

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ – UTFPR
CURSO SUPERIOR DE TECNOLOGIA EM ANÁLISE E DESENVOLVIMENTO DE
SISTEMAS

GABRIELLE KARINE CANALLE

DESAFIOS E PERSPECTIVAS DA WEB SEMÂNTICA

TRABALHO DE DIPLOMAÇÃO

MEDIANEIRA

2013

GABRIELLE KARINE CANALLE

DESAFIOS E PERSPECTIVAS DA WEB SEMÂNTICA

Trabalho de Diplomação apresentado à disciplina de Trabalho de Diplomação, do Curso Superior de Tecnologia em Análise e Desenvolvimento de Sistemas – COADS – da Universidade Tecnológica Federal do Paraná – UTFPR, como requisito parcial para obtenção do título de Tecnólogo.

Orientador: Prof. Dr. Everton Coimbra de Araújo

MEDIANEIRA

2013



TERMO DE APROVAÇÃO

Desafios e Perspectivas da Web Semântica

Por

Gabrielle Karine Canalle

Este Trabalho de Diplomação (TD) foi apresentado às 15:50 h do dia 20 de agosto de 2013 como requisito parcial para a obtenção do título de Tecnólogo no Curso Superior de Tecnologia em Análise e Desenvolvimento de Sistemas, da Universidade Tecnológica Federal do Paraná, *Campus* Medianeira. Os acadêmicos foram argüidos pela Banca Examinadora composta pelos professores abaixo assinados. Após deliberação, a Banca Examinadora considerou o trabalho aprovado com louvor e mérito.

Prof. Dr. Everton Coimbra de Araújo
UTFPR – *Campus* Medianeira
(Orientador)

Prof. Me. Juliano Rodrigo Lamb
UTFPR – *Campus* Medianeira
(Convidado)

Prof. Me. Nelson Miguel Betzek
UTFPR – *Campus* Medianeira
(Convidado)

Prof. Me. Juliano Rodrigo Lamb
UTFPR – *Campus* Medianeira
(Responsável pelas atividades de TCC)

À meus pais que são a razão pela qual tento me superar a cada dia.
Palavras ou ações nunca serão o suficiente para agradecê-los.

AGRADECIMENTOS

Agradeço em primeiro lugar a Deus pela vida. Em seguida, agradeço aos meus pais e familiares que sempre me incentivaram e apoiaram quando necessitei.

Agradeço a minha prima Lucirene, por estar ao meu lado nos piores e melhores momentos da minha vida, e ter me ensinado a diferença entre o “querer” e o “merecer”.

Também agradeço ao meu orientador, professor Everton Coimbra, por acreditar em mim, e sempre me passar sua auto-confiança.

Não poderia deixar de agradecer a todos os professores que participaram do meu aprendizado no curso.

Por último, agradeço aos amigos conquistados durante a graduação, com os quais compartilhei tantos momentos importantes.

“A web não está concluída, é apenas a ponta do iceberg. As novas mudanças irão balançar o mundo ainda mais.”

Tim Berners-Lee

RESUMO

CANALLE, Gabrielle Karine. Desafios e Perspectivas da Web Semântica. 2013. Trabalho de Conclusão de Curso (Tecnologia em Análise e Desenvolvimento de Sistemas), Universidade Tecnológica Federal do Paraná. Medianeira. 2013.

A Web se constitui de um emaranhado de informações desorganizadas, dificultando as tarefas a serem realizadas na rede, como buscas por informação. A Web Semântica é uma tentativa de solução para essa desorganização e outros problemas da web, surgiu como um novo conceito de internet, que visa permitir uma melhor interação entre máquina e usuário. Tem como principal objetivo dar significado às informações, fazendo com que computadores e pessoas trabalhem melhor em cooperação. Para isso, se baseia em três pilares, a linguagem XML (*Extensible Markup Language*), a sintaxe RDF (*Resource Description Framework*) e as Ontologias. Embora seja considerada o futuro da Web, a Web Semântica ainda possui muitos desafios a superar. Esse trabalho de conclusão tem como objetivo apresentar os desafios e as perspectivas da Web Semântica, identificando os problemas da web, e o que poderá mudar com o uso da Web Semântica em conjunto com outras tecnologias associadas.

Palavras-chave: Problemas da Web. Ontologias. Tecnologias Associadas.

ABSTRACT

CANALLE, Gabrielle Karine. Desafios e Perspectivas da Web Semântica. 2013. Trabalho de Conclusão de Curso (Tecnologia em Análise e Desenvolvimento de Sistemas), Universidade Tecnológica Federal do Paraná. Medianeira. 2013.

The Web is now a tangle of information is disorganized, making it difficult to perform tasks on the network, how to search for information. The Semantic Web is an attempt at solution to this clutter and other web problems. Emerged as a new internet concept that aims to allow greater interaction between machine and user. Its main objective is to give meaning to information, causing computers and people to work in better cooperation. To do this, is based on three pillars, the language XML (Extensible Markup Language), the syntax RDF (Resource Description Framework) and Ontologies. Although it is considered the future of the Web, the Semantic Web still has many challenges to overcome. This final project aimed to present the challenges and prospects of the Semantic Web, identifying the current web problems, and what you can change with the use of the Semantic Web in conjunction with other associated technologies.

Keywords: Web Problems. Ontologies. Associated Technologies.

LISTA DE FIGURAS

FIGURA 1 - EXEMPLO DE TRIPLAS RDF.....	26
FIGURA 3 - INTERFACE DO FREEBASE.....	33
FIGURA 4 - INTERFACE DO TRIPIT.....	34
FIGURA 5 - INTERFACE DO SITE BOOHA.....	35
FIGURA 6 - ARQUITETURA DAS PÁGINAS WEB QUE UTILIZAM AS TECNOLOGIAS DA WEB SEMÂNTICA.....	36
FIGURA 7 - CONJUNTOS DE DADOS COM <i>LINKED DATA</i> - 2007.....	41
FIGURA 8 - CONJUNTOS DE DADOS COM <i>LINKED DATA</i> - 2008.....	41
FIGURA 9 - CONJUNTOS DE DADOS COM <i>LINKED DATA</i> - 2009.....	42
FIGURA 10 - CONJUNTOS DE DADOS COM <i>LINKED DATA</i> - 2011.....	42

LISTA DE QUADROS

QUADRO 1 - EXEMPLO DE RDF COM XML.....25

LISTA DE SIGLAS

CERN	Laboratório Europeu de Física de Partículas
DARPA	Agência de Projetos de Pesquisa avançada dos Estados Unidos
FTP	File Transfer Protocol
HTML	Hypertext Markup Language
HTTP	Hypertext Transfer Protocol
IMPs	Interface Message Processor
MIT	Instituto Tecnológico de Massachusetts
NCP	Network Control Protocol
NCSA	Centro Nacional para Aplicações de Supercomputação
OWL	Web Ontology Language
RDF	Resource Description Framework
RDFS	RDF Schema Language
RFC	Request for Comments
SGML	Standard Generalized Markup Language
TCP/IP	Transmission Control Protocol/Internet Protocol
URI	Uniform Resource Identifiers
WWW	World Wide Web
W3C	World Wide Web Consortium
XML	Extensible Markup Language

SUMÁRIO

1 INTRODUÇÃO.....	12
1.1 OBJETIVO GERAL.....	13
1.2 OBJETIVOS ESPECÍFICOS	13
1.3 JUSTIFICATIVA	13
1.4 ESTRUTURA DO TRABALHO	15
2 A HISTÓRIA DA <i>WORLD WIDE WEB</i> (WWW).....	16
2.1 AS ORIGENS DA INTERNET	16
2.2 A CRIAÇÃO DO W3C (WORLD WIDE WEB CONSORTIUM).....	18
2.3 OS PROBLEMAS ATUAIS DA WEB.....	19
3 WEB SEMÂNTICA	21
3.1 CONCEITO E OBJETIVO	21
3.2 TECNOLOGIAS ASSOCIADAS	22
3.2.1 XML (EXTENSIBLE MARKUP LANGUAGE).....	23
3.2.2 RDF (RESOURCE DESCRIPTION FRAMEWORK) E RDFS (RDF SCHEMA LANGUAGE).....	24
3.2.3 ONTOLOGIAS	27
3.2.4 OWL (WEB ONTOLOGY LANGUAGE).....	28
3.2.5 AGENTES	29
3.2.6 ASSINATURA DIGITAL.....	30
3.3 ARQUITETURA EM CAMADAS DA WEB SEMÂNTICA	31
3.4 O USO DA WEB SEMÂNTICA EM CASOS REAIS.....	33
3.5 ARQUITETURA DAS PÁGINAS WEB QUE UTILIZAM WEB SEMÂNTICA.....	36
4 DESAFIOS E PERSPECTIVAS DA WEB SEMÂNTICA	38
5 <i>LINKED DATA</i>	40
6 CONSIDERAÇÕES FINAIS.....	44
6.1 CONCLUSÃO.....	44
6.2 TRABALHOS FUTUROS/CONTINUAÇÃO DO TRABALHO	44
REFERÊNCIAS BIBLIOGRÁFICAS	45

1 INTRODUÇÃO

Surgida no início dos anos 90, a *World Wide Web* (WWW), ou simplesmente Web, é atualmente tão popular e ubíqua, que não raro, no imaginário dos usuários, confunde-se com a própria balzaquiana Internet – a infraestrutura de redes, servidores e canais de comunicação que lhe dá sustentação (SOUZA; ALVARENGA, 2004).

Segundo Souza e Alvarenga (2004), a Internet surgiu como proposta de um sistema distribuído de comunicação entre computadores para possibilitar a troca de informações. Ao implantar esse conceito, o projeto buscava oferecer interfaces amigáveis e intuitivas para a organização do acesso ao crescente repositório de documentos que a Internet vinha se tornando. Entretanto, o crescimento acelerado (além das expectativas) do alcance e tamanho desta rede, além da ampliação das possibilidades de utilização, fazem com que seja necessária uma nova filosofia, com suas tecnologias subjacentes, assim como a ampliação da infraestrutura tecnológica de comunicação.

