



AS POTENCIALIDADES DA WEB SEMÂNTICA E WEB 2.0 PARA A CIÊNCIA DA INFORMAÇÃO E OS NOVOS FORMATOS DE PUBLICAÇÕES ELETRÔNICAS PARA A PESQUISA ACADÊMICO-CIENTÍFICA

Déborah Motta Ambinder¹ e Carlos Henrique Marcondes²

¹Mestranda do PPGCI – Universidade Federal Fluminense (UFF) – Bibliotecária Documentalista da UFF/NDC/BEE – Brasil

²Doutor em Ciência da Informação – Departamento de Ciência da Informação – Universidade Federal Fluminense (UFF) – Brasil

RESUMO

Mesmo com a facilidade de acesso ao texto completo dos artigos de periódicos científicos, o formato utilizado no meio eletrônico ainda é o textual, (legível por pessoas) o que impossibilita a sua recuperação em uma forma legível por programas. As tecnologias semânticas propõem avanços na questão da recuperação e processamento semântico de conteúdos nos ambientes de *e-Science*¹. Este artigo apresenta algumas experiências e propostas de publicações semânticas que exploram no ambiente digital, as possibilidades abertas pelas tecnologias da Web Semântica, e Web 2.0, que têm em comum, o fato de proporem novos formatos que ultrapassem a mera cópia digital do impresso, ampliando o potencial de compreensão e recuperação do conteúdo semântico e de interação entre autores e leitores de artigos científicos digitais.

Palavras-Chave: Periódico Científico; Comunicação Científica; Publicações Semânticas; Web Semântica; Web 2.0.

ABSTRACT

Even with easy access to the full-text journal articles scientific, the format used in the electronic environment is still textual (human-readable), hindering their recovery in a form readable by programs. The semantic technologies offer advances on the issue of recovery and semantic processing of contents in *e-Science* environments. This paper presents some experiments and propositions of semantic publications That exploit, in the digital environment, the possibilities opened by Semantic Web technologies and Web 2.0, which have in common the fact of offering new formats that exceed the mere digital print copy, Enlarging the potential for understanding and retrieval of semantic content and interaction between authors and readers of scientific digital articles.

Keywords: Scientific Journal; Scientific Communication; Publications Semantic; Semantic Web; Web 2.0.

1 INTRODUÇÃO

Os avanços das Tecnologias de Informação e Comunicação (TICs) e o surgimento da web mudaram a forma e o conteúdo dos documentos. Sua interatividade possibilitou uma mudança fundamental no modo de fazer ciência e no esquema clássico da comunicação, desencadeando uma série de transformações culturais, econômicas, sociais e política.

Dentre as mudanças mais evidentes, destacamos o crescente valor da informação e das técnicas relacionadas à pesquisa, coleta, armazenamento e difusão da informação.

A busca do conhecimento científico se constitui como um grande e permanente empreendimento, um fator determinante para a ampliação da capacidade de assimilação e desenvolvimento de novas tecnologias.

A construção do edifício do saber remonta-se a milhares de anos, sendo que dia após dia, adicionam-se tijolos à sua estrutura. Descobrir coisas, ou ainda, saber coisas à frente baseando-se, para isso, em observações, são apenas algumas das justificativas que levam os pesquisadores a realizarem pesquisas científicas (GUMIEIRO, 2009, p.9).

“O progresso de uma sociedade está intimamente relacionado com o acesso que ela tem a informação” (GUMIEIRO, 2009, p.1). O grande diferencial de competitividade passou a ser o domínio de técnicas para o processamento e gerenciamento do conhecimento.

O tratamento das questões de comunicação do conhecimento nos remete ao compromisso fundamental e crítico da Ciência da informação com a sociedade: “[...] a tarefa massiva de tornar mais acessível um acervo crescente de conhecimento” (SARACEVIC, 1996, p.60).

Diante desta perspectiva, o cientista da Informação deve se envolver no cenário atual, de acordo com a visão de Saracevic (1995), que nos alerta para o fato de que “[...] a explosão da informação eletrônica exige um novo olhar para o desenvolvimento dos tradicionais serviços de informação: seleção, tratamento e recuperação da informação”.

Cabe aqui recordar, que o processo de recuperação de informações é uma atividade que tem recebido a atenção da Ciência da Informação ao longo do tempo.

A Ciência da Informação é um campo dedicado às questões científicas e à prática profissional voltadas para os problemas da efetiva comunicação do conhecimento e de seus registros entre seres humanos, no contexto social, institucional ou individual do uso e das necessidades de informação. No tratamento destas questões são consideradas de particular interesse as vantagens das modernas tecnologias informacionais (SARACEVIC, 1996, p.47).

Atualmente, a *World Wide Web* é o ambiente para qual estão voltados os maiores esforços de desenvolvimento na área de recuperação de informação devido a grande quantidade de informações disponível em toda a *Web*.

O maior desafio do profissional da informação no Século XXI é, contudo, gerenciar e buscar de alguma maneira, melhorar a qualidade de recuperação dos milhões de informações hoje disponíveis no ambiente *Web*.

“Diante deste novo cenário, em que a própria informação se torna o produto do processo produtivo, garantir o acesso ao conhecimento científico, gerado no país e no mundo, é estratégico para promover o desenvolvimento do indivíduo e da sociedade” (MARTINS, 2006, p.13).

Kuramoto (2006, p.91) lembra que “[...] a informação científica é o insumo básico para o desenvolvimento científico e tecnológico de um país. Esse tipo de informação, resultado das pesquisas científicas, é divulgado à comunidade por meio de revistas”

Sendo assim, com a expansão da rede de computadores e o acesso à internet, torna-se indispensável à discussão, o entendimento do principal meio ou instrumento utilizado no processo de comunicação científica: os periódicos científicos.

De acordo com Garvey e Griffith (1979, p.299) a comunicação científica pode ser entendida como:

Todo espectro de atividades associadas com a produção, disseminação e uso da informação, desde a busca de uma idéia para a pesquisa, até a aceitação da informação sobre os resultados dessa pesquisa como componente do conhecimento científico.

