



UNIVERSIDADE
ESTADUAL DE LONDRINA

Trabalho de Conclusão de Curso

Fabício Ferracioli

Web Semântica no Domínio de Informática Aplicada à Saúde

Londrina

2008

Fabício Ferracioli

Web Semântica no Domínio de Informática Aplicada à Saúde

Trabalho apresentado à Universidade Estadual de Londrina, como parte do requisito para obtenção do título de Bacharel em Ciência da Computação, sob orientação da Prof^aDoutora Maria Angélica de Oliveira Camargo Brunetto.

Orientadora: Maria Angélica de Oliveira Camargo Brunetto

Universidade Estadual de Londrina

Londrina

2008

Fabício Ferracioli

Web Semântica no Domínio de Informática Aplicada à Saúde

Prof^a. Dr^a. Maria Angélica de Oliveira Camargo Brunetto
Universidade Estadual de Londrina

Prof^o. Ms^o. Elieser Botelho Manhas Jr.
Universidade Estadual de Londrina

Prof^a. Ms^a. Veronice de Freitas
Universidade Estadual de Londrina

Londrina de 05 de Dezembro de 2008

À Guiomar Lopes de Souza

Agradecimentos

À Deus pelo desafio da vida.

À meus pais, Mônica e Francisco, por absolutamente tudo.

À minha irmã, Camila, por acreditar e apoiar.

À professora Maria Angélica por ter me auxiliado a crescer como profissional e como pessoa, além de todo o resto.

Aos amigos e amigas, pelos bons e maus momentos, “fitas erradas” e principalmente pela amizade.

À Ayn Rand por Atlas Shrugged.

Ao capitalismo, por nunca me deixar satisfeito.

À falta de dinheiro, por não me fazer desistir.

Ao dinheiro por abrir as portas.

À música pelo que ela é capaz.

À Edson Arantes do Nascimento e Ayrton Senna por terem trazidos momentos felizes ao Brasil.

À Pedro de Lara por seu humor.

Ao Coringa por seu espírito de vida inspirador.

Ao álcool, a cafeína e a pizza por diversos momentos.

If you're good at something, never do it for free.
The Joker - Batman the Dark Knight (2008)

Resumo

Este trabalho apresenta uma análise da Web Semântica aplicada ao domínio de cuidados com saúde que, como será visto, é uma das maiores responsáveis pelo desenvolvimento dela. Desde o projeto da World Wide Web as funcionalidades da Web Semântica já estavam previstas, entretanto, ainda não havia recursos o suficiente para sua implantação. Com o desenvolvimento da Web pavimentado pelo W3C desde a publicação da Web como um serviço através da Internet, foram sendo desenvolvidas tecnologias que, em 2001 possibilitaram o início do desenvolvimento da Web Semântica. Neste ano, o criador da Web e da Web Semântica, Tim Berners-Lee, publicou um artigo que mostrava uma nova gama de serviços inteligentes disponíveis na Web, prevendo inclusive a sua vasta utilização pela área de cuidados com saúde. Sabe-se atualmente que a área de cuidados com saúde se encontra na vanguarda da utilização das tecnologias de Web Semântica, sendo uma das maiores responsáveis pelos seus avanços conceituais e tecnológicos. Diversas aplicações surgem com diferentes escopos e soluções, mostrando o grande potencial e heterogeneidade da Web Semântica não somente em uma área restrita, mas em toda a Web. Neste trabalho apresentamos os conceitos gerais da Web, abordando seu desenvolvimento a partir de seu projeto. São mostradas as linguagens base da Web, que após um estágio de desenvolvimento, culminaram na base da Web Semântica. Esta também possui seus conceitos gerais apresentados, assim como sua base de linguagens. São abordadas algumas possibilidades de utilização da Web Semântica assim como áreas em que ela se aplica. Também é apresentada a estrutura da Web Semântica em conjunto com suas linguagens base. Então, o domínio de cuidados com saúde é explorado. Sua importância é apresentada e as atividades que já foram desenvolvidas nesse campo. Entre elas temos a padronização da troca de dados e representação de conhecimento, o que gera interoperabilidade. Iniciativas de governos em atividades que envolvem ambas as áreas também são mostradas, como o caso brasileiro, e aplicações que utilizam recursos semânticos de maneira efetiva na área de cuidados com saúde. A partir da análise pode-se concluir que a Web Semântica possui grande potencial, não somente para a área de cuidados com saúde, mas para a web como um todo. Entretanto, muito do que foi previsto ainda não está implementado, o que deve-se ao fato de muitas partes estruturais ainda estarem em desenvolvimento e a falta de consenso em alguns pontos. Mas, acreditamos que a área de cuidados com saúde deve continuar o desenvolvimento acelerado da Web Semântica, o que é uma boa prática para todos pois, através do desenvolvimento de soluções localizadas, a Web Semântica cresce de maneira global.

Palavras-chaves: Web Semântica, Informática na Saúde, Web, Internet

Abstract

This work presents an analysis of Semantic Web applied to healthcare domain that, as will be seen, is one of the most responsible for its development. Since the World Wide Web project the Semantic Web functionalities was predicted, however, not yet had enough resources for its implementation. With the evolution of Web assisted by W3C from its publication as a Internet service, technologies that were been developed, in 2001, enabled the beginning of Semantic Web development. In this year, the World Wide Web and Semantic Web creator, Tim Berners-Lee, published an article that showed new intelligent services through the Web, predicting even a greater use in the healthcare area. Today we know that healthcare area is in the forefront of use Semantic Web technologies, being one of the greater responsible for conceptual and technology advances. Various applications born with different scopes and solutions, shows the big Semantic Web potential and heterogeneity, not applied in one restricted area, but in all Web. In this work we present the general concepts of Web, showing his development from his project. The web base languages are showed, that after a development stage, culminated in the Semantic Web base. Semantic Web also have its general concepts presented, like its base languages. We show some use possibilities of Semantic Web and areas that can be applied. Its structure are presented together in their base languages. So, the healthcare area is explored, its importance is presented and the activities that was developed in this area. Between them we have the data exchange and knowledge representation standardization, that generates interoperability. Government initiatives in activities that evolves the two areas are presented too, like the Brazilian case, and applications that uses semantic resources in a effective way in the healthcare area. From this analysis we can conclude that semantic web has a big potential, not only for the healthcare area, but for all the web. However, much of what were predicted was not developed yet, which is due to the fact that many structural parts are still in development and the lack of consensus on some points. But, we believe that healthcare area must continue the accelerate development of Semantic Web, that is a good practice for all because, through development of localized solutions, the Semantic Web grows in a global way.

Keywords: Semantic Web, Healthcare, Web, Internet

Sumário

Lista de Siglas e Abreviaturas	p. 10
Lista de Figuras	p. 12
Introdução	p. 13
1 World Wide Web	p. 15
1.1 O World Wide Web Consortium	p. 17
1.2 O despertar para a Web	p. 19
1.3 Desenvolvimento Tecnológico da Web	p. 20
1.3.1 HTML	p. 21
1.3.2 XML	p. 21
1.3.3 XHTML	p. 22
1.3.4 Um antigo problema volta a aparecer	p. 24
1.3.5 Web Semântica	p. 24
2 Web Semântica	p. 26
2.1 Faces da Web Semântica	p. 26
2.2 Problemas que a Web Semântica pode resolver	p. 28
2.3 Esclarecimentos sobre o conceito	p. 29
2.3.1 Estrutura da Web Semântica	p. 30
2.3.2 As Camadas de bolo da Web Semântica	p. 32
2.4 O Semantic Web Activity	p. 34

2.4.1	Linguagens	p. 34
2.4.2	RDF e RDFS	p. 35
2.4.3	Ontologia	p. 37
2.4.4	SPARQL	p. 41
2.4.5	RIF	p. 43
2.5	Recursos Semânticos	p. 43
2.5.1	Web Services	p. 44
2.5.2	Agentes Inteligentes	p. 44
3	Web Semântica Aplicada à Saúde	p. 46
3.1	O Domínio de Cuidados com Saúde	p. 47
3.1.1	Padrões de troca de dados	p. 48
3.1.2	Aplicações	p. 53
3.2	Resultados e Atividades da Web Semântica Aplicada a Saúde	p. 61
	Conclusão	p. 65
	Referências Bibliográficas	p. 67

Lista de Siglas e Abreviaturas

ADL	-	Archetype Definition Language
ANS	-	Agência Nacional de Saúde
ANSI	-	American National Standards Institute
ASCII	-	American Standard Code for Information Interchange
AVC	-	Acidente Vascular Cerebral
CDA	-	Clinical Document Architecture
CEN	-	European Committee for Standardization
CERN	-	European Organization for Nuclear Research
CNS	-	Cartão Nacional de Saúde
COAS	-	Clinical Observation Access Specification
CPG	-	Collaborative Practice Guidelines
DAML	-	DARPA Agent Markup Language
DARPA	-	Defense Advanced Research Projects Agency
DATASUS	-	Departamento de Informática do SUS
DCMI	-	Dublin Core Metadata Initiative
DOM	-	Document Object Model
DTD	-	Document Type Definitions
EAV/CR	-	Entidade-Atributo-Valor com Classes e Relacionamentos
GEHR	-	Good European Health Record
HCLSIG	-	Health Care and Life Sciences Interest Group
HIV	-	Human Immunodeficiency Virus
HL7	-	Health Level Seven
HTML	-	HyperText Markup Language
HTTP	-	HyperText Transfer Protocol
IA	-	Inteligência Artificial
ICD	-	International Classification of Diseases
ICT	-	Information and Communication Technologies
IETF	-	Internet Engineering Task Force
IHC	-	Intermountain Health Care
ISO	-	International Organization for Standardization
LHC	-	Large Hadron Collider
MIT/LCS	-	Massachusetts Institute of Technology/Laboratory of Computer Science
MS	-	Ministério da Saúde
OIL	-	Ontology Interface Layer
ONG	-	Organização Não Governamental
OWL	-	Web Ontology Language
P2P	-	Peer-to-Peer
PIDS	-	Person Identification Specification

PRC	-	Padronização de Registros Clínicos
PURL	-	Persistent Uniform Resource Locator
RDF	-	Resource Description Framework
RDFa	-	Resource Description Framework - in - attributes
RDFS	-	RDF Schema
REST	-	Representational State Transfer
RIF	-	Rule Interchange Format
RSS	-	RDF Site Summary
SBIS	-	Sociedade Brasileira de Informática na Saúde
SGML	-	Standard Generalized Markup Language
SPARQL	-	Simple Protocol and RDF Query Language
SUS	-	Sistema Único de Saúde
TI	-	Tecnologia da Informação
TISS	-	Troca de Informações em Saúde Suplementar
UDDI	-	Universal Description, Discovery and Integration
URI	-	Uniform Resource Identifier
URL	-	Uniform Resource Locator
XHTML	-	eXtensible HyperText Markup Language
XML	-	eXtensible Markup Language
W3C	-	World Wide Web Consortium
WHO	-	World Health Organization
WWW	-	World Wide Web
WSDL	-	Web Service Definition Language

Lista de Figuras

1.1	Arquitetura Geral da Web (HERMAN, 2008a)	p. 17
2.1	Recursos e links com tipos na Web Semântica (Adaptação de (KOIVUNEN; MILLER, 2001))	p. 31
2.2	“Camadas do bolo da Web Semântica” (Adaptação de (CONSORTIUM,)) .	p. 32
2.3	Exemplo de uma tripla	p. 35

Introdução

Uma idéia aparentemente simples, e com objetivo de solucionar um problema localizado hoje é responsável por grandes mudanças em campos variados da sociedade. A *World Wide Web* (WWW) surgiu como uma solução de Tim Berners-Lee para o problema de gerência de documentos do CERN (*European Organization for Nuclear Research*). Na maioria dos casos, quando um documento era necessário não podia ser encontrado pois não existiam dados sobre onde encontrá-lo, responsável pelo documento, entre outros dados sobre aquele documento que se procurava. Além da solução, Berners-Lee também desenvolveu todo um conjunto estrutural que permitiu o crescimento da quantidade de documentos sem comprometimento do sistema, além da possibilidade de acessar informação de qualquer ponto.

A solução proposta foi tão eficiente e completa que acabou se tornando um meio para diversos fins, mas desta vez em escala mundial. Hoje sabemos que a WWW foi responsável por várias transformações, não somente no campo tecnológico, mas até mesmo no social.

Muitas possibilidades foram abertas a partir desse projeto, e a partir da percepção disso pela comunidade, a WWW passou a crescer de maneira incrível. Antes era possível apenas publicar, compartilhar e localizar documentos, atualmente é possível utilizar a WWW para os mais diversos fins: compras, trabalho, diversão, pesquisa, ensino, entre outros. Inclusive, existe a tendência da conexão entre recursos disponíveis na Web, fazendo com que tenhamos que realizar cada vez menos tarefas para chegar a um objetivo complexo.

Esse tipo de serviço já pode ser tido como realidade a partir de algumas implementações existentes da Web Semântica. Quando apresentada em 2001 (BERNERS-LEE; HENDLER; LASSILA, 2001), ela mostrava alguns cenários de utilização como o agendamento de uma consulta clínica para uma pessoa, que teve que fornecer somente os sintomas e seu plano de saúde. Um agente inteligente, capaz de realizar operações semânticas na web, foi responsável por encontrar um médico com a especialidade necessária para os sintomas, em um horário disponível na agenda da pessoa, também encontrada pelo agente, em um local próximo a residência dela. Bastaria, então a pessoa com os sintomas, selecionar a melhor opção fornecida pelo agente. Esse tipo de operação ainda não é possível, mas já caminhamos para estágios intermediários.

Muitos dos avanços da Web Semântica aconteceram justamente na área da saúde (SHADBOLT; HALL; BERNERS-LEE, 2006), que é a área que mais vem contribuindo para os avanços gerais daquela. Isso se deve a quantidade de dados padronizados existentes e ao próprio inte-

resse dos profissionais da área nos resultados que podem ser alcançados através disso. Desde o surgimento do *eXtensible Markup Language* (XML), a área da saúde vem trabalhando em sistemas interoperáveis e baseados em padrões semânticos. Essa iniciativa, ferramentas implementadas e resultados alcançados pela área da saúde são objeto de estudo deste trabalho.

Para isso, apresentamos no capítulo 1 a parte histórica da WWW, a partir do projeto de Tim Berners-Lee. Ao chegar a dias mais atuais, no capítulo 2 informações detalhadas sobre o conceito de Web Semântica, possíveis cenários de utilização e informações gerais sobre tecnologias a serem utilizadas são apresentadas. No capítulo 3, a Web Semântica é apresentada sob o foco das ciências da saúde, uma das áreas que possuem os primeiros sucessos na utilização da Web Semântica (FEIGENBAUM et al., 2007). Para isso, são apresentados estudos realizados sobre padrões de codificação e representação de dados, iniciativas governamentais e algumas aplicações estudadas durante o desenvolvimento deste, que obtiveram sucesso ao implementar funcionalidades semânticas.

1 World Wide Web

Em 1989 surgia uma das maiores ferramentas utilizadas pela comunidade para a disponibilização, busca e troca de informações: a *World Wide Web* (WWW). Inicialmente foi considerada “vaga, mas excitante” (BERNERS-LEE,) pelo seu próprio criador, que buscava resolver um problema de gerenciamento de informação sobre projetos do *European Organization for Nuclear Research* (CERN), seu instituto de pesquisa, e principalmente os relacionados ao *Large Hadron Collider* (LHC). Devido a quantidade de informação gerada nos projetos do instituto, dados complexos eram perdidos prejudicando as pesquisas (BERNERS-LEE, 1989). A solução se baseava num sistema de texto distribuído não linear, conhecido como hipertexto. Mais especificamente, o problema consistia de um grande número de pessoas que necessitava do compartilhamento de informações entre elas. Como toda organização, existe uma hierarquia entre as pessoas, novas pessoas ingressam em novos projetos diariamente, e ao observar isso, foi notada uma organização das pessoas conectadas como uma teia. As pessoas que ingressavam perdiam muito tempo para saber o que havia sido realizado antes delas, logo a informação era constantemente perdida, apesar de estar gravada, não era encontrada.

Alguns exemplos das questões sem resposta, devido a esse problema de gerenciamento de informação eram:

- Onde este módulo é utilizado?
- Quem escreveu este código?
- Que documentos existem sobre este conceito?
- Que documentos referenciam este?

No próprio projeto, já havia uma previsão de que isso aconteceria com o mundo todo em alguns anos, sendo o CERN, durante a época do projeto, uma pequena amostra do problema. Inclusive já haviam previsões acerca de projetos comerciais que resolveriam esse tipo de problema, baseados na solução apresentada. Alguns conceitos chave, apresentados no projeto, eram os nós e os *links*. Nós representam ou descrevem algo em particular, já os *links* interligam dois nós. Exemplos de nós podem ser:

- Pessoas;
- Módulos de software;
- Conceitos;
- Documentos.