Conforme Lima e Carvalho (2004), a web oferece uma enorme quantidade de informações criadas por uma grande variedade de diferentes organizações, comunidades e indivíduos. Ela também dispõe de uma variedade de serviços que facilitam a vida das pessoas, tais como: comércio, turismo, notícias, ensino e movimentações bancárias, entre outras. Tudo isso, unido a sua simplicidade de uso, é o que fez a WWW tão popular. Todavia, o crescimento do número de páginas disponíveis tem dificultado, cada vez mais, a localização, o acesso e a manutenção de informações nesta rede. Estes problemas vêm do fato que usuários, com diferentes níveis de conhecimento, disponibilizarem uma vasta quantidade de recursos, sem nenhuma padronização. Surge daí, a insatisfação dos próprios usuários, pois quando estes realizam uma pesquisa na Web, geralmente demoram muito para obter o resultado esperado.

Atualmente, a Web caracteriza-se, principalmente, pela forte presença de páginas dinâmicas, pelo grande volume de aplicações colaborativas e pela crescente facilidade de processamento da informação por parte dos usuários. Como parte deste cenário, surgiu a Web Semântica, que tem como objetivo prover meios para a criação de uma infraestrutura onde o significado (semântica) da informação esteja disponível de maneira estruturada e explícita. Dessa forma, espera-se facilitar tarefas muito comuns na Web, como: busca, acesso,

integração e reutilização de informações, aumentando, conseqüentemente, a eficiência dos motores de busca e oferecendo melhorias na disponibilização de serviços na Web (SALGADO; LÓSCIO, 2001).

A Web Semântica é uma extensão da Web atual, na qual a informação é dada com significado bem definido, permitindo que computadores e pessoas trabalhem melhor em cooperação (BERNERS-LEE; MILLER, 2002).

A Web Semântica, portanto, não se trata de uma nova rede de informações, ela pode ser vista como um projeto para aplicar sobre a Web atual conceitos inteligentes, onde cada informação vem com um significado bem definido, organizando o conteúdo e permitindo assim uma melhor interação entre seus usuários.

1.1 OBJETIVO GERAL

Identificar os desafios e perspectivas na implantação da web semântica para resolução dos problemas existentes na web atual, ou clássica.

1.2 OBJETIVOS ESPECÍFICOS

- Apresentar um histórico da WWW;
- Apresentar a Web Semântica e seus objetivos;
- Identificar os problemas da Web e o que poderá mudar com a Web Semântica;
- Apontar os desafios e perspectivas da Web Semântica;
- Apresentar casos reais que utilizam Web Semântica.

1.3 JUSTIFICATIVA

Um dos objetivos originais da Web era a troca de informação entre pessoas, mas de forma de que os computadores pudessem participar da comunicação, ajudando os usuários. Os computadores na Web, atualmente, têm papel somente no direcionamento e entrega de informações, não tendo acesso ao conteúdo das páginas, pois essa informação está estruturada

para utilização pelas pessoas e não por máquinas. Hoje, tem-se uma Web de documentos e não de informações. Por isso, os computadores oferecem ajuda limitada no acesso e processamento da informação, deixando as funções de extração e interpretação dessa informação a cargo dos usuários (DIAS; SANTOS, 2001).

A Web Semântica visa resolver este problema, estruturando o conteúdo das páginas Web de forma que a informação possa ser interpretada pelas máquinas. A proposta não é a de uma Web separada da atual, mas uma extensão da mesma, baseada em documentos – as ontologias (serão melhor explicadas em um tópico próprio) - descrevendo relacionamentos entre objetos e contendo informação semântica dos mesmos para automatizar o processamento pelas máquinas (DIAS; SANTOS, 2001).

Embora a Web apresente muitas deficiências, como por exemplo, limitações de segurança e qualidade de serviço, dificuldades no processo de busca, tornando-a pouco eficaz e grande quantidade de informações impertinente (como será visto mais a frente, no tópico: Os problemas atuais da Web), ela tem um papel muito importante para a sociedade como um todo. Portanto, seu desenvolvimento constitui um desafio para a comunidade científica, que se esforça na tentativa de que a integração, o intercâmbio e o entendimento semântico das informações ocorram com sucesso o mais breve possível. Várias iniciativas, como as desenvolvidas pelo Consórcio W3C (*World Wide Web Consortium*), buscam, por intermédio da criação de padrões, arquitetura de metadados, serviços de inferências e ontologias, uma melhor forma de tornar as informações processáveis pelas máquinas. O uso de metadados proporciona a descrição abrangente de recursos na Web, sendo uma solução adequada para promover a recuperação de serviços com mais eficiência (LIMA; CARVALHO, 2004).

A Web Semântica pode ser vista então como sendo uma solução prodigiosa para resolver os problemas da Web atual, mas ela é mais do que isto. Espera-se que ela facilite a publicação de dados em vários formatos. Desta maneira, mais pessoas irão querer publicar seus dados com semântica, gerando-se um efeito dominó. Com isto, imagina-se que um grande número de aplicações da Web Semântica possam ser desenvolvidas (LIMA; CARVALHO, 2004).

1.4 ESTRUTURA DO TRABALHO

Este trabalho possui cinco capítulos, sendo que o primeiro capítulo trata da Introdução do tema abordado, a justificativa e objetivos do projeto. O segundo apresenta a história da *World Wide Web*, bem como a criação do W3C Consortium e os problemas da web atual. O terceiro fornece os conceitos e objetivos da Web Semântica, as tecnologias associadas, e sua arquitetura. O quarto busca apresentar os desafios e perspectivas da Web Semântica. Por fim, o quinto capítulo é composto pelas considerações finais.

2 A HISTÓRIA DA *WORLD WIDE WEB* (WWW)

Este capítulo apresenta um histórico da *World Wide Web*, identificando os principais acontecimentos que desencadearam em seu surgimento. Apresenta também como se deu a criação do W3C (*World Wide Web Consortium*), e quais são os principais problemas existentes na web.

2.1 AS ORIGENS DA INTERNET

John Carl Robnett Licklider, do MIT (Massachusetts Institute of Technology), em agosto de 1962, propôs pela primeira vez uma rede global de computadores. Ele imaginou um conjunto de computadores interligados globalmente por meio do qual todos pudessem acessar rapidamente dados e programas de qualquer local. Em essência, o conceito foi muito parecido com a Internet de hoje (LEINER et al., 2012).

Leonard Kleinrock, do MIT estudou e pesquisou sobre a teoria de troca de pacotes, tendo seu primeiro trabalho sobre o tema lançado em julho de 1961, e o primeiro livro, também sobre esse assunto, em 1964. Com isso, ele convenceu Lawrence G. Roberts que existia a possibilidade de as comunicações utilizarem pacotes ao invés de circuitos, o que foi um grande passo no caminho em direção a redes de computadores. Outro grande passo foi fazer com que dois computadores interligados trocassem informações. Em 1965, juntamente com Thomas Merrill, Roberts conectou um computador TX-2 em Massachusetts com um Q-32 da Califórnia, através de uma conexão *dial-up* de baixa velocidade, sendo essa a primeira rede de computadores construída. Com esse trabalho, houve a constatação de que circuitos não eram adequados para isso, sendo necessária a troca de pacotes (LEINER et al., 2012).

Em 1966 Roberts começou a trabalhar no DARPA (Defense Advanced Research Projects Agency) e desenvolveu a ARPANET (Advanced Research Projects Agency Network). A Internet, então conhecida como ARPANET, foi colocada *online* em 1969, conectando inicialmente quatro computadores localizados em grandes universidades do sudoeste dos EUA. Depois disso, foram adicionados à rede várias outras universidades e órgãos, formando uma rede bem maior do que o esperado (HOWE, 2012).

Em agosto de 1968, depois de Roberts e o grupo do DARPA terem refinado a estrutura e as especificações para a ARPANET, foi realizada uma seleção para o desenvolvimento de um dos componentes chaves, o pacote chamado *Interface Message Processors* (IMPs). O grupo liderado por Frank Heart Bolt Beranek and Newman (BBN), venceu a licitação. A equipe BBN trabalhou no Imps juntamente com Bob Kahn. Em outubro de 1972 foi realizada a primeira demonstração pública da nova tecnologia de rede, a ARPANET. Em 1972 também, foi quando a primeira aplicação de correio eletrônico foi introduzido. Depois de Ray Tomlinson na BBN escrever o e-mail básico para enviar e ler apenas, foi Roberts quem expandiu sua utilidade. Ele escreveu o primeiro programa utilitário de e-mail para listar, ler, arquivar, encaminhar e responder mensagens. A partir daí o e-mail foi considerado como a maior aplicação de rede por mais de uma década (LEINER et al., 2012).

O protocolo TELNET, que permite realizar *logon* (ou autenticação) em um computador remoto, foi publicado como um *Request for Comments* (RFC) em 1972. RFC é um meio de compartilhar o trabalho de desenvolvimento em toda comunidade. O protocolo FTP, que permite a transferência de arquivos entre os sites de Internet, foi publicado como um RFC, em 1973. Como os comandos para e-mail, FTP e TELNET foram padronizados, tornou-se muito mais fácil para as pessoas não-técnicas usar a rede. A Internet cresceu muito na década de 70, como resultado do TCP/IP, protocolo proposto pela primeira vez por Bob Kahn na BBN e desenvolvido por Kahn e Vint Cerf, da Universidade de Stanford e outros. Esse protocolo foi adotado pelo Departamento de Defesa em 1980, substituindo o *Network Control Protocol* (NCP), e universalmente adotado em 1983 (HOWE, 2012).

