Aplicação da Web Semântica no jornalismo

Iuri Lammel

Universidade Federal de Santa Maria - iuri.lammel@gmail.com
Graduado em Jornalismo pela UFSM. Mestre pelo Programa de Pós-Graduação
em Comunicação Midiática da UFSM. Professor do curso de Jornalismo
do Centro Universitário Franciscano (Santa Maria - RS).

Luciana Mielniczuk

Universidade Federal de Santa Maria - luciana.mielniczuk@gmail.com
Doutora em Comunicação e Cultura Contemporâneas pela Universidade Federal da Bahia.
Professora da Faculdade de Biblioteconomia e Comunicação da
Universidade Federal do Rio Grande do Sul.

Resumo

A web e as bases de dados são consideradas plataformas tecnológicas fundamentais para o desenvolvimento do jornalismo contemporâneo em redes digitais. No início da década de 2000, surge a ideia da Web Semântica, que é uma proposta de expansão da atual web, para torná-la mais automatizada e eficiente. A partir da observação do BBC Wildlife, um site que utiliza tecnologias da Web Semântica para gerenciar e publicar conteúdos editoriais sobre o mundo natural, são analisados aspectos que contribuem na potencialização de características do jornalismo estruturado em bases de dados. A interoperabilidade automatizada e a ativação da memória são apontados como possíveis rupturas para produtos jornalísticos que venham a explorar as tecnologias semânticas.

Palavras-chave

Web semântica, base de dados, jornalismo digital

Abstract

The web and the databases are considered critical technologies to the development of contemporary journalism in digital networks. In the early 2000s, the concept of the Semantic Web arises, which is a proposed expansion to the current web, to make it more automated and efficient. From an observation of the BBC Wildlife website, that uses Semantic Web technologies to the management and publishing of editorial content about the natural world, this study analyzes some aspects that contribute to the enhancement of the characteristics of database journalism. Interoperability and automated activation of memory are indicated as possible ruptures to journalistic products that use semantic technologies.

Keywords

Semantic Web, Data base, online journalism

Artigo recebido em 20/03/2012

Aprovado em 12/04/2012

O jornalismo, como uma atividade que trabalha essencialmente com informações, não tem como escapar da realidade apresentada pelas tecnologias digitais. Os conhecimentos dos campos do jornalismo e da ciência da computação se entrelaçam e esboçam um cenário emergente. O jornalismo digital é baseado em plataformas tecnológicas que passam por constantes mutações, então a produção jornalística praticada neste ambiente e os produtos dela oriundos também passam por redefinições (PAVLIK, 2000).

Sem defender o pensamento reducionista (e tentador) de que novas tecnologias tendem a melhorar a prática jornalística – pois, como afirma Palacios, corremos o perigo de instaurar “um pensamento guiado por uma lógica evolucionista de caráter simplista” (PALACIOS, 2003, p. 16) – consideramos que a contemporaneidade é marcada, entre outros fenômenos, pelo surgimento de um ciberespaço que redefine práticas sociais e profissionais (LEMO; LÉVY, 2010). As mudanças tecnológicas têm potencial para transformações nos modos de produzir/consumir informação jornalística.

Este artigo sintetiza os principais aspectos de uma pesquisa realizada sobre a web semântica aplicada ao jornalismo digital ao fazer o estudo de produtos desenvolvidos para as redes digitais pela empresa BBC. Ao observar o exemplo da BBC Wildlife, o presente texto sistematiza questões fundamentais para o desenvolvimento do jornalismo digital: a web semântica e as bases de dados aplicadas ao jornalismo.

A web e o jornalismo

Desenvolvida entre os anos de 1989 a 1991, pelo cientista britânico Tim Berners-Lee, a web foi fundamental para entrada do jornalismo nas redes digitais. Foi a partir desta tecnologia, que tornava acessível para leigos a navegação no ciberespaço, que as empresas jornalísticas despertam sua atenção para as possibilidades que a internet poderia oferecer ao jornalismo.

No ano de 2011, a web completou 20 anos desde o lançamento público e, no decorrer deste período, apresentou atualizações em especificações técnicas importantes (como as atualizações do HTML publicadas pela W3Cⁱ), além de ter sido enriquecida com o surgimento de tecnologias paralelas, como *plug-ins* para multimídia e linguagens

de *script*. Porém, na essência, a organização da web continuou funcionando com base no mesmo conceito de sua origem: como uma rede de documentos conectados. No início da década de 2000, Berners-Lee apresentou um conceito mais avançado para esta rede. Tal proposta ele denominou “Web Semântica”: uma rede que funcionaria não apenas como um sistema de associações de documentos criados prioritariamente para a leitura humana, mas como uma rede de dados, em que os computadores também seriam capazes de identificar os significados dos conteúdos publicados nas páginas (BERNERS-LEE et al., 2002).

