A Evolução da Comunicação e Interação Humana: da Web à Web Semântica

Wilson Castello Branco Neto[1], Fernando Álvaro Ostuni Gauthier[2]

[1]Depto. de Ciências Exatas e Tecnológicas - Universidade do Planalto Catarinense.
Av. Castelo Branco, 170 - 88509-900 - Lages – SC - Brasil

[2]Departamento de Informática e de Estatística – Universidade Federal de Santa Catarina.
Campus Universitário Trindade – Florianópolis-SC-Brasil

castello@uniplac.net, gauthier@inf.ufsc.br

Resumo. *Este artigo apresenta conceitos relacionados à Web Semântica, abordando as mudanças que ela pode causar em alguns seguimentos após seu completo desenvolvimento. Para fundamentar esta discussão, inicialmente descreve-se como a Web revolucionou áreas como Comércio e Negócio Eletrônico, Ciência Eletrônica, Gestão do Conhecimento e Educação a Distância, para então, na seqüência, apresentar os benefícios que a Web Semântica ainda pode trazer para estas mesmas áreas. O artigo não pretende expor nenhuma ferramenta ou tecnologia desenvolvida, mas sim disseminar o potencial uso da Web Semântica na comunidade científica.*

1. Introdução

A aplicação mais conhecida da Internet é a *World Wide Web* (WWW), ou simplesmente *Web*, nome pelo qual ficou conhecido o serviço de publicação de conteúdo através da criação de hipertextos. Este serviço tem sua força baseada em algumas características que lhe são intrínsecas, como a facilidade para qualquer pessoa publicar informações e associá-las a outros conteúdos já existentes através de *hiperlinks.* Isto é possível já que a Internet possui uma estrutura distribuída, não existindo um órgão regulador responsável por catalogar as informações nela disponibilizadas.

Todavia, esta liberdade tem seu preço. Com a ausência de controle sobre o desenvolvimento, publicação e vinculação de *sites*, é muito difícil encontrar alguma informação neste emaranhado sem o auxílio de um mecanismo automatizado de busca. Mesmo com o auxílio destes mecanismos, é notória a dificuldade para encontrar informações úteis, tendo em vista a quantidade de resultados gerados por eles, sendo grande parte irrelevantes.

O problema supracitado deve-se ao fato da *Web* que se conhece atualmente ter sido desenvolvida para ser utilizada apenas por humanos. Como os documentos e serviços por ela disponibilizados são escritos em linguagem natural, é necessária inteligência em um grau que apenas os humanos possuem, para entender e utilizar completamente seus recursos (DACONTA, OBRST e SMITH, 2003).

Web Semântica é o nome dado a um projeto sob coordenação do W3C (*World Wide Web Consortium*), que, segundo Berners-Lee, Hendler e Lassila (2001), disponibiliza "uma nova forma de conteúdo *Web* que possui significado para os

computadores". Ela busca meios para representar o conhecimento humano de uma forma que ele possa ser facilmente entendido e manipulado pelo computador. Além de facilitar a busca de informações, a *Web* Semântica permitirá que os computadores manipulem informações de forma autônoma, expandindo as possibilidades de aplicações da *Web* atual.

Este artigo apresenta de que maneira a *Web* Semântica pode contribuir para melhorar os serviços existentes atualmente na *Web*, através de exemplos nas áreas de Comércio e Negócio Eletrônico, Ciência Eletrônica, Gestão do Conhecimento e Educação a Distância. Para isto, apresenta na seção 2 um breve histórico sobre *Web* e *Web* Semântica. A seção 3 descreve resumidamente a estrutura da *Web* Semântica, enquanto a seção 4 expõe algumas considerações sobre os benefícios que a *Web* trouxe para as áreas previamente citadas, juntamente com conjecturas sobre as novas mudanças que a *Web* Semântica pode causar. Por fim, na seção 5 são tecidas as considerações finais.