Já os *links* podem representar:

- dependência;
- uma parte de;
- referência;
- utilização.

Os requisitos levantados para atender as necessidades do CERN eram:

1. Acesso remoto através da rede;
2. Heterogeneidade;
3. Descentralização;
4. Acesso a dados existentes;
5. *Links* privados;
6. Armazenamento e exibição de texto em ASCII em telas 24x80;
7. Análise automatizada de dados;
8. *Links* vivos.

Nesta etapa já era desejável a exibição de gráficos, mas ainda era um requisito que possuía pouca penetração. Em 1990, Tim Berners-Lee já possuía todas as ferramentas necessárias para que a Web funcionasse. Foi implementado o *HyperText Transfer Protocol* (HTTP) (FIELDING et al., 1999), as *Uniform Resource Identifier* (URI) (BERNERS-LEE; FIELDING; MASINTER, 2005), o *HyperText Markup Language* (HTML) (HOURS; JACOBS, 1999), um *Web Browser* chamado *WorldWideWeb* (COMMUNICATIONS, b) e um servidor Web (COMMUNICATIONS, a), além das primeiras páginas web que descreviam o projeto ¹. Em Agosto de 1991 a *World Wide Web* foi publicada como um serviço disponível na Internet.

¹<http://www.w3.org/History/19921103-hypertext/hypertext/WWW/TheProject.html>

Desde então a web foi crescendo, com a implementação dos primeiros *browsers* gráficos e que funcionassem em outras plataformas, e não apenas na NeXT, a única suportada inicialmente, como o WorldWideWeb. Surgiu o primeiro *browser* comercial, o Netscape Navigator, derivado do Mosaic. O primeiro *browser* da Microsoft foi chamado de Cello.

1.1 O World Wide Web Consortium

Fundado em 1994 por Tim Berners-Lee, o *World Wide Web Consortium* (W3C) é um consórcio internacional onde as organizações membro, equipes fixas e o público trabalham em conjunto para desenvolver padrões para Web, chamados de Recomendações. Criado no Instituto de Tecnologia de Massachussets, Laboratório para Ciência da Computação, MIT/LCS, em colaboração com a Organização Européia para Pesquisa Nuclear, CERN, sua missão é levar a WWW a todo seu potencial desenvolvendo protocolos e guias que permitam o crescimento da Web, sempre levando em conta a interoperabilidade (JACOBS, 2008b). Para isso, foi tomada a decisão de fornecer as tecnologias gratuitamente, sem patentes ou dívidas de *royalties*, podendo assim ser adotadas por todos. O consórcio possui escritórios em todo o mundo, inclusive no Brasil. Esses escritórios possuem o objetivo de promover as tecnologias do W3C e encorajar a participação internacional nas atividades do W3C (JACOBS, 2008a). Seus objetivos, em conformidade com sua missão, são criar uma Web, para todos, em qualquer lugar tendo como papel ser uma base de conhecimento confiável. A figura 1.1 mostra a visão estrutural da Web², foco de trabalho do W3C.

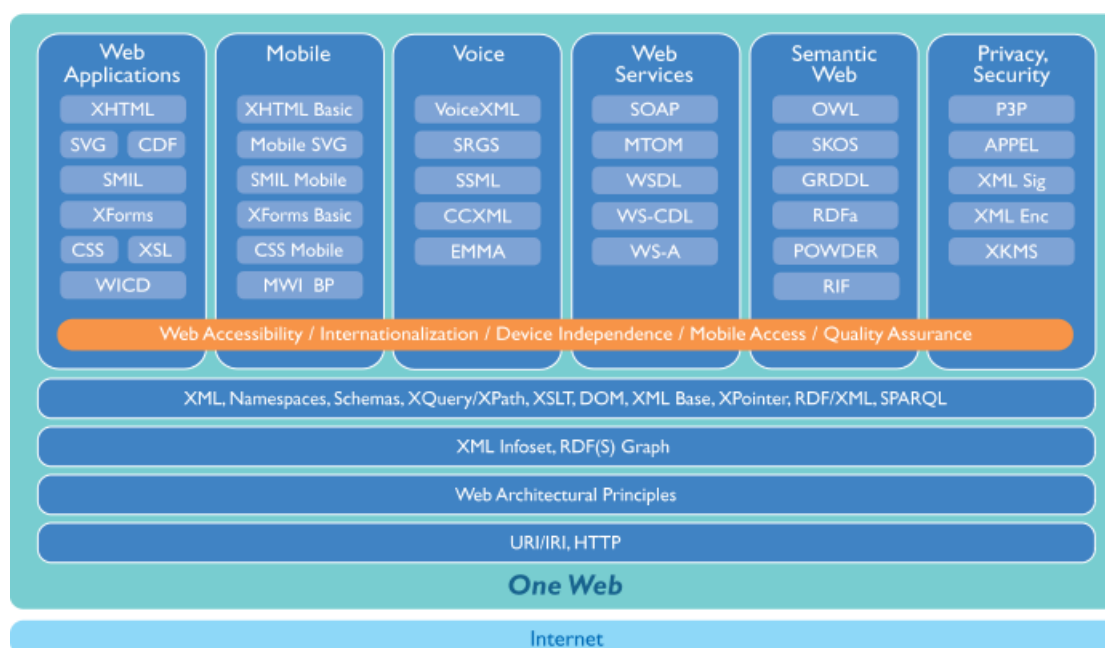


Figura 1.1: Arquitetura Geral da Web (HERMAN, 2008a)

²<http://www.w3.org/Consortium/technology>

A maior parte do trabalho do W3C está centralizada no desenvolvimento de padrões para web. Para isso, existe um processo realizado para promover o desenvolvimento de alta qualidade baseado no consenso dos *stakeholders*. O objetivo final de todo trabalho, é chegar ao nível de Recomendação, o equivalente a um Padrão Web. O processo para isso ocorre da seguinte maneira:

1. Pessoas que possuem interesse em um assunto particular expressam-no através de um formulário de submissão de membros. Então um conjunto de monitores trabalham dentro e fora do W3C por sinais de interesse. O W3C também pode organizar *Workshops* para recrutar pessoas com interesse no assunto e discutir tópicos de interesse;
2. Quando existe interesse suficiente, o diretor anuncia o desenvolvimento de uma proposta para uma nova atividade ou uma proposta de Grupo de Trabalho. Uma Proposta de Atividade descreve o escopo, duração e outras características do trabalho pretendido e inclui as propostas de Grupos de Trabalho, Interesse e Coordenação para realizar o trabalho. Membros do W3C revisam cada Proposta de Atividade e os Grupos de Interesse associados. Quando há suporte dentro do W3C para investir recursos no tópico de interesse, o diretor aprova as atividades e os grupos iniciam seus trabalhos;
3. Os tipos de participantes dos Grupos de Trabalho do W3C são três: Membros Representantes, Especialistas Convidados e Representantes de Grupo;
4. Os Grupos de Trabalho geralmente criam especificações e guias que passam por ciclos de revisão até avançarem para o estágio de recomendação. Para chegar a esse estado, o W3C exige que os relatórios técnicos sejam amplamente revisados pelos membros e pelo público, e o grupo de trabalho deve atender requisitos de implementação e interoperabilidade.

Até que um documento chegue a um consenso e seja publicado como uma recomendação do W3C, ele passa pelas seguintes fases (JACOBS, 2005):

1. Publicação do primeiro rascunho de trabalho público;
2. Anúncio de última chamada;
3. Chamada para implementação;
4. Chamada para revisão de uma proposta de recomendação;
5. Publicação de uma recomendação.

Atualmente, diversos padrões são mantidos ou apoiados pelo W3C, inclusive os que são a sua base, HTTP, HTML e URI. Novas versões desses e outros padrões já estabelecidos são

desenvolvidos, e analisados pelos grupos de trabalho do consórcio. Além disso, novas recomendações de padrões são desenvolvidas conforme as necessidades levantadas pelos membros.

1.2 O despertar para a Web

Apesar de seu grande potencial, a web ainda não era muito disseminada até meados de 1996, até porque seu custo ainda era elevado. Neste ano, empresas perceberam a possibilidade de publicar informação mundialmente de forma instantânea e, a grande possibilidade de comércio pela Web. A partir desse momento, a web iniciou um grande crescimento, tanto na quantidade de informações e serviços disponibilizados, quanto no número de utilizadores (FERRACIOLI; CAMARGO-BRUNETTO, 2007). Um dos acontecimentos mais marcantes dessa época foi a chamada guerra dos *browsers*, travada principalmente entre o Microsoft Internet Explorer e Netscape Navigator (GROMOV,). Durante essa época, os padrões web eram desenvolvidos e disponíveis como hoje, entretanto, eram praticamente inexistentes nas implementações dos *web sites*, que utilizavam tecnologias proprietárias presentes, na maioria das vezes, em apenas um dos *browsers*. Essa época foi marcada por grandes dificuldades para o desenvolvimento de sistemas web interoperáveis, visto que os padrões web praticamente não eram seguidos e tecnologias proprietárias acabaram se difundindo (FERRACIOLI; CAMARGO-BRUNETTO, 2007).

Talvez isso se deva a necessidade de web sites com cada vez mais recursos, de maneira ágil. Visto que a definição de um padrão é um processo lento, as desenvolvedoras de *browsers* desenvolviam suas próprias soluções rapidamente, oferecendo uma alternativa ao mercado (FERRACIOLI; CAMARGO-BRUNETTO, 2007). Isso atraía mais usuários para os *sites*, entretanto aumentava o esforço de implementação e os custos. Devido à difusão do sistema operacional Microsoft Windows, que já vinha com o navegador Microsoft Internet Explorer por padrão, o Netscape acabou perdendo mercado e, em 1998 abriu seu código fonte, o que deu origem aos navegadores Mozilla, presentes até hoje no mercado.

A preocupação com a semântica de documentos começa a se tornar um conceito mais evidente a partir desse período. Durante a guerra dos navegadores foi percebido pelos desenvolvedores que a utilização dos padrões Web, principalmente os fornecidos pelo W3C, poderiam melhorar a produtividade, diminuir custos, aumentar a interoperabilidade e facilitar a manutenção. Também havia o benefício relacionado aos motores de busca, visto que um documento semanticamente correto possui mais visibilidade nos mesmos. Entretanto, o principal motivo da adoção dos padrões foi justamente a interoperabilidade, visto que um código semanticamente correto possuía mais chances de funcionar corretamente em todos os *browsers*, evitando assim a necessidade de diversas implementações para um mesmo site. A única coisa que não garantia esse comportamento, era a conformidade dos *browsers* com os padrões

web e o desenvolvimento de tecnologias proprietárias nos mesmos, o que vivenciamos até os dias atuais. A principal mudança originária dessa época foi a preocupação com a separação do conteúdo, apresentação e comportamento. Para isso, deveriam ser utilizadas as linguagens apropriadas para cada um desses recursos, o que caracteriza uma semântica mais correta na utilização de elementos e um aumento na preocupação relacionada ao conteúdo. Essa tendência foi aplicada no desenvolvimento do HTML 4.1, que eliminou ou tornou depreciados diversos elementos que tinham fins de apresentação, alguns inclusive criados durante a guerra dos *browsers* (FERRACIOLI; CAMARGO-BRUNETTO, 2007).

Entre 1999 e 2001 ocorreu o chamado “boom .com”, onde foram realizados grandes investimentos de capital em novos negócios na web. Este “boom” é comparável a outros já ocorridos na sociedade, como o das estradas de ferro, rádio, transistor, entre outros. Segundo o Miniwatts Marketing Group, em dezembro de 2000, o número de usuários chegou a 360.985.492 de pessoas³. Após esse período, o crescimento da web continuou, agregando cada vez mais novos usuários e serviços (GROMOV,). Entretanto, pouco tempo depois, a própria web se reinventou. O usuário final possuía um papel muito passivo, apenas como um simples consumidor. O aumento da interatividade dos sistemas com o usuário e a possibilidade de criação de conteúdo são premissas da chamada Web 2.0, uma forte tendência atual. Possui grande cunho social e reinventa o conceito de usuário, sendo este também responsável pela criação da própria aplicação que utiliza, o que proporciona uma melhor experiência para o mesmo. É interessante notar que o conceito além de cunho social, possui grande fundamentação tecnológica baseada em padrões para Web, promovendo assim a sua utilização (O'REILLY, 2005). Atualmente, segundo o Miniwatts Marketing Group, o número de usuários continua crescendo em grandes porcentagens, chegando a 1.463.632.361 de usuários, um crescimento de 305% em relação ao ano 2000. A próxima revolução tende a ser a Web Semântica, conceito que apareceu pela primeira em 2001 com um artigo publicado na *Scientific American* pelo criador da WWW, Tim Berners-Lee (BERNERS-LEE; HENDLER; LASSILA, 2001), assunto que será abordado posteriormente no capítulo 2.

1.3 Desenvolvimento Tecnológico da Web

As linguagens para marcação para Web são grandes responsáveis pela sua disseminação e essenciais para o funcionamento da mesma. Mantidas em sua maioria pelo W3C, possuem toda documentação com acesso fácil e por serem padrões abertos, não impedem a sua utilização devido a restrições de licença. Nesta seção descrevemos as principais linguagens de marcação de conteúdo para web disponíveis atualmente.

³<http://www.internetworldstats.com/stats.htm>

1.3.1 HTML

Das páginas simples, às mais avançadas, possuem a maior parte de sua codificação em HTML, a linguagem de marcação essencial para o funcionamento da Web. Como já mencionado, sua primeira versão surgiu em 1990, junto com a disponibilização da WWW ao público. É uma linguagem derivada do *Standard Generalized Markup Language* (SGML) utilizada para publicar informação globalmente. Por isso é necessário que tanto autores quanto vendedores de páginas web compartilhem as mesmas convenções para o HTML. Entretanto, esse consenso não é simples de ser atingido, o que pode ser exemplificado pela não adoção do HTML 3.0, durante a guerra dos *browsers*, visto que um consenso entre as empresas interessadas nunca acontecia (HOURS; JACOBS, 1999).

A linguagem foi popularizada com a criação do *browser* Mosaic, e teve sua segunda versão especificada como um padrão da *Internet Engineering Task Force* (IETF) (BERNERS-LEE; CONNOLLY, 1995). Como já mencionado, o HTML 3.0 nunca foi lançado como um padrão web, devido à falta de consenso. Apenas a versão 3.2 atingiu esse estado, entretanto não foi de grande representatividade, devido a tecnologias proprietárias presentes na especificação e a baixa adoção dos padrões web na época (HOURS; JACOBS, 1999).

Atualmente o HTML se encontra na versão 4.1, publicada pelo W3C como uma recomendação em 1999 estendendo a versão 4.0. A partir dessa versão foi possível adicionar folhas de estilo, *scripts*, *frames*, melhor suporte a texto com direções diferenciadas, tabelas mais ricas, melhorias em formulários, maior suporte a internacionalização e melhorias na parte de acessibilidade para pessoas com deficiências (HOURS; JACOBS, 1999).

Encontra-se em desenvolvimento a nova versão do HTML, que chegará à sua versão 5. Atualmente ele se encontra em estágio de *Working Draft*, e possui como um dos pilares de desenvolvimento a interoperabilidade definindo uma conformidade clara para os agentes de usuário (HICKSON; HYATT, 2008).

1.3.2 XML

Apesar do HTML ser uma linguagem muito útil, ele possui um conjunto de regras muito restrito e com um objetivo bastante específico. Surgiu a necessidade de criar documentos de maneira semelhante, entretanto de maneira mais aberta, permitindo publicar documentos mais específicos e em larga escala. Devido a essa necessidade surgiu o *eXtensible Markup Language* (XML), também derivado do SGML. Entretanto, devido a sua grande portabilidade, o XML também é largamente utilizado como uma ferramenta para fornecer interoperabilidade, visto que pode ser utilizado para troca de dados em qualquer ambiente. Sua primeira versão recomendada surgiu em 1998, e atualmente encontra-se na versão 1.1 (Fourth Edition). O

XML fornece um mecanismo para impor restrições sobre o *layout* de armazenamento e estrutura lógica (BRAY et al., 2006).

Os objetivos a serem alcançados durante seu desenvolvimento eram:

- Ser diretamente utilizado na Web;
- Suportar uma grande variedade de aplicações;
- Compatibilidade com o SGML;
- Ser fácil de escrever programas que interpretem o XML;
- Recursos adicionais deveriam ser minimizados, idealmente nenhum;
- Possibilidade de leitura e compreensão por humanos com clareza;
- O projeto do XML deve ser feito rapidamente, formal e conciso;
- Documentos XML devem ser fáceis de criar;
- Marcações concisas são de importância mínima.

Neste ano, comemora-se os 10 anos de lançamento da versão inicial do XML, que obteve um grande sucesso, visto que é largamente utilizado em diversos tipos de aplicações, além de ser a base para diversas outras linguagens de grande importância, como MathML, XHTML, SOAP, entre outras. Suas linguagens derivadas proporcionam uma maneira otimizada de utilização do XML em campos mais específicos.