Atualmente existem diferentes propostas de soluções tecnológicas para o funcionamento da Web Semântica. Tratamos aqui da proposta apresentada por Berners-Lee (et al., 2002) e recomendada pela W3C, em que a Web Semântica tem como base a própria estrutura da atual web (HTTPⁱⁱ, URIⁱⁱⁱ e HTML^{iv}), porém articulada com outras tecnologias, como a linguagem *Resource Description Framework* (RDF), os vocabulários e as ontologias. Basicamente, o que a Web Semântica realiza é a identificação dos significados presentes na rede (como pessoas, lugares ou objetos), através da linguagem RDF, e as utiliza para relacionar as informações presentes na web (como palavras, textos, páginas ou imagens) a coleções de conceitos que são disponibilizadas e compartilhadas na própria web. Desta forma, os sites e serviços da web podem associar seus dados aos mesmos conceitos, e por isso os computadores recebem uma capacidade maior de identificação de conteúdos da web.

Além da associação de dados a significados com o uso de RDF, a Web Semântica também aproveita as funcionalidades de um recurso denominado ontologia, que permite aos computadores realizarem inferências sobre os conceitos identificados pelas máquinas. Para isso, as ontologias oferecem um modelo lógico de relacionamentos entre conceitos, como em uma taxonomia, e, além disso, institui regras de relacionamentos entre estes conceitos, tais como pertencimento, igualdade ou distinção. As ontologias possibilitam uma maior automatização no tratamento de dados, pois as inferências dão autonomia às máquinas.

Em outras palavras, na Web Semântica as informações publicadas na rede são preparadas para serem compreendidas tanto por humanos quanto por máquinas, o que resultaria em uma web mais eficiente e autônoma na busca e na associação de

informações. Para Berners-Lee et al. (2002), passaríamos do paradigma de **web de documentos** para a de **web de dados**, estruturados e adaptados para a interpretação das máquinas. As vantagens de um sistema semântico global alcançariam diversas áreas que trabalham com a organização e o compartilhamento de dados, além da automação em operações que envolvem o gerenciamento dos mesmos, tais como na ciência da computação (BERNERS-LEE et al, 2002; SHADBOLT et al, 2006; KASHYAP et al, 2008), na ciência da informação (SOUZA E ALVARENGA, 2004) e também no jornalismo (BERTOCCHI, 2010), devido à natureza informativa dessa prática profissional. Esse cenário abre caminho para o desenvolvimento de produtos mais complexos e integrados aos conteúdos publicados na rede, pois na Web Semântica a estruturação de dados é universalmente padronizada, permitindo seu compartilhamento.

Jornalismo de Dados

A web começou como sistema de documentos digitais estáticos, ou seja, sem o uso de bases de dados para o armazenamento dos conteúdos. Com o tempo, as BDs se consolidaram como uma forma mais eficiente de armazenamento de dados na web. Uma das razões para essa consolidação foi a disseminação das ferramentas de publicação e dos sistemas gerenciadores de conteúdos (*Content Management System* ou CMS), que são sistemas direcionados “à administração e gerenciamento do conteúdo, voltado para publicação, para os processos de seleção, aprovação e edição” dos mesmos (SCHWINGEL, 2009, p. 2).

As funcionalidades das BDs em modelos relacionais^v têm um potencial de criação significativo, justamente devido às possibilidades de associações e combinações de dados digitais, mesmo nos casos em que os dados se encontram em formatos diferentes, pois apresentam natureza bastante flexível nas combinações. Para se compreender essa natureza dos dados, Manovich (2001) lista em seu o texto *The Language of New Media* cinco princípios das “novas mídias” que se aplicam aos conteúdos digitais:

- **representação numérica:** qualquer mídia digital, independente de ser originalmente criada no computador ou convertida de um suporte analógico, é

composta por códigos digitais que são representados numericamente; logo, todas as mídias digitais podem ser manipuladas matematicamente;