Historicamente, a comunicação científica formal, representada por artigos organizados em periódicos científicos, teve sua origem em 1665, com a criação do *Journal de Sçavans* na França e do *Philosophical Transactions of the Royal Society* na Inglaterra, quando a comunidade científica passa a exigir evidências baseadas na

observação e na experiência empírica, para que os conhecimentos resultantes possam ser considerados científicos (MEADOWS, 1999).

O periódico científico “[...] surge como uma evolução do papel das cartas pessoais trocadas entre cientistas e das atas de memórias de reuniões científicas” (BOORSTIN, 1989) e passa a assumir o papel de principal canal de publicação no processo de comunicação da ciência, sendo considerado ainda hoje, como um dos veículos mais importantes para legitimar a autoria das descobertas científicas, e o meio mais utilizado pelos pesquisadores para tornar pública as suas pesquisas.

Desde então, o periódico científico vem passando por transformações significativas, sendo que a mais importante no seu processo de evolução ocorreu a partir dos Anos 90, com o estabelecimento de uma nova categoria: a comunicação científica eletrônica.

A coexistência dos meios impresso e eletrônico constitui de fato, o novo paradigma no sistema de comunicação científica (COSTA, 2005, p.1).

“Com a chegada da informação eletrônica e uso de redes, ocorreu uma grande ruptura no modo de editoração e disseminação de informações, principalmente da produção científica” (FACHIN, 2002, p.18).

“A evolução da comunicação científica e o uso das tecnologias de informação e comunicação contribuíram na mudança do suporte tradicional do periódico científico - do impresso, para um novo meio: o on-line” (MEDEIROS; FACHIN; RADOS, 2008, p.416).

A mudança de suporte modificou o espaço-tempo da informação. O meio eletrônico, possibilitou ao periódico científico uma grande oportunidade de expansão, principalmente no que diz respeito à redução de custos, disponibilização de conteúdos e troca de informações entre os pares.

Atualmente uma grande quantidade de artigos científicos é disponibilizada na web em formato digital. Em fevereiro de 2011 o Portal de Periódicos da CAPES oferecia acesso a 26.449 publicações periódicas internacionais e nacionais e a renomadas bases de dados de resumos em todas as áreas do conhecimento.

No entanto, um dos maiores obstáculos encontrados por pesquisadores e grupos de pesquisa que trabalham para estruturar a crescente quantidade de conteúdos digitais disponibilizados na web, é o fato dos mesmos apesar de digitais, estarem em formato textual não estruturado.

Mesmo com a facilidade de acesso ao texto completo dos artigos de periódicos proporcionados pelas bibliotecas digitais, o formato utilizado no meio eletrônico ainda é o textual, (legível por pessoas) o que impossibilita a sua recuperação em uma forma legível por programas, ou seja, o periódico científico ainda não utiliza todo o potencial fornecido pela tecnologia de informação.

O periódico científico se firmou como publicação acadêmica em ambiente digital, porém continua calcado no modelo impresso, de modo a ser lido e interpretado apenas por seres humanos.

Uma pesquisa realizada por Tenopir (2003) mostra que os artigos de periódicos são tidos pelos cientistas americanos, como a fonte de informação mais importante para suas pesquisas, visto que os pesquisadores sempre utilizaram diversos artigos científicos ao mesmo tempo, para ler, analisar e anotar fragmentos do seu conteúdo. Esta prática chamada por Renear e Palmer (2009, p.828) de “leitura estratégica”, vem sendo intensificada no ambiente digital com o objetivo de explorar o maior número de artigos em menor tempo possível.

Contudo, para dar suporte e intensificar a efetividade desta prática de leitura, se faz necessário ampliar a interação entre autores e leitores de artigos científicos digitais, com o acesso direto ao conteúdo semântico e às afirmações contidas no texto de artigos científicos.

Recentes avanços em TICs estão trazendo novas oportunidades e novos modelos de publicação acadêmica que irão possibilitar aos pesquisadores pesquisarem e divulgarem os resultados de sua investigação de um modo ainda mais eficaz.

A comunidade acadêmica atenta às possibilidades trazidas pelas tecnologias semânticas vem buscando conhecer novas ferramentas para otimizar a pesquisa acadêmico-científica e a leitura estratégica que foi intensificada no ambiente web.

Diante disto surgem as seguintes questões:

- Como tem sido a evolução recente do periódico científico eletrônico à luz das novas tecnologias de informação e Comunicação?
- Que novas potencialidades existem para ir além do modelo impresso?
- Que potencialidades as TICs, em especial, as tecnologias semânticas e as da Web 2.0, oferecem e que não estão sendo utilizadas pelos periódicos científicos eletrônicos?

- Que experiências inovadoras estão tentando usar essas potencialidades?

2 REFERENCIAL TEÓRICO

O periódico científico é o principal canal de publicação no processo de comunicação da ciência, um dos veículos mais importantes para legitimar a autoria das descobertas científicas, e o meio mais utilizado pelos pesquisadores para tornar pública as suas pesquisas. Desde 1665 (época do seu surgimento) o mesmo vem passando por transformações significativas, sendo que a mais importante no seu processo de evolução ocorreu a partir dos anos 90, com o estabelecimento de uma nova categoria: a comunicação científica eletrônica.

Marcondes (2011, p.83) recorda que, “[...] antes do surgimento da Web o acervo de conhecimento científico validado da humanidade era difuso, e armazenado de forma dispersa em coleções de periódicos em bibliotecas”.

“A evolução da comunicação científica e o uso das tecnologias de informação e comunicação contribuíram na mudança do suporte tradicional do periódico científico - do impresso, para um novo meio: o on-line” (MEDEIROS; FACHIN; RADOS, 2008, p.416).

“Com a chegada da informação eletrônica e uso de redes, ocorreu uma grande ruptura no modo de editoração e disseminação de informações, principalmente da produção científica” (FACHIN, 2002, p.18).

A mudança de suporte modificou o espaço-tempo da informação. O meio eletrônico, possibilitou ao periódico científico uma grande oportunidade de expansão, principalmente no que diz respeito à redução de custos, disponibilização de conteúdos e troca de informações entre os pares.