Todo o potencial do XML foi atingido com o lançamento da recomendação do XML Schema em 2001, uma linguagem de marcação para definição de regras de formação de documentos XML. As construções são utilizadas para restringir e documentar o significado, relacionamentos e as partes constituintes de um documento XML. Ou seja, o XML Schema pode ser utilizado para definir, descrever e catalogar vocabulários XML para classes de documentos XML (MCQUEEN; THOMPSON, 2008).

1.3.3 XHTML

O *eXtensible HyperText Markup Language* é uma reformulação do HTML em XML, e de suas três *Document Type Definitions* (DTD), que possuem as definições da linguagem HTML 4.0, visto que o XHTML 1.x é derivado da última versão do HTML. Esses documentos são desenvolvidos para trabalhar em conjunto com agentes de usuário⁴ baseados em tecnologias

⁴http://en.wikipedia.org/wiki/User_agent

XML. Ele pretende ser utilizado como linguagem para conteúdo que esteja em conformidade tanto com XML, e se simples guias forem seguidos, operar em conformidade com HTML 4 em agentes de usuário. Os benefícios ao se utilizar o XHTML são:

- Documentos XHTML são documentos em conformidade com XML. Dessa maneira, eles são rapidamente lidos, editados, e validados com ferramentas padrão XML;
- Documentos XHTML podem ser escritos para operar como ou melhor do que era obtido anteriormente com agentes de usuário HTML 4;
- Documentos XHTML podem utilizar aplicações que são invocadas no HTML *Document Object Model* ou XML *Document Object Model* (DOM);
- Como a família do XHTML evolui, documentos em conformidade com XHTML irão ser mais interoperáveis dentro dos e entre vários ambientes XHTML.

A necessidade de se utilizar XHTML existe pois é mais simples adicionar novos elementos a documentos XML e também ao se pensar em interoperabilidade entre os diferentes agentes de usuário, mais facilmente alcançável com XML, que já provou ser uma grande ferramenta no auxílio à interoperabilidade (PEMBERTON et al., 2002).

O XHTML foi publicado como recomendação em 2000, e atualmente possui a versão 1.1 em estado de *Working Draft* no W3C (MCCARRON et al., 2007). Também está sendo desenvolvida em paralelo uma nova versão, a XHTML 2.0, que também se encontra no estado de *Working Draft*. Essa versão pretende ser a mais genérica possível em termos de XML; ter mais estruturas de conteúdo, menos de apresentação; oferecer maior usabilidade e acessibilidade; melhor suporte a internacionalização; maior independência de dispositivo; menor utilização de *scripts* e integração com Web Semântica. XHTML 2.0 essencialmente consiste em empacotar diversas partes que atualmente são independentes, mas estão em processo de recomendação (AXELSSON et al., 2006):

- RDFa;
- XForms;
- Access;
- Role;
- XML Events.

É interessante notar que não existe uma conformidade da versão 5 do HTML com a versão 2 do XHTML, sendo as duas linguagens distintas para definição de documentos de apresentação

na Web. Ambos padrões estão em fase de desenvolvimento, possuem abordagens diferentes e coexistem, e desde já estão em busca de público (JONGE, 2007). Essa questão é algo contraditório, pois nem mesmo o órgão que regulamenta os padrões web trabalha com apenas um padrão para a próxima versão do (X)HTML, o que pode gerar resultados como os que aconteciam nos anos onde apesar de existir padrões, não havia a sua utilização.

1.3.4 Um antigo problema volta a aparecer

Devido à descentralização, a facilidade de publicação e simplicidade, o número de informações disponíveis *online* cresceu rapidamente. Isso trouxe possibilidades excitantes para acesso à informação e comércio eletrônico, por exemplo (FENSEL et al., 2003). Apesar do projeto original da web tentar solucionar um problema de localização de informação, o mesmo problema voltou a surgir devido às características da própria web. Entretanto, o problema possuía um agravante: muitas vezes a informação encontrada não era confiável ou era irrelevante. O usuário é que deveria filtrar as informações de um documento para descobrir se realmente tratava de um assunto de seu interesse.

Essa tarefa é frustrante, pois, na maioria dos casos, necessita-se da informação de maneira rápida, sendo o processo de filtragem de dados um tempo desperdiçado. Por que um computador não pode realizar essa tarefa? Esse problema pode ser ilustrado através do mais clássico exemplo, os mecanismos de busca. Mesmo com as várias melhorias em seus algoritmos, continuam a retornar resultados com baixa precisão, baixa sensibilidade ao vocabulário e resultados individuais de páginas (BREITMAN, 2005).

Foi então que começaram as atividades para descrição das informações disponíveis na web, ou informação sobre a informação.

A solução encontrada foi a utilização de Metadados. Através deles, é possível a um computador encontrar a informação e processá-la, ou então ao encontrá-los é possível compreender de que assunto trata um determinado recurso (BREITMAN, 2005). Um grupo responsável por Metadados foi criado pelo W3C, entretanto, ele rapidamente foi substituído por um grupo chamado *Semantic Web Activity* (SWICK, 2002).

1.3.5 Web Semântica

Metadados formam a base para uma web mais inteligente e confiável, prevista desde seu projeto original, mas ainda indisponível. Ela se tornou possível a partir do momento que o XML e o XML Schema se tornaram linguagens maduras e amplamente difundidas. O *Semantic Web Activity* é o grupo de trabalho responsável pela Web Semântica. Ela fornece um *framework* comum permitindo que dados sejam compartilhados e reutilizados em toda aplicação, empresa

e comunidade (HERMAN, 2008b).

A Web Semântica é a Web dos dados, que atualmente são controlados pelas aplicações, cada uma mantendo seus próprios. A Web Semântica trata de dois pontos (HERMAN, 2008b):

- Formatos comuns para integração e combinação de dados retirados de diversas fontes, onde a web original concentra-se principalmente na troca de documentos.
- Também trata de uma linguagem para gravar como os dados se relacionam com objetos do mundo real. Isso permite a uma pessoa ou máquina, iniciar num banco de dados e então se mover através de um conjunto de bancos de dados conectados não por fios, mas por possuírem o mesmo conceito.

A Web Semântica é tida como solução para os problemas que enfrentamos na Web atual, como serÃj abordado no capítulo 2.

2 Web Semântica

A Web Semântica é uma extensão da web já existente, onde a informação ganha melhores significados, proporcionando aos humanos trabalhar melhor em conjunto com os computadores (BERNERS-LEE; HENDLER; LASSILA, 2001). Verifica-se, que desde o projeto de Tim Berners-Lee quando inventou a WWW em 1989 (BERNERS-LEE, 1989), existia a preocupação de que ela pudesse ser compreendida tanto por humanos quanto por agentes inteligentes (FENSEL et al., 2003). Novamente sua importância foi mostrada ao público por Tim Berners-Lee durante a primeira conferência da *World Wide Web* em 1994 (SHADBOLT; HALL; BERNERS-LEE, 2006). Entretanto, somente em 2001 surgiu a primeira publicação específica sobre o assunto. Em um artigo, também escrito por Tim Berners-Lee na revista *Scientific American*, foi abordado um novo mundo de possibilidades na web, que até o momento eram desejadas, mas tidas como inviáveis (BERNERS-LEE; HENDLER; LASSILA, 2001).

Através da Web Semântica, de modo geral, deverá ser possível localizar dados em qualquer lugar, tornando-os acessíveis e compreensíveis, tanto a computadores quanto a máquinas (PASSIN, 2004), o que resgata a parte não implementada do projeto original (FENSEL et al., 2003). Desde o surgimento da Web, o caminho para Web Semântica veio sendo pavimentado ao longo dos anos. Ao ter seu grupo de trabalho oficializado no W3C em 2001, a web semântica já possuía uma base de tecnologias disponíveis para seu desenvolvimento: XML e RDF (BERNERS-LEE; HENDLER; LASSILA, 2001).

2.1 Faces da Web Semântica

É possível encarar a Web Semântica através de diversas perspectivas, devido às suas diversas possibilidades. Podemos encará-la como (PASSIN, 2004):

- Visão dos dados compreensíveis por máquinas: a web semântica é uma visão, da ideia de ter dados na web definidos e associados de maneira que possam ser utilizados por máquinas e não apenas para fins de apresentação, mas também automação, integração e reutilização de dados entre várias aplicações;
- Visão de sistemas inteligentes: fazer com que a web atual seja mais compreensível por

máquinas, permitindo a agentes inteligentes recuperar e manipular informação pertinente;

- Visão de bancos de dados distribuídos: transformar o processamento de informação provendo um meio comum para acesso dos dados, interligados juntos e compreensíveis. Para transformar a web de um grande livro ligado em um grande banco de dados ligado;
- Visão de infra estrutura automatizada: visto que ela não é uma aplicação, mas sim uma infra-estrutura para elas;
- Visão de servo da humanidade: permitir aos computadores livrar-nos da carga de localizar recursos na Web que são relevantes para nossas necessidades e, então, extrair, integrar e indexar a informação;
- Visão de melhores metadados: suprir a web que conhecemos atualmente com anotações expressas de uma forma processável por máquinas e interligadas;
- Visão de melhores buscas: acessar recursos web pelo seu conteúdo, e não apenas por palavras-chave;
- Visão de *web services*: não fornecer apenas documentos com informações úteis, mas também serviços que tenham comportamento adequado.

Cada uma dessas perspectivas possui aplicações em diversos ramos, facilitando ou melhorando resultados de tarefas realizadas atualmente. Além disso, novas perspectivas podem ser vislumbradas. Um exemplo clássico do que a web semântica é capaz de realizar pode ser exemplificado através dos motores de busca. As buscas tradicionais, realizadas atualmente, trazem seus resultados baseados em palavras chave fornecidas pelo usuário, ficando a cargo deste realizar a filtragem da informação, verificando se ela realmente atende suas necessidades. Através de uma pesquisa com motores semânticos, é possível realizar uma pergunta ao motor de busca que, compreendendo a semântica de cada conceito, suas relações e combinando-os, interpreta o significado da pergunta e busca a resposta mais efetiva possível (FEIGENBAUM et al., 2007), como disponível na ferramenta de busca True Knowledge⁵. A utilização de tecnologias de Web Semântica está se difundindo, e sendo adotada como diferencial mercadológico, como no caso do Yahoo! que anunciou a integração de semântica a suas buscas⁶.

Apesar dos casos de sucesso e de todo potencial que possui, a Web Semântica não será algo visível aos usuários. Isso se deve ao fato dela ser uma camada adicional integrada que melhora a web atual (FEIGENBAUM et al., 2007). A interação do usuário com funcionalidades semânticas será totalmente transparente, auxiliando inclusive na utilização dos sistemas.

⁵<http://www.trueknowledge.com/>

⁶<http://www.ysearchblog.com/archives/000527.html>

2.2 Problemas que a Web Semântica pode resolver

Através da web, novas possibilidades de acesso a informação e comércio eletrônico surgiram, o que se deve principalmente à sua simplicidade. Entretanto, a mesma facilidade impede um crescimento ainda maior. A área de representação de conhecimento, por exemplo, devido à globalização geográfica de organizações, centraliza documentos na web. Para esse campo é possível fornecer definições estruturais e semânticas de documentos que permitirão possibilidades completamente novas: buscas inteligentes, ao invés de busca por palavras-chave, busca por perguntas ao invés de recuperação de informação, troca de documentos entre departamentos através de mapeamentos de ontologias, como visto na seção 2.4.3, e definição de visões customizadas de documentos (FENSEL et al., 2003).

Para área do comércio eletrônico, a web representa uma extensão de seus modelos de negócios, reduzindo custos e estendendo os canais de distribuição existentes e traz, em alguns casos, novas possibilidades de distribuição. Também consegue fornecer modelos de negócio completamente novos, ou dar mais importância do que possuíam sem semântica associada. Atualmente, temos disponíveis os serviços de agregadores de comércio eletrônico, que buscam informações sobre determinado produto em lojas pré-determinadas. Apesar de não utilizarem somente a busca por palavra-chave, essa tecnologia possui algumas limitações. O esforço de implementação para desenvolver uma ferramenta desse tipo é muito grande, e qualquer mudança no *layout* de uma loja assistida consome grande tempo de manutenção. A qualidade da informação extraída não é das melhores, limitada normalmente a uma terminologia, passível de erros e incompleta. Para solucionar esses tipos de problema seriam necessárias semânticas processáveis por máquinas para a informação fornecida. A programação em baixo nível desses robôs baseados em extração de texto será substituída por mapeamentos semânticos que traduzem diferentes formatos utilizados para representar produtos e poderão ser utilizadas para navegar e encontrar automaticamente a informação requerida (BREITMAN, 2005).

O comércio entre empresas também se beneficiará. Para que as transações sejam realizadas entre as empresas, é necessário que elas entrem em acordo para utilização de um padrão de troca de informações. Com o surgimento do XML, isso se tornou mais simples através da web, entretanto ele não fornece estruturas de dados e terminologias padrão para descrever processos de negócio e troca de produtos. A tecnologia da Web Semântica irá ocupar papéis importantes no comércio eletrônico em XML. Linguagens com modelos de dados melhores definidos permitirão definição, mapeamento e troca de dados de produtos, ontologias cobrirão várias áreas de negócio, traduções eficientes de serviços serão necessárias em áreas em que tecnologias padrão não existem. Esse serviço de tradução cobrirá diferenças semânticas estruturais e de linguagem (FENSEL et al., 2003).

2.3 Esclarecimentos sobre o conceito

Devido a algumas de suas características, é possível inferir dados sobre a Web Semântica que não traduzem a sua realidade.

Um dos mais comuns é pensar que a Web Semântica é uma Web separada, o que aconteceu similarmente com a chamada Web 2.0. Assim como a Web 2.0, a Web Semântica será uma extensão da Web atual, onde a informação terá significado definido através de linguagens de marcação semântica. Essas linguagens serão acrescentadas às páginas atuais, segundo uma arquitetura definida (BREITMAN, 2005). Entretanto, esses recursos serão praticamente invisíveis aos usuários, que sentirão somente seus efeitos, trabalhando melhor em conjunto com os computadores, um dos objetivos de seu projeto (BERNERS-LEE; HENDLER; LASSILA, 2001).

Outro erro comum é pensar que a Web Semântica é Inteligência Artificial (IA). Isso é tido como falso já no projeto da Web Semântica, em um dos casos de uso citados no artigo de Tim Berners-Lee (BERNERS-LEE; HENDLER; LASSILA, 2001). Nele, um agente que visita páginas e traz informações através de inferência, não fazendo a utilização de Inteligência Artificial, mas apenas através da utilização da semântica fornecida corretamente na página (BERNERS-LEE; HENDLER; LASSILA, 2001). Não é necessário que os computadores passem a entender a linguagem das pessoas, as pessoas que deverão realizar o esforço de codificar a informação em representações onde seja possível o processamento automático (BREITMAN, 2005). Esse pensamento talvez seja em decorrência da necessidade de utilização de técnicas e ferramentas derivadas dos 50 anos de pesquisa da Inteligência Artificial na Web Semântica (SHADBOLT; HALL; BERNERS-LEE, 2006). Apesar disso, não será necessário o mesmo nível de inteligência desejado pelos pesquisadores de IA, visto que a Web Semântica pretende auxiliar humanos, não superá-los.

Mas afinal, o que é a Web Semântica? Podemos buscar diversas definições para a Web Semântica, pois, como já abordado anteriormente, ela pode variar conforme o foco tido como objetivo. Dentre as várias definições que se pode assumir, a que mais possui embasamento é a de seu precursor, Tim Berners-Lee. Para ele, a Web Semântica irá envolver a web atual, fazendo com que informações e serviços sejam compreendidos por humanos e máquinas (BERNERS-LEE; HENDLER; LASSILA, 2001), ou criar uma Web para as máquinas (GLOBE, 2003).

Para que ela seja possível, alguns requisitos devem ser atendidos.

- Desenvolvimento de linguagens para expressar meta-informação de maneira compreensível por máquinas, desenvolver terminologias utilizando essas linguagens e disponibilizá-las na web;
- Desenvolver ferramentas e novas arquiteturas que usem essas linguagens e terminologias

provendo suporte na localização, acesso, apresentação e manutenção das origens de informação.

2.3.1 Estrutura da Web Semântica

Visto que a web deverá ser estendida pela Web Semântica, é importante que seu projeto e arquitetura sejam compatíveis com a estrutura atual da web. A web é projetada sobre recursos, endereços padronizados desses recursos e uma pequena quantidade de comandos amplamente compreendidos. Também faz parte de seu projeto que funcione em redes de maneira descentralizada.