Em 1989, Tim Berners-Lee e outros no CERN (Laboratório Europeu de Física de Partículas) propuseram um novo protocolo para a distribuição de informações. Este protocolo que se tornou a *World Wide Web* em 1991 foi baseado em hipertexto, sistema de incorporação de links em um texto, onde o acesso a outro texto se dá por meio de hiperlinks. Em 1993, Marc Andreessen e sua equipe do Centro Nacional para Aplicações de Supercomputação (NCSA), criaram o navegador gráfico Mosaic, que foi o grande responsável por impulsionar o crescimento da web. (HOWE, 2012).

2.2 A CRIAÇÃO DO W3C (WORLD WIDE WEB CONSORTIUM)

Em 1994 foi criado o *World Wide Web Consortium* (W3C), que é um consórcio internacional no qual organizações filiadas, uma equipe em tempo integral e o público trabalham juntos para desenvolver padrões para a Web. Liderado pelo inventor da web, Tim Berners-Lee e o CEO Jeffrey Jaffe, o W3C tem como missão conduzir a *World Wide Web* para que atinja todo seu potencial, desenvolvendo protocolos e diretrizes que garantam seu crescimento de longo prazo (W3C, 2011).

Tim Berners-Lee e Jeffrey Jaffe são apoiados por uma equipe de técnicos que ajudam a coordenar o desenvolvimento dos padrões e gerenciar as operações do Consórcio. Os membros do W3C e especialistas convidados por eles auxiliam os grupos de desenvolvedores responsáveis por escrever os padrões da Web. Os padrões da web são recomendações do W3C destinadas a orientar os desenvolvedores a utilizarem boas práticas tornando a web acessível para todos. Organizações localizadas em todo o mundo e de diferentes campos se unem ao W3C para participar de um fórum neutro que trata sobre a criação de padrões Web. Os membros do W3C e uma dedicada equipe de técnicos em tempo integral ganharam reconhecimento internacional por suas contribuições para na web, que inclui:

- ligações com organizações internacionais, nacionais e regionais, para ajudar a manter uma cultura global no desenvolvimento da *World Wide Web*;
- traduções dos padrões da web e outros materiais da comunidade W3C, em outros idiomas além do inglês;
- palestras pelo mundo, em vários idiomas destinadas as pessoas envolvidas na criação de padrões para a web (W3C, 2011).

Além de poder participar de eventos, workshops e palestras sobre os padrões para a web, as pessoas podem contribuir para o projeto. O W3C disponibiliza várias maneiras para aqueles que querem ajudar, como por exemplo ser voluntário para traduções de relatórios técnicos, ou quem escreve código e trabalha com software pode contribuir realizando experiências de implementação com as especificações lançadas (W3C, 2011).

Ao promover a interoperabilidade e incentivar um fórum aberto para discussão, o W3C se compromete a liderar a evolução técnica da web. Em pouco mais de sete anos, o W3C desenvolveu mais de 40 especificações técnicas para a infra-estrutura da Web. No entanto, a Web ainda é nova e há muito trabalho a fazer (ERCIM NEWS, 2003).

2.3 OS PROBLEMAS ATUAIS DA WEB

No surgimento da Internet, seus desenvolvedores visavam desenvolver uma rede entre universidades, onde os usuários seriam os estudantes. Esses usuários, podia-se dizer que eram confiáveis, e tinham conhecimentos técnicos sobre a rede. Desde a sua origem, a Internet tem crescido aceleradamente, tanto em quantidade de informação quanto de usuários. Segundo o site *Internet World Stats*(.), o número de usuários da Internet no mundo todo atualmente já passou de 2 bilhões. No Brasil, em junho de 2012, o número de usuários da Internet era de 88.494,756 de usuários.

A realidade é bem diferente de quando a Internet surgiu. Usuários do mundo todo, com variados tipos de formação, ou não, utilizam a Internet, portanto, começaram a surgir algumas limitações no aspecto de segurança e qualidade de serviço.

Um dos objetivos originais da Web era a troca de informação entre pessoas, mas de forma que os computadores pudessem participar da comunicação, ajudando os usuários. Os computadores na Web, atualmente, têm papel somente no direcionamento e entrega de informações, não tendo acesso ao conteúdo das páginas, porque essa informação está estruturada para utilização pelas pessoas e não por máquinas. Hoje, tem-se uma Web de documentos e não de informações. Por isso, os computadores oferecem ajuda limitada no acesso e processamento da informação, deixando as funções de extração e interpretação dessa informação a cargo dos usuários (DIAS; SANTOS, 2003).

Existe na Web uma quantidade muito grande de informações¹ impertinentes, que mesmo assim são fornecidas pelos buscadores ao realizar uma pesquisa. As ferramentas de busca encontram dificuldades ao executar pesquisas entre documentos que não estão diferenciados em termos de assunto, qualidade e relevância. A tecnologia atual não é capaz de diferenciar uma informação comercial de uma educacional, ou informação entre idiomas, culturas e mídia. É necessário haverem informações de qualificação da própria informação, chamada de metadados, para ser possível classificá-las e tornar os processos de busca mais eficazes (DIAS; SANTOS, 2003).

¹ Informações são dados tratados. O resultado do processamento de dados são informações. As informações tem significado, podem ser tomadas decisões ou fazer afirmações considerando informações. Os dados representam um ou mais significados que isoladamente não podem transmitir uma mensagem ou representar algum conhecimento.

A Web sofreu um processo de popularização muito grande nos últimos anos, o que resultou em um considerável crescimento. Segundo Souza e Alvarenga (2004), o enorme crescimento, além das expectativas, do alcance e tamanho desta rede, além da ampliação das possibilidades de utilização, fazem com que seja necessária uma nova filosofia, com suas tecnologias subjacentes, além da ampliação da infraestrutura tecnológica de comunicação. Esse crescimento infelizmente não aconteceu de maneira ordenada e controlada. A facilidade de se construir páginas HTML possibilita que todos os dias novos documentos sejam disponibilizados na internet sem controle de conteúdo, permitindo com isso que os mesmos documentos sejam alterados com facilidade.

Embora tenha sido projetada para possibilitar o fácil acesso, intercâmbio e a recuperação de informações, a Web foi implementada de forma descentralizada e quase anárquica; cresceu de maneira exponencial e caótica e se apresenta hoje como um imenso repositório de documentos, que deixa muito a desejar quando se precisa recuperar aquilo de que tem-se necessidade. 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, é baseada primariamente em palavras-chave contidas no texto dos documentos originais, o que é pouco eficaz. A dificuldade de determinar os contextos informacionais tem como consequência a impossibilidade de se identificar, de forma precisa a atenção dos documentos. Além disso, a ênfase das tecnologias e linguagens atualmente utilizadas nas páginas Web, focaliza os aspectos de exibição e apresentação dos dados, de forma que a informação seja pobremente descrita e pouco passível de ser consumida por máquinas e seres humanos (SOUZA; ALVARENGA, 2004).

Segundo Dias; Santos (2003), a Web Semântica visa resolver este problema, estruturando o conteúdo das páginas Web de forma que a informação possa ser interpretada pelas máquinas.

3 WEB SEMÂNTICA

Este capítulo apresenta a Web Semântica, seus conceitos e objetivos, bem como as tecnologias associadas a ela. Apresenta também a arquitetura da Web Semântica, a arquitetura das páginas Web que utilizam Web Semântica e alguns exemplos dessas páginas.

3.1 CONCEITO E OBJETIVO

Segundo Berners-Lee e Miller (2002), a Web Semântica é uma extensão da web atual, onde a informação possui um significado bem definido, permitindo que computadores e pessoas trabalhem melhor em cooperação.

A Web Semântica tem como um dos seus principais objetivos prover meios para organizar dados e permitir que estes possam ser interpretados pelos computadores. Assim, espera-se melhorar os resultados das buscas e facilitar a automação de tarefas relevantes, evitando, sempre que possível desperdício de tempo e fazendo com que os dados sejam utilizados de forma mais inteligente (SALGADO; LÓSCIO, 2001).

A ideia é introduzir estrutura e significado ao conteúdo já disponível na web, transformando assim uma rede de documentos em uma rede de dados, que pode ser compreensível tanto para humanos quanto para computadores. Dessa maneira, os computadores ajudarão mais durante a realização de tarefas.

A Web Semântica visa resolver os problemas da Web atual, estruturando o conteúdo das páginas de forma que a informação possa ser interpretada pelas máquinas. A proposta não é a de uma web separada da atual, mas uma extensão da mesma, baseada em documentos que descrevem os relacionamentos entre os objetos e contém informação semântica dos mesmos para automatizar o processamento pelas máquinas (DIAS; SANTOS, 2003).

Para que a Web Semântica possa ser implantada, é necessário que os dados publicados na web estejam interligados entre si. Para que isso aconteça de forma organizada e padronizada existe o estilo *Linked Data* (o conceito será abordado em um tópico próprio) de publicação de dados linkados na web.

O projeto que visa implantar a Web Semântica é coordenado pelo W3C e Tim Berners-Lee, juntamente com empresas do ramo tecnológico, e traz ideias revolucionárias, possibilitando uma conotação nova para a web atual, com novos usos, práticas e tecnologias. A semântica é utilizada no âmbito de estabelecer relação dos significados com a realidade, segundo a ordem semântica. Baseia-se na ideia de oferecer dados bem definidos na web e interligá-los de acordo com seu significado semântico, de maneira que possam ser usados para descoberta, automação, integração e reuso de várias aplicações (SOUZA, 2009).