- **modularidade:** todas as mídias digitais são formadas pelas mesmas estruturas modulares, independente da escala em que se encontram: a foto é formada por pixels, o vetor é formado por curvas e linhas etc. Tais mídias podem ser combinadas, mas podem manter suas estruturas modulares independentes umas das outras, como no caso de uma animação em Flash, que combina áudio, imagens, textos e vídeos. Mesmo que exista combinação, cada mídia mantém sua estrutura mínima. Nas páginas HTML, ocorre o mesmo;
- **automação:** os princípios da representação numérica e da modularidade permitem que certas operações sejam automatizadas na criação, na manipulação e no acesso das mídias, removendo em parte a participação humana no processo de criação;
- **variabilidade:** como consequência dos princípios da representação numérica e da modularidade, as novas mídias podem existir em diferentes (potencialmente infinitas) versões. Ao invés de variabilidade, seria possível utilizar também os termos “mutável” ou “líquido”;
- **transcodificação:** considerada a consequência mais substancial da computadorização da mídia, o princípio diz que as novas mídias, quando digitalizadas, passam a ser codificadas tanto em um formato com organização estrutural compreensível por humanos (como o significado simbólico de uma imagem a partir das linhas, curvas etc), quanto em uma organização estrutural “compreensível” pelas máquinas a partir de convenções estabelecidas (como a cor RGB dos pixels, a dimensão da foto, o tamanho do arquivo). Com a transcodificações, os computadores podem relacionar diferentes tipos de arquivos (textos, áudios, vídeos etc) a partir destas convenções.

Ao identificar a organização estrutural reconhecida por humanos como “camada cultural” (*cultural layer*) e as convenções dos computadores como “camada computacional” (*computer layer*), e ao considerar que as novas mídias são criadas,

distribuídas, armazenadas e arquivadas em computadores, Manovich (2001) acredita que a camada computacional deverá influenciar a camada cultural.

A prática do jornalismo digital está inserida em um cenário bastante diversificado em termos tecnológicos. Com o crescimento vertiginoso na produção e no consumo de dados, paralelamente à web, uma tecnologia específica se destaca: a base de dados^{vi} (BD). Mais do que mera ferramenta de armazenamento, a BD passa a ser a tecnologia fundamental na organização, estruturação e apresentação das informações, e, por isso, define funcionalidades e estética dos produtos informacionais, e passa a ser considerada formato cultural de nossa época (MACHADO, 2006; MANOVICH, 2001).

Para Barbosa (2008), o que marca a transição entre a terceira e a quarta geração do jornalismo digital é uma fase caracterizada pela exploração das bases de dados como um recurso estruturante, responsável não apenas pelo armazenamento e pela gestão dos dados, mas também pela estruturação e apresentação do conteúdo. Para a autora, a relevância desta tecnologia na construção de produtos jornalísticos passou a ser fundamental, e isso a levou a desenvolver o conceito do **Jornalismo Digital em Base de Dados (JDBD)**, que não apenas marca esta transição de gerações, como também pode ser considerado o atual paradigma que define as características dos produtos jornalísticos da quarta geração. Em estudos de identificação das características do paradigma JDBD, Barbosa elencou oito categorias que representam áreas com contribuições relevantes das funcionalidades das BDs. São elas: dinamicidade^{vii}, automatização^{viii}, flexibilidade^{ix}, inter-relacionamento/hiperlincagem^x, densidade informativa^{xi}, diversidade temática^{xii}, visualização e convergência^{xiii}.

Enquanto as BDs servem como suporte para armazenamento e como estrutura para conteúdos da web, a Web Semântica pode funcionar de forma concomitante, como um sistema inteligente de publicação e gerenciamento destes mesmos conteúdos. As BDs são estruturas que tornam a publicação e o gerenciamento de dados muito mais dinâmicos e flexíveis, além de possibilitar certa automatização, como na categorização e recuperação de dados. Já a Web Semântica oferece recursos que tornam as ações de publicação e gerenciamento mais inteligentes, pois as possibilidades de identificação de significados e de geração de inferências têm o potencial de maximizar a automatização nos processos realizados pelas máquinas. Em outras palavras, a Web Semântica pode

funcionar como uma camada lógica em um sistema que se utiliza das BDs para gerenciar dados.

Em pesquisa realizada por Lammel (2011), foram descritos dois casos de produtos jornalísticos produzidos pela British Broadcasting Corporation (BBC) que utilizam bases de dados e também tecnologias da Web Semântica para um processo denominado pelos desenvolvedores dos produtos como “publicação dinâmica e semântica” (RAYFIELD, 2010). Neste processo, as bases de dados funcionam como estruturas de armazenamento de conteúdos, enquanto que as tecnologias semânticas atuam como um sistema automatizado, que decide de forma autônoma como os conteúdos devem ser publicados e como devem ser convergidos. Um desses casos, o BBC Wildlife, é apresentado no presente artigo.