Um recurso representa qualquer idéia que possa ser referenciada. Sua noção é flexível o bastante para atingir qualquer tipo de recurso. Um recurso pode mudar durante o tempo e ainda assim é considerado o mesmo recurso, endereçado pela mesma URI. Alternativamente, uma URI pode apontar para uma versão específica do mesmo documento que nunca muda. Um recurso nunca é acessado diretamente, mas apenas a sua representação. Por fim, um recurso pode ser algo que nem exista, uma referência a algo fictício. Tudo isso pode ser identificado como um recurso, através de uma URI na Web Semântica. Padronização desse endereçamento é necessária, de maneira que qualquer informação que possua uma URI possa ser colocada em um *hyperlink* e ser associada.

O protocolo HTTP utiliza um conjunto pequeno de instruções, conhecidos universalmente por servidores web, agentes de usuário e componentes intermediários. Ainda é uma questão aberta se a Web Semântica deverá se restringir a apenas esse pequeno conjunto de instruções, presente na arquitetura atual da web.

A Web precisa operar em uma grande rede com uma grande quantidade de *web sites*, e conforme o crescimento da rede, deve continuar operando. Isso se deve primeiro à sua descentralização e segundo, pelo fato de cada transação possuir todos os dados necessários para que seja completada. Nenhum dado deve ser guardado no servidor de uma requisição para outra, por isso as transações na web são consideradas *stateless*.

Visto que a web não possui controles centrais, é considerada aberta. Ela é incompleta pois nada pode garantir que todo link irá funcionar ou que toda a informação estará disponível. Também é inconsistente, visto que qualquer um pode publicar qualquer coisa em uma página web, o que facilmente gera contradições entre diferentes páginas.

A partir dessas características estruturais da web é que surgirá a Web Semântica. Os aspectos que ela deve seguir são (BERNERS-LEE; HENDLER; LASSILA, 2001), (PASSIN, 2004), (SHADBOLT; HALL; BERNERS-LEE, 2006), (KOIVUNEN; MILLER, 2001):

- Utilizar o endereçamento baseado em URL;

- Possuir noções de recursos endereçáveis e não endereçáveis;
- Utilizar protocolos com um conjunto de comandos pequeno e universalmente conhecidos;
- Manter o atributo de *stateless*;
- Ser o mais descentralizada possível;
- Funcionar em larga escala;
- Permitir *cache* local de informações diminuindo a carga do servidor e reduzindo custos de rede;
- Ser capaz de operar com *links* perdidos ou informação inconsistente.

Complementando, também é necessário que tanto recursos quanto *links* tenham tipos, de maneira que uma máquina possa compreender. Essa necessidade é bem exemplificada pela figura 2.1.

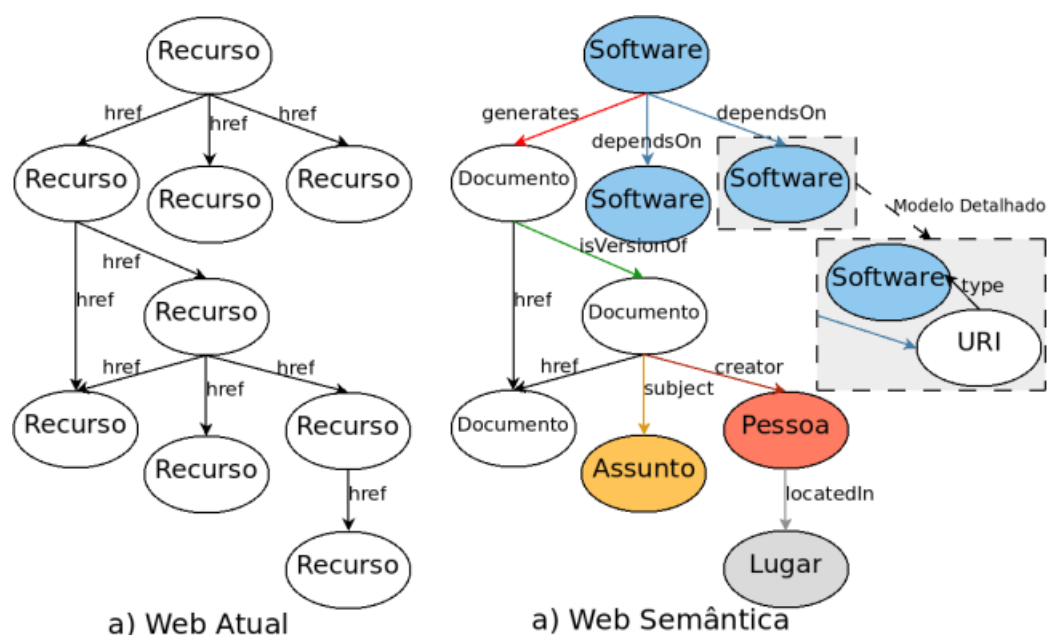


Figura 2.1: Recursos e links com tipos na Web Semântica (Adaptação de (KOIVUNEN; MILLER, 2001))

Note que esse é um dos requisitos pensados no projeto da web (BERNERS-LEE, 1989).

Todos esses princípios fazem parte das camadas da Web Semântica, que pode ser vista na seção 2.3.2.

2.3.2 As Camadas de bolo da Web Semântica

Proposta por Tim Berners-Lee durante a conferência de XML de 2000, a arquitetura em camadas da Web Semântica também é conhecida como “Camadas do Bolo da Web Semântica” (BREITMAN, 2005). A idéia por trás desse modelo é ao invés de propor uma arquitetura totalmente nova para a Web, utilizar a já existente e construir a Web Semântica em cima do que já existe. Todos os princípios de Web Semântica são implementados nessas camadas de tecnologias Web e padrões (KOIVUNEN; MILLER, 2001), conforme ilustrado pela figura 2.2.

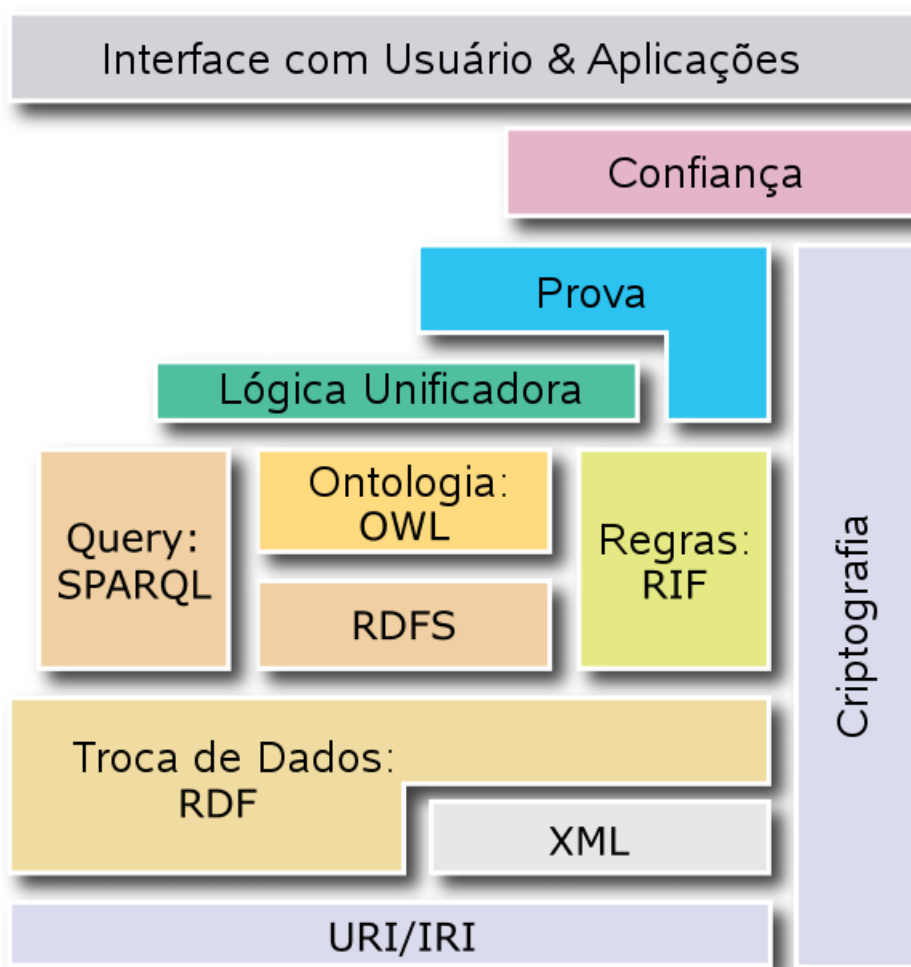


Figura 2.2: “Camadas do bolo da Web Semântica” (Adaptação de (CONSORTIUM,))

Construir sobre o que já existia era preciso, pois uma quebra com a arquitetura passada seria uma mudança muito drástica, devido à grande quantidade de páginas existentes. Modelar em pequenas camadas era necessário, pois pequenas mudanças alcançam consenso mais rapidamente (BREITMAN, 2005).

As diferentes camadas estão em estágios de evolução distintos, mas progredem rapidamente. Tal progresso é mensurado numa escala que começa com a concepção de uma idéia, trabalhando para sua aceitação como uma tecnologia comum e torna-se um padrão intero-

perável, terminando com o mercado global baseando-se nela (FENSEL et al., 2003). Cada camada é vista como construção da camada anterior, sendo cada camada progressivamente mais especializada e complexa que as abaixo dela. Uma camada de baixo não é dependente de nenhuma camada acima, por isso, as camadas podem se desenvolver e operar independentemente (PASSIN, 2004). Visto que a Web Semântica ainda não possui todos os seus padrões e tecnologias estabelecidos, essa visão mudou e continua mudando constantemente. A versão mais atual do Grupo de Trabalho da Web Semântica⁷ no W3C pode ser conferida em(CONSORTIUM,).

Em versões anteriores, o HTML aparecia explicitamente como a primeira camada. Devido à sua simplicidade, ele originou a necessidade do desenvolvimento de linguagens mais sofisticadas, pois ele descreve apenas unidades genéricas e sua ordem, o que tem extremo sucesso apenas quando a informação é destinada apenas a pessoas. Com XML, é possível descrever a estrutura da informação de outras maneiras, escolhendo o tipo de estrutura mais apropriada para uma informação em particular. Por esse motivo, o XML é tido como a camada de fundação da Web Semântica. O XML Schema fornece a habilidade de especificar estruturas e tipos de dados a serem usados por um documento XML particular (PASSIN, 2004).

Na próxima camada encontra-se o *Resource Description Framework* (RDF), que tem o papel de fornecer um modelo formal de dados e sintaxe para codificar metadados para serem processados por máquinas. O princípio do RDF é fornecer interoperabilidade entre aplicativos que trocam informações na rede (BREITMAN, 2005). Para isso, ele define os significados em dois níveis: atribuindo propriedades e relações entre elas. O RDF Schema descreve essas propriedades, o que são, a quais recursos elas podem estar atribuídas, entre outros.

A camada de ontologia leva esse passo mais além: não apenas realiza a descrição de propriedades e termos que podem ser usados mas também descrevem os relacionamentos entre elas. Essa camada e a camada anterior são utilizadas para descrever ou representar conhecimento (PASSIN, 2004). Também é a camada mais pesquisada dentre as outras (FREITAS, 2003).

A camada *Simple Protocol and RDF Query Language* (SPARQL) pode ser usada para fazer buscas entre diversas origens de dados, se os dados estiverem guardados nativamente em RDF ou vistos como RDF através de um *middleware*. Possui capacidade de buscar por padrões de grafos obrigatórios e opcionais juntamente com suas conjunções e disjunções (PRUD'HOMMEAUX; SEABORNE, 2008).

A camada *Rule Interchange Format* (RIF) permite a criação de regras que podem ser utilizadas em diversos sistemas. Funciona como uma interlíngua que estabelece uma nova regra de idioma que pode ser mapeada, permitindo que as regras de uma aplicação sejam publicadas,

⁷<http://www.w3.org/2001/sw/>

compartilhadas e reutilizadas em outras aplicações e motores de regra (HAWKE, 2005).

A camada de lógica permite escrever regras enquanto a camada de prova executa as regras e as avalia em conjunto com o mecanismo da camada de confiança para aplicações, verificando se a prova é correta ou não (KOIVUNEN; MILLER, 2001). Para que essas camadas entrem em operação é necessário que as demais estejam bem fundamentadas, o que ainda não aconteceu (FREITAS, 2003).

2.4 O Semantic Web Activity

O *Semantic Web Activity* é o grupo de trabalho do W3C responsável pela Web Semântica. Em sua página⁸ é possível encontrar informações em diferentes formatos e com diferentes exigências de conhecimento sobre a Web Semântica. Devido a isso, caracteriza-se como uma das fontes mais completas sobre o assunto. Seus subdomínios são organizados em áreas mais específicas, oferecendo informação mais focada.

Na página é possível encontrar as especificações correntes da Web Semântica, publicações em formatos variados e de diferentes fontes, apresentações de eventos internacionais pertencentes a membros do W3C sobre o assunto e relacionados, informações sobre os subgrupos ativos, completados, passados e outros grupos do W3C relacionados a Web Semântica. Também é possível encontrar casos de uso, validadores, últimas atividades e notícias entre outras informações.

Um dos recursos mais importantes são as especificações das linguagens que fazem a Web Semântica facilmente acessíveis, juntamente com seus casos de uso em organizações que já as utilizam. Dentro dos grupos de trabalho específicos encontram-se casos de uso mais focados, como no caso do *Health Care and Life Sciences Interest Group* (HCLSIG), um dos focos de estudo do capítulo 3. Da mesma maneira que as linguagens que originaram a Web, espera-se que as linguagens que darão suporte a Web Semântica possuam a mesma robustez, possibilitando fácil utilização para aplicações com variados níveis de complexidade.

2.4.1 Linguagens

Por que não XML e XML Schema?

Apesar do XML permitir a qualquer um criar suas próprias *tags*, para que as mesmas sejam utilizadas é necessário conhecer o que é utilizado em cada *tag* (BERNERS-LEE; HENDLER; LASSILA, 2001). Apenas é possível, através do XML, escolher o tipo de estrutura mais apropriada para uma informação em particular. Entretanto, XML não fornece semântica

⁸<http://www.w3.org/2001/sw/>

explicitamente, visto que as *tags* são apenas etiquetas interpretadas pelas aplicações (GLOBE, 2003), oferecendo apenas estrutura (FENSEL et al., 2003). Mais ainda, o XML não fornece estruturas de dados e terminologias para descrever processos de negócio e troca de produtos, apesar de ter um grande papel com relação a interoperabilidade de dados (FENSEL et al., 2003). Mas, justamente por suas características principais que ele é tido como a camada de fundação da Web Semântica (PASSIN, 2004), sendo inclusive base para linguagens da Web Semântica.

2.4.2 RDF e RDFS

O *Resource Description Framework* é uma linguagem de representação de informação sobre recursos na *World Wide Web*, com intenção particular de representar metadados sobre recursos Web (MANOLA; MILLER; MCBRIDE, 2004). De maneira mais técnica, trata-se de uma aplicação XML que adiciona um modelo de dados simples sobre este. Este modelo de dados fornece três elementos (FENSEL et al., 2003):

- Sujeito (recurso);
- Predicado (propriedade);
- Objeto (valor).

Possui como visão o fornecimento de uma estrutura de representação de conhecimento minimalista para Web (SHADBOLT; HALL; BERNERS-LEE, 2006). Uma linguagem com o papel do RDF deve ser capaz de descrever a maior parte dos dados que serão disponíveis, projeto estrutural de conjuntos de dados e os relacionamentos entre dados (PASSIN, 2004). Deve ser utilizado em situações onde a informação precisa ser processada por aplicações, e não apenas apresentada as pessoas. Visto que ele fornece um *framework* comum para expressar a informação, ela pode ser trocada entre aplicações sem perda de significado (MANOLA; MILLER; MCBRIDE, 2004).

O RDF é escrito em frases do tipo Recurso, Propriedade e Valor, compreendidos como sujeito, predicado e objeto de uma sentença (BREITMAN, 2005). Uma declaração RDF é chamada de tripla, pois possui três partes.

(pessoa –1, nome , ‘ ‘Nome Sobrenome ’ ’)

Figura 2.3: Exemplo de uma tripla

A figura 2.3 possui uma tripla, onde cada valor representa uma parte de uma declaração RDF:

- Sujeito: pessoa-1;
- Predicado: nome;
- Objeto: Nome Sobrenome.

Os valores de um objeto são chamados de literais, e os das propriedades podem ser um literal ou outro recurso. Já os sujeitos não podem assumir valores literais. Para identificar recursos, o RDF utiliza uma referência de URI, que é uma URI mais caracteres opcionais, chamados identificadores de fragmentos, contidos após o caractere '#' na URI (PASSIN, 2004).

O RDF tornou-se uma recomendação do W3C em 2004, mas teve sua primeira versão pública em 1999. Atualmente, a linguagem já se encontra mais madura e é utilizada em muitas aplicações, como o 11g da Oracle ⁹, Photoshop, entre outros (FEIGENBAUM et al., 2007). Um padrão de metadado muito utilizado, chamado vCard (DAWSON; HOWES, 1998) também possui implementação em RDF, bastando utilizar um *namespace* definido (IANNELLA, 2001). Outro padrão bastante conhecido que pode ser implementado em RDF é o RSS, que quando implementado em RDF significa *RDF Site Summary* (BEGED-DOV et al., 2000), bastando também declarar o *namespace* adequado para sua utilização.

