorrendo à estruturação e descrição dos mesmos (separação entre informação e conteúdo), permite a correcta adequação às diferentes aplicações, independentemente do idioma ou sistema operativo. A tecnologia XML é extremamente eficaz na extracção e pesquisa de informação, recorrendo à boa capacidade de análise sintáctica. O facto de ser auto-descritiva e extensível permite ao utilizador a liberdade para anotar os dados de forma ajustada ao contexto do assunto em causa. Esta acção, denominada de *tagging*, valida os dados para os motores de pesquisa possibilitando, pela análise dos dados, reconhecer relações intrínsecas[12].

Quem já procurou extrair informação da Internet para constituir um arquivo por certo enfrentou dificuldades técnicas na catalogação dos dados – os ficheiros HTML apenas descrevem o aspecto de um documento num navegador, mas sem possibilidade de relacionar os dados inclusos. A alternativa recorrente é a catalogação manual numa base de dados convencional, que apesar de não ser um sistema menos válido, é, não obstante, mais moroso e mais permissivo ao erro.

As recentes transformações na *Web* têm escapado aos olhares de esmagadora maioria dos usuários. A proliferação de documentos XML e a crescente adopção tecnologia RSS em domínios mais alargados no universo da *Internet* abrem boas perspectivas para a elaboração imediata e automática de bases de dados, sem intervenção humana, na penumbra do código (Hogan, s.d.). Um *feed* de uma agência noticiosa, por exemplo, contém anotações sobre o título, autor, data, imagem, categoria, breve descrição e um *link* para o desenvolvimento da notícia, elementos bastantes para alimentar automaticamente uma base de dados que responda ao "quê", "quando", "onde" e "como". A anotação com informação semântica – *o código oculto* - permite que uma aplicação (seja uma base de dados ou um motor de

[12] Cf. algumas "killer applications" recentes como o *Flickr*, o *Feedster*, ou ainda, o *Technorati* ("Tags: The real-time web, organized by you").

pesquisa de meta-informação) organize os dados lendo a sua descrição e a sua estrutura a partir do próprio ficheiro original. Será sempre mais fácil relacionar conhecimento, evitando a "infopollution" (Orihuela, 2005), e aproximando-se da ideia original da *Semantic Web* de Tim Berners-Lee na emergência de um meio universal de partilha de informação atribuindo significado ao conteúdo dos documentos na Web, de um modo perceptível pela máquina, mas fazendo deste um meio colaborativo entre seres humanos (Berners-Lee, 1997).

Bibliografia

Berners-Lee, Tim (1997) 'Realising the full potential of the Web', *World Wide Web Consortium (W3C)*, [http://www.w3.org/1998/02/Potential.html] (Outubro 2005)

Hammersley, Ben (2005a), 'RSS 2.0', *Developing Feeds with RSS and Atoms*, Cap. IV, Sebastopol: O'Reilly Media, Inc., p. 25-50, [http://www.oreilly.com/catalog/deveoprssatom/chapter/ch04.pdf] (Outubro.2005)

Hammersley, Ben (2005b) 'What Is Atom', *XML.com, O'Reilly Media, Inc.*, [http://www.xml.com/pub/a/2005/10/26/what-is-atom.html] (Outubro.2005)

Hogan, Mike (s.d.) 'XML: The Foundation for the Future', *XML.org, OASIS*, [http://www.xml.org/xml/xml_foundation_future.shtml] (Outubro.2005)

Levine, Irving (s.d.) 'NewsML – Reuters Advances Multi-Media Publishing Standard', *REUTERS - NewsML Showcase*, [http://about.reuters.com/newsml/reutersandnewsml.asp] (Outubro.2005)

Orihuela, José Luís (2005) 'Dez chaves para entender o impacto dos weblogues e a loucura dos blogueiros', *eCuaderno.com, Universidad de Navarra*, [http://mccd.udc.es/orihuela/covilha/jlori_covilha/sld001.htm] (Outubro, 2005)

Pilgrim, Mark (2002) 'What Is RSS', *XML.com, O'Reilly Media, Inc.*, html, [http://www.xml.com/pub/a/2002/12/18/dive-into-xml.html] (Outubro.2005)

Ramalho, José Carlos & **Henriques**, Pedro Rangel (2002) *XML & XSL. Da teoria à prática*, Lisboa: FCA

Ramalho, José Carlos & **Simões**, Alberto (eds.) (2004) *XATA - XML, Aplicações e Tecnologias Associadas. Actas da 1ª Conferência Nacional*. Porto: Faculdade de Engenharia da Universidade do Porto

Ramalho, José Carlos, **Simões**, Alberto & **Lopes**, João Correia (eds.) (2005) *XATA2005 - XML: Aplicações e Tecnologias Associadas. Actas da 3ª Conferência Nacional*. Braga: Universidade do Minho

SGML Users' Group (1990) 'A Brief History of the Development of SGML', *Charles F. Goldfarb's SGML Source*, [http://www.sgmlsource.com/history/sgmlhist.htm] (Outubro 2005)

Winer, Dave (2005) 'RSS 2.0 Specification', *Berkman Center for Internet & Society, Harvard Law School*, html, [http://blogs.law.harvard.edu/tech/rss] (Outubro.2005)