3.2 TECNOLOGIAS ASSOCIADAS

O projeto da Web Semântica, em sua essência, é a criação e implantação de padrões tecnológicos para permitir este panorama, que não somente facilite as trocas de informações entre agentes pessoais, mas principalmente estabeleça uma linguagem franca para o compartilhamento mais significativo de dados entre dispositivos e sistemas de informação de uma maneira geral. Para atingir tal propósito, é necessária uma padronização de tecnologias, de linguagens e de metadados descritivos, de forma que todos os usuários da Web obedeçam a determinadas regras comuns e compartilhadas sobre como armazenar dados e descrever a informação armazenada e que esta possa ser "consumida" por outros usuários, humanos ou não, de maneira automática e não ambígua. Com a existência da infra-estrutura tecnológica comum da Internet, o primeiro passo para este objetivo está sendo a criação de padrões para descrição de dados e de uma linguagem que permita a construção e codificação de significados compartilhados (SOUZA; ALVARENGA, 2004).

Sendo assim, para inserir semântica na web, a adoção de algumas tecnologias, propostas na arquitetura em camadas da Web Semântica são necessárias. Para facilitar essa adoção, o W3C tem trabalhado no intuito de apresentar padrões para estas tecnologias. Dentre estes padrões, destacam-se as propostas para a especificação de metadados e ontologias, bem como a especificação de uma linguagem de consultas para tais informações.

Uniform Resource Identifiers (URIs) são um componente fundamental da Web e são também uma base da Web Semântica. A *Extensible Markup Language* (XML) é também um componente fundamental para apoiar a Web Semântica. O XML fornece uma base sintática interoperáveis em que a questão mais importante de representar relações e significados podem

ser construídas. URIs fornecem a capacidade para identificar exclusivamente recursos, bem como as relações entre os recursos. A *Resource Description Framework* (RDF) família de padrões utiliza URIs e XML para fornecer um conjunto de etapas de funcionalidade para representar essas relações e significados. As ontologias também são parte importante dentro da Web Semântica, pois atuam fazendo os relacionamentos entre os termos através da linguagem de manipulação de ontologias OWL (Web Ontology Language).

Outra tecnologia associada à Web Semântica são os agentes, que são programas inteligentes que podem ser considerados assistentes para o usuário. Já existem, mas, com a implantação da Web Semântica passarão a ser bem mais comum, com o objetivo de ajudar o usuário a realizar suas tarefas.

Por fim, a assinatura digital, que também é considerada uma tecnologia que está associada a Web Semântica. As perspectivas são de que com a implantação da Web Semântica cada informação disponibilizada na web contenha uma assinatura digital. Dessa maneira se passará maior segurança para os usuários, pois será possível identificar se a fonte da informação é confiável ou não.

3.2.1 XML (EXTENSIBLE MARKUP LANGUAGE)

Embora não seja por si só suficiente, a adoção da linguagem XML como um padrão para representação e troca de dados na Web pode ser considerado o primeiro passo em direção à Web Semântica, uma vez que o uso de marcadores extensíveis permite associar algum significado aos dados de um documento XML. Entretanto, o uso de XML para prover semântica para os dados é muito restrito. Não existe um significado associado a um conjunto de elementos aninhados em um documento XML, de forma que cada aplicação deverá atribuir a sua própria interpretação a estes aninhamentos (SALGADO; LÓSCIO, 2001).

A linguagem XML permite que qualquer um possa criar suas próprias *tags*. Os scripts ou programas podem fazer uso destas *tags* segundo as suas necessidades. Em outras palavras, XML permite que usuários adicionem uma estrutura arbitrária em seus documentos, mas sem nada a dizer sobre o que a estrutura significa (LIMA; CARVALHO, 2004).

O XML, o RDF (*Resource Description Framework*) e o RDFS (*RDF Schema Language*) são utilizados juntos. O XML oferecendo os meios necessários para a criação de

marcações (*tags*) para descrição de informações sobre domínios específicos. Por meio das definições de *namespaces* e *schema*, permite a ligação das definições da Web Semântica com outros padrões baseados em XML, o RDFS, que é uma extensão de RDF para a criação de vocabulários, diferentemente do RDF, que permite apenas a definição de sentenças sobre recursos específicos, o RDFS possibilita a criação de conceitos mais gerais, como a definição de hierarquia de classes e propriedades (SALGADO; LÓSCIO, 2001).

3.2.2 RDF (RESOURCE DESCRIPTION FRAMEWORK) E RDFS (RDF SCHEMA LANGUAGE)

Apesar de prover uma maior semântica para os dados, por meio da definição de marcadores específicos para o domínio de aplicação que se deseja representar, o XML não possui todos os requisitos necessários para descrever adequadamente a semântica de uma informação. A forma como os elementos devem estar relacionados, bem como a definição e o significado dos mesmos fica a cargo do desenvolvedor. O RDF foi criado como uma possível solução para esta limitação do XML. Geralmente referenciado como uma linguagem, o mais correto seria afirmar que RDF é um modelo que possibilita a definição de afirmações, chamadas sentenças, sobre um recurso. (SALGADO; LÓSCIO, 2001).

O RDF utiliza a linguagem XML para representar os itens na Web em forma de triplas <objeto, atributo, valor>. A utilização da sintaxe XML é apenas uma possível sintaxe para se escrever o RDF, podem surgir outras sintaxes para a sua representação. O significado da tripla <objeto, atributo, valor> é de que o objeto X tem o valor Y para um certo atributo Z (MORAIS; SOARES, 2004).

O RDF tem por objetivo definir um mecanismo de representação de metadados para descrever recursos não vinculados a um domínio específico de aplicação, resultado do trabalho em conjunto desenvolvido por várias comunidades. O RDF recebeu a influência de várias fontes diferentes. As principais influências vieram das comunidades de padronização da web (HTML, XML e SGML), da Biblioteconomia (metadados de catalogação), da representação do conhecimento (ontologias), da programação orientada a objetos, da linguagem de modelagem, entre outras (DZIEKANIAK; KIRINUS, 2004).

De acordo com DIAS e SANTOS (2001, p. 10)

O modelo RDF é utilizado para identificação de equivalência de significado, já que duas ou mais expressões em RDF são equivalentes se, e somente se, a representação de seus modelos de dados forem similares. Essa definição de equivalência permite a variação sintática em algumas expressões sem alterar seu significado. Esse modelo de dados é representado através de um DLG (*Directed Labeled Graphs*) e consiste de três tipos de objetos:

- *Resource* (Recurso): tudo que é descrito através de expressões RDF, podendo ser tanto um documento HTML, quanto um elemento XML de um documento; uma coleção de páginas ou um site inteiro. Um recurso pode também ser objeto que não seja acessado diretamente pela Web, tal como um livro impresso. Recursos são sempre nomeados por uma URI, o que permite a criação de identificadores para qualquer entidade imaginável;
- *Property* (Propriedade): é uma característica, atributo ou relação utilizado para descrever um recurso. Cada propriedade possui um significado específico, define seus próprios valores permitidos, tipos de recursos a que podem ser aplicados e seus relacionamentos com outras propriedades;
- *Statement* (Declaração): é composto da associação de um recurso específico, uma propriedade e o valor da propriedade para esse recurso. Essas três partes individuais da declaração são denominadas, respectivamente, de sujeito, predicado e objeto, onde o objeto pode ser um outro recurso ou um literal, ou seja, um recurso especificado por uma URI ou uma cadeia de caracteres ou outro tipo de dados definido por XML.

O quadro 1 mostra um exemplo de RDF utilizando XML.

```
<?xml version="1.0"?>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02
            22-rdf-syntax-ns#"
  xmlns:s="http://description.org/schema/">
  <rdf:Description
    about="http://www.abc.com/Canalle">
    <s:Creator>
      Gabrielle Canalle
    </s:Creator>
  </rdf:Description>
</rdf:RDF>
```

Quadro 1. Exemplo de RDF com XML.

O exemplo mostra que o recurso identificado pela tag `<rdf:Description about=http://www.abc.com/Canalle>` tem a propriedade (atributo) *creator*, identificada pela tag `<s:Creator>`, com o valor Gabrielle Canalle

Um recurso, em RDF, pode estar relacionado com dados, ou também com outros recursos por meio das sentenças, que são estruturadas no formato sujeito+predicado+objeto, onde:

- **Sujeito:** Possui como valor o recurso sobre o qual se quer escrever uma sentença.

- **Predicado:** Especifica um relacionamento entre um sujeito e um objeto. É especificado por meio de propriedades, que são relações binárias geralmente nomeadas por um verbo e permitem relacionar um recurso a dados ou a outros recursos.
- **Objeto:** Especifica o recurso ou dado que se relaciona com o sujeito. O valor de um objeto pode ser um recurso ou um literal, que pode ser um valor numérico ou uma cadeia de caracteres (CUNHA; LÓSCIO; SOUZA, [200-]).

A Figura 1 mostra alguns exemplos de triplas RDF.

Sujeito	Predicado	Objeto
http://www.w3c.org/RDF/Validator/run/p91002043177	http://uni.org/uni/elements/1.1/nomeDocente	"Berna Farias"
http://www.w3c.org/RDF/Validator/run/ck120	http://uni.org/uni/elements/1.1/nomeDisciplina	"Banco de Dados I"
http://www.w3c.org/RDF/Validator/run/ck120	http://uni.org/uni/elements/1.1/EnsinadoPor	http://www.w3c.org/RDF/Validator/run/p91002043177

Figura 1 - Exemplo de triplas RDF.

Fonte: Cunha; Lóscio; Souza, [200-].

A primeira tripla informa que o recurso "p91002043177" possui nome Berna Farias. A segunda tripla descreve um segundo recurso "CK120", cujo nome é "Banco de Dados I". A terceira tripla é responsável por criar um relacionamento entre os dois primeiros recursos. Então, a afirmação "A disciplina Banco de Dados I é ensinada por Berna Farias" é descrita nessa tripla RDF. Várias triplas de diferentes fontes podem ser combinadas facilmente para formar um único grafo (CUNHA; LÓSCIO; SOUZA, [200-]).

Segundo Salgado; Lóscio, (2001), "Devido ao seu formato, uma sentença RDF é frequentemente chamada de tripla. Sendo assim, um documento RDF pode ser visto como um conjunto de triplas, onde estas descrevem informações sobre os recursos envolvidos no domínio de interesse".