O RDF Schema (RDFS), publicado como recomendação do W3C em 2004, define primitivas de modelagens adicionais sobre o RDF. Ele toma a especificação básica do RDF e a estende para suportar a expressão de vocabulários estruturados (BRICKLEY; GUHA; MCBRIDE, 2004). Ele fornece uma linguagem de representação de ontologia mínima que foi largamente adotado pela comunidade de pesquisadores (SHADBOLT; HALL; BERNERS-LEE, 2006). Permite a definição de classes, hierarquias de herança para classes e propriedades, domínio e restrições de propriedades. Somente devido as semânticas externas que o RDFS é útil (FENSEL et al., 2003).

Através do RDF Schema é possível, a qualquer pessoa ou comunidade, criar seus próprios vocabulários RDF. Uma das implementações mais utilizadas de um RDF Schema é a *Dublin Core Metadata Initiative* (DCMI) (INITIATIVE, 2008). Ele não fornece as classes e propriedades, mas sim um *framework* para descrevê-las. As classes em RDF Schema são muito semelhantes as obtidas com o conceito de orientação a objeto (BRICKLEY; GUHA; MCBRIDE, 2004).

O RDFS vem sendo criticado como linguagem de representação de ontologias, devido a falta de expressividade de seus construtores, que fazem seu poder de expressão limitado. Para atingir o nível desejado, foi necessária a criação de uma camada de ontologia sobre o RDFS (BREITMAN, 2005), vista a seguir.

⁹<http://www.oracle.com/technology/products/database/oracle11g/index.html>

2.4.3 Ontologia

Ontologia é um conceito multidisciplinar, possuindo várias definições conforme a área de utilização. Para Web Semântica, ontologia é uma especificação formal e explícita de uma conceitualização compartilhada (BREITMAN, 2005). Em filosofia, uma ontologia é uma teoria sobre a natureza da existência, sobre que tipos de coisas existem. Inteligência artificial e pesquisadores da web têm usado o termo para seus jargões, e para eles uma ontologia é um documento ou arquivo que define formalmente as relações entre termos. O tipo mais comum de ontologia para a web possui uma taxonomia e um conjunto de regras de inferência.

A taxonomia define classes de objetos e relações entre eles. Classes, subclasses e relações entre entidades são ferramentas poderosas para a web. Pode-se expressar um grande número de relações entre entidades atribuindo propriedades para classes e permitindo às subclasses herdar essas propriedades. Regras de inferência em ontologias fornecem mais poder. Um programa pode deduzir que, por instância, o endereço da Universidade Estadual de Londrina está no Londrina, deve ser no estado Paraná, que está no Brasil, e então deve formatá-lo para padrões brasileiros. O computador não entende realmente nenhuma informação, mas agora pode manipular os termos mais efetivamente de maneira útil e significativa para usuários humanos. Com páginas ontológicas na web soluções para problemas terminológicos começam a emergir (BERNERS-LEE; HENDLER; LASSILA, 2001).

Ontologias podem melhorar o funcionamento da web de diversas maneiras. Podem ser utilizadas para uma melhor efetividade de buscas na web. Aplicações mais avançadas utilizarão ontologias para relatar informação em uma página a estruturas de conhecimento associadas e regras de inferência. Em adição, essa marcação faz com que seja muito mais simples desenvolver programas que abordam questões complexas cujas respostas não se encontram em apenas uma página web (FREITAS, 2003).

Uma linguagem de ontologia deve preencher três requisitos (FENSEL et al., 2003):

1. Deve ser intuitiva para o ser humano;
2. Deve possuir uma semântica formal bem definida com propriedades de razão estabelecidas em termos de completude, corretude e eficiência;
3. Deve possuir uma ligação com linguagens para Web existentes, como XML e RDF, garantindo interoperabilidade.

Algumas linguagens falharam em atender esses requisitos. As linguagens a serem abordadas atendem perfeitamente esses requisitos, como poderá ser visto a seguir. O W3C coloca algumas necessidades que linguagens de ontologia devem atender, que são (HEFLIN, 2004):

- Ontologias como recursos distintos;

- Conceitos não ambíguos referenciados com URIs;
- Extensão explícita de ontologias;
- Compromisso com outras ontologias;
- Metadados ontológicos;
- Versionamento de informação;
- Primitivas de definição de classes e propriedades;
- Tipos de dado;
- Equivalência de classes e propriedades;
- Equivalência individual;
- Anexar informação em declarações;
- Classes e instâncias;
- Restrições de cardinalidade;
- Sintaxe XML;
- Etiquetas com significados para humanos;
- Suporte a modelos de caracteres;
- Suportar apenas strings Unicode.

Boa parte desses requisitos foram atendidos pela linguagem RDFS, a primeira linguagem de definição de ontologias para Web Semântica. Entretanto, ele não possui expressividade suficiente (BREITMAN, 2005), e devido a isso foram criadas outras linguagens, abordadas a seguir.

OIL

A *Ontology Interface Layer* (OIL) é uma linguagem de ontologia patrocinada por um consórcio da Comunidade Européia. Teve como principal requisito a facilidade na adoção por parte dos programadores (FREITAS, 2003). Sua grande vantagem era definir uma linguagem de representação de conhecimento que combina ao mesmo tempo (BREITMAN, 2005):

- Lógica de descrição, ou seja, semântica formal e suporte a inferência;

- Sistemas baseados em *Frames*, ou seja, primitivas de modelagem epistemológica;
- Linguagens da Web, ou seja, compatível com XML e RDF.

Um fato que ajudou na sua adoção foi a sua modularidade, visto que OIL era dividida em camadas de acordo com a complexidade desejada com seu uso, permitindo um uso mais simples (FREITAS, 2003). O conhecimento de uma ontologia OIL é dividido em três níveis (BREITMAN, 2005):

- Recipiente: responsável por armazenar metadados sobre a própria ontologia;
- Definição: responsável pela definição dos conceitos da ontologia;
- Objeto: local onde são armazenadas as instâncias.

DAML

Paralelamente ao OIL, a agência americana *Defense Advanced Research Projects Agency* (DARPA), em conjunto com o W3C, estava desenvolvendo a linguagem *DARPA Agent Markup Language* (DAML), acrescentando construtores mais expressivos ao RDF. Muitos aspectos dessa linguagem foram herdados do OIL, possuindo funcionalidades relativamente similares (FENSEL et al., 2003).

DAML + OIL

DAML + OIL é a combinação das duas linguagens, que apresentavam características similares. A linguagem é dividida em duas partes:

- Domínio dos objetos: objetos que são membros de classes definidas na ontologia DAML;
- Domínio do tipo de dados: valores importados do modelo XML.

Essa separação foi realizada para permitir a implementação de mecanismos de inferência, pois realizar inferência sobre tipos concretos de dados não é possível (BREITMAN, 2005). A linguagem é composta por Elementos de Classe, Expressões de Classe e Propriedades.

Posteriormente, a DAML + OIL foi acrescida de recursos de internacionalização, apresentação e documentação, gerando a linguagem OWL (FREITAS, 2003), próximo assunto abordado.

OWL

Em 2004, a *Web Ontology Language* (OWL), linguagem para publicação e compartilhamento de ontologias na WWW, foi lançada como uma recomendação do W3C. Como já citado, ela é uma derivação da linguagem DAML + OIL e desenvolvida como uma extensão do RDF (BECHHOFFER et al., 2004). Foi projetada para atender aplicações para Web Semântica, resumidas em (BREITMAN, 2005):

- Construção de ontologias;
- Explicitar fatos sobre um determinado domínio;
- Racionalizar sobre ontologias e fatos.

OWL é dividida em três linguagens, de acordo com sua expressividade (PASSIN, 2004):

- OWL Full: linguagem completa;
- OWL DL: suporta descrição lógica, que aplica certas restrições cuidadosamente escolhidas ao tipo de coisas que podem ser ditas de maneira a ganhar vantagens computacionais. Isso permite saber que uma descrição lógica é computável, ou “Completa e Decidível”;
- OWL Lite: constitui a OWL DL com mais restrições. A idéia é fazer com que seja fácil começar e implementar processadores, permitindo às pessoas começar utilizando OWL Lite facilmente e ir se graduando para usos mais complexos.

A linguagem OWL utiliza-se de RDF para se expressar, o que também funciona de maneira reversa, logo, qualquer grafo RDF também é um OWL Full, apesar do OWL não atribuir nenhum significado especial para os recursos. Um processador que reconheça OWL pode deduzir um pouco do que um processador RDF plano não pode. Por exemplo, ao se afirmar que uma propriedade *contains* é transitiva, o que é realizado declarando para *contains* o tipo *owl:transitiveProperty*. A partir daí um processador pode reconhecer que se A contém B e B contém C, A deve conter C também (PASSIN, 2004).

Os seguintes elementos são básicos da linguagem OWL (BREITMAN, 2005):

- Namespaces: declarações localizadas entre tags *rdf:RDF*, no formato de XML *namespaces*, que permitem uma interpretação dos identificadores a serem utilizados nas ontologias sem ambigüidades;
- Cabeçalhos: coleção de metadados sobre a ontologia agrupados pela sentença *owl:Ontology*, cuja responsabilidade é registrar comentários, controle de versão e inclusão de conceitos e propriedades de outras ontologias;

- **Classes:** fornecem um mecanismo de abstração para agrupamento de recursos com características similares. Toda classe está associada a um grupo de indivíduos, chamados de extensão da classe. Os indivíduos nas extensões de classe são chamados de instâncias da classe. Uma classe tem significado intencional relacionado, mas não igual a sua extensão de classe, por isso, duas classes podem possuir a mesma extensão de classe e permanecerem classes diferentes (BECHHOFFER et al., 2004);
- **Indivíduos:** Indivíduos são definidos com axiomas individuais, chamados fatos. São objetos do mundo, pertencem as classes e são relacionados a outros indivíduos;
- **Propriedades:** Descrevem fatos em geral. Podem se referir a todos os membros que pertencem a uma classe. Os tipos de propriedades são duas: Propriedades do tipo *object*, indivíduos a indivíduos e Propriedades do tipo *datatype*, relacionando indivíduos a valores de dados (BECHHOFFER et al., 2004);
- **Restrições:** Define limites para indivíduos que pertencem a uma classe. Podem ser de três tipos, que utilizam quantificadores, de cardinalidade e do tipo *hasValue*.

O projeto da linguagem OWL se beneficiou de diversas gerações de linguagens de ontologia anteriores, uma base teórica forte, e uma determinação na parte que muitos projetistas criaram a linguagem para atender os usos da Web Semântica. Devido a essas características, e a grande aceitação pela comunidade, certamente irá continuar a evoluir (PASSIN, 2004).

2.4.4 SPARQL

Como RDF e RDFS se estabeleceram, a necessidade por repositórios que pudessem guardar conteúdo RDF cresceu. Esses, chamados de *Triple-Stores*, variam em suas capacidades. Alguns focam em fornecer mecanismos ricos para raciocinar sobre as triplas, enquanto outros focam em grandes quantidades de dados. Alguns operam como *plugins* para navegadores existentes e outros como sistemas que podem operar com um grande número de bancos de dados de terceiros (SHADBOLT; HALL; BERNERS-LEE, 2006).

Com o desenvolvimento dos armazenadores, surgiu a necessidade para acesso aos dados dentro do RDF de maneira mais confiável e padronizada. *Simple Protocol and RDF Query Language* (SPARQL) tornou-se uma recomendação em 2008 e pode ser utilizada para expressar solicitações através de diversas bases de dados, enquanto os dados são armazenados nativamente como RDF ou visualizados como tal. Ele contém capacidades para solicitar padrões de grafo requeridos e opcionais em conjunto com suas conjunções e disjunções. Os resultados das solicitações podem ser vistos como conjuntos de grafos RDF (PRUD'HOMMEAUX; SEABORNE, 2008).

A linguagem possui como requisitos de projeto as seguintes características (CLARK, 2005):

- *Matching* de padrões de grafo RDF - Conjunção;
- Resultados vinculados a variáveis;
- Teste extensível de valores;
- Resultados de subgrafos;
- Solicitações locais;
- *Match* opcional;
- Suporte a tipos de dados limitados;
- Limites de resultados;
- Resultados em *streaming*;
- *Matching* de padrões de grafo RDF - Disjunção;
- Protocolo *Web Service Definition Language* (WSDL).

Como objetivos de projeto, possui as seguintes características (CLARK, 2005):

- Sintaxe amigável para humanos;
- Integração e agregação de dados;
- Solicitação pela existência de triplas;
- Protocolo com utilização de banda eficiente;
- Busca por literais;
- Solicitações sim-não;
- Resultados de solicitações endereçáveis;
- Ordenação de resultados.

Apesar de ser um padrão recente, SPARQL já é utilizada em algumas aplicações, como será mostrado mais adiante no capítulo 3.

2.4.5 RIF

O *Rule Interchange Format* (RIF) é uma linguagem em fase de desenvolvimento que possui um grupo de trabalho com o objetivo de produzir um núcleo de regras de linguagem e extensões que em conjunto permitem regras serem traduzidas entre linguagens de regras e assim transferi-las entre sistemas. Não se sabe ainda se será um formato ou uma linguagem, o que está a cargo do grupo de trabalho definir. Deverá funcionar como uma interlíngua que estabeleça uma nova regra de idioma que pode ser mapeada, permitindo que as regras de uma aplicação possam ser publicadas, compartilhadas e reutilizadas em outras aplicações e motores de regra.

Devido a grande variedade de linguagens de regras e tecnologias de regras, esse formato comum tomará forma de linguagem principal a ser utilizada em conjunto com extensões padrão ou não. Uma das utilizações de uma tecnologia como essa é para o suporte a inferência em ontologias independentemente de domínio (SHADBOLT; HALL; BERNERS-LEE, 2006).

Essa missão faz parte dos objetivos do W3C em compartilhar informação de forma que seja possível o processamento de máquina. As próprias regras representam um formato de informação que ainda não possuem um formato padrão de troca. Regras fornecem uma poderosa representação de lógica de negócio, como regras de negócio, em muitos sistemas de informação modernos. Regras são muitas vezes a tecnologia de escolha para criação de adaptadores sustentáveis entre sistemas de informação. Como parte da arquitetura da Web Semântica, regras podem estender ou complementar a linguagem OWL para cobrir cuidadosamente um conjunto mais amplo de aplicações, com conhecimento sendo codificado em OWL, regras ou ambas.

São objetivos do grupo de trabalho utilizar padrões e tecnologias existentes, mesmo que isso torne o projeto mais difícil. O grande desafio em estabelecer uma linguagem de regras padrão é grande quantidade de abordagens existentes no mercado. Interoperação entre as mais difundidas será crucial para obtenção do desejo de padronização (HAWKE, 2005).

2.5 Recursos Semânticos

Alguns tipo de aplicações tem muito a ganhar com a proposta da Web Semântica. Dois exemplos em que isso foi previsto desde os primeiros passos da mesma são os *Web Services* e os Agentes de Software (BERNERS-LEE; HENDLER; LASSILA, 2001), abordados nesse tópico.

2.5.1 Web Services

Um Web Service é um software identificado por uma URI, cujas interfaces públicas são definidas e descritas utilizando XML. Outros sistemas podem interagir com *Web Services* da maneira prescrita em sua definição, usando mensagens baseadas em XML veiculadas por protocolos de Internet (PASSIN, 2004). Os elementos de *Web Service* podem ser mapeados em quatro categorias:

- Troca de dados de maneira que possam ser realizadas transações;
- Invocação do serviço;
- Descrição do serviço e como invocá-lo;
- Encontrar o serviço correto.

O Objetivo básico de desenvolver *Web Services* Semânticos é levar a Web a todo seu potencial. Como a tecnologia de *Web Services* traz uma dimensão dinâmica à utilização da Internet, a Web Semântica pode trazer benefícios no tratamento da informação, facilitando a busca, extração, representação, interpretação e manutenção de dados.

A abordagem atual dos Web Services, baseados em WSDL e UDDI, não é orientada rumo ao estilo de funcionamento da Web Semântica e não é madura o suficiente para fornecer semântica dos dados, lógica de negócio e definição de uma seqüência de mensagens (BREITMAN, 2005). Existem possibilidades de adaptar esses e outros serviços, baseados em padrões diferentes, através de RDF. Isso se tornará mais simples quando for seguida a abordagem *Representational State Transfer* (REST) (FIELDING, 2000), onde estágios de Web Services se transformam em recursos (PASSIN, 2004). Acredita-se que a inclusão de descrições semânticas em Web Services atuais, principalmente através de ontologias, permitirá um melhor anúncio e descoberta de serviços na rede, fornecendo soluções mais elaboradas para seleção, composição e interoperabilidade de serviços heterogêneos (BREITMAN, 2005).