BBC Wildlife: um caso de aplicação da Web Semântica

O BBC Wildlife é um portal que reúne uma grande produção de conteúdos sobre o mundo natural, mais especificamente biológico, como animais selvagens, plantas, fungos e seres pré-históricos. O site armazena e organiza o conteúdo (textos, imagens, áudios e vídeos) sobre a natureza como se fosse uma enciclopédia multimídia, com destaque para os milhares de clipes de vídeo produzidos em mais de 30 anos de produção televisiva, e utiliza esta base de conhecimento em matérias jornalísticas sobre o tema. Neste caso, as tecnologias semânticas são aplicadas para um tipo de conteúdo conhecido no jornalismo como *soft news*, que se aproximam de produtos como reportagens de revista, documentários e produtos informativos para educação e entretenimento.

Além do rico repositório de mídias, outro grande destaque do site é a organização das páginas. Para cada uma das mais de mil espécies, há uma página única gerada e atualizada de forma dinâmica, que agrega informações, áudios e vídeos sobre a espécie em questão. Além das páginas para as espécies, o site gera outras centenas de páginas para reunir conteúdos sobre animais que compartilham das mesmas características. Essas características são: habitat (ex.: floresta, deserto, marinho, urbano etc), comportamento/adaptação (ex.: se é carnívoro, se voa, se é noturno etc) e nível da classificação biológica (domínio, reino, filo, superclasse, classe, superordem, ordem,

subordem, superfamília, família, gênero e espécie). Como exemplo, podemos citar uma página que agrega vídeos e informações apenas sobre animais da classe dos insetos, outra apenas sobre animais e plantas com comportamento carnívoro, outra só com animais voadores, ou ainda uma página somente sobre animais e plantas que vivem no habitat urbano. As páginas apresentam diversos links que remetem umas para as outras, de forma dinâmica, de acordo com os tipos de relacionamento entre os conceitos.

A **navegação** do site não funciona com um menu central permanente, como ocorre em sites tradicionais. Há somente um menu na página inicial, que não é mostrado nas outras páginas. O menu aparece somente na página inicial porque é apenas porta de entrada para uma série de páginas internas que, por sua vez, apresentam diversas listas de links, servindo como menus contextualizados para o conteúdo que mostram.

Tanto o menu da página inicial quanto as listas de links encontrados nas páginas internas se encaixam em três grandes categorias de navegação: por espécies, por comportamentos/adaptações e por habitats. Além destas três categorias, as páginas internas do site ainda podem oferecer duas outras categorizações que possibilitam outra forma de navegação: por biorregião ou pelos outros níveis da classificação dos seres vivos além das espécies (domínios, reinos, filos, superclasses, classes, superordens, ordens, subordens, superfamílias, famílias e gêneros), que geram centenas de páginas agregadoras de conteúdo.

É importante notar: o site Wildlife constrói uma **página dinâmica** para cada habitat, cada comportamento/adaptação, cada biorregião e cada nível da classificação biológica (espécies, domínios, filos etc). Como é possível observar na Figura 1, as páginas são construídas a partir de um *template* padrão que divide o espaço da tela em duas grandes áreas. Na parte superior, são disponibilizados os conteúdos relacionados ao assunto em questão (título, texto descritivo e mídias) oriundos de diferentes fontes internas e externas à BBC. Já na parte inferior, são apresentadas várias listas de links, como se fossem novos menus para conteúdos relacionados ao tema da página.

As páginas apresentam informações em diferentes formatos, como textos, cliques de áudio, estatísticas e mapas. Destes, o tipo de conteúdo que se destaca são os cliques de vídeos, que, no Wildlife, são trechos curtos obtidos de documentários televisivos (aproximadamente 3 minutos), mas que apresentam um sentido completo. Para uma

descrição detalhada, apresentamos a página das espécies na Figura 1, com as indicações dos elementos que formam o *layout*.

Parte superior

Parte inferior

A **B** **C** **D** **E** **F** **G** **H** **I** **J** **K** **L** **M** **N**

Figura 1 – Página das espécies, com marcações indicativas. Disponível em: <<http://www.bbc.co.uk/nature/life/Lion/>>. Acesso em: 22 dez. 2011.