A liberdade dos usuários e a facilidade de se publicar na internet fizeram com que um enorme volume de informações se proliferasse desordenadamente, transformando a internet em um imenso repositório de documentos.

Esse cenário interativo, descentralizado, aberto e favorável ao crescimento da rede, posteriormente desencadeou um grave problema [...] o processo de explosão informacional [...] Um usuário, ao efetuar uma pesquisa na internet, tende a ficar mais frustrado pelo excesso de “respostas” que pela falta delas. Isto porque o excesso de informações não representa uma solução, mas um problema: a desinformação (SCHONS, 2007, p.2).

Marcondes e Campos (2008, p.109), reforçam que “[...] encontrar a informação adequada passa a ser o principal problema cultural, econômico e científico da atualidade”.

O fato é que, ainda existe muita dificuldade para encontrar informações relevantes através dos buscadores *Web*. “A coexistência dos meios impresso e eletrônico se constitui de fato, o novo paradigma no sistema de comunicação científica” (COSTA, 2005, p.1).

Atualmente uma grande quantidade de artigos científicos é disponibilizada na web em formato digital. Para Gardin (2001), artigos publicados na Web são “bases de conhecimento”. No entanto, um dos maiores obstáculos encontrados por pesquisadores e grupos de pesquisa que trabalham para estruturar a crescente quantidade de conteúdos digitais disponibilizados na web, é o fato dos mesmos apesar de digitais, estarem em formato textual não estruturado.

A Web se transformou num meio de comunicação, informação e compartilhamento de conhecimento entre as pessoas. Este novo ambiente informacional é constituído por uma enorme quantidade de documentos e recursos. Contudo, o que parece ser o seu maior poder de atração, é ao mesmo tempo, um de seus fatores mais críticos.

Por serem criados de forma autônoma, sem preocupação com regras de estruturação, catalogação e descrições de suas propriedades, essas informações são difíceis de serem abrangidas pelos mecanismos de pesquisa, ocasionando demora e ineficácia na localização de informações (DIAS; SANTOS, 2001, p.3).

As páginas da Web foram construídas com semânticas locais (silos de informações) que não se falam. Este fato se constitui como o maior obstáculo para o seu desenvolvimento. Pensar em organizar o caos informacional disponível na Web, se tornou imperativo para possibilitar novas formas de acesso, e recuperação da informação virtual.

A internet como um vasto conjunto de redes interligadas disponibiliza informações através de servidores espalhados em todo o mundo, No entanto, sua organização e recuperação são feitas de forma ineficiente, levando a buscas mal sucedidas com resultados desinteressantes para o usuário (SILVA, 2007, p.iv).

Mesmo com a facilidade de acesso ao texto completo dos artigos de periódicos proporcionados pelas bibliotecas digitais, o formato utilizado no meio

eletrônico ainda é o textual, (legível por pessoas) o que impossibilita a sua recuperação em uma forma legível por programas. De acordo com esta perspectiva, o periódico científico ainda não utiliza todo o potencial fornecido pela tecnologia de informação.

O potencial das TICs tem sido aplicado a sistemas modernos de informação bibliográfica para melhorar a comunicação científica, proporcionando o acesso rápido e notificação imediata e acesso ao texto integral de artigos científicos. Mas esse potencial não tem sido usado para processar diretamente o conhecimento incorporado no texto de artigos científicos (MARCONDES, 2011, p.83).

O periódico científico se firmou como publicação acadêmica em ambiente digital, porém continua calcado no modelo impresso, de modo a ser lido e interpretado apenas por seres humanos. Até que o conhecimento contido nos textos seja incorporado aos estoques de conhecimento humano, este passa obrigatoriamente por um longo processo social, que por sua vez compreende leitura crítica, avaliação e citação do artigo.

Para os artigos científicos e a comunicação científica em geral, a *Web* foi bem sucedida como uma plataforma de divulgação de trabalhos científicos, contudo, muita informação ainda permanece trancada em documentos distintos e não interligados entre si no ambiente digital. O formato digital dos textos dos artigos científicos tem ampliado as potencialidades de acesso e recuperação dessas “bases de conhecimento”.

De acordo com estudos realizados por (TENOPIR; KING, 2009), pesquisadores precisam ler uma quantidade cada vez maior de artigos e dispensam cada vez menos tempo a essas leituras. Na verdade, pesquisadores e cientistas utilizam diversos artigos científicos ao mesmo tempo, para ler, analisar e anotar fragmentos do seu conteúdo. Esta prática chamada por Renear e Palmer (2009, p.828) de “leitura estratégica”, é motivada pelo grande número de artigos que um pesquisador precisa ler para se manter atualizado. Esta prática vem sendo intensificada no ambiente digital valendo-se de recursos de indexação e citações como indicadores de relevância, revisões de literatura para se ligar a outros documentos, e redes sociais que funcionam como um serviço de alerta pessoal. O objetivo principal, desta estratégia, é mover-se rapidamente através da literatura buscando acessar e explorar o artigo em menor tempo possível.

Para dar suporte e intensificar a efetividade desta prática de leitura(cada vez mais presente no ambiente *Web*), se faz necessário ampliar a interação entre autores e leitores de artigos científicos digitais, com o acesso direto ao conteúdo semântico e às afirmações contidas no texto de artigos científicos.

Buscas por palavras-chave ligadas pelos operadores booleanos não dão conta da expressividade e precisão necessária para a recuperação de conteúdo semântico contido no crescente número de artigos científicos e outras fontes de informação na *Web* (MARCONDES, 2011, p.84).

As tecnologias semânticas estão sendo utilizadas pela comunidade acadêmica como uma nova ferramenta para otimizar a pesquisa acadêmico-científica e a leitura estratégica no ambiente *Web*. As tecnologias da *Web* semântica propõem avanços na questão da recuperação e processamento semântico de conteúdos nos ambientes de *e-Science*.