2.5.2 Agentes Inteligentes

Agentes Inteligentes são agentes de software flexíveis, de comportamento autônomo capaz de reagir apropriadamente ao seu ambiente e tomar iniciativas para atender seus objetivos. Também possui capacidades sociais, de maneira que um agente possa interagir eficientemente com outros agentes, e até mesmo com pessoas (PASSIN, 2004).

Na Web Semântica, os agentes se comunicarão através do compartilhamento de ontologias de domínio. Elas são importantes por representar entidades importantes do domínio, sendo papel das ontologias, no contexto da Web Semântica, não somente explicitar uma teoria de

domínio, mas servir como base para troca de informações entre agentes. Aplicados à Web Semântica, os agentes terão o papel de realizar a parte mais trabalhosa de uma tarefa, e então exibir os resultados alcançados para os humanos, para que estes possam tomar as suas decisões (BREITMAN, 2005).

O poder real da Web Semântica será realizado somente quando as pessoas criarem muitos programas que coletem conteúdo web de diversas origens, processem a informação e troquem resultados com outros programas. A efetividade desses agentes de software aumentará exponencialmente quanto mais conteúdos lidos por máquinas e serviços automatizados, incluindo outros agentes, se tornarem disponíveis. A web semântica promove essa sinergia: até mesmo agentes que não são expressamente modelados para trabalhar em conjunto podem transferir dados entre si quando os dados possuem semântica.

Uma importante faceta dos agentes é a troca de provas escritas na linguagem unificada da Web Semântica, que nesse caso pode ser interpretada como a RIF.

Outra funcionalidade fundamental são as assinaturas digitais, que são blocos criptografados de dados que computadores e agentes podem utilizar para verificar que a informação anexa foi fornecida por uma origem confiável. Os agentes são céticos quanto as informações que eles encontram na Web Semântica até que eles tenham checado as origens das informações. Logo, agentes terão um papel importante para a desejada “Web of Trust”.

O processo, chamado descoberta de serviço, só pode acontecer quando há uma linguagem comum para descrever um serviço de maneira que permita aos agentes compreender a função oferecida e como tirar vantagem disso. Alguns esquemas de descoberta de serviço de baixo nível estão atualmente disponíveis. Entretanto, eles atacam os problemas em seu nível estrutural ou sintático dependendo altamente de uma padronização de um determinado conjunto de descrições de funcionalidades.

Na Web Semântica, muito mais flexível, os agentes produtor e consumidor podem chegar a uma compreensão compartilhada pela troca de ontologias, que fornece o vocabulário necessário para discussão. Um processo típico envolverá a criação de um valor em cadeia em que os subconjuntos de informação serão passados de um agente para outro, cada um adicionando valor, para construir o produto final requisitado pelo usuário (BERNERS-LEE; HENDLER; LASSILA, 2001).

3 Web Semântica Aplicada à Saúde

A área da saúde adotou a informática como grande parceira para melhoria dos seus processos, e a web foi uma grande auxiliadora dos trabalhos de profissionais da saúde, visto que é possível acessar dados de qualquer lugar a qualquer momento. Entretanto, algumas necessidades foram surgindo durante a evolução da Web e outras se tornaram mais evidentes. Entre elas temos:

- Interoperabilidade;
- Codificação de dados;
- Confiabilidade da informação;
- Recuperação de dados;
- Auxílio aos profissionais da saúde.

Diferentes abordagens podem ser utilizadas na resolução desses requisitos. Uma das primeiras ferramentas que surgiram e tiveram grande representatividade para a obtenção desses requisitos foi o XML (SCHOROEDER; NEUMANN, 2006). Com ele foi possível superar algumas dificuldades, o que fez várias aplicações de sucesso emergirem. Entretanto, a área da saúde ainda necessitava de algo mais robusto. Pouco tempo depois surgiu a Web Semântica, prometendo recursos e soluções nunca vistos anteriormente.

O ramo das ciências da vida possui grande importância para a área da Web Semântica, sendo inclusive citado num dos cenários de utilização previstos no artigo que apresenta a Web Semântica (BERNERS-LEE; HENDLER; LASSILA, 2001).

Apesar de seu potencial multidisciplinar, os casos de sucesso da Web Semântica deverão aparecer inicialmente em sistemas com domínios específicos, onde um controle central poderá impor a utilização padrões e tecnologias (SILVA; COSTA; JATOBÁ, 2007). Isso vem ocorrendo há algum tempo no domínio da saúde, que desde o surgimento do XML vem adotando tecnologias de metadados para diversos fins (SHADBOLT; HALL; BERNERS-LEE, 2006), como pode ser observado na literatura. Com o surgimento da Web Semântica, a área da saúde

percebeu seu potencial para a resolução de deficiências encontradas em sistemas utilizados na área. Devido a isso, a aceitação da utilização de padrões acontece com maior facilidade, não sendo necessário um trabalho de conscientização dos ganhos possíveis com a utilização dos mesmos (GLOBE, 2003). Isso pode ser percebido pelo próprio desenvolvimento de padrões por profissionais de saúde, ajudando no crescimento não somente de suas áreas específicas, mas também na adoção de padrões web internacionais e confiáveis e da própria Web Semântica (SCHOROEDER; NEUMANN, 2006).

Outro fato que evidencia o interesse da área da saúde pela Web Semântica é a existência do grupo de trabalho no *Semantic Web Activity* específico da área, chamado *Health Care and Life Sciences Interest Group* (HCLSIG). Por possuir diversos domínios de conhecimento mapeados em padrões, é de grande interesse para os profissionais da área, expressar a semântica de seus domínios (DOGAC et al., 2006), o que os auxiliaria em diversas ocasiões, como na tomada de decisão, recuperação de informações sobre uma determinada doença ou paciente (DOGAC et al., 2006), (BOULOS; ROUDSARI; CARSON, 2002). A utilização de tecnologias semânticas já ocorre com sucesso na área, sendo uma das mais avançadas atualmente (FEIGENBAUM et al., 2007), o que confirma as previsões iniciais de Tim Berners-Lee.

O Grupo de interesse em semântica para saúde e ciências da vida, HCLSIG, desenvolve e apoia o uso de tecnologias da Web Semântica e práticas para melhorar a colaboração, pesquisa e desenvolvimento, e adoção de inovações nesses domínios. Sua missão é (PRUD'HOMMEAUX, 2008):

“desenvolver, advogar e apoiar o uso de tecnologias de Web Semântica para ciências biológicas, medicina translacional e cuidados de saúde. Esses domínios têm a ganhar inúmeros benefícios pela adoção de tecnologias de Web Semântica, uma vez que dependem da interoperabilidade da informação de vários domínios e processos eficientes de apoio a decisão”

Visto que o objetivo deste trabalho está focado apenas em cuidados com saúde, apresentaremos apenas as estratégias especificadas para esta área.

As aplicações de Web Semântica destinadas a área da saúde tem grande importância no desenvolvimento de tecnologias semânticas de escopo geral. Isso pode ser feito através da análise de diversos casos de uso encontrados (SHADBOLT; HALL; BERNERS-LEE, 2006).

3.1 O Domínio de Cuidados com Saúde

A demanda de aplicações mais inteligentes que possam capturar dados clínicos expressivos, armazenar conhecimento clínico de qualidade e capazes de fornecer suporte a decisão robusto

vem aumentando. Através disso, espera-se aumentar a qualidade dos cuidados de saúde, da pesquisa clínica e a integração de dados (SHADBOLT; HALL; BERNERS-LEE, 2006). Um dos focos de atividade do HCLSIG é justamente esse domínio, sendo sua tarefa indicar as forças da Web Semântica que atendem as necessidades dessa área. Entre seus objetivos temos o trabalho com os esforços já em curso para padronizar e harmonizar a aquisição e a troca de dados médicos fornecidos por instituições de padronização, como o *Health Level 7* (HL7), para que esses padrões possam ser utilizados plenamente com tecnologias de Web Semântica. Também pretende colaborar com esforços focados na construção de ontologias formais para medicina clínica e investigações expressas em linguagens de Web Semântica, como OWL e RDFS. Uma das áreas com maior atividade sendo desenvolvidas é justamente a construção ontologias (SILVA; COSTA; JATOBÁ, 2007). Por fim, pretende explorar a documentação que permita a interoperabilidade por meio de mapeamentos entre terminologias. O grupo também possui interesse na construção de ferramentas de suporte a decisão clínica inovadoras dentro dos registros de pacientes. Nesta área pretendem identificar as melhores práticas para representação de diretrizes clínicas de maneira que sistemas de raciocínio baseados em padrões e origens de conhecimento possam ser aproveitados em sistemas de registro dos pacientes (PRUD'HOMMEAUX, 2008).

Neste domínio existem formas de trabalho em ramos diferentes do HCLSIG. São encontradas ações na área de padronização da codificação, interoperabilidade, além de aplicações finais, tanto privadas quanto governamentais. A utilização de padrões de codificação está diretamente ligada à confiabilidade da informação e a interoperabilidade. A confiabilidade das informações é garantida pois, através da sua codificação utilizando padrões e terminologias médicas, faz com que elas estejam definidas dentro de um padrão aceito internacionalmente e codificadas de maneira específica a área que desejam representar. Ainda com esse tipo de garantia, a confiabilidade das informações é uma questão aberta (SHADBOLT; HALL; BERNERS-LEE, 2006). Por outro lado, visto que os padrões atuais utilizam tecnologias de Web Semântica, a interoperabilidade pode ser garantida (MCDONALD et al., 2003).

3.1.1 Padrões de troca de dados

Considerando que um dos pontos chave da Web Semântica é a representação do conhecimento do domínio ao qual pertence o conteúdo que está sendo trabalhado, no nível da *World Wide Web*, a vasta heterogeneidade de domínios constitui um grande desafio (SILVA; COSTA; JATOBÁ, 2007). Muito já foi desenvolvido nessa área, entretanto, devido à necessidade de integração de dados de diferentes fontes, uma necessidade maior surgiu: padrões de troca de dados em diferentes formatos de representação.

Os padrões mais utilizados para troca de informações na saúde são: *Health Level Seven* (HL7), *European Committee for Standardization* (CEN) TC/251, *International Organization*

for Standardization (ISO) TC215 e *openEHR* (PETRY; LOPES; WANGENHEIM, 2006).

O HL7 é um padrão para troca de dados entre aplicações de saúde (DOGAC et al., 2006) bastante utilizado. Trata-se de um padrão proprietário desenvolvido por uma organização sem fins lucrativos denominada *Health Level Seven*. As estruturas dos dados das mensagens representam informações clínicas, administrativas e financeiras em um ambiente hospitalar (PETRY; LOPES; WANGENHEIM, 2006). Além disso, o HL7 é um padrão aprovado pela *American National Standards Institute* (ANSI). Através de sua utilização é possível fornecer um conjunto rigoroso e bem definido de procedimentos operacionais, garantindo o consenso, abertura e equilíbrio de interesses (ORGANIZATION, a). Uma de suas especificações mais utilizadas é o *Clinical Document Architecture* (CDA), que especifica a estrutura e semântica de documentos clínicos de forma que possam ser lidos por humanos e sistemas e com o propósito de intercâmbio (DOLIN et al., 2008). Alguns de seus problemas são a falta de compatibilidade entre as versões, dificultando a interoperabilidade entre diferentes padrões de codificação, falta de definição de critérios de segurança e complexidade (PETRY; LOPES; WANGENHEIM, 2006). Apesar disso, é implementado em diversas aplicações e aparelhos clínicos, e foi escolhido como padrão em diversos países.

O CEN TC/251 define metadados sobre o Registro Eletrônico de Saúde através de componentes significativos (DOGAC et al., 2006). Trata-se de um padrão regional, no caso União Européia, e seu foco é exclusivamente tecnologia de conteúdo, e não de comunicação, como no caso do HL7. A especificação pretende fornecer padrões interoperáveis e implementáveis que permitam trocas de dados seguras. Esses padrões também colaboram para um *framework* técnico comum e uma terminologia para desenvolvimento de aplicações *eHealth* (NORMALISATION,).

Já a ISO TC/215 tem como principal objetivo contribuir para a melhoria e manutenção da saúde produzindo o próprio padrão, que a comunidade internacional considera como necessária para utilização de ICT no domínio da saúde. Esse padrão pretende padronizar a prática clínica da medicina, definir uma estrutura padronizada de prestação de serviços com cuidados de saúde, padronizar conhecimento médico, desempenho de cuidados com saúde e operação interna de sistemas e dispositivos (ORGANIZATION, b).

Em Setembro de 2006, os chefes do TC/251, ISO TC 215 e HL7 firmaram um acordo para

“avançar com planos compartilhados, coordenando e colaborando na administração de normas globais que permitam capacidades interoperáveis no domínio da saúde” (NORMALISATION,)

Entretanto, apesar do acordo firmado, a interoperabilidade entre os padrões citados ainda não é uma realidade. Apenas alguns recursos de um padrão são mapeados nos demais. Esse fato

pode ser considerado uma deficiência quando são necessárias funcionalidades como suporte a decisão, descoberta de conhecimento ou inferência de dados. Isso pois, quanto mais dados padronizados uma aplicação tiver acesso, melhor pode ser seu desempenho nessas tarefas (BERNERS-LEE; HENDLER; LASSILA, 2001), por isso uma maior interoperabilidade entre os padrões ou um consenso entre eles é de grande interesse. Para contornar essa deficiência, existe a possibilidade de mapeamento manual dos termos entre um padrão e o outro, o que é algo muito custoso. Outra opção, neste caso viável, é a utilização dos padrões do openEHR. O projeto, além de definir padrões para saúde, produz as especificações também com bons princípios de modelagem e engenharia. Os padrões são mantidos pela openEHR Foundation, que também é responsável por outras ações, como a implementação de arquétipos.

Arquétipos são especificações formais, expressas em termos de restrições sobre um modelo de referência, sendo que qualquer um desses pode ser utilizado. O modelo de referência do openEHR fornece uma base para quase todos arquétipos. Eles podem ser combinados através de *templates*, utilizados em tempo de execução para criar dados, solicitações inteligentes, e para apoio a transformação de dados. No openEHR são implementados utilizando a linguagem Archetype Definition Language (ADL), também utilizada pelo CEN 13606, mas também podem ser utilizados em XML.

As especificações dos padrões openEHR além de interoperabilidade, pretendem fornecer especificações baseadas em registro eletrônico de saúde que tenham as seguintes características (FOUNDATION,):

- manutenção de software reduzida;
- validade dos dados aumentada;
- alta interoperabilidade;
- baseadas em padrões;
- integração com sistemas já existentes;
- segurança e privacidade.

Deve-se observar que alguns dos itens são os mesmos do projeto da Web Semântica. Sua ligação com a área é facilmente observada, visto que são oferecidos diversos recursos das especificações no formato XML ou XML Schema no site da fundação. Suas principais vantagens são a apresentação de um mecanismo formal de validação de documentos, possibilidade de utilização de diversas terminologias, integração com o serviço de mensagens do HL7 v2 e possui licença *open source* (PETRY; LOPES; WANGENHEIM, 2006).

Apesar dos esforços de implementação dos grupos mencionados, foi constatado que de 80 a 90% dos conceitos clínicos são representados genericamente, sendo o restante definidos por

organizações internacionais, como HL7, CEN, ISO e openEHR (PETRY; LOPES; WANGENHEIM, 2006).

Situação Brasileira

No Brasil são encontradas três iniciativas para padronização: Elaboração de Padronização de Registros Clínicos (PRC), desenvolvimento do Sistema Cartão Nacional de Saúde e um padrão para Troca de Informações em Saúde Suplementar (TISS). Todas são desenvolvidas pelo DATASUS¹⁰, a divisão de informática do Ministério da Saúde. Nos últimos anos, a Sociedade Brasileira de Informática na Saúde (SBIS) está envolvida na definição de padrões e no suporte à construção de uma estratégia nacional de informática médica em conjunto com o DATASUS.

O DATASUS controla toda a infraestrutura nacional de informática e possui como objetivos (LEAO et al., 2001):

- Coordenar a implementação do sistema nacional de informação em saúde;
- Ajudar estados e cidades no processo de utilização de TI em suas atividades;
- Ser o mantenedor das bases nacionais de informação em saúde;
- Fazer que a informação esteja disponível ao público, gestores, pesquisadores e sociedade civil;
- Definir padrões para troca de informação em saúde em sistemas públicos e privados.

Seus principais usuários são os departamentos e conselhos de saúde federal, estadual e municipal, funcionários de unidades de saúde, universidades e centros de pesquisa, departamentos de auditoria, instituições de justiça ou polícia, sociedade civil, ONGs, jornais e cidadãos.

No domínio de cuidados com saúde, o DATASUS é o maior repositório web no país também sendo reconhecido como um dos mais completos repositórios de informação no mundo. Seu principal desafio atualmente é estabelecer interoperabilidade entre os diversos softwares desenvolvidos ao longo de 25 anos (LEAO et al., 2001).