Na parte superior da Figura 1, são apresentados os seguintes conteúdos:

- A) O título e a descrição do conteúdo mostrado no momento. No caso do leão, como mostrado na figura, o texto da descrição apresenta links para algumas características da espécie, como uma região onde o animal é encontrado (África) e um comportamento (carnívoro), além do nível da classificação biológica a qual o leão se refere (espécie). Cada um destes links remete o usuário a uma página com estrutura semelhante, porém com conteúdos relacionados ao respectivo link.
- B) Espaço para a reprodução das mídias (clipes de vídeo produzidos pelas dezenas de programas de televisão da BBC).
- C) Lista horizontal dos vídeos que a página agrega. Ao clicar em uma das miniaturas, o vídeo é aberto no espaço demarcado com a letra B, e é mostrada a descrição do vídeo no espaço A (título do clipe, texto descritivo e nome do programa de origem da BBC em que o vídeo foi produzido, linkado para o site do respectivo programa). A lista de miniaturas pode ser deslizada para o lado, para se revelar mais miniaturas (a página da espécie leão continha 32 miniaturas no momento da pesquisa).

Na parte inferior da Figura 1, são apresentados os seguintes conteúdos:

- D) Mapa com destaque às biorregiões onde a espécie é encontrada.
- E) Lista de habitats em que a espécie em questão pode ser encontrada. Este espaço serve como um menu para outros conteúdos do site. Cada imagem é um link que remete o usuário para uma página semelhante a esta, porém que agrega animais encontrados no habitat em questão.
- F) Lista de comportamentos (e adaptações) que a espécie demonstra, tais como “nômade”, “noturno”, “social” ou “territorial”. Mais uma vez, o espaço serve como um menu para o conteúdo do site, neste caso para páginas que agregam vídeos de animais com estes mesmos comportamentos.
- G) Status da conservação da espécie (ex.: vulnerável, ameaçado de extinção, extinto).
- H) Texto linear com informações mais detalhadas sobre o assunto em questão.

- I) Nível da classificação biológica em que o assunto da página se localiza. Como no caso o leão é uma espécie, então também são mostrados os outros níveis anteriores, como gênero, família etc. A lista da classificação se torna um menu para conteúdos do site.
- J) Mídias sonoras relacionadas ao tema, oriundos de programas da BBC. No caso da Figura 1, são disponibilizadas gravações de rugidos de leões.
- K) Campo para pesquisa no site.
- L) Lista de notícias da BBC sobre o tema da página. Os links podem remeter o usuário para diferentes páginas da BBC.
- M) Lista com coleções especiais de vídeos em que o assunto em questão é mencionado.
- N) Links para páginas externas à BBC que tratem sobre o assunto em questão.

Como já explicado, cada página monta a sua estrutura de navegação automaticamente, de acordo com os tipos de relacionamentos que possuem com os conteúdos. Dessa forma, o sistema cria automaticamente uma malha de páginas interlinkadas, rica em relacionamentos. As diferentes maneiras de se categorizar os conteúdos permitem que o site formule e distribua pelas páginas internas várias listas de links, que convidam o usuário a continuar a navegação pelo site de acordo com o contexto, como se as próprias páginas internas fosse um grande menu de navegação.

Embora não se trate de material sobre acontecimentos factuais, a BBC aproveita essa base de conhecimento para enriquecer seus conteúdos jornalísticos, tanto nos sites especializados sobre as questões ambientais e científicas quanto no site principal de notícias. Em outras palavras, o enorme conjunto de páginas e de suas interligações formuladas de acordo com os tipos de relacionamentos entre conceitos tornam o Wildlife uma base de conhecimento sobre um domínio específico do mundo (a vida natural) que servem de **complemento aos produtos jornalísticos** da BBC.

A partir desse repositório, a BBC cria conexões entre o conhecimento ali organizado e os materiais jornalísticos dos outros sites da emissora. Essa conexão pode ser apresentada pelos dois lados: se há links para notícias nas páginas do Wildlife, pode haver links do Wildlife nas respectivas páginas de notícias.

O funcionamento do sistema empregado na publicação e no gerenciamento do site BBC Wildlife adota várias tecnologias que caracterizam a proposta de Web Semântica, tais como o RDF e as ontologias, além da técnica de *tagging* manual, que enriquece o repositório de metadados utilizados nas operações semânticas (LAMMEL, 2011). Embora fosse possível o desenvolvimento das páginas de espécies do Wildlife apenas com o uso de BDs relacionais, o diferencial do sistema semântico ocorre por trás dos bastidores. As tecnologias semânticas maximizam a automatização nos processos de publicação e gerenciamento das informações, pois a ontologia do BBC Wildlife possibilita às máquinas realizarem inferências e decidirem em quais páginas os conteúdos textuais e multimídia devem ser publicados. Além disso, outra contribuição das tecnologias semânticas foi o reaproveitamento de dados originados em sites externos, tais como descrições e estatísticas sobre espécies. Essa interoperabilidade de dados entre sites ocorre principalmente devido à padronização dos significados adotados por ambos, através dos vocabulários compartilhados. Como exemplo, podemos citar a página de uma espécie: se a página em questão é um agregador de conteúdos sobre a espécie leão, então as tecnologias semânticas possibilitam que o BBC Wildlife reaproveite conteúdos externos que foram associados a este mesmo conceito através dos vocabulários compartilhados.