Embora a *Web* tenha sido projetada com o objetivo de facilitar o acesso, o intercâmbio e a recuperação de informações, Souza e Alvarenga (2004, p.133) afirmam que:

Não há nenhuma estratégia abrangente e satisfatória para a indexação dos documentos nela contidos, e a recuperação das informações, possível por meio dos motores de busca (*search engines*), é baseada primariamente em palavras-chave contidas no texto dos documentos originais, o que é muito pouco eficaz.

“Apesar da quantidade de informações recuperadas pelos mecanismos de busca ser massiva, apenas uma parte da *web* é pesquisada, enquanto uma parte considerável do conteúdo fica inacessível através dos buscadores (*Web Oculta*)²”.

Estes argumentos apontam para uma necessidade de evolução da atual estrutura da rede mundial. Isto porque a tecnologia utilizada na *Web* atual apresenta limitações para manipulação das informações disponíveis.

A este respeito, Breitman (2005, p.3) nos diz que na *Web* atual (*Web Sintática*), “[...] os computadores fazem apenas a apresentação da informação, enquanto o processo de interpretação fica a cargo dos seres humanos”, ou seja, a informação continua a ser avaliada, classificada e selecionada pelos seres humanos.

Grande parte das páginas disponíveis na *Web* são direcionadas para outras pessoas e não para serem processadas por programas de software [...] As páginas não contêm informações sobre si mesmas, ou seja, que tipo de conteúdo está descrito e a que assunto a página se refere (BREITMAN, 2005, p.2).

A proposta da *Web Semântica* é justamente incorporar sentido às informações de maneira que as máquinas possam compreender a linguagem humana, ou seja, fornecer estruturas e dar significado semântico ao conteúdo das páginas Web. Seu maior desafio é permitir que a Web passe de uma plataforma de divulgação para uma plataforma de informação. Oferecendo vantagens, principalmente na área de busca e automatização de tarefas, a Web semântica irá facilitar a obtenção, classificação e organização de informações na Web.

De acordo com Berners-Lee, Hendler e Lassila (2001), a Web Semântica

[...] não é uma Web separada, mas uma extensão da Web atual, na qual é dada à informação um significado bem definido, permitindo que computadores e pessoas trabalhem em cooperação. É uma fonte para recuperar informações a partir da web (utilizando as aranhas web a partir de arquivos RDF) e acessar os dados através de agentes da Web Semântica.

A *Web 2.0* se caracteriza por uma concepção mais descentralizada. Nesta Web, o usuário participa ativamente da criação, seleção e troca de informações, podendo não só acessar, mas também alterar o conteúdo a qualquer momento.

Blattman e Silva (2007, p.211) define a *Web 2.0* como sendo “[...] um novo espaço para acessar, organizar, gerenciar, tratar e disseminar a informação, conhecimentos e saberes” e recordam que:

Se antes a web era estruturada por meio de sites que colocavam todo o conteúdo on-line, de maneira estática, sem oferecer a possibilidade de interação aos internautas, agora é possível criar uma conexão por meio das comunidades de usuários com interesses em comum, resultado do uso da plataforma mais aberta e dinâmica (BLATMANN; SILVA, 2007, p.199).

A *Web 2.0* vem impulsionando a comunicação, com seus princípios colaborativos e participativos nas mais diversas atividades humanas, inclusive na atividade científica. Pesquisadores de diversas áreas vêm utilizando as ferramentas da *Web 2.0* com finalidades científico-acadêmicas, caracterizando a web 2.0 como um objeto de pesquisa da Ciência da Informação. Galdo (2010, p.45) destaca abaixo, algumas ferramentas da *Web 2.0* com potencial de uso científico e acadêmico:

- *Weblogs* – Ferramenta para publicação de informações, opiniões e ideias, com espaços para comentários de outros usuários da Internet. Somam o poder noticiador dos grupos de discussão às informações organizadas nas

páginas *web*. Os *weblogs* ou *blogs* são personalizados pelo autor/autores e podem conter textos, imagens, vídeos, ferramentas de busca, *links* para outros *blogs*, estatísticas de acesso, “nuvem de *tags*”, entre outros recursos.

- *Wikis* – Ambiente em que cada usuário redige e comenta um determinado termo acessível a todos os outros, que o lêem, e podem também contribuir com alterações. Os *wikis* permitem a criação coletiva de conteúdo na *web* e possuem formas de regulação da produção da coletividade. O exemplo mais conhecido é a Wikipédia. Ferramentas *wiki* têm um grande potencial para a construção colaborativa de trabalhos acadêmicos.
- *Sites* de Redes Sociais – *Site* que foca a publicização da rede social dos atores (usuários da Internet). Representam processos dinâmicos em consequência dos processos de interação entre esses atores.
- *Folksonomias* – Ferramentas de classificação, recuperação e compartilhamento da informação, na qual os usuários colaboram livremente na classificação da informação. As “nuvens de *tags*”, uma das formas de navegar pelas informações classificadas espontaneamente pelos usuários, modificam-se em tempo real, em interação constante entre os usuários e a informação, modificando também a relação de tempo entre a classificação da informação e o seu uso.
- Compartilhamento de vídeos – Usuários da Internet compartilham vídeos criados por outros ou criam seus próprios vídeos e os compartilham livremente. Alguns vídeos bem como alguns autores de vídeos têm se tornado mundialmente acessados sem que haja nenhuma motivação financeira direta. Vários tipos de vídeos têm sido produzidos, desde vídeos humorísticos a vídeos educativos. O *Youtube*, ferramenta mais popular de compartilhamento de vídeos, tem um grande número de vídeos com aulas expositivas, palestras, demonstração de experiências, entre outros de interesse acadêmico.
- Compartilhamento de apresentações/slides – Autores compartilham apresentações e slides de palestras e/ou aulas. O autor define a permissão de uso, cópia ou distribuição.