2. Web: O Sonho e a Realidade

Conforme descrito em Berners-Lee (1990), a proposta inicial do idealizador da *Web* visava criar uma forma para armazenar informações sobre grandes projetos no CERN (*Nuclear Research European Center*), evitando assim a perda e facilitando o acesso às mesmas. Alguns requisitos deste projeto, tais como o acesso remoto através de redes, heterogeneidade, não centralização, acesso a dados já existentes, possibilidade de criar *links* particulares e analisar dados, definiram muitas das características da *Web* atual.

Para Fensel *et. al.* (2003), "A *Web* que se vê hoje, como um glorificado canal de televisão é apenas uma parte do sonho original". Embora ela tenha sido concebida pela necessidade de organização do trabalho em grandes grupos, foi fora do ambiente de trabalho, mais especificamente no ramo da publicação de informações, que a *Web* teve seu maior crescimento. A segunda parte deste sonho, que é ver a *Web* sendo usada tão genericamente que ela se torne um espelho real de como os humanos trabalham, se divertem e se relacionam, ainda não aconteceu, mas existem sinais e planos que deixam os pesquisadores confiantes.

O primeiro passo em direção a esta nova *Web* consiste na criação de informações sobre informações. Será necessária, também, uma mudança de paradigma quando se considera o dado como secundário em relação ao processamento. É importante que sejam desenvolvidos novos padrões para o armazenamento de dados que contenham algum nível de inteligência, retirando um pouco da responsabilidade dos programas (processamento).

Esta evolução da importância e *inteligência* dos dados pode ser observada através de uma rápida retrospectiva feita por Daconta, Obrst e Smith (2003) sobre a forma de armazenamento de dados. Nos estágios iniciais da computação os dados eram proprietários das aplicações que os manipulavam, sendo armazenados em arquivos (texto ou binário) e após em bases de dados.

A importância começou a migrar da aplicação para os dados após a criação da linguagem XML, a qual tornou os dados independentes das aplicações dentro de um domínio específico. Em um terceiro estágio desta evolução encontram-se as taxionomias, que permitem a composição de dados de diferentes domínios. Ao final

desta linha estão as ontologias, a partir das quais novos dados podem ser inferidos a partir dos existentes, seguindo regras lógicas, gerando os *dados inteligentes.*

É através da criação dos *dados inteligentes* que se pretende expandir as possibilidades da *Web* e trazê-la ao seu completo potencial de aplicação, criando para isto linguagens capazes de gerar documentos para uma *Web* projetada para ser processada por máquinas, além de humanos. A *Web* de documentos para humanos está sendo incorporada a uma *Web* de dados entendíveis pelas máquinas e o potencial do trabalho cooperativo entre humanos e máquinas trabalhando juntos e se comunicando através da *Web* pode ser imenso (BERNERS-LEE, 2004).

3. Web Semântica: Conceitos e Estrutura

"A *Web* Semântica é uma extensão da *Web* atual, na qual é dado um significado bem definido à Informação, permitindo que os computadores e pessoas trabalhem em cooperação." (HENDLER, BERNERS-LEE e MILLER, 2002). Ela fornece uma infra-estrutura que habilita, não apenas *homepages*, mas também bases de dados, serviços, programas, sensores, dispositivos pessoais e até eletrodomésticos a consumir e a produzir informações para a *Web*. Agentes de software poderão usar estas informações para buscar, filtrar e preparar informações de novas formas para ajudar as pessoas na execução das mais variadas tarefas (HENDLER, BERNERS-LEE e MILLER, 2002).

A *Web* Semântica foi estruturada em níveis com diferentes funções (figura 1).