Memória e interoperabilidade: possíveis rupturas para o jornalismo digital

No jornalismo digital, Web Semântica e bases de dados relacionais podem ser complementares, da mesma forma que os documentos hipertextuais em HTML continuaram existindo com o surgimento das BDs. A partir do caso BBC Wildlife, foi observado que as tecnologias semânticas contribuem com avanços em determinadas funções desempenhadas pelos atuais sistemas em bases de dados.

O processo de publicação dinâmica e semântica da BBC demonstrou que a Web Semântica enriquece os produtos digitais estruturados em base de dados, principalmente devido à automatização, que potencializa as categorias do JDBD: tornou o produto ainda mais dinâmico, flexibilizou a produção ao proporcionar a integração automática de conteúdos produzidos em sites externos (por equipes descentralizadas), aumentou o

inter-relacionamento, a hiperlincagem e a diversidade informativa com a criação automática de listas de links para as páginas das espécies, convergiu diferentes tipos de conteúdos e ainda contribui na diversidade temática, ao elencar listas de links em diferentes categorias (habitat, comportamento etc) relacionados a um conceito específico (ex.: a espécie tratada na página em questão). Das oito categorias do JDBD, a única que não apresentou resultados relevantes foi a da visualização, pouco explorada pelos desenvolvedores da BBC.

Além da potencialização nas categorias do JDBD, acreditamos que a Web Semântica apresenta contribuições que irrompem com práticas até então dominantes no jornalismo digital; ou seja: pode vir a representar um salto ainda maior do que uma potencialização de características até então exploradas. Esta possível ruptura^{xiv} seria a interoperabilidade automatizada. Ela permite que diferentes sites (que estejam na lógica da Web Semântica) troquem entre si dados e informações de maneira automatizada, a partir de associações de conceitos definidos por vocabulários ou ontologias em comum. Tal situação resulta em pelo menos três grandes contribuições da interoperabilidade automatizada: 1) diversifica as fontes externas de dados, ocasionando o enriquecimento dos produtos informacionais em diferentes categorias, como no inter-relacionamento/hiperlincagem, na diversidade informativa e na diversidade temática; 2) oportuniza o reaproveitamento de dados e informações produzidos por terceiros ou por outras equipes da mesma organização; 3) permite o compartilhamento de informações entre diferentes sites que sigam a lógica da Web Semântica, forma complementar uns aos outros, auxiliando na formação de uma base de conhecimento compartilhada e colaborativa. Em relação a este último item, é possível citar como exemplo o projeto Linked Data^{xv}, uma rede com centenas de sites que produzem informações e que as compartilham de forma aberta entre eles, utilizando padrões da Web Semântica, formando assim uma grande rede de dados ligados, abertos e, muitas vezes, complementares uns aos outros.

Também acreditamos que a Web Semântica potencializa uma característica em especial do jornalismo digital: a memória. Segundo Palacios (2003), a memória pode ser considerada uma ruptura, pois, pela primeira vez na história, o jornalismo pode se aproveitar de um espaço praticamente ilimitado, disponível tanto ao produtor quanto ao

consumidor da informação. No caso do BBC Wildlife, a Web Semântica funcionou como um sistema ativador da memória, pois oportunizou o desenvolvimento de um novo produto (o próprio site) através da exploração de um grande repositório de vídeos produzidos pela BBC que, até então, estavam arquivados. Em outras palavras, tanto a interoperabilidade de dados quanto a automatização das máquinas na identificação de significados e na geração de inferências permitiram que milhares de vídeos fossem associados a conteúdos produzidos na web e que fossem integrados em um mesmo sistema, gerando, assim, um novo produto. O caso demonstra que a Web Semântica é um facilitador no processo de construção de bases de conhecimento, que podem vir a ser exploradas pelos sites jornalísticos.