- Leitor de *Really Simple Syndication Feeds*(RSS) – Agregador de notícias, amplamente utilizado pela comunidade dos *blogs*, para compartilhar as suas últimas novidades ou textos completos e até mesmo arquivos multimídia.
- Serviços de *microblogs* – A primeira ferramenta de *microblog* e, ainda, a mais popular é o *Twitter*. Foi criado inicialmente com a pergunta: “O que você está fazendo?” para que cada usuário respondesse, compartilhando com pessoas que o acompanham, ao mesmo tempo em que receberia curtas mensagens das pessoas que optasse por acompanhar (seguidos e seguidores). Entretanto, os usuários descobriram outras funções e criaram aplicativos para serem usados em conjunto com o *Twitter*, como programas que condensam os endereços *web* em poucos caracteres. Com isso, a ferramenta passou a ser utilizada com maior frequência para compartilhar informações por meio dos *links*. A ferramenta é utilizada, principalmente, pela comunidade acadêmica.

Cabe aqui ressaltar que as tecnologias semânticas estabelecem uma estreita relação com a representação do conhecimento, por isso a Ciência da Informação investiga o atual projeto da Web Semântica objetivando contribuir não só para a o processamento semântico de informações por computadores, mas também para a organização e modelagem de domínios de conhecimento.

Este artigo apresenta algumas experiências e propostas inovadoras de publicações semânticas, que exploram no ambiente digital integrado que constitui a Web, as possibilidades abertas pelas tecnologias da Web Semântica, e Web 2.0, ampliando o potencial de compreensão e recuperação do conteúdo semântico e de interação entre autores e leitores de artigos científicos digitais.

Procurou-se identificar na literatura da área e na própria Web, projetos e experiências que têm em comum, o fato de proporem novos formatos que ultrapassassem a mera cópia digital do artigo impresso, e que inovam em termos de modelos de negócio, livre acesso, etc.

“Publicações semânticas” é uma área de pesquisa emergente, como demonstra a realização, neste ano do I SEPUBLICA, *Workshop on Semantic Publications*³. Esta Conferência propõe reunir pesquisadores e profissionais que lidam com diferentes aspectos de tecnologias semânticas na indústria editorial.

De acordo com Meadows (1999), encontros, workshops e conferências são fundamentais para o surgimento de uma nova área, daí a importância de se sistematizar estas experiências inovadoras.

Já existem algumas iniciativas (identificadas como resultados parciais ou finais deste estudo), no sentido de ampliar as potencialidades do periódico científico eletrônico com o uso das tecnologias semânticas para formalizar os textos dos artigos de periódicos eletrônicos em resumos estruturados, e textos formatados em XML.

3 PROCEDIMENTOS METODOLÓGICOS

Este estudo possui um caráter exploratório, onde projetos e experiências são apresentados e analisados qualitativamente. Para realizar este levantamento foram pesquisados os seguintes temas: publicações eletrônicas, publicações semânticas, *Web Semântica*, *Web 2.0*, processamento semântico. Até onde conseguimos identificar, não foi encontrado um artigo que identificasse e fizesse uma revisão dessas experiências.

Para desenvolvimento e investigação desta pesquisa, o método utilizado foi buscado dentro dos Métodos específicos às Ciências Sociais: comparativo, tipológico e estruturalista. Dentre estes, o método comparativo foi considerado o mais adequado, por possibilitar a exploração dos fenômenos, e identificar características comuns e diferenças nas experiências a serem comparadas.

Método comparativo – Ocupa-se da explicação de fenômenos, permitem analisar o dado concreto, deduzindo do mesmo os elementos constantes, abstratos e gerais. Constitui uma verdadeira “experimentação indireta”. É empregado em estudos quantitativos e qualitativos e pode ser utilizado em todas as fases e níveis de investigação: num estudo descritivo, pode averiguar a analogia entre ou analisar os elementos de uma estrutura; nas classificações, permite a construção de tipologias; e finalmente, em termos de explicação, pode, até certo ponto, apontar vínculos causais entre os fatores presentes e ausentes (MARCONI; LAKATOS, 2004, p.92-96).

Trata-se de uma pesquisa de natureza:

- Bibliográfica - Porque a leitura e a reflexão da literatura irão permitir focalizar com maior precisão o problema a ser investigado;

- Aplicada - pois objetiva gerar conhecimentos para a aplicação prática voltada à solução do problema;
- Exploratória - Na medida em que não se visa verificar apenas teorias, e sim, a aplicabilidade do problema com vista a torná-lo explícito.
- Qualitativa - Uma vez que o foco é a compreensão dos fatos, considerando o processo e seu significado como enfoques principais de abordagem;

A pesquisa qualitativa se ocupa nas Ciências Sociais, com um nível de realidade que não pode ou não deveria ser quantificado. Ou seja, ela trabalha com o universo dos motivos, das aspirações, das crenças, dos valores e das atitudes. Esse conjunto de fenômenos humanos é entendido aqui como parte da realidade social, pois o ser humano se distingue não só por agir, mas por pensar sobre o que faz e por interpretar suas ações dentro e a partir da realidade vivida e partilhada com seus semelhantes (MINAYO, 2007, p.21).

4 RESULTADOS PARCIAIS OU FINAIS

Como resultados parciais, foram identificados na literatura e na própria *web*, alguns projetos e experiências de periódicos semânticos, os quais são apresentados e comentados a seguir, a partir de uma ordem cronológica.

O surgimento da linguagem XML (*Extensible Markup Language*), a precursora dos formalismos de representação semântica de informações da Web Semântica, influenciou diretamente propostas pioneiras de superar o formato textual para marcação e publicação de artigos científicos na Web. Existem diferentes propostas, como a pioneira *Chemical Markup Language (CML)* (MURRAY-RUST; RZEPA, 1999), a *System Biology Markup Language (SBML)* (HUCKA; FINNEY; BOLORI, 2003), a *Mathematical Markup Language (MathML)* e também enfoques mais gerais como a *Scientific Technical and Medical Markup Language (STMML)* (MURRAY-RUST; RZEPA, 2002).

Gardin (2001) com sua proposta de escrita logicista, que consiste em numa forma de escrita que explicita as partes semânticas do raciocínio científico constantes do texto de artigos científicos, foi um dos pioneiros do tratamento semântico de artigos científicos. Este autor considera que artigos são bases de conhecimento, cujos conteúdos poderiam ser mais bem apreendidos se

formalizados através dos métodos da escrita logicista. O projeto *Arkeotek*, <http://www.arkeotek.org>, voltado para uma publicação na área de arqueologia, mostra a concretização das idéias pioneiras de Gardin, iniciadas com o sistema *SYMTOOL* ainda na década de 60.