Para realizar consultas nos grafos RDF, e recuperar informações, é utilizada a SPARQL, linguagem padrão recomendada pelo W3C. O SPARQL possui uma estrutura semelhante ao SQL, e possui três blocos principais:

- *Select*: Utilizada para especificar a forma final em que os dados serão apresentados ao usuário.
- *From*: Utilizada para declarar as fontes que serão consultadas.
- *Where*: Utilizada para impor restrições na consulta.

3.2.3 ONTOLOGIAS

Em filosofia, o significado de ontologia é uma teoria sobre a existência da natureza, sobre que tipos de coisas existem. Os pesquisadores da inteligência artificial e da Web converteram este termo para o seu próprio jargão, sendo que, para eles, ontologia é um documento ou arquivo que formalmente define as relações entre termos.

Quando se pensa em dois bancos de dados que possuem diferentes identidades para um mesmo conceito e que se tem um programa que queira comparar ou utilizar informações dos dois bancos de dados, este programa necessita saber que existem diferentes termos que possuem o mesmo significado. Para isso, o programa deve ter um caminho para descobrir os significados destes termos quando for manipulá-los. Uma solução para este problema é a criação de coleções de informações, chamada de ontologias (LIMA; CARVALHO, 2004).

Segundo Gruber (1993), “Uma ontologia é uma especificação explícita de uma conceituação”. Salgado e Lóscio (2001, p. 3) analisando a definição de Gruber pensam que,

Dizer que uma ontologia é uma conceituação significa que uma ontologia é um modelo abstrato que descreve objetos, conceitos e outros elementos, associados a algum domínio geralmente restrito, além de relacionamentos entre os mesmos. Uma especificação explícita significa que os elementos e relacionamentos no modelo abstrato recebem nomes explícitos, expressos em alguma linguagem formal, e possuem uma semântica associada.

As ontologias são como um modelo de relacionamento de entidades e suas interações, em algum domínio particular do conhecimento ou específico a alguma atividade. O objetivo da criação das ontologias é a necessidade de um vocabulário compartilhado para realizar a troca de informações entre os membros de uma comunidade, sejam eles humanos ou agentes inteligentes. Existem diversos padrões e linguagens sendo criados para a construção e compartilhamento de ontologias na web, sendo que, todos são baseados em XML. Como exemplo pode-se citar o SHOE8, a *Ontology Exchange Language*, a *Ontology Markup Language* e a *Resource Description Framework Schema Language* (SOUZA; ALVARENGA, 2004).

Essa tecnologia agregará valor ao funcionamento da Web, pois podem ter várias aplicações diferenciadas. A aplicação mais simples seria aumentar a precisão dos mecanismos de busca, onde eles pesquisariam somente em páginas que fizessem referência a um conceito pré-definido e não em todas as páginas que contenham palavras-chave. As aplicações mais

avançadas utilizariam as ontologias com o objetivo de relacionar o conteúdo das páginas às suas estruturas existentes de conhecimento e regras de inferência (DIAS; SANTOS, 2003).

3.2.4 OWL (WEB ONTOLOGY LANGUAGE)

A OWL foi criada para prover uma forma comum de processar o conteúdo da Web (e não apenas divulgá-lo), e para ser lida por computadores (e não por humanos). O objetivo da construção de OWL foi torná-la um padrão aceito e largamente utilizado da Web Semântica com o compromisso de ser processada por mecanismos de inferência eficientes e, ao mesmo tempo, ser suficientemente expressiva para permitir a representação de uma grande gama de ontologias e conhecimentos (SALGADO; LÓSCIO, 2001).

Essa linguagem foi elaborada para ser utilizada por aplicações que necessitem processar o conteúdo de informações, ao invés de somente apresentar a visualização destas informações. Ela pode ser usada para explicitamente representar o significado de termos em vocabulários e os relacionamentos entre esses termos (LIMA; CARVALHO, 2004).

Segundo Salgado e Lóscio (2001, p.25)

Considerando a eficiência de raciocínio e o poder de expressão citados como necessários para uma linguagem de manipulação de ontologias, OWL foi definida como três diferentes sublinguagens (ou dialetos), a saber:

- OWL *Full* é a versão mais expressiva do OWL. Permite o uso de todas as funcionalidades disponíveis. Porém, tamanha expressividade compromete a eficácia na extração e inferência de informações bem como a decidibilidade da ontologia. Uma ontologia decidível é aquela capaz de ser processada em tempo hábil por um motor de inferência (*reasoner*).
- OWL DL é um subconjunto de OWL *Full*, restringindo o uso de alguns operadores OWL e RDF para garantir eficiência computacional. Esta restrição foi feita no sentido de garantir que a linguagem corresponda à lógica descritiva (*Description Logic*). OWL DL permite uma inferência eficiente mas não é completamente compatível com RDF.
- OWL *Lite* é a menos expressiva das versões de OWL. Possui um subconjunto dos construtores de OWL DL, excluindo, por exemplo, classes enumeradas, disjunção e cardinalidade arbitrária. Tem sido muito usada para definição de domínios simples. Pode ser uma boa alternativa para migração de tesouros. A principal vantagem é a facilidade de entender e de se implementar uma ontologia usando esta versão. A desvantagem é a restrição na sua expressividade.

Fica a cargo dos desenvolvedores escolherem a versão da linguagem que irão utilizar de acordo com as suas necessidades. A escolha consiste em balancear o nível de expressividade com decidibilidade e eficiência na inferência de informações, por que quanto

maior o nível de expressividade utilizado na descrição de uma ontologia, maior é a complexidade dos algoritmos utilizados (SALGADO; LÓSCIO, 2001).

3.2.5 AGENTES

Os agentes possuem um papel muito importante na Web Semântica. São programas que buscam e capturam conteúdos de várias fontes na Web, processam estas informações e fazem intercâmbio desses resultados com outros programas. Possuem um certo grau de autonomia e são capazes de realizar tarefas que auxiliem o usuário no desempenho de suas atividades, de acordo com seus interesses (DIAS; SANTOS, 2003)

Segundo Berners-Lee; Lassila (2001),

O verdadeiro poder da Web Semântica será realizado quando as pessoas criarem muitos programas que colem o conteúdo da Web a partir de diversas fontes processam a informação e trocam os resultados com outros programas. A eficácia de tais agentes de software vai aumentar exponencialmente à medida que o conteúdo da Web se torne mais legível pela máquina e por serviços automatizados (incluindo outros agentes). A Web Semântica promove esta sinergia: mesmo os agentes que não foram expressamente concebidos para trabalhar em conjunto podem transferir dados entre si quando os dados vêm com semântica.

São sistemas computacionais que interagem autonomamente para atingir os objetivos estipulados pelo seu criador. Possuem algumas características como reatividade, independência, tem comportamento colaborativo, possuem objetivos, são flexíveis, sociáveis e possuem capacidade de aprender. A perspectiva é de que a Web Semântica tenha vários agentes interagindo entre si, compreendendo, trocando ontologias, obtendo novas capacidades e formando cadeias que facilitem a comunicação e a ação humana (MORAIS; SOARES, 2004).

De acordo com Dias; Santos (2003),

A eficácia dos programas, baseados em agentes, tende a aumentar na medida em que houver mais conteúdo na Web estruturado de maneira que possa ser utilizado pelos computadores. Os agentes seriam responsáveis por captar as necessidades do usuário, pesquisar e disponibilizar os resultados esperados de forma interativa.

Embora não exista uma definição universal para o termo "agente", em computação, pode-se dizer que agentes são como assistentes de tarefa, entidades de software que

empregam técnicas de inteligência artificial com o objetivo de auxiliar o usuário na realização de uma tarefa, agindo de forma autônoma e utilizando a metáfora de um assistente pessoal. Com a utilização dos agentes, homem e computador se tornam parceiros, cooperando para o alcance dos objetivos que foram traçados. Desse modo, pode-se dizer que em um futuro próximo os usuários apenas delegarão tarefas aos computadores, e estes, por sua vez, com certo grau de autonomia, realizarão as tarefas, auxiliando os usuários no desempenho de suas atividades (SOUZA; ALVARENGA, 2004).

Segundo Wooldridge; Jennings (1995),

- apresenta-se um conjunto de propriedades desejáveis a um agente, saber:
- autonomia, de modo a agir sem qualquer tipo de intervenção, possuindo controle sobre suas ações e estado interno;
 - sociabilidade, de modo a interagir com outros agentes (artificiais ou humanos) por meio de algum tipo de linguagem de comunicação;
 - reatividade, de modo a perceber alterações em seu ambiente, reagindo a tempo;
 - proatividade, de modo a estar apto a tomar iniciativas, em vez de simplesmente atuar em resposta ao ambiente;
 - continuidade temporal, ou seja, está sendo executado continuamente, ativamente ou em *background*, possivelmente captando informações sobre o usuário e sobre o ambiente, para melhor desempenhar suas funções;
 - orientação para objetivos, por ser capaz de interagir e desempenhar uma série diversa de ações isoladas, com objetivo de executar uma tarefa mais complexa.

“Os agentes de software poderão usar os seus conhecimentos para buscar as informações requisitadas, filtrar informações irrelevantes e estudar estas informações, na tentativa de recuperar os dados exatos requisitados pelos usuários, e ainda tentar inferir novos conhecimentos” (LIMA; CARVALHO, 2004).

3.2.6 ASSINATURA DIGITAL

A assinatura digital é um código que é incluído em um documento, mensagem ou texto, que tem como objetivo identificar o remetente da mensagem ou o escritor do texto, por exemplo. A assinatura digital também é aplicada para dados RDF. As pessoas vão criar seus documentos, como sendo dados RDF, e enviá-los a um destinatário utilizando a assinatura digital (LIMA; CARVALHO, 2004).