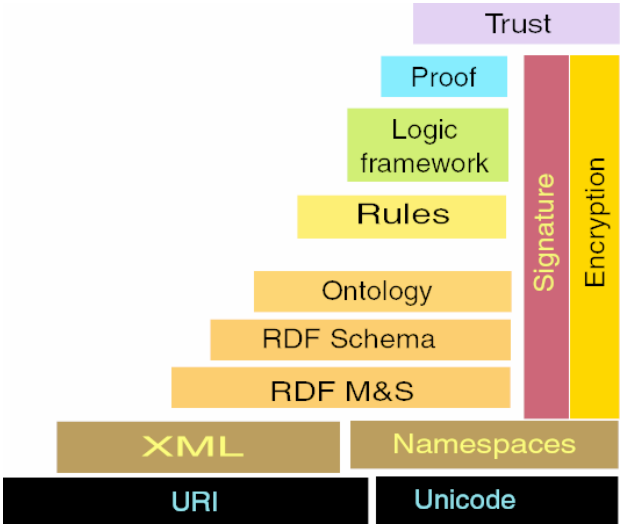

Figura 1: Níveis da Web Semântica

Enquanto os níveis inferiores representam a estrutura básica para o transporte de dados, os superiores contemplam definições semânticas que possibilitam o processamento automatizado dos mesmos pelo computador.

As camadas inferiores (URI e UNICODE) são responsáveis por possibilitar, respectivamente, a identificação de forma unívoca dos recursos e padronizar os códigos para representar caracteres de forma independente de plataforma.

Em um segundo nível estão a linguagem XML, que possibilita a associação de descrições semânticas simples as conteúdos, e os *namespaces*, que se constituem em um mecanismo para criar nomes únicos para elementos e atributos de documentos XML.

No terceiro nível encontra-se a linguagem RDF, que possibilita a criação de descrições semânticas mais elaboradas que a linguagem XML, representadas por sentenças de três elementos: recurso, propriedade e valor (KLYNE e CARROL, 2004).

O RDF é uma linguagem construída sobre o XML (BECKETT, 2004) que é utilizada para fazer afirmações sobre entidades e documentos como um todo. Enquanto o XML serve para adicionar metadados a partes de um documento, com o RDF pode-se associar metadados ao documento como uma entidade única (DACONTA, OBRST e SMITH, 2003).

Ao analisar a colocação de Fensel *et. al.* (2003), onde é explicitado que o desenvolvimento da *Web* Semântica requer:

- Linguagens para expressar meta-informação sobre documentos que seja processável pelo computador;
- Terminologias padrões sobre os mais diversos domínios;
- Ferramentas e arquiteturas que usem tais linguagens e terminologias para fornecer suporte a busca, acesso, apresentação e manutenção de informações;

cabe destacar que as linguagens XML e RDF possibilitam a associação de meta-informação aos documentos, satisfazendo, desta forma, o primeiro requisito. Entretanto, embora elas possibilitem a descrição semântica dos conteúdos disponíveis na *Web*, não são capazes de garantir que diferentes pessoas ou organizações utilizem os mesmos termos para referenciar entidades diferentes ou utilizem termos diferentes para referenciar uma mesma entidade. Para resolver este problema é necessária a criação de esquemas que padronizem os termos a serem empregados por uma comunidade, de forma a satisfazer o segundo requisito listado por Fensel , *et al*. (2003). Tanto o Esquema RDF (RDFS), quanto as ontologias são utilizadas para esta tarefa, embora as últimas tenham um poder de representatividade maior.

Segundo Gruber (1993), ontologia é uma especificação de uma conceitualização e por conceitualização entende-se uma visão simplificada e abstrata do mundo que se deseja representar para algum propósito. Embora ainda existam dúvidas sobre o que é uma ontologia na engenharia do conhecimento, os seus objetivos nesta área já estão consolidados como sendo uma forma de possibilitar a formalização, reutilização e o compartilhamento do conhecimento.