O sistema *HyBrow*, <http://www.hybrow.org>, considerado por seus autores (RACUNAS et al., 2004) ainda como um protótipo, se propõe a criar um ambiente para formulação e teste de validade de novas hipóteses em biologia. Integra diversas fontes de informação biológicas, uma ontologia de processos biológicos em diferentes níveis de detalhe, um banco de dados e programas para formulação e avaliação de novas hipóteses com conhecimento já existente, armazenado no banco de dados. Este é criado a partir de programas que processam e analisam a literatura da área, formatando-a segundo as entidades usadas na ontologia de processos biológicos.

A proposta do *MachineProse* (DINAKARPADIN et al., 2006) tem como ponto de partida a concepção do papel essencial do teste de hipóteses na pesquisa biomédica. É um ambiente para a formulação e codificação formal de afirmações científicas, consideradas como a unidade fundamental de conhecimento. Estas afirmações, na forma de relações entre conceitos biomédicos, são baseadas numa ontologia de relações que serve de guia para a formulação consistente de hipóteses. As afirmações assim formuladas são usadas para sumarizar e anotar o conteúdo semântico da literatura biomédica, e indexar esta literatura, provendo assim meios semanticamente mais ricos para sua recuperação.

A *Semantic Web Application in Neuromedicine (SWAN)* (GAO et al., 2006), <http://swan.mindinformatics.org/>, é um projeto que se propõe a criar um ambiente Web para a comunidade que pesquisa a doença de Alzheimer, com facilidades para acesso integrado a diferentes bases de conhecimento. Integra dados pessoais, laboratoriais, de fontes externas, permitindo apoio à geração e teste de hipóteses a partir destas diferentes fontes sua utilização na produção e formatação de textos para publicação.

A *Ontology for Experiment Self-Publishing* é um trabalho coletivo proposto e coordenado pela *Scientific Publishing Task Force* (2006), do *Semantic Web Health Care and Life Science Interest Group (W3C)*, com o objetivo de formalizar todas as

etapas de um experimento científico em ciências biológicas com vistas a sua publicação como artigo científico.

O trabalho de Soldatova e King (2006) evolui diretamente da proposta anterior. Sua proposta de Ontologia para Experimentos Científicos (EXPO) –, têm também com o objetivo de formalizar e padronizar todas as etapas de um experimento científico, desde a coleta de dados, procedimentos experimentais, teste de hipóteses, etc., incluindo também, mas sem ter como único foco como a proposta anterior, a publicação final dos resultados.

O grupo editorial Elsevier, tradicional editor científico, desenvolve junto com o periódico Cell, o projeto “*The article of the future*”, no qual diferentes facilidades são agregadas aos artigos deste periódico. Artigos são apresentados segundo um novo formato de apresentação, que conta com um resumo gráfico, uma seção de “*highlights*” resumindo as principais conclusões do artigo, a possibilidade de assinalar comentários a artigos.

O sistema *Information Hyperlinked over Proteins (Ihop)*, <http://www.ihop-net.org/> cria um hipertexto digital de sentenças com menções a genes e proteínas, encontrados e extraído de resumos da base de dados Medline. Pesquisadores podem navegar por este hipertexto de sentenças, avaliando as afirmações que lhes possam interessar que contenham relações entre genes e proteínas. As sentenças são mostradas dentro do resumo de onde foram extraídas e relacionadas à referência bibliográfica do artigo correspondente, preservando assim o contexto completo da sentença. O foco diretamente nas afirmações de relações entre genes e proteínas e nas sentenças que as contenham é uma tentativa de viabilizar uma busca semântica no interior dos textos dos artigos, permitindo uma leitura direta, “estratégica” dos artigos, nas palavras de Renear e Palmer (2009).

O sistema Textpresso, <http://www.textpresso.org/> é um exemplo do uso e integração de ontologias biomédicas na formatação e recuperação de artigos científicos. O sistema realiza mineração de textos biomédicos previamente tratados e decompostos no nível das suas sentenças, cujas palavras ou expressões são marcados com termos de várias categorias de ontologias biomédicas como a GO. Os resultados da busca por uma palavra-chave são listas de sentenças ordenadas por ordem de relevância contendo os termos identificados da ontologia “iluminados” dentro de artigos, também ordenados por relevância.

Shotton *et al.* (2009) relatam a experiência de ampliarem o potencial semântico da revista *PLoS Neglected Tropical Diseases*. Entre as facilidades oferecidas aos leitores está a possibilidade de assinalar comentários a artigos, a de consultar índices de citação do artigos, o fato das referências bibliográficas serem anotadas através de uma ontologia, a CitO, onde é explicitado, através das categorias da ontologia, o motivo da citação.

O sistema *Utopia Document* (ATTWOOD *et al.*, 2009) permite a leitores navegar de forma semanticamente expressiva pelos artigos da publicação experimental *Semantic Biochemical Journal*. O sistema integra ferramentas de visualização, análise de dados e anotações/comentários do leitor, permitindo “linkar” um artigo, suas referências, tabelas e gráficos a outros recursos disponíveis na Web, como ontologia biomédicas, bancos de dados genéticos e outros recursos/serviços externos, disponíveis na Web, transformando um artigo num ponto de acesso e integração, de forma direta e transparente, a estes recursos, sem necessitar do uso de outros programas, como navegadores, programas estatísticos, etc. O leitor, através do Utopia, pode também assinalar comentários a um artigo e disponibilizá-los para outros leitores.