A certificação digital tem trazido inúmeros benefícios para os cidadãos e para as instituições que a adotam. Com a certificação digital é possível utilizar a Internet como meio

de comunicação alternativo para a disponibilização de diversos serviços com uma maior agilidade, facilidade de acesso e substancial redução de custos. A tecnologia da certificação digital foi desenvolvida graças aos avanços da criptografia nos últimos 30 anos (GDF, [200-]).

Assinaturas digitais que são blocos de dados criptografados que os computadores e os agentes podem usar para verificar se as informações anexas foram fornecidas por uma fonte confiável. Os agentes devem ser céticos em relação a afirmações que lêem na Web Semântica após terem verificado as fontes de informação (BERNERS-LEE; HENDLER; LASSILA, 2001).

Segundo Pfützenreuter (2004), “A assinatura digital garante a integridade dos dados e a comprovação da procedência dos recursos”.

Os computadores e a Internet são largamente utilizados para o processamento de dados e para a troca de mensagens e documentos entre cidadãos, governo e empresas. No entanto, estas transações eletrônicas necessitam da adoção de mecanismos de segurança capazes de garantir autenticidade, confidencialidade e integridade às informações eletrônicas. A certificação digital é a tecnologia que provê estes mecanismos (GDF, [200-]).

Com a aplicação da Web Semântica, a assinatura digital passa a ser um recurso ainda mais importante, pois passa a verificar se a informação disponibilizada na web foi realmente inserida por uma fonte segura. Assim, a web passará aos usuários mais credibilidade, pois passará a apresentar informações mais confiáveis.

3.3 ARQUITETURA EM CAMADAS DA WEB SEMÂNTICA

A arquitetura da Web Semântica ainda está em construção pela W3C, com o objetivo de mapear as complexas relações semânticas, lógicas, de sintaxe e de apresentação dos documentos nesse novo espaço.

A arquitetura proposta para a Web Semântica se divide em várias camadas, como mostra a Figura 2.

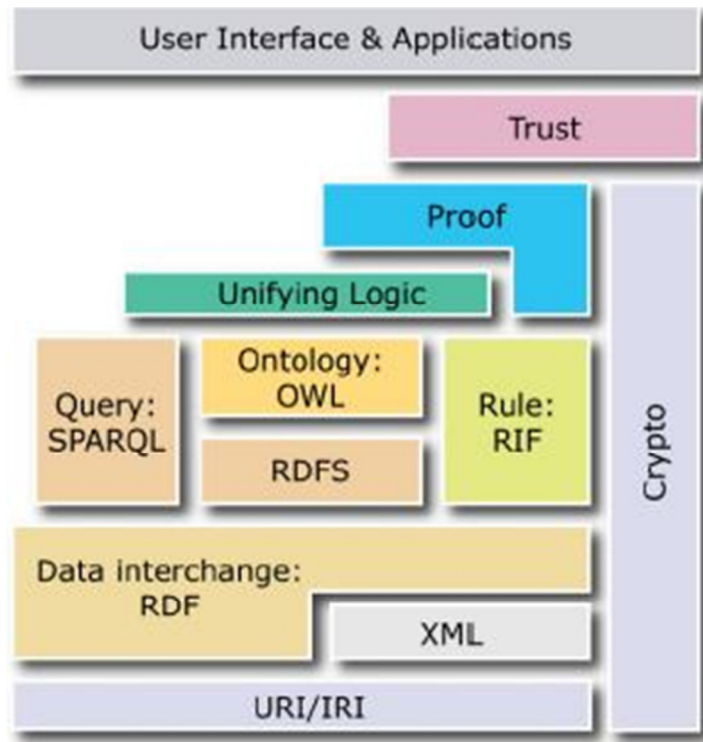


Figura 2 - Arquitetura em Camadas da Web Semântica.

Fonte: Salgado; Lóscio (2001, p. 5)

A base da arquitetura é a camada URI (*Uniform Resource Identifier*), que é responsável por prover meios para a identificação única de recursos da Web. A camada XML oferece meios para a criação de *tags* para descrição de informações. A camada RDF provê um modelo de descrição lógica de dados. A camada seguinte é a RDFS, extensão de RDF para criação de vocabulários, possibilitando a criação de conceitos mais gerais. A camada superior a RDFS é a camada de Ontologia, que estende a camada RDFS possibilitando um maior nível de expressividade para a definição de vocabulários, sendo a OWL a linguagem padrão para definição de ontologias. Junto com RDFS e OWL localizam-se as camadas de Consulta (Query:SPARQL) e Regras (Rule:RIF). A camada de consulta provê meios para a realização de consultar em documentos RDF e ontologias, e a camada de regras provê um mecanismo de criação de relações entre recursos que não podem ser descritas diretamente na ontologia. As próximas camadas são Lógica (*Unifying Logic*), Prova (Proof) e Confiança (Trust), que permitem o raciocínio e a execução de inferências lógicas a partir da semântica previamente descrita, e oferecem mecanismos para avaliar o nível de confiabilidade das fontes de recursos e informações. As últimas camadas, de Aplicação e Interface do Usuário (*User Interface & Applications*) permite a interação entre o usuário e as aplicações da Web Semântica, e a

camada de criptografia (Crypto) possui o objetivo de garantir a segurança e a privacidade das informações (SALGADO; LÓSCIO, 2001).

3.4 O USO DA WEB SEMÂNTICA EM CASOS REAIS

Apesar de ser uma nova tecnologia, de acordo com MacManus (2008), já existem várias aplicações e sites utilizando Web Semântica, pode-se citar como exemplo:

- Freebase: Base de dados aberta, onde as pessoas criam dados e conectam esses dados entre si, e são classificados por temas e sessões. Possui como objetivo ser um grande repositório de conhecimento humano, seguindo o estilo da Wikipédia. A Figura 3 mostra a interface do Freebase.

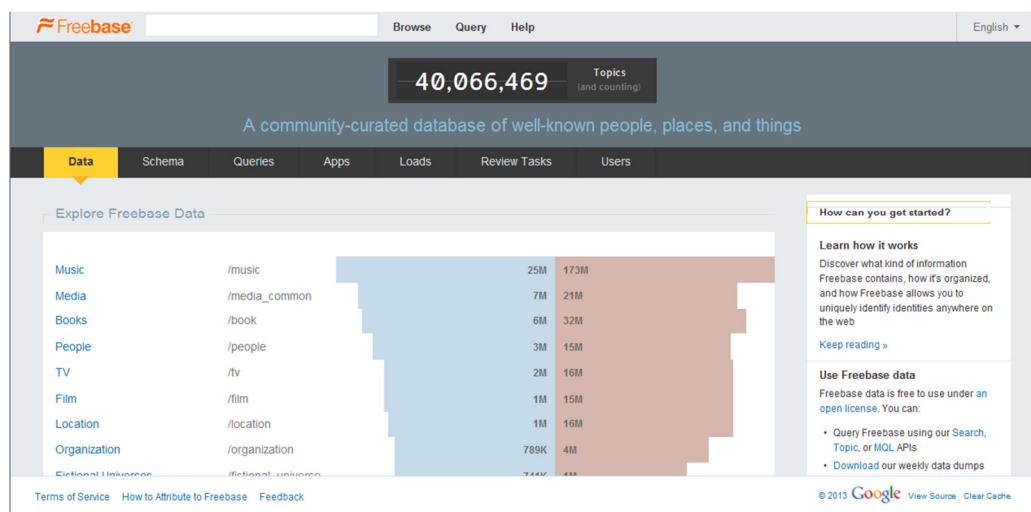


Figura 3 - Interface do Freebase.

Fonte: Print Screen site Freebase.

O Freebase possui uma API no Google Developers, sendo apenas é necessário ter uma conta do Google para utilizar os serviços. Os conteúdos são separados por tipos, o que ajuda a filtrar a pesquisa. Como é uma base de dados aberta qualquer pessoa pode adicionar conteúdo ou classificar conteúdos com novos tipos. Cada conteúdo possui uma série de *tags* e *links* para

interligar os dados. Já existem até o momento, mais de 39 milhões de tópicos que correspondem a, por exemplo, pessoas, lugares e coisas.

A base de dados do Freebase é utilizada por várias aplicações, como por exemplo, o Powerset. Essa aplicação faz a busca com perguntas complexas, o que no google retornaria várias páginas, no Powerset retorna apenas uma resposta.

- Hakia: Apontado como um dos mais promissores concorrentes do Google, faz a análise do termo buscado antes de realizar a busca;
- TripIt: Sistema que permite que o usuário planeje suas viagens. Permite ainda que o usuário procure por hotéis, vôos e aluguel de carros. A Figura 4 mostra a interface do TripIt.



Figura 4 - Interface do TripIt.

Fonte: Print Screen do site TripIt.

O TripIt disponibiliza uma aplicação para aparelhos móveis como iOS, Android, BlackBerry e Windows Phone, sendo possível acessar os dados sobre sua viagem de qualquer lugar. Para utilizar os serviços é necessário fazer um cadastro prévio.

- Swotti: Reúne informações e análise sobre produtos a partir de fóruns, blogs e sites especializados, com o objetivo de ajudar o usuário a fazer a melhor opção de compra.

- BooRah: Processa comentários e análises sobre os restaurantes dos EUA. A Figura 5 mostra a interface do site BooRah.



Figura 5 - Interface do site BooHa.

Fonte: Print Screen do site BooHa.

O Boorah abrange as cidades mais populares dos EUA, como Las Vegas, Nova York, San Francisco, Los Angeles, Boston e Miami. O site está crescendo, e com isso adicionando novas cidades. Já possui mais de 1,5 milhões de comentários sobre os restaurantes das cidades onde atua, ajudando as pessoas a escolherem os lugares onde irão frequentar.