Entre outras linguagens para construção de ontologias destaca-se a OWL, por ser a linguagem recomendada pelo W3C para o desenvolvimento da *Web* Semântica (SMITH, WELTY e McGUINNESS, 2004). A OWL deve ser utilizada quando a informação contida em documentos precisa ser processada por aplicações, em contraposição às situações onde o conteúdo apenas deve ser mostrado aos humanos. A OWL pode ser utilizada para representar explicitamente o conteúdo e a semântica de termos em vocabulários e os relacionamentos entre estes termos.

Acima das ontologias estão as camadas que definem as regras lógicas e os mecanismos de prova e confiança requeridos pela *Web* Semântica. Entretanto, ainda são poucos os resultados consolidados nestas camadas, sendo grande parte das aplicações desenvolvidas apenas com as tecnologias das camadas inferiores.

4. Aplicações da *Web* Semântica

Tomando alguns exemplos bem conhecidos, é possível entender como velhos serviços foram completamente remodelados com base nos recursos fornecidos pela *Web*. Também, com estes mesmos exemplos, pode-se visualizar como eles ainda podem ser consideravelmente melhorados, com os recursos da *Web* Semântica.

4.1 Comércio Eletrônico

O relacionamento entre clientes e fornecedores é, sem sombra de dúvida, um dos ramos mais influenciados pelo surgimento da *Web*. Os termos comércio eletrônico (*e-commerce*) e negócio eletrônico (*e-business*) são conhecidos, praticamente, por todas as pessoas que têm acesso a um computador.

Entretanto, apesar dos expressivos resultados já alcançados, muitos pesquisadores ainda apontam o comércio eletrônico como um ramo que está longe de atingir todo seu potencial. Segundo Fensel *et. al.* (2003) a automatização do comércio eletrônico é seriamente prejudicada pelo modo com que o conteúdo é atualmente apresentado. O fato dos agentes responsáveis por buscar informações e realizar negócios necessitarem de heurísticas e *crawlers* (Componente de um mecanismo de busca que percorre a *Web* armazenando URLs e indexando palavras e textos das páginas encontradas) para extrair informações de documentos textuais fracamente estruturados é um dos grandes obstáculos a ser superado. Embora a área de processamento de linguagem natural tenha evoluído, nem sempre a informação obtida é suficiente para tomar uma decisão correta.

De acordo com Euzenat (2001) a *Web* Semântica é capaz de solucionar grande parte destes problemas, criando padrões para descrição de produtos, serviços e outros tipos de documentos. Mais importante que a criação destes padrões é a utilização de ontologias para fazer os mapeamentos entre eles, relacionando diferentes terminologias e possibilitando o desenvolvimento de um comércio eletrônico mais flexível.

4.2 Ciência Eletrônica e *Grids* de Conhecimento

Da mesma forma que o comércio eletrônico e o negócio eletrônico refletem o processo de globalização e as alterações ocorridas na forma de efetuar transações comerciais, o termo *e-science* (Ciência Eletrônica) denota as mudanças ocorridas na forma de se fazer pesquisa. Segundo Euzenat (2001), *e-science* indica o uso de recursos eletrônicos, como instrumentos, sensores, bases de dados, computadores, por cientistas que trabalham cooperativamente e de forma distribuída em grandes projetos.

Atualmente, a *e-science* tem como seu maior expoente o projeto do genoma humano. Devido à imensa quantidade de informação manipulada, nem sempre é possível ter em um mesmo laboratório todas as bases de dados necessárias ou os supercomputadores capazes de processar estes dados. Devido a isto, pesquisadores de todas as partes do mundo estão se unindo e compartilhando os dados e os recursos necessários para a execução de um projeto comum.

É denominado *grid* de conhecimento, a infra-estrutura que faz com que a *e-science* seja possível. Sua fusão com a *Web* Semântica é plenamente possível, através da integração de recursos como *Web* Services, RDF e ontologias aos *grids* existentes. Para Newhouse *et. al.* (2002) “O desenvolvimento destes serviços orientados a ontologias

permitirão que serviços compatíveis construam de forma autônoma os grandes e complexos ambientes computacionais distribuídos que constituirão o *grid*".