De Waard e colegas (2009), no projeto *Hypotheses, Evidence and Relationships (HypER)*, propõe, como vários outros projetos, a visão do conhecimento biomédico contido em artigos científicos poder ser representado como triplas de entidade-relacionamento-entidade. Os autores, no entanto, vão além desta representação semântica, ao proporem agregar a ela todo o contexto onde tal afirmação foi feita, em especial, as intenções do autor e os mecanismos retóricos usados por este para convencer seus leitores das afirmações feitas ao longo do artigo; assim, papel privilegiado é dado à hipótese formulada no artigo, às evidências de sua validade apresentadas pelo autor e às hipóteses de outros na quais o autor se baseia. Marcondes (2011) propõe um modelo semântico de publicações científicas digitais, no qual as conclusões, como parte privilegiada do conteúdo semântico do artigo, têm um papel essencial de sintetizarem este conteúdo. A proposta, para a qual existe um protótipo em funcionamento, tira partido do momento em que autores submetem seus artigos para publicações eletrônicas, para solicitar a estes que entrem com a(s) conclusão(ões) do artigo; o texto das conclusões é então processado linguisticamente, formatada segundo uma relação,

codificado em RDF e agregado aos metadados bibliográficos tradicionais, ampliando assim o potencial de recuperação semântica do artigo.

5 CONSIDERAÇÕES FINAIS

O periódico científico se firmou como publicação acadêmica em ambiente digital, porém continua calcado no modelo textual, de modo a ser lido e interpretado apenas por seres humanos, o que impossibilita o processamento de seu conteúdo por programas. Os recentes avanços em tecnologias de informação e comunicação mudaram a nossa capacidade de reproduzir, distribuir, controlar e publicar informações. A proposta da *Web Semântica* é incorporar sentido às informações de maneira que as máquinas possam compreender a linguagem humana, ou seja, fornecer estruturas e dar significado ao conteúdo das páginas *Web*. A *Web 2.0* além de facilitar a comunicação interpessoal, e compartilhar informações, vem se destacando pela colaboração científica. As plataformas *Web* vêm sendo desenvolvidas para a comunidade científica e os periódicos científicos tradicionais passam a adotar as ferramentas colaborativas como os blogs em seus websites. Este trabalho discute projetos, experiências e propostas de publicações semânticas, bem como os novos formatos de artigos científicos digitais, tomando como base as possibilidades abertas pelas tecnologias de informação e comunicação, em especial pela *Web Semântica* e *Web 2.0*, contribuindo para não só para o processamento semântico de informações por computadores, mas também para a organização e modelagem de domínios de conhecimento.

REFERÊNCIAS

BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. The semantic web: a new form of web content that meaningful. **Scientific American**, v.284, n.5, p.34-43, 2001. Disponível em: <<http://www.scientificamerican.com/2001/0501issue/0501berners-lee.htm>>. Acesso em: 24 nov. 2010.

BLATTMANN, U.; SILVA, F. C. C. da. Colaboração e interação na Web 2.0 e Biblioteca 2.0. **Revista ABC: Biblioteconomia em Santa Catarina, Florianópolis**, v.12, n.2, p.191-215, jul./dez. 2007.

BRANSKI, R. M. Recuperação de informações na Web. **Perspectivas em Ciência da Informação**, Belo Horizonte, v.9, n.1, p.70-87, jan./jun. 2004. Disponível em: <<http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/351/160>>. Acesso em: 24 nov. 2010.

- BREITMAN, K. K.. **Web semântica: a internet do futuro**. Rio de Janeiro: LTC, 2005.
- COSTA, S. **Impactos de uma “filosofia aberta” na comunicação científica hoje: as mudanças que o acesso aberto/livre (open access) provocam**. In: Palestra proferida na 57ª Reunião Anual da SBPC, julho de 2005. Disponível em: <<http://reposcom.portcom.intercom.org.br/handle/1904/18469>>. Acesso em: 16 maio 2006.
- DIAS, T. D.; SANTOS, N. **Web semântica: conceitos e tecnologias associadas**. Disponível em: <www.ime.uerj.br/cadernos/cadinf/vol23/cadime-inf-vol23-completo.pdf>. Acesso em: 27 out. 2010.
- DINAKARPADIAN, D.; LEE, Y.; VISHWANATH, K.; LINGAMBHOTLA, R. Machine prose: An ontological framework for scientific assertions. **Journal of the American Medical Informatics Association**, v.13, n.2, p.220-232, Mar./Apr. 2006. DOI 10.1197/jamia.M1910.
- FACHIN, G. R. B. **Modelo de avaliação para periódicos científicos on-line: proposta de indicadores bibliográficos e telemáticos**. Florianópolis: UFSC, 2002. 210f. Dissertação (Mestrado) – Programa de Pós-Graduação em Engenharia de Produção – UFSC.
- GALDO, A. M. R. **Web 2.0 e colaboração científica: análise do uso científico-por docentes de Pós-Graduação Stricto Sensu em Ciência da Informação no Brasil**. Florianópolis: UFSC, 2010. Dissertação (Mestrado) – Programa de Pós-Graduação em Ciência da Informação – UFSC.
- GAO, Y. *et al.* SWAM: a distributed knowledge infrastructure for Alzheimer disease research, **Journal of Web Semantics**, v.4, n.3, 2006. Disponível em: <<http://www.websemanticsjournal.org/ps/pub/2006-17>>. Acesso em: 12 dez. 2010.
- GARDIN, J. V. Vers un remodelage des publication savants: ses rapports avec science de l'informatio. In: CHAUDRION, ?; FLUHR, ? (Eds.). **Filtrage et résumé automatique de l'information sur les reseaux** : actes du 3ème Colloque du Chapitre Français de l' ISKO, 2001.
- HUCKA, M.; FINNEY, A.; BOLORI, H. System Biology Markup Language (SBML) Level 1: structures and facilities for basic model definitions. 2003. Disponível em: <<http://www.sbml.org/specifications/sbml-level-1/version-2/sbml-level-1-v2.pdf>>. Acesso em: 02 nov. 2005.
- MARCONDES, C. H. Um modelo semântico de publicações eletrônicas. **Liinc em Revista**, Rio de Janeiro, v.7, n.1, p.82-103, mar., 2011. Disponível em: <<http://www.ibict.br/liinc>>. Acesso em: 05 maio 2011.
- _____; CAMPOS, M. L. de A. Ontologia e web semântica: o espaço da pesquisa em Ciência da Informação. **Ponto de Acesso**, Salvador, v.2, n.1, p.107-136, jun./jul., 2008. Disponível em: <<http://www.pontodeacesso.ici.ufba.br>>. Acesso em: 15 out. 2010.
- MARCONI, M. de A.; LAKATOS, E. M. **Fundamentos de metodologia científica**. 6.ed. São Paulo: Atlas, 2005. 315p.
- MEADOWS, A. J. **A comunicação científica**. Brasília: Briquet de Lemos, 1999.
- MEDEIROS, G. M. de; FACHIN, G. R. B.; RADOS, G. J. V. Padronização de periódicos científicos on-line da área de Biblioteconomia e Ciência da Informação: adequação as normas ISSO. **Revista ACB: Biblioteconomia em Santa Catarina**, Florianópolis, v.13, n.2, p.415-438, jul./dez.2008. Disponível em: <<http://revista.acbsc.org.br/index.php/racb/article/viewarticle/539>>. Acesso em: 09 fev. 2011.
- MINAYO, M. C. de S. (Org.). **Pesquisa social: teoria, métodos e criatividade**. 25.ed. Petrópolis:Vozes, 1993. 108p.