Uma das estratégias nacionais para informática na saúde é o Sistema Nacional de Informações em Saúde, definido em 1990. Ele possui como missão melhorar a qualidade dos cuidados através do uso adequado de TI, qualidade e acessibilidade de serviços de saúde para pacientes e público. Um dos resultados esperados com a sua implementação é a melhoria da qualidade e abrangência da informação. O processo de construção desse sistema demanda um longo

¹⁰<http://www.datasus.gov.br>

prazo e investimento contínuo, além de flexibilidade para adoção de novas metodologias de desenvolvimento de software, modelos e padrões (LEAO et al., 2001).

O *International Classification of Diseases*, ICD-10, é o padrão nacional para diagnóstico, desde Janeiro de 1999. Trata-se de um padrão internacional de diagnóstico e classificação epidemiológica, propostas de gerenciamento de saúde e uso clínico. Passou a ser endossado como padrão pelo *World Health Assembly* em 1990 e a ser utilizado pelos Estados Membro da *World Health Organization* (WHO) desde 1994 (ORGANIZATION, c). O Brasil, integrante da WHO, aderiu ao ICD por décadas sempre em sua última versão.

Para o PRC foi definido o principal objetivo, de estabelecer padrões para construção de prontuários informatizados através de um processo aberto com a participação de todos os atores envolvidos. Espera-se viabilizar a interoperabilidade entre diferentes sistemas de informação no país, possibilitando assim a integração dos mesmos (DATASUS, b).

A padronização de alguns dados serviu como base para definições do registro de atendimento do Cartão Nacional de Saúde (CNS) (PETRY; LOPES; WANGENHEIM, 2006), financiado pelo Banco Inter-Americano de Desenvolvimento. Ele possui como base a identificação individualizada do paciente, acompanhamento do registro do atendimento em saúde, independente do local onde foi realizado, e a possibilidade da construção do repositório nacional de atendimentos.

Sua implantação embasará transformações do modelo e a reorganização dos sistemas e redes de atenção à saúde no Brasil. Para atender às expectativas geradas em torno do CNS, ele deverá ser capaz de (SAÚDE,):

- Integrar os diversos Sistemas de Informações em Saúde de base nacional existentes, tornando factível o intercâmbio de dados e ensejando a rediscussão das metodologias de coleta dos dados e de geração das informações de saúde;
- Construir uma base de dados de caráter nacional gerada, diretamente, a partir das unidades de saúde desvinculada do faturamento;
- Promover uma vinculação entre atendimento em estabelecimento de saúde do SUS e usuário de serviços de saúde, permitindo a identificação da clientela do sistema e a individualização dos procedimentos.

São utilizados padrões já estabelecidos em sua implementação, sendo que os padrões em uso no projeto são:

- vocabulário;
- conteúdo e estrutura;

- comunicação;
- privacidade, confidencialidade e segurança.

Além disso, ele utiliza o XML como padrão de mensagens, possibilitando assim uma maior interoperabilidade entre sistemas (DATASUS, a). Trata-se do projeto mais desafiador do DATASUS.

Suas expectativas são que através do acesso a essa informação melhores cuidados serão desenvolvidos, assim como uma maior eficiência na administração de serviços de saúde poderá ser atingida. Adicionalmente, a informação crucial estará disponível a construtores de políticas públicas e times de pesquisa. Em sua primeira fase ele pretende identificar 13 milhões de pessoas e 100.000 profissionais de 44 cidades, construir uma infraestrutura nacional para conectar todas as unidades envolvidas no projeto, secretarias municipais e estaduais ao Ministério da Saúde (MS), fornecer treinamento e educação em todos os níveis, interoperabilidade em todos os sistemas legados e no CNS e oferecer uma avaliação de todo o processo. Até o momento, 6 milhões pessoas foram registradas (LEAO et al., 2001).

A partir da implementação do CNS, um grande passo será dado em direção à construção de um sistema de informações em saúde capaz de atender as demandas do Sistema Único de Saúde e da comunidade científica (SAÚDE,).

Já o TISS estabelece um padrão para registro e intercâmbio de dados entre operadoras de planos privados de assistência à saúde e prestadoras de serviço de saúde. É possível obter informações através do TISS sobre avaliações clínicas, epidemiológicas e gerenciais, orientações para decisões e planejamentos e além disso, serve como base para estatísticas da Agência Nacional de Saúde (ANS) e outros órgãos governamentais. A proposta utiliza padrões já existentes, permitindo uma maior interoperabilidade (SUPLEMENTAR, 2007).

Apesar dos esforços do governo brasileiro, os padrões ainda não são muito difundidos. Além disso possuem algumas deficiências, como a não disponibilização de acesso ao modelo formal de validação dos dados, estruturas muito simples sem padronização de alguns dados, longo prazo de implementação, entre outros. A interoperabilidade dos sistemas é razoável, visto que todos utilizam padrões já estabelecidos. Porém, o TISS se destaca perante os demais e é considerado inovador pela sua capacidade operacional. Porém, abrange um nicho muito restrito de aplicações (PETRY; LOPES; WANGENHEIM, 2006).

3.1.2 Aplicações

O critério de seleção para estudo de aplicações foi a utilização de tecnologias da Web Semântica. As aplicações estudadas atuam de diferentes maneiras, em objetivos distintos, entretanto, utilizam tecnologias semelhantes, como RDF, RDF Schema, XML, XML Schema,

OWL e SPARQL. Algumas implementam sistemas de auxílio a decisão, outras trabalham com recuperação de dados, entre outras funcionalidades. As que mais se destacam são mostradas abaixo.

Artemis

O Artemis é um *Web Service* que define ontologias baseadas em um domínio de conhecimento definidas em padrões de informática na saúde. É um projeto turco do Centro de Pesquisa e Desenvolvimento de Software da *Middle East Technical University*. Visto que as ontologias não são globalmente aceitas, ele implementa um componente mediador para reconciliar diferenças entre projetos. O objetivo do projeto Artemis é permitir que as organizações mantenham seus projetos proprietários expondo sua funcionalidade através de *web services*. Também propoem uma descrição baseada em ontologias dos padrões de troca de dados, o que deixa os projetos partipantes livres para definir ontologias próprias. Entretanto, para uma maior efetividade e facilidade de agregação, recomenda-se que as ontologias sejam baseadas em padrões médicos estabelecidos. Um dos objetivos com o uso de ontologias é a redução conceitual e de diferenças terminológicas entre os padrões de troca de dados de saúde através da mediação semântica. Para promover interoperabilidade, são mostradas aplicações de saúde em conjunto como *web services* semanticamente enriquecidos.

Sua arquitetura ajuda a reconciliar diferenças semânticas entre institutos de saúde através de um componente mediador. Para prover escalabilidade e a descoberta de outros mediadores, possui uma arquitetura de comunicação P2P, que resolve diferenças bilateralmente. Entre as tecnologias e padrões de codificação utilizadas durante sua implementação, podemos destacar RDF, OWL, HL7, CEN e GEHR.

Como resultados, o Artemis conseguiu interoperabilidade entre *Web Services* que utilizam ontologias codificadas em diferentes padrões. Mesmo com o alto nível de interoperabilidade obtido, foram necessárias algumas operações de agregação semântica para junções mais complexas de *web services* (DOGAC et al., 2006).

HealthCyberMap

O HealthCyberMap é um projeto do *Centre for Measurement and Information in Medicine* da *School of Informatics* de Londres que pretende mapear recursos de informações sobre saúde na Internet em novos caminhos para uma melhor recuperação e navegação. Ele realiza essa tarefa através de coletas de metadados de recursos de forma não ambígua, preservando a semântica. Para alcançar o objetivo, foi modelado um conjunto de metadados ontológicos baseados no padrão Dublin Core, além de elementos extra que informam a qualidade do recurso e suas informações geográficas, adicionados ao Dublin Core pela aplicação. Outra característica

chave no sucesso da aplicação foi a utilização de códigos clínicos para descrever os temas de recursos médicos, o que melhorou a qualidade dos metadados e ofereceu categorização superior de tópicos desses recursos. Os mesmos códigos clínicos também são um pivô na provisão de problemas clínicos para serviços de ligação de conhecimento.

Entre as tecnologias utilizadas encontram-se RDF, RDFS, padrões de metadados Dublin Core, além de padrões internacionais de codificação de dados como o ICD-9. Como resultados, foi obtida uma base de metadados de recursos médicos RDF, além de uma navegação semântica entre coleções de recursos relatados a saúde na web, de maneira gráfica e intuitiva, baseada na metáfora do corpo humano. Os códigos clínicos utilizados, descreveram confiavelmente os recursos médicos, estabeleceram ligações semânticas e automatizaram a categorização de tópicos. É recomendada a utilização de ontologias fundacionais implementadas como serviços através de servidores de terminologia. Isso permite recuperação de conceitos relatados e sinônimos, busca e mapeamento de múltiplas terminologias simultaneamente. A combinação de metadados com ontologias foi de grande importância para relevância dos resultados obtidos pela aplicação (BOULOS; ROUDSARI; CARSON, 2002).

KAT

O KAT é um projeto do Intermountain Health Care (IHC), uma organização sem fins lucrativos que atende as necessidades médicas das cidades de Utah e Idaho nos Estados Unidos. Possui uma rede integrada de 21 hospitais, diversas clínicas, mais de 450 médicos e um plano de saúde. O KAT atua na área de suporte a decisão, utilizando conceito de colaboração dos usuários com especialidades médicas, com diferentes níveis e áreas de especialidade, na definição de conhecimento. Estes usuários são os mais recomendados para manter a base de conhecimento, necessária para suprir o suporte a decisão, da aplicação atualizada, visto que eles estarão melhores fundamentados clinicamente. Para construção de metadados foi utilizado XML e XML Schema, considerada uma escolha chave na integração com ambientes de gerenciamento distribuído de conhecimento. Visando uma maior efetividade em buscas e versionamento, os documentos desenvolvidos foram divididos em cabeçalho e corpo, sendo que o cabeçalho possui os metadados e o corpo o documento em si. Numa busca por um documento, verifica-se as informações de cabeçalho como dados do autor, título do documento, categoria do documento, entre outras, avaliando se o conteúdo será apropriado para o uso desejado. O KAT foi desenvolvido para maximizar a reutilização de conteúdo durante o processo de autoria. Essa foi uma funcionalidade chave por duas razões: primeiro, para reduzir a carga de trabalho sobre os autores prevenindo a necessidade de múltiplos documentos com o mesmo conteúdo e, segundo, para melhorar a colaboração e o compartilhamento de conhecimento entre os grupos de autores. Também utilizaram o CDA do HL7, que no caso dessa aplicação era a escolha adequada pois possibilitava um desenvolvimento iterativo.

Um resultado obtido com o KAT foi a possibilidade do desenvolvimento iterativo, por parte dos especialistas, de bases de conhecimento baseadas em XML disponíveis em um repositório global. Este repositório está acessível a outros interessados podendo ser reutilizado na construção de novas bases. Também é possível computá-lo, e ainda assim ser lido e compreendido pelo clínicos. Um resultado interessante obtido nesta aplicação foi o aumento no uso e colaboração dos clínicos após a implementação desses recursos, mantendo as bases de conhecimento mais atuais e sempre em expansão. Um resultado que prova a interoperabilidade de XML é que os documentos gerados no KAT são utilizados em outras três aplicações no IHC, do mesmo grupo de trabalho do KAT, e por outro grupo, chamado *Collaborative Practice Guidelines* (CPG) do IHC, que está recriando sua biblioteca de referência de padrões de cuidados interdisciplinares (HULSE et al., 2005).

SHAPPIRE

O SHAPPIRE foi desenvolvido para uma melhor detecção, análise e resposta de problemas de saúde pública emergentes no centro de Ciências da Saúde da Universidade do Texas em Houston. Para isso, ele integra uma vasta gama de dados de provedores de saúde locais, hospitais, agências de proteção do ambiente e literatura científica. Permite oficiais de saúde acessarem a informação através de diferentes visões, como o monitoramento da disseminação da gripe ou tratamentos de casos de HIV. A cada 10 minutos ele recebe relatórios originários de diferentes fontes, com diferentes escopos. Então uma integração semântica é realizada mostrando a situação na área. Um dos conceitos chaves utilizados em sua implementação é a ontologia de doenças não identificadas, que apresentam sintomas semelhantes a outras já conhecidas.

Além disso, existe a funcionalidade de geração de relatórios automáticos, antes realizados por enfermeiras. Agora, elas estão mais livres para desempenhar seu papel, e os relatórios menos sujeitos a erros e entregues mais rapidamente.

Outro caso de utilização ocorreu durante o furacão Katrina. Através dele foi possível identificar surtos gastrointestinais, respiratórios e de conjuntivite nos sobreviventes do desastre em tempo muito menor do que seria possível sem sua utilização. Os dados eram coletados por oficiais públicos em *handhelds* através de questionários de saúde. As respostas dos refugiados eram carregadas no sistema e integradas com os dados das clínicas de emergência, abrigos e relatórios de vigilância dos epidemiologistas de campo do Departamento de Saúde e Serviços Humanos de Houston.

A flexibilidade do SHAPPIRE mostra que uma vez que um sistema de Web Semântica esteja configurado para um problema geral, ele poderá rapidamente ser adaptado para uma variedade de situações dentro daquele campo. O SHAPPIRE obteve sucesso pois pode unificar

a informação de diferentes origens e então utilizá-las para objetivos variados, o que é inclusive um dos grandes objetivos gerais da Web Semântica (FEIGENBAUM et al., 2007).

PropeRWEB

O PropeRWEB é um sistema utilizado em um ambiente com comunicação extremamente heterogênea em Maastricht na Holanda. Atua em um projeto de cuidados com AVC, onde a fase de reabilitação do paciente pode ser realizada em sua própria residência. As informações trocadas pela equipe de reabilitação, formada por um fisioterapeuta, um fonoaudiólogo e uma enfermeira comunitária além do coordenador de cuidados em AVC, podem vir do hospital de diversas maneiras: carta, diretamente, e-mail, fax, telefone, entre outras. Cada uma das formas de comunicação é utilizada em um momento específico e com diferentes objetivos de comunicação. Devido a isso, a comunicação entre o pessoal de reabilitação e o hospital não é ótima, principalmente quando a reabilitação é feita na casa do paciente. Este é o motivo do desenvolvimento do sistema PropeRWEB.

Um dos requisitos mais importantes de desenvolvimento era a flexibilidade, visto que o sistema deve acomodar diferentes domínios de conhecimento, no caso AVC e hematologia. Também era necessário que o sistema fosse compatível com diferentes padrões. O padrão utilizado para desenvolvimento do sistema foi o HL7, implementado suas definições de PIDS e COAS. O servidor PIDS foi dividido em dois, sendo um responsável pelos dados dos pacientes e o outro pelos usuários profissionais de saúde, o que proporciona maior segurança. Um conceito interessante utilizado na aplicação para suportar diferentes domínios é a separação do software e do banco de dados através da utilização de arquétipos, conceito definido pelo projeto openEHR. Um arquétipo pode ser descrito como um template para um conceito de domínio médico. Sua utilização é recomendada pois são definidos não por apenas um usuário, mas por um grande grupo representativo. Este consenso na definição previne variações. Como são definidos por especialistas médicos, normalmente serão mais facilmente aceitos por eles. O XML é utilizado para escrever dados do PIDS, COAS, arquétipos e na camada de apresentação da aplicação.

Após uma análise do PropeRWEB pelos seus possíveis usuários, acredita-se que ele esteja pronto para uso diário, sendo que um dos resultados mais estimulantes atingidos foi o encontro das informações rapidamente. Uma experiência que é possível aproveitar é a reutilização de módulos open source já existentes, como aconteceu na implementação dos servidores COAS e PIDS (LINDEN et al., 2004).

HealthFinland

O portal nacional semântico de saúde da Finlândia *Health-Finland*, integrante do projeto nacional Finlandês de Web Semântica FinnONTO, provê aos cidadãos buscas inteligentes e serviços de navegação para a localização de dados de saúde confiáveis e atuais, criados por várias organizações na Finlândia. O sistema é baseado em *schemas* de metadados semânticos compartilhados, ontologias e serviços de *mash-up* de ontologias. Como as informações de saúde na web são fornecidas por organizações diferentes e independentes com níveis variados de confiabilidade, direcionadas para leigos e peritos, disponíveis em diversos formatos, escritas em diversos idiomas, a dificuldade de encontrar informações confiáveis e relevantes neste tipo de ambiente heterogêneo cria um obstáculo para cidadãos preocupados com sua saúde. O *HealthFinland* discute os problemas que atingem portais de informação quando publicam informações sobre saúde na web para os cidadãos, considerando tanto o ponto de vista dos editores, quanto dos usuários finais. A solução desenvolvida foi um modelo de publicação de conteúdo distribuído de Web Semântica para organizações de saúde, baseados em um esquema de metadados compartilhados, ontologias e serviços de *mash-up* de ontologias, através do qual o conteúdo é criado com efetividade de custos por produtores de conteúdo independentes em localizações diferentes. O sistema agrega e faz o conteúdo semanticamente interoperável para ser reutilizado em diferentes aplicações sem que seja necessário modificá-lo.