- Siri: O objetivo é desenvolver um assistente que possui uma nova maneira de interagir e utilizar a Internet. Esse assistente seria capaz de se adaptar ao usuário e aprender com as ações realizadas.
- Evri: Site de busca que possui uma grande quantidade de conteúdo semanticamente interligado, principalmente sobre filmes, músicas e artistas.

Esses são alguns exemplos de aplicações que estão utilizando os conceitos da Web Semântica. Além dos exemplos apresentados anteriormente, estão surgindo cada vez mais aplicações sendo criadas seguindo os padrões propostos pelo W3C.

3.5 ARQUITETURA DAS PÁGINAS WEB QUE UTILIZAM WEB SEMÂNTICA

Segundo Santos; Carvalho (2007),

Para atingir os seus objetivos, as aplicações semânticas necessitam de uma Web semanticamente estruturada. Esta estruturação pode ser feita a partir do uso das tecnologias da Web Semântica. Dessa forma, as Aplicações de Suporte à Web Semântica podem auxiliar na manipulação de informações definidas de acordo com as tecnologias da Web Semântica e no suporte ao desenvolvimento de aplicações semânticas.

A Figura 6 mostra a arquitetura das aplicações que utilizam as tecnologias propostas pelo W3C para o desenvolvimento de páginas web.

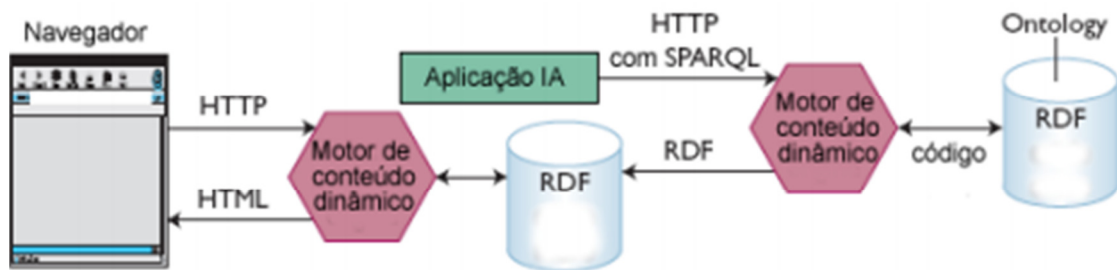


Figura 6 - Arquitetura das páginas web que utilizam as tecnologias da Web Semântica.

Fonte: Santos; Carvalho (2007, p. 9)

Na figura 6, o usuário faz uma requisição HTTP através do navegador. O motor de conteúdo dinâmico (servidor de aplicações) recebe a requisição, acessa a base de dados RDF e retorna a página HTML para o usuário. Um Agente inteligente, faz uma requisição HTTP com a linguagem de consulta SPARQL, o motor de conteúdo dinâmico recebe a requisição, manipula a ontologia e grava informações em RDF, que poderão ser utilizadas pelos usuários em outras requisições futuramente (SANTOS; CARVALHO, 2007).

O real poder que a Web Semântica possui só poderá ser constatado quando existirem muitas aplicações que sejam capazes de obter informações, vindas de diversas fontes da Web, de processar estas informações e de trocar os resultados deste processamento com outras aplicações (LIMA; CARVALHO, 2004).

No entanto, a transição da Web para a Web Semântica será feita aos poucos, transformando os dados desorganizados da web em dados estruturados, para que seja possível que as aplicações semânticas compreendam as informações contidas neles e os significados

desses dados. Depois que essa estruturação acontecer será possível criar aplicações semânticas inovadoras, melhorando a vida das pessoas seja na busca eficiente por informações, no avanço científico ou na evolução do conhecimento humano (SANTOS; CARVALHO, 2007).

4 DESAFIOS E PERSPECTIVAS DA WEB SEMÂNTICA

Sabe-se que a web atualmente é uma das mais ricas fontes de informação. Nos últimos anos uma quantidade muito grande de documentos foi colocada na web de forma desorganizada. Para o usuário, essas milhares de informações “jogadas” na rede, não trazem muitos benefícios. Segundo Lima; Carvalho (2004),

A principal propriedade da WWW é a sua universalidade, onde “qualquer coisa pode ser ligada a qualquer coisa”. As tecnologias Web, portanto, não discriminam entre um texto bem escrito e um rascunho, entre informações comerciais e acadêmicas ou entre culturas e línguas diferentes. Além disso, estas informações podem estar expressas em diversos formatos. Alguns destes formatos podem ser mais adequados ao entendimento humano e outros ao processamento por máquinas.

Ao realizar uma busca qualquer, para chegar ao resultado esperado, o usuário tem um considerável trabalho, para filtrar as informações desejadas. Isso se dá por que os buscadores atualmente fazem as buscas por meio de palavras-chave, retornando várias páginas de documentos que podem ou não ser o que o usuário procura. Cabe ao usuário pesquisar essas páginas retornadas e encontrar o que for de seu interesse. Com a web semântica, por exemplo, os buscadores iriam pesquisar, analisar, interpretar e compreender a busca, e mostrar a resposta esperada pelo usuário. Atualmente, na rede, a semântica existente é entendida apenas pelos humanos. O objetivo da web semântica é transformar essa semântica já existente para que ela possa ser compreendida também pelos computadores, já que, atualmente eles servem apenas para direcionar e entregar as informações, pois não possuem conhecimento do conteúdo das páginas. De acordo com Sabino (2007),

Web semântica permitirá que as pessoas e computadores trabalhem em cooperação na exploração do conhecimento, uma vez que pressupõe a atribuição de significado aos conteúdos publicados na Internet e o desenvolvimento de tecnologias e linguagens que colocam esse significado ao alcance das máquinas. Ao mesmo tempo, o estabelecimento de significados e de linguagens compatíveis garante ambientes de maior interoperabilidade entre serviços. Tendo em conta essas características, não é de estranhar que os computadores possam, no futuro próximo, ser instruídos para fazer pesquisas tão complexas como apresentar uma lista de preços de televisores de alta definição, com écran superior a 70 cm, resolução de 1080p, à venda em lojas da cidade mais próxima, abertas até às 20H00 durante os fins-de-semana.

A Web Semântica pressupõe uma geração de serviços baseados na Internet, os quais se suportam no que pode-se chamar “Web inteligente”, isto é, um tipo de Web que, se baseia numa maior capacidade do software em interpretar os conteúdos em rede, devolvendo resultados mais objetivos e personalizados a cada vez que se fizer uma pesquisa (SABINO, 2007).

O W3C trabalha para que a Web Semântica seja disponibilizada universalmente, para que se possa desenvolver um ambiente onde a informação seja expressa de maneira a possibilitar a automatização de tarefas e a otimização de uso dos recursos por parte dos usuários. Outro objetivo da W3C também é a criação de uma Web confiável, possibilitando que as pessoas assumam a autoria e responsabilidade por suas publicações (DIAS; SANTOS, 2003). Dias; Santos (2003) apontam que

O potencial da Web Semântica será realmente compreendido quando forem desenvolvidos programas que sejam capazes de efetuar buscas de informação de diferentes fontes disponíveis na Web, as processem e compartilhem os resultados com outros programas. A eficácia dos programas, baseados em agentes, tende a aumentar na medida em que houver mais conteúdo na Web estruturado de maneira que possa ser utilizado pelos computadores. Os agentes seriam responsáveis por captar as necessidades do usuário, pesquisar e disponibilizar os resultados esperados de forma interativa.

A Web Semântica não é uma realidade em curto prazo, mas do mesmo modo que o surgimento da Web surpreendeu a todos, o surgimento desta nova Web poderá surpreender também. É imprescindível que, principalmente, as empresas busquem conhecimento sobre essa nova tecnologia e mantenham-se atentas ao início de ofertas de softwares agentes no mercado para que não sejam as últimas a reestruturarem seus sites Web (DIAS; SANTOS, 2003).

Aliada a Inteligência Artificial, a Web Semântica terá um papel fundamental na tecnologia do futuro. Muitos avanços acontecerão, onde a rede de computadores ficará mais organizada e inteligente.

5 LINKED DATA

Para tornar a Web Semântica uma realidade, é importante ter uma quantidade significativa de dados em formato padrão disponíveis na web, acessíveis e gerenciáveis pelas ferramentas da Web Semântica. Essas coleções de conjuntos de dados inter-relacionados, disponíveis na web, também podem ser chamados de *Linked Data* (W3C, 2013).

Linked Data se refere a um estilo de se disponibilizar e interligar dados estruturados na Web. Quanto mais um dado for interligado com outro dado, maior é o seu valor e utilidade na rede.

Segundo Berners-Lee et al. (2006), *Linked Data* resume-se em quatro princípios básicos:

1. Utilizar URIs como nome para recursos.
2. Utilizar URIs e HTTP para que as pessoas possam encontrar esses nomes.
3. Quando alguém procurar por uma URI, garantir que informações úteis possam ser obtidas através dessa URI, as quais devem estar representadas no formato RDF.
4. Incluir links para outras URIs de forma que outros recursos possam ser descobertos.

O primeiro princípio defende o uso de referências URI para identificar documentos, conteúdos digitais e também objetos do mundo real e conceitos abstratos. O segundo princípio defende o uso de URIs HTTP para identificar os objetos e os conceitos abstratos definidos pelo primeiro princípio. O terceiro princípio defende que exista um acordo sobre o formato padrão para disponibilização dos dados, esse padrão é o RDF. Com ele é possível descrever o significado de recursos, habilitando agentes de software a explorar os dados automaticamente. E por fim, o quarto princípio diz respeito ao uso de hiperlinks para conectar não apenas os documentos na web, mas qualquer tipo de recurso (CUNHA; LÓSCIO; SOUZA, [200-]).

O objetivo do projeto *Linked Data*, é identificar os dados existentes disponíveis sob licenças abertas, convertê-los para RDF de acordo com os princípios *Linked Data*, e publicá-los na web. Como uma questão de princípio, o projeto sempre foi aberto a quem publica dados de acordo com os princípios *Linked Data* (HEATH, T.; BIZER, C., 2011).