MURRAY-RUST, P.; RZEPA, H. S. STXML. A markup language for scientific, technical and medical publishing. **Data Science Journal**, v.1, n.2, p.128-193, 2002. Disponível em: <http://journals.eecs.qub.ac.uk/codata/journal/contents/1_2/1_2pdfs/ds121.pdf>. Acesso em: 18 set. 2005.

MURRAY-RUST, P.; RZEPA, H. S. Chemical markup, XML and the worldwide web. I: basic principles. **Journal of Chemical Information and Computer Science**, v.39, p.928-942, 1999.

ORGANIZAÇÃO MUNDIAL DA SAÚDE; ORGANIZAÇÃO PANAMERICANA DE SAÚDE; BIREME. Centro Latino-Americano e do Caribe de Informação em Ciências da Saúde. Biblioteca Virtual em Saúde. *Blogs se afirmam como meios de comunicação científica. Newsletter BVS 087 27*, São Paulo, fev. 2009. Disponível em: <<http://espacio.bvsalud.org/boletim.php?newsletter=20090227&newsLang=pt&newsName=Newsletter%20BVS%20087%2027/fevereiro/2009&articleId=02171933200950>>. Acesso em: 09 nov. 2010.

PALAZZI, D. C. **QDAontology**: abordagem para o desenvolvimento de ontologias em e-Science: um estudo de caso em Biologia. Dissertação (Mestrado) Programa de Pós-Graduação em Ciência da Computação – Universidade Federal de Juiz de Fora. Disponível em: <http://www.ufjf.br/mmc/files/2010/06/dissertacao_daniele_palazzi.pdf>. Acesso em: 12 fev. 2010.

PUBLIC LIBRARY OF SCIENCE. **Definition of open access**. Disponível em: <<http://www.plos.org/about/openaccess.html>>. Acesso em: 3 jan. 2011.

RACUNAS, S. A.; SHAH, N. H.; I. ALBERT, I; FEDOROV, N. V. HyBrow: a prototype system for computer-aided hypothesis evaluation. **Bioinformatics**, v.20, n.1, p.257-264, 2004.

RENEAR, A. H.; PALMER, C. L. Strategic reading, ontologies, and the future of scientific publishing. **Science**, n.325, p.828-832, 14 Aug. 2009.

SCHONS, C. H.. O volume de informações na internet e sua desorganização: reflexões e perspectivas. **Informação & Informação**, Londrina, v.12, n.1, jan./jun. 2007. Disponível em: <<http://www.uel.br/revistas/uel/index.php/informacao/article/viewFile/1748/1497>>. Acesso em: 17 out. 2010.

SOUZA, R. R.; ALVARENGA, L. A web semântica e suas contribuições para a Ciência da informação. **Ciência da Informação**, Brasília, v.33, n.1, p.132-141, jan./abr.2004.

SILVA, J. W. de M. S. **Recuperação de informação através de recursos da web semântica**: modelagem e implementação. João Pessoa: CEFETPB-GEET-CST, 2007. 33p. (Relatório Técnico)

SOLDATOVA, L. D; KING, R. D. An ontology of scientific experiments. **Journal of the Royal Society Interface**, v.3, n.11, p.795-803, 2006. Disponível em: <<http://journals.royalsociety.org/content/u552845783800t73/fulltext.pdf>>. Acesso em: 5 fev. 2008.

TENOPIR, C.; HITCHCOCK, B.; PILLOW, A. **Use and users of electronic library resources**: an overview and analysis of recent research studies. Washington, DC: Council on Library and Information Resources, 2003.

_____; KING, D. W. Electronic journals and changes in scholarly article seeking and reading patterns. **Aslib Proceedings: New Information Perspectives**, v.61, n.1, p.5-32, 2009. Disponível em: <<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.156.2701&rep=rep1&type=pdf>>. Acesso em: 06 jun. 2011.

WORKSHOP SEPUBLICA SEMANTIC PUBLICATION,1., 2011, Grécia. Disponível em: <<http://sepublica.mywipaper.org>>. Acesso em: 12 abr. 2011.

NOTAS

- ¹ O termo *e-Science* geralmente é empregado para descrever o desenvolvimento de infra-estruturas de serviços de software capazes de prover acesso a facilidades remotas, recursos computacionais distribuídos, armazenamentode informações em bancos de dados dedicados, disseminação e compartilhamento de dados, resultados e conhecimento. (PALAZZI, 2010, p.1)
- ² Web oculta é uma parte importante da Web na qual os mecanismos de busca tradicionais não podem ou não querem incluir em seus bancos de dados.Sendo assim, estes sites não aparecem nos resultados apresentados por estas ferramentas de busca. Estima-se que esta parte oculta da Web tenha mais que o dobro do tamanho da parte visível e seu conteúdo é bastante relevante. Há, basicamente, duas razões para estes sites estarem fora dos bancos de dados de grande parte dos buscadores: questões técnicas que impedem o acesso dos spiders a alguns tipos de sites [...] decisão dos administradores dos mecanismos de busca. (BRANSKI, 2004, p.82)
- ³ Disponível em: <<http://sepublica.mywikipaper.org/drupal>>.