Nestes últimos anos, as máquinas vêm desempenhando um papel substancial no gerenciamento da informação. Tomamos como comprovação dessa afirmativa a proliferação de sistemas estruturados em bases de dados. As máquinas se tornam ferramentas que liberam o potencial criativo do humano, pois assumem em nosso lugar as operações mecânicas e repetitivas, dignas de uma máquina. A proposta da Web Semântica é de se afirmar como uma solução nesta linha: deixar para as máquinas a tarefa tediosa de buscar e recuperar grandes quantidades de dados e informações. O humano, essencialmente racional e criativo, usa os dados e as informações para alcançar um nível muito mais elevado, onde encontramos a criatividade e a geração de conhecimento.

Referências Bibliográficas

BARBOSA, Suzana. **Jornalismo Digital em Base de Dados (JDBD)** - Um paradigma para produtos jornalísticos digitais dinâmicos. Tese de doutorado. Facom/Ufba, Salvador, 2007.

BARBOSA, Suzana. **As bases de dados no curso da convergência jornalística**: uma análise preliminar a partir do modelo JDBD. 2008b. Disponível em:
<http://grupojol.files.wordpress.com/2011/05/2008_barbosa_base_de_dados.pdf>. Acesso em: 14 jan. 2012.

BERNERS-LEE, Tim; HENDLER, James; LASSILA, Ora. **The Semantic Web**. Scientific American Special Online Issue, abril de 2002, p. 24-30. Disponível em:
<http://cms.brookes.ac.uk/modules/notes/112_SemWeb.pdf>. Acesso em: 28 nov. 2010.

BERTOCCHI, Daniela. **Narrativas jornalísticas no contexto da web semântica**. 2010. Anais do II Seminário de Ciberjornalismo do Mato Grosso do Sul.

KASHYAP, Vipul; BUSSLER, Christoph; MORAN, Matthew. **Semantic Web: Semantics for Data and Services on the Web**. Berlin, Heidelberg: Springer-Verlag Berlin Heidelberg, 2008.

LAMMEL, Iuri. **Organização e gerenciamento de conteúdos jornalísticos na Web Semântica**. Dissertação de Mestrado. Programa de Pós-Graduação em Comunicação Midiática, Universidade Federal de Santa Maria, Santa Maria, 2011.

LEMOS, André; LÉVY, Pierre. **O futuro da internet**: Em direção a uma ciberdemocracia planetária. São Paulo: Paulus, 2010.

MACHADO, Elias. **O jornalismo digital em base de dados**. Florianópolis: Calandra, 2006.

MANOVICH, Lev. **The Language of New Media**. 2001. Disponível em:
<<http://ucsd.academia.edu/LevManovich/Papers>>. Acesso em: 14 nov. 2011.

OLIVIERO, Carlos A. J. **Faça um aplicativo: Banco de dados cliente/servidor com Delphi 6 – Orientado a projeto**. São Paulo: Érica, 2002.

PALACIOS, Marcos. Ruptura, Continuidade e Potencialização no Jornalismo Online: o Lugar da Memória. In: MACHADO, Elias & PALACIOS, Marcos (Orgs). **Modelos do Jornalismo Digital**, Salvador: Calandra, 2003.

PAVLIK, J. **The Impact of Technology on Journalism**. Journalism Studies, V. 1, Nº 2, 2000, p. 229–237. Disponível em: <<http://www.tandfonline.com/doi/abs/10.1080/14616700050028226>>. Acesso em: 8 fev. 2012.

RAYFIELD, J. **BBC World Cup 2010 dynamic semantic publishing**. 2010. Disponível em:
<http://www.bbc.co.uk/blogs/bbcinternet/2010/07/bbc_world_cup_2010_dynamic_sem.htm>. Acesso em: 26 jan. 2012.

SHADBOLT, N.; BERNERS-LEE, T.; HALL, W. **The Semantic Web Revisited**. IEEE Intelligent Systems, vol. 21, n. 3, maio/junho de 2006, p. 96-101. Disponível em:
<http://eprints.ecs.soton.ac.uk/12614/1/Semantic_Web_Revisited.pdf>. Acesso em: 26 jan 2012.

SCHWINGEL, Carla. **A produção de conteúdos no ciberespaço**: sistemas de gerenciamento de conteúdos. Artigo apresentado no VII Encontro Nacional de Pesquisadores em Jornalismo. São Paulo, 2009.

SOUZA, Renato Rocha; ALVARENGA, Lúcia. **A Web Semântica e suas contribuições para a ciência da informação**. Ci. Inf., Brasília, v. 33, n. 1, abril de 2004. Disponível em:
<http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-1962004000100016&lng=en&nrm=iso>. Acesso em: 27 mar. 2011.