O exemplo mais visível de adoção e aplicação dos princípios *Linked Data* tem sido o projeto *Linking Open Data*, que foi fundado em janeiro de 2007 e apoiado pelo W3C *Semantic Web Education and Outreach Group*. O principal objetivo do projeto *Linked Open Data* é identificar conjuntos de dados disponíveis sob licenças abertas e convertê-los para RDF de acordo com os princípios *Linked Data* (CUNHA; LÓSCIO; SOUZA, [200-]).

Fonte: Health; Bizer (2011).

A Figura 10 mostra os conjuntos de dados que foram publicados e interligados pelo projeto até o ano de 2011. Coletivamente, os conjuntos de dados (295) consistem de mais de 31 bilhões de triplas RDF, que são interligadas por cerca de 504 milhões de ligações RDF (W3C SWEO *Community Project*, 2013).

6 CONSIDERAÇÕES FINAIS

6.1 CONCLUSÃO

Com a implantação da Web Semântica, o trabalho do usuário seria desnecessário, já que os buscadores realizariam toda a tarefa e retornariam o resultado esperado. Isso acontecerá por que os computadores serão capazes de compreender a semântica da web, que atualmente é compreendida apenas pelos humanos. A web passará a ser mais inteligente onde os softwares terão maior capacidade em interpretar as informações da rede, tornando assim os resultados mais objetivos.

Para transformar a web em uma Web Semântica, é necessária a adoção de algumas tecnologias. Para facilitar esse processo, o W3C apresenta alguns padrões para essas tecnologias, como, por exemplo, a utilização da linguagem XML junto com RDF, a uso da linguagem OWL e também das Ontologias.

A Web Semântica é um novo passo no desenvolvimento da Internet, marcado pela organização do conteúdo e pela cooperação de usuários e computadores na realização de tarefas.

6.2 TRABALHOS FUTUROS/CONTINUAÇÃO DO TRABALHO

Este trabalho teve como objetivo apontar os conceitos em torno da Web Semântica e das tecnologias associadas a ela, mostrando os desafios e perspectivas dessa nova tecnologia. Como ideia para trabalhos futuros, para exemplificar os conceitos abordados, pode-se criar uma aplicação ou site utilizando tecnologias como XML, RDF e Ontologias. Pode-se também realizar uma pesquisa aprofundada sobre *Linked Data* já que no presente trabalho o assunto foi abordado conceitualmente.

REFERÊNCIAS BIBLIOGRÁFICAS

BERNERS-LEE, T. *Semantic Web Road map*. Disponível em:

<<http://www.w3.org/DesignIssues/Semantic.html>>. Acesso em: 27 de jun. de 2013.

BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. The Semantic Web. *Scientific American*.

Mai. 2001. Disponível em: <<http://www.scientificamerican.com/article.cfm?id=the-semantic-web>>. Acesso em: 16 de jun. de 2013.

BERNERS-LEE, T.; MILLER, E. The Semantic Web lifts off. *Ercim News*. N° 51, Out. 2002.

Disponível em: <http://www.ercim.eu/publication/Ercim_News/enw51/berners-lee.html>.

Acesso em: 09 de maio de 2013.

BERNERS-LEE, T. et al. *Tabulator: Exploring and Analyzing linked data on the Semantic*

Web. Semantic Web User Interaction Workshop. Athens, EUA. Nov. 2006.

CUNHA, D. R. B.; LÓSCIO, B. F.; SOUZA, D. *Linked Data: da Web de Documentos para a*

Web de Dados. Disponível em: <<http://die.ufpi.br/ercemapi2011/minicursos/MC2.pdf>>.

Acesso em: 8 de jul. de 2013.

DIAS, T.D.; SANTOS, N. Web Semântica: Conceitos Básicos e Tecnologias Associadas.

Cadernos do IME – Série Informática, Rio de Janeiro, v. 14, Jun, 2003. Disponível em:

<<http://magnum.ime.uerj.br/cadernos/cadinf/vol14/7-neide.pdf>> Acesso em 09 de maio de 2013.

DZIEKANIAK, G. V.; KIRINUS, J. B. Web Semântica. *Encontros Bibli – Revista Eletrônica*

de Biblioteconomia e Ciências da Informação, Florianópolis, v. 9, n°. 18, 2004. Disponível

em: <[http://www.periodicos.ufsc.br/index.php/eb/article/view/1518-](http://www.periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2004v9n18p20/5471)

2924.2004v9n18p20/5471>. Acesso em: 05 de jul. de 2013.

ERCIM NEWS. About World Wide Web Consortium. *Ercim News*. N°. 52, Jan., 2003.

Disponível em: <http://www.ercim.eu/publication/Ercim_News/enw52/w3c.html>. Acesso

em: 6 de jun. de 2013.

GDF - SECRETARIA DE ESTADO DE FAZENDA DO DISTRITO FEDERAL. *O que é*
Certificação Digital? Disponível em:

<http://www.fazenda.df.gov.br/arquivos/pdf/O_que_e_certificado_digital.pdf>. Acesso em:

15 de jul. de 2013.

GRUBER, T. R. *A Translation Approach to Portable Ontology Specifications*. Knowledge Acquisition. V. 5. 199-220. 1993.

HEATH, T.; BIZER, C. *Linked Data: Evolving the Web into a Global Data Space*. 1a Edição. Berlin: Morgan & Claypool, 2011.136 p.

HOWE, W. *A Brief History of the Internet: An anecdotal history of the peoples and communities that brought about the Internet and the Web*. Disponível em: <<http://www.walthowe.com/navnet/history.html>>. Acesso em 13 de maio De 2013.

INTERNET WORLD STATS. *Internet Usage Statistics*. Disponível em: <<http://www.internetworldstats.com/stats.htm>>. Acesso em: 19 de jun. de 2013

KOIVUNEN, M. R.; MILLER, E. *W3C Semantic Web Activity*. Disponível em: <<http://www.w3.org/2001/12/semweb-fin/w3csw>>. Acesso em: 23 de maio de 2013.

LEINER, B. M. et al. *Brief History of the Internet*. Disponível em: <<http://www.internetsociety.org/internet/what-internet/history-internet/brief-history-internet#Authors>>. Acesso em 13 de maio de 2013.

LIMA, J. C.; CARVALHO, C. L. *Uma Visão da Web Semântica*. Disponível em: <http://www.inf.ufg.br/sites/default/files/uploads/relatorios-tecnicos/RT-INF_001-04.pdf>. Acesso em 09 de maio de 2013.

LINKING OPEN DATA – W3C SWEO Community Project. *Project Description*. Disponível em: <http://www.w3.org/wiki/SweoIG/TaskForces/CommunityProjects/LinkingOpenData#Project_Description>. Acesso em: 3 de jul. de 2013.

MACMANUS, R. *10 More Semantic Apps to Watch*. Disponível em: <http://readwrite.com/2007/11/28/10_semantic_apps_to_watch#awesm=~obGv8njgwx9pjb>. Acesso em: 15 de jul de 2013.

MORAIS, E. F.; SOARES, M. B. *Web Semântica para Máquinas de Busca*. Disponível em: <<http://homepages.dcc.ufmg.br/~nivio/cursos/pa03/seminarios/seminario7/seminario7.pdf>>. Acesso em 27 de jun. de 2013.

PFÜTZENREUTER, E. *Aplicabilidade e desempenho do protocolo de transporte SCTP*. Dez, 2004. 122 f. Dissertação (Mestrado em Ciências da Computação) – Universidade Federal de Santa Catarina, Santa Catarina, 2004.

RAMALHO, R. A. S.; VIDOTTI, S. A. B. G.; FUJITA, M. S. L. Web Semântica: uma investigação sob o olhar da Ciência da Informação. *DataGramaZero - Revista de Ciência da Informação*, Marília, SP, v. 8, n.º 6, dez. 2007. Disponível em: <http://www.dgz.org.br/dez07/Art_04.htm>. Acesso em: 11 de jul. de 2013.

SABINO, J. *Web 3.0 e Web Semântica – do que se trata?* Disponível em: <http://www.cin.ufpe.br/~hsp/Microsoft-web.pdf>. Acesso em 08 de jul. de 2013.

SALGADO, A. C.; LÓSCIO, B. F. *Web Semântica*. Centro de Informática – UFPE. Recife, PE. 2012.

SANTOS, F. C.; CARVALHO, C. L. de. Aplicações de Suporte à Web Semântica. Disponível em: <http://www.inf.ufg.br/sites/default/files/uploads/relatorios-tecnicos/RT-INF_004-07.pdf>. Acesso em: 11 de jun de 2013.

SOUZA, M. A. A Web Semântica e o Ensino a Distância. *e-Civitas*, MG, v. 2, n. 1. 2009. Disponível em: <<http://revistas.unibh.br/index.php/dcjpg/issue/view/2>>. Acesso em: 21 de jun. de 2013.

SOUZA, R. R.; ALVARENGA, L. A. Web Semântica e suas contribuições para a ciência da informação. *SciELO - Scientific Electronic Library Online*, Brasília, v. 33, n.º 1, jan/abr. 2004. Disponível em <http://www.scielo.br/scielo.php?pid=S0100-19652004000100016&script=sci_arttext>. Acesso em 09 de maio de 2013.

SWICK, R. R.; LASSILA, O. *Resource Description Framework (RDF) Model and Syntax Specification*. Disponível em: <<http://www.w3.org/TR/PR-rdf-syntax/>>. Acesso em 23 de maio de 2013.

WOOLDRIDGE, M.; JENNINGS, N. Intelligent Agents: Theory and Praticce. *Cambridge Journals*, Cambridge, v. 10, Issue 02, p 115-152. 07 jul. 2009.

W3C. *World Wide Web Consortium*. Disponível em: <<http://www.w3.org/>>. Acesso em: 21 de jun. de 2013.