4.3 Gestão do Conhecimento

A importância da gestão do conhecimento, atualmente, para o sucesso de qualquer empreendimento é um consenso entre os estudiosos da área, devido à grande quantidade de informação disponível e aos avanços das tecnologias de informação e comunicação.

É primordial que uma pessoa em seu ambiente de trabalho saiba tirar proveito dos benefícios das ferramentas automatizadas que visam agilizar o processo de geração, armazenamento, busca, seleção e manutenção de informações durante o processo de tomada de decisão e gerência de uma organização.

Apesar destas tarefas terem sido facilitadas com o advento da *Web* e sua estrutura distribuída, muitos aspectos são ainda falhos e fazem com que os usuários percam um tempo considerável na execução de tarefas que poderiam ser feitas automaticamente, caso a forma com que as informações são organizadas e armazenadas fosse melhorada. De acordo com Fensel *et. al.* (2001) as ferramentas existentes para gestão do conhecimento no mercado apresentam as seguintes fraquezas:

- Busca de Informação: os algoritmos baseados em palavras-chave utilizados atualmente podem trazer informações irrelevantes ou ignorar alguma informação útil, devido aos termos sinônimos e homônimos;
- Extração de Informação: este processo requer a participação de um humano, pois os agentes automatizados não possuem o conhecimento necessário para extrair informações de documentos textuais e para relacioná-las com informações obtidas de outras fontes.
- Manutenção de Informações: manter as fontes de dados atualizadas, quando elas se tornam muito grandes, requer um grande esforço e tempo disponível devido à falta de estruturação das mesmas;
- Geração automática de documentos: considerando a forma pela qual os dados são organizados e armazenados atualmente, não é possível criar um sistema capaz de processar estes dados e gerar informações automaticamente.

Estes e outros problemas relacionados à gestão do conhecimento podem ser mais facilmente tratados se os dados forem armazenados de forma estruturada e com seu significado semântico associado a ele. As tecnologias utilizadas para o desenvolvimento da *Web* Semântica, como metadados e ontologias serão fundamentais para aprimorar este tipo de aplicação.

4.4 Educação a Distância

São inúmeros e bem conhecidos os benefícios que a *Web* trouxe para a Educação a Distância (EAD). Como descrito por Nagy (2004), "Estas novas tecnologias trouxeram novas oportunidades para o aluno não tradicional, assim como para as instituições de ensino tradicionais". Todavia, também são inúmeros os problemas enfrentados, tanto pelos alunos participantes, quanto pelas pessoas que oferecem este tipo de curso.

Do ponto de vista do aluno, pode-se destacar a dificuldade que o mesmo tem para encontrar um curso com o conteúdo adequado às suas necessidades. Além, ou em decorrência disto, os alunos sentem-se desmotivados para continuar seus estudos devido à forma com que a informação é apresentada e as atividades são propostas e desenvolvidas.

Analisando sob a ótica de quem oferece o serviço, encontra-se a reutilização de conteúdos e a vinculação com outros sistemas de EAD como as maiores dificuldades enfrentadas.

Segundo Nilsson (2001), estes e outros problemas existentes na EAD serão resolvidos, ou pelo menos amenizados, com o desenvolvimento da *Web* Semântica. Ele ainda destaca que a *Web* Semântica permitirá a construção de agentes inteligentes para auxiliar os alunos a encontrarem e utilizarem recursos de aprendizagem globalmente distribuídos e tornará possível a criação de cursos à distância de forma distribuída e colaborativa, a partir da reutilização de objetos de aprendizagem já existentes, devido as descrições semânticas associadas a eles.

5. Considerações

Através destes exemplos, pode-se observar que está ocorrendo outra grande transformação relacionada à forma com que os serviços são acessados e executados. No século XXI, a *Web* Semântica mudará a maneira pelas quais muitas tarefas são realizadas, da mesma forma que a *Web* fez com que as pessoas revissem seus papéis dentro das organizações e estas dentro da sociedade como um todo no final do século passado.