ⁱ A W3C é um grupo de especialistas e de empresas que desenvolvem as principais tecnologias e padrões da Web. Segundo o site da W3C: “The World Wide Web Consortium (W3C) is an international community that develops standards to ensure the long-term growth of the Web”. Em tradução livre: “A World Wide Web Consortium (W3C) é uma comunidade internacional que desenvolve padrões que asseguram o crescimento da Web em longo prazo”. Disponível em: <<http://www.w3.org>>. Acesso em: 23 jun 2010.

ⁱⁱ HTTP (HyperText Transfer Protocol) é um protocolo de transferência de dados entre computadores; permite que as máquinas se comuniquem utilizando “a mesma língua”.

ⁱⁱⁱ URI (Uniform Resource Identifier) é um esquema único de nomes para localização de recursos da rede, como os endereços de páginas que começam com o “www”.

^{iv} HTML (HyperText Markup Language) é a linguagem de marcação utilizada para a montagem de páginas da web. É formada por códigos padronizados (*tags*) que executam comandos de formatação ao conteúdo, como negrito e itálico, e que permitirem a inserção de hiperlinks e metadados nas páginas.

^v Bases de dados relacionais são modelos baseados em tabelas, em que cada tabela possui dados estruturados em colunas e linhas, que podem ser relacionados a outras tabelas da base de dado. É amplamente utilizado na web.

^{vi} Uma base de dados (BDs), ou banco de dados, é um “mecanismo capaz de manipular, armazenar e organizar informações de modo que possam ser recuperadas rapidamente e a qualquer momento” (OLIVIERO, 2002, p. 26). Logo, as BDs não são apenas estruturas para armazenamento de dados, elas também servem para gerenciá-los de forma mais eficiente.

^{vii} É a característica básica das BDs de dinamizar os conteúdos apresentados em produtos da web. Ao contrário do conteúdo estático dos sites produzidos apenas em HTML, os conteúdos oriundos das BDs são dinâmicos porque podem mudar seu estado sem a intervenção direta de um programador no código-fonte do site em que tal conteúdo é apresentado.

^{viii} Ocorre quando os dados são manipulados de forma automática pela máquina, ou seja, quando não há a necessidade da intervenção humana direta para que ocorra uma mudança de estado. Há três tipos básicos de automatização: a parcial (aplicada apenas a algumas etapas do processo de produção jornalística), a procedimental (quando mais etapas do processo jornalístico ocorrem de forma automatizada) e a total (quando o produto jornalístico funciona de forma totalmente automatizada).

^{ix} A tecnologia das BDs traz certas facilidades à produção jornalística, pois assegura maior agilidade, qualidade e flexibilidade à produção. Com elas, os sistemas de apuração se tornam menos hierarquizados, os conteúdos são mais facilmente recuperados e o trabalho dos jornalistas se torna mais autônomo e descentralizado.

^x Considerado pela autora como um dos grandes potenciais das BDs, é a “capacidade de identificar padrões combinatórios e inter-relacionamentos diversos entre as informações” (BARBOSA, 2007, p. 238). A tecnologia tem o poder de vasculhar rapidamente grandes quantidades de dados e identificar quais deles podem ser inter-relacionados, de acordo com o contexto.

^{xi} É a quantidade de informações presente em um conteúdo. Geralmente, uma notícia é inicialmente apresentada com uma baixa densidade, mas com o decorrer do tempo a densidade é elevada com a inserção de novas informações, na medida em que a notícia é complementada, alterada, corrigida, contextualizada ou aprofundada.

^{xii} A categoria representa a diversidade de tematizações além das mais comuns (como política, economia, esportes, cultura, ciência, saúde e tecnologia).

^{xiii} São as diferentes maneiras de se representar na tela as informações jornalísticas armazenadas nas BDs.

^{xiv} Ruptura é o grande distanciamento que uma nova prática toma em relação às práticas adotadas por suportes anteriores. Segundo Palacios, “[...] as características do Jornalismo na web aparecem, majoritariamente, como Continuidades e Potencializações e não, necessariamente, como Rupturas com relação ao jornalismo praticado em suportes anteriores” (2003, p. 22). Contudo, para o autor, é possível apontar algumas rupturas e a principal delas é a memória.

^{xv} Linked Data é um projeto que reúne iniciativas que seguem recomendações de boas práticas na construção de repositórios de dados em conformidade com a Web Semântica.