Toda a discussão previamente apresentada sobre os benefícios da *Web* Semântica e suas aplicações parece um tanto distante para aqueles que ainda não tomaram conhecimento dos projetos que estão em desenvolvimento, ou pelo menos que se pretende desenvolver, para que a *Web* Semântica torne-se uma realidade.

O fato de nenhuma grande aplicação com retorno financeiro significativo ter sido desenvolvida também contribui para que a *Web* Semântica ainda seja vista apenas como um projeto de laboratório, com pouca, ou até mesmo sem aplicação prática. Mas, certamente, as pessoas que pensam desta forma ainda não conhecem os planos de trabalho coordenados pelo W3C, que definem a *Web* Semântica como um projeto de longo prazo que está sendo desenvolvido em partes.

Um grande passo para a consolidação da *Web* Semântica já está sendo dado, com as diversas iniciativas que visam a elaboração de padrões de metadados nos mais variados domínios do conhecimento. Após a consolidação dos mesmos e a conseqüente representação através de ontologias, será possível descrever os conteúdos disponíveis na *Web* através de uma terminologia padrão, por meio das linguagens XML e RDF apresentadas, para que se alcance os benefícios descritos na seção 4.

Referências

BECKETT, D. (2004). “RDF/XML Syntax Specification (Revised)”. http://www.w3.org/TR/rdf-syntax-grammar.

BERNERS-LEE, T. (2004) “The World Wide Web: A Very Short Personal History”.

http://www.w3.org/People/Berners-Lee/ShortHistory.html, Maio.

BERNERS-LEE, T. (1990) “Information Management: A Proposal”. http://www.w3.org/History/1989/proposal.html.

BERNERS-LEE, T.; HENDLER, J;. e LASSILA, O. (2001). “The Semantic Web”**.** Scientific American.

DACONTA, M. C.; OBRST, L. J.; SMITH, K. T. (2003) The Semantic Web**.** A Guide to the Future of XML, Web Services, and Knowledge Management. Wiley Publishing.

EUZENAT, J.(Editor). (2001). “Research Challenges and Perspectives of the Semantic Web”. Report of the European Commision and US National Science Foundation Strategic Workshop. Sophia Antipolis (França).

FENSEL, D. *et al*. (2003). “Spinning the Semantic Web: Bringing the World Wide Web to its Full Potential. Cambridge”. MIT Press.

FENSEL, D. *et al*. (2001). “On-To-Knowledge: Semantic Web Enabled Knowledge Management”. http://www.ontoknowledge.org.

GRUBER, T. R. (1993). “A Translation Approach to Portable Ontologies”. In: Knowledge Acquisition. v. 5, n. 2. pages 199 – 220.

HENDLER, J.; BERNERS-LEE, T.; MILLER, E. (2002). “Integrating Applications on the Semantic Web”**.** In: Journal of the Institute of Electrical Engineers of Japan. v. 122 n.10, pages 676 - 680.

KLYNE, G., CARROL J.. (2004). “Resource Description Framework (RDF): Concepts and Abstract Syntax”. http://www.w3.org/TR/rdf-concepts.

NAGY, A. (2004). “e-Learning”. E-Content Report 6. ACTeN – Antecipating Content Technology Need. http://www.acten.net

NEWHOUSE, S. et. al. (2002). “Laying the Foundations for the Semantic Grid”. http://www-icpc.doc.ic.ac.uk/components/papers/AISB2002.pdf.

NILSSON, M. (2001). “The Semantic Web: How RDF will change learning technology standards”. http://www.cetis.ac.uk/content/20010927172953/printarticle.

SMITH, M. K.; WELTY, C.; McGUINNESS, D. L. (2004) “OWL Web Ontology Language Guide”. http://www.w3.org/TR/owl-guide/.