A primeira idéia do modelo é minimizar trabalho redundante e duplicado além dos custos na criação de conteúdo no nível nacional, produzindo-o apenas uma vez e por apenas uma organização, e fazendo com que seja possível reutilizar o conteúdo produzido em diferentes aplicações web por outras organizações, não somente o próprio portal da organização.

A segunda é tentar minimizar os custos de manutenção de portais deixando que o computador tome conta da manutenção semântica de *links* e agregação de conteúdo de diversos publicadores. Essa possibilidade é também baseada em metadados semânticos compartilhados e ontologias. Conteúdo novo relevante de um tópico pode ser publicado a qualquer momento por qualquer provedor de informação, e o sistema deve ser capaz de colocar o novo trecho de informação no contexto correto no portal, e associá-lo automaticamente com informações relacionadas.

A terceira idéia é fornecer ao usuário final serviços inteligentes para encontrar informações corretas baseadas em sua própria visão conceitual de saúde, e para navegar por conteúdos baseados em relações semânticas. As visões e vocabulários utilizados na interface com o usuário final podem ser independentes da perspectiva organizacional dos provedores de conteúdo, e são baseadas no vocabulário “leigo” que é diferente do vocabulário especializado médico utilizado pelos provedores de conteúdo na indexação deste.

O HealthFinland possui uma estrutura ontológica baseada em dois itens:

1. Um *schema* de metadados, compartilhado por todas as organizações que criam conteúdo, garantindo interoperabilidade semântica de conteúdo;
2. Um conjunto de vocabulários ontológicos cujos conceitos são utilizados para preencher os valores dos *schemas* de metadados. As ontologias também são compartilhadas entre as organizações, e sua utilização garante a interoperabilidade semântica do conteúdo.

O esquema de metadados utilizado é codificado em RDF e baseado no *Dublin Core Element Set*, com refinamentos introduzidos no *Dublin Core Metadata Initiative Terms*.

A interoperabilidade semântica no *HealthFinland* é obtida através da utilização de um conjunto de ontologias compartilhadas para preencher os valores dos metadados. As ontologias incluem uma ontologia de mídia, uma ontologia de audiência, representando tipo de pessoas, ontologia de fuga, ontologia de tipos de gênero, ontologia de tipos DCMI para tipos de mídia e uma ontologia de tempo. As três ontologias chave, utilizadas para descrever o assunto dos conteúdos web, tem origens em diferentes organizações e são codificadas em RDF e OWL.

Resultados muito interessantes foram obtidos pelo *HealthFinland*. Para os cidadãos é possível realizar buscas semânticas entre portais, baseadas em facetas definidas pelos usuários, o que também facilita a agregação de conteúdo. Os *links* obtidos como resultados são alterados automaticamente, assim que o conteúdo que eles apontam é modificado. Visto que somente produtoras de conteúdo confiáveis são adicionadas, a qualidade dos dados é assegurada. E a facilidade de navegação é otimizada pela apresentação dos conceitos indexados por terminologias médicas profissionais em vocabulário adequado ao conhecimento do cidadão leigo. Já para os criadores de conteúdo, a duplicação de criação deste foi minimizada, através da sua agregação entre portais, o que também prova que a reutilização de conteúdo é possível. O gerenciamento de links internos e externos é facilitado pelo sistema de recomendação dinâmica e semântica e pela agregação de conteúdo (HYVONEN; VILJANEN; SUOMINEN, 2007).

SenseLab

Um dos desafios enfrentados pelo grupo de interesse em Web Semântica para Ciências da Vida e Cuidados com Saúde é a conversão de bancos de dados relacionais no formato da Web Semântica. As questões e passos envolvidos em uma conversão ainda não foram bem documentadas. Uma aplicação que obteve sucesso nesse ponto foi o *SenseLab*, uma coleção de bancos de dados relacionais de pesquisa neurocientífica do Laboratório Shepherd, da Universidade Yale no Connecticut, Estados Unidos. A conversão desses bancos de dados em RDF/OWL é um passo importante em direção a realização de benefícios da Web Semântica em integrar-se com pesquisa neurocientífica.

O SenseLab é dividido em um número especializado de bancos de dados, os quais foram

convertidos em árvores para formatos de Web Semântica. Juntos, eles permitem ao pesquisador solicitar informação e executar simulações pertinentes a função de neurônios em estados de saudáveis e doentes. Todos os bancos de dados contém extensivas referências literais e fragmentos de textos tem sido utilizados para vigiar as entradas do banco de dados. Os bancos de dados são baseados no esquema “entidade-atributo-valor com classes e relacionamentos” (EAV/CR).

Duas versões em paralelo foram desenvolvidas, uma com os bancos de dados sendo codificados em RDF e outra em OWL DL. Um conversor em Java foi desenvolvido para solicitar os bancos de dados do SenseLab. A conversão para RDF foi feita automaticamente, já a versão em OWL necessitou de edição manual. Devido a grande dependência da estrutura original do banco de dados, algumas ontologias OWL resultantes foram inconsistentes e possuíam algumas deficiências.

Devido a essas deficiências da primeira versão OWL, uma segunda foi desenvolvida, baseada nos resultados da primeira. Na segunda versão, a representação ontológica foi focada em representações diretas de objetos físicos e processos, não em suas abstrações, como na primeira versão.

Na segunda versão foi criada uma ontologia com hierarquias de classes básicas e relações, baseada na estrutura do banco de dados existente. Essa, foi criada manualmente, baseada em ontologias fundacionais, para uma maior interoperabilidade. A partir dessa ontologia, os dados do banco de dados foram automaticamente convertidos para OWL, utilizando programas em Java e Python.

Para mapeamentos para bancos de dados externos de bioinformática que ainda não ofereciam URIs estáveis para referência em Web Semântica, foi utilizado um esquema de URI para gravar identificadores de bancos de dados estabelecidos pelo Science Commons. As entradas do banco de dados foram conectadas com as representações ontológicas de entidades do mundo real através de relações, utilizando outras ontologias. Os *namespaces*/localizações de ontologias foram alterados para URIs baseadas em Persistent URL (PURL), que são mais fáceis de manter quando as configurações do servidor mudam ou o servidor original está indisponível e as ontologias precisam ser servidas em localizações diferentes. A crescente estabilidade de PURLs encoraja a reutilização de entidades em ontologias desenvolvidas por outros grupos, o que é um fator chave na criação de uma Web Semântica Coerente. Por fim, um ponto final SPARQL para as ontologias SenseLab foi implementado. Um ponto final SPARQL é um serviço que permite a clientes solicitar um dado RDF com a linguagem de requisição SPARQL através de simples requisições HTTP GET.

Resultados obtidos com o SenseLab foram:

- Integração de dados do SenseLab com ontologias desenvolvidas por outros projetos

facilitada pelo OWL, visto que este não necessita de mediadores centrais, reduzindo custo de desenvolvimento e manutenção;

- Modularização de ontologias, através de declarações de relações explícitas, o que encoraja a reutilização de ontologias de outros grupos;
- Bom nível de descrição atingido através das ontologias OWL;
- Utilização de raciocinadores para identificação de erros e contradições reais em conjuntos de dados enviados;
- Utilização de ontologias para representação direta de realidades biológicas, sem a necessidade da introdução de abstrações.

Algumas práticas também são recomendadas:

- criação de ontologias em OWL DL, ao invés de RDF(S), visto que este possui algumas limitações;
- reutilização de entidades e propriedades de ontologias existentes;
- basear a ontologia em desenvolvimento em ontologias fundacionais;
- valide sua ontologia em um validador RDF;
- cheque a consistência de ontologias OWL periodicamente;
- utilize URIs purl para as ontologias;
- entre outras.

Neste caso, a utilização de tecnologias de Web Semântica trouxe benefícios na integração dos dados do SenseLab com outros bancos de dados neurocientíficos de maneira consistente, flexível e descentralizada (SAMWALD et al., 2008).

3.2 Resultados e Atividades da Web Semântica Aplicada a Saúde

Algumas necessidades da área da saúde coincidem com requisitos gerais da Web Semântica. Entre elas temos:

- Interoperabilidade;

- Representação de conhecimento através de padrões de codificação de dados;
- Recuperação de informações confiáveis codificadas em padrões;
- Auxílio a tomada de decisão dos profissionais.

A Web Semântica promete interoperabilidade de dados, possibilidade de trabalhar com inferência sobre os mesmos, sistemas inteligentes, ter a web toda como um grande banco de dados de informações e uma melhora significativa na busca de dados. O entusiasmo da área da saúde na web semântica foi previsto pelo criador na sua apresentação dela ao mundo (BERNERS-LEE; HENDLER; LASSILA, 2001). Durante os anos posteriores à previsão, pode-se verificar realmente esse interesse através de grandes resultados não somente para a área de saúde, mas para a web semântica como um todo (SHADBOLT; HALL; BERNERS-LEE, 2006), (FEIGENBAUM et al., 2007).

Uma das áreas em que é possível perceber o entusiasmo do pessoal da saúde é no desenvolvimento de ontologias, onde muitos médicos ajudam no mapeamento de conceitos em ontologias, sendo que estas são desenvolvidas a partir de um consenso entre muitos profissionais da área, que trabalham de maneira a alcançar uma maior integração e interoperabilidade de dados (SHADBOLT; HALL; BERNERS-LEE, 2006). O resultado disso é uma das maiores bases de conhecimento mapeadas em padrões semânticos dentre todas as áreas, o que é extremamente importante para os resultados alcançados por aplicações da área (DOGAC et al., 2006).

Durante a análise realizada no desenvolvimento desse trabalho, foi constatada a melhoria de sistemas como Web Services, que quando enriquecidos semanticamente podem realizar a agregação de outros serviços e dados (DOGAC et al., 2006). E para isso não é extremamente necessário realizar uma mudança de tecnologia de representação de dados em sistemas já existentes, o que mostra a flexibilidade da web semântica, que pretende funcionar baseada nos alicerces da web que conhecemos (BREITMAN, 2005). Experiências interessantes também foram alcançadas ao se falar de colaboração entre pessoas, que através de tecnologias semânticas podem representar seu conhecimento e disponibilizá-lo para uso e expansão por outras pessoas, como visto em (LINDEN et al., 2004) e (HULSE et al., 2005). Um fato interessante notado na maioria das aplicações é justamente a reutilização de componentes e conhecimento já existente. Uma vez codificados em padrões, basta utilizá-los para uma necessidade específica, algo que era quase impossível na web convencional, pois um dado poderia ser codificado de diversas maneiras em diferentes locais. A flexibilidade, agilidade e efetividade de sistemas de web semântica foi colocada a prova e obteve grandes resultados em diferentes campos de conhecimento. Um sistema pode se adaptar a uma base de conhecimento nova bastando que essa esteja padronizada, como mostrado em (FEIGENBAUM et al., 2007). Conceitos bem mapeados, oriundos de diversos locais diferentes também auxiliam na abstração

de processos, melhorando a atuação de profissionais em uma área. Um exemplo disso pode ser visto em (BOULOS; ROUDSARI; CARSON, 2002). Um exemplo de integração de dados pode ser visto na conversão de bancos de dados relacionais para linguagens de web semântica em (SAMWALD et al., 2008), que alcançou grandes resultados, melhorando a aplicação em questão através da migração da maneira de como o conhecimento era representado.

A web semântica possui grande força que faz parte até mesmo de programas de governo em diversas áreas, assim como a saúde. Isso pode ser visto nos exemplos apresentados do governo brasileiro (LEAO et al., 2001) e finlandês (HYVONEN; VILJANEN; SUOMINEN, 2007). Em ambos os casos percebeu-se a necessidade de uma estrutura nacional de saúde que deveria ser baseada em tecnologias semânticas. É importante notar o trabalho conjunto dos governos com os profissionais de saúde e organizações que desenvolvem padrões. Na Europa, por exemplo, há o CEN, o HL7 é utilizado como padrão em muitos países e, não faltam casos de colaboração entre governos e entidades a serem analisados, com diferentes exemplos e resultados. Um desses casos é um padrão de classificação de doenças definido pela Organização Mundial da Saúde, chamado ICD-10, utilizado em diversos países integrantes da WHO, inclusive o Brasil. Os dados, definidos através de XML, possuem grande difusão e podem certamente ser utilizados para fornecer interoperabilidade entre países.

A tabela 3.1 mostra um resumo das iniciativas estudadas.

Tabela 3.1: Resumo das implementações semânticas estudadas

Aplicação	Necessidades Resolvidas	Área de Atuação
Artemis	Agregação semântica entre Web Services	Web Services Semanticamente enriquecidos
HealthCyberMap	Recuperação de Dados	Base de Dados Semânticos, Recuperação de Informação
KAT	Compatilhamento de bases de conhecimento	Suporte a Decisão
SHAPPIRE	Agregação de Dados	Suporte a Decisão
PropeRWEB	Recuperação e compartilhamento de dados	Recuperação de informação
HealthFinland	Recuperação, compartilhamento e agregação de dados	Recuperação de dados com confiabilidade
SenseLab	Representação de conhecimento e recuperação de dados	Mapeamento ontológico de bases de dados
CNS	Representação de conhecimento, integração de dados	Recuperação da informação

É importante notar que todas as aplicações fornecem interoperabilidade de dados, por isso esse dado não foi explicitado na tabela 3.1. Nota-se que as necessidades da área da saúde já

vem sendo atingidas por essas aplicações, inclusive algumas já são capazes de fornecer dados confiáveis, um dos grandes objetivos da Web Semântica. Também é interessante notar a adoção de diferentes tecnologias e padrões de representação de dados. Todas as tecnologias estudadas neste trabalharam foram utilizadas em pelo menos uma das aplicações, o que pode explicar a interoperabilidade alcançada por todos.

Através destes exemplos é possível perceber que a área da saúde realmente vem implementando diferentes esforços em variadas áreas baseados na Web Semântica. Como já mencionado, alguns desses esforços representam grande avanço para a Web Semântica em termos gerais, podendo ser aplicados a qualquer área de conhecimento. Sabemos que todo o potencial da tecnologia ainda não foi atingido, mas a partir desses casos de sucesso, acreditamos que algum dia será possível utilizar as funções previstas por Tim Berners-Lee (BERNERS-LEE; HENDLER; LASSILA, 2001) e também chegar a um estado satisfatório para as aplicações de saúde, que terão suas necessidades atendidas, melhorando diversos aspectos dos cuidados com saúde.

Conclusão

A saúde é um direito de todos, reconhecido nacional e universalmente. Qualquer tipo de melhoria relacionada à prática da saúde é bem vinda, pois otimiza o que já existe ou, em melhores casos, passa a tratar de uma área anteriormente não assistida. Como se sabe, a informática vem ajudando a área da saúde desde seu surgimento, tanto na área de pesquisa quanto na prática. Através dela, foi possível realizar grandes descobertas importantes para toda a humanidade, possibilitando curas de doenças, processos realizados com maior precisão, melhor gerenciamento de dados, disseminação de práticas saudáveis a população, maior fidelidade durante o ensino, entre diversas outras.

Devido a esse potencial encontrado pelos profissionais de saúde na informática, seu entusiasmo para que novas tecnologias se desenvolvam para utilização em suas áreas é grande. Isso pode ser notado pela ajuda desses profissionais para o desenvolvimento de novas soluções, trabalhando em conjunto com profissionais de tecnologia. O resultado desse trabalho em conjunto pode ser percebido com robôs que auxiliam em cirurgias, mapeamento do genoma humano e no desenvolvimento da web semântica.

As contribuições da web para a área da saúde também constituíram um grande avanço. Desde seu surgimento, os vários casos de uso encontrados na área da saúde servem como exemplo para a web como um todo. A possibilidade de armazenamento e troca de dados entre instituições, em qualquer lugar a qualquer momento é algo extremamente útil para a área em questão. Entretanto, com o crescimento da utilização da web, algumas necessidades surgiram, e não podia ser diferente na saúde. Entre essas necessidades havia a interoperabilidade de dados, que estavam sendo codificados da maneira que fosse mais conveniente a quem os criava, um auxílio maior dos sistemas baseados nos dados fornecidos, entre outras.

Acreditamos que grandes resultados já foram alcançados pela área da saúde através da web semântica, o que caracteriza grandes resultados da própria web semântica, que cresce a medida que novas aplicações são desenvolvidas nas diversas áreas onde é possível. Esse entusiasmo da área da saúde é responsável pelos melhores resultados da web semântica até o momento (FEIGENBAUM et al., 2007). Autores acreditam que a partir da grande utilização da web semântica mudanças ocorrerão até mesmo no modo como as pessoas vivem, e que a pesquisa científica será extremamente facilitada (BERNERS-LEE; HENDLER; LASSILA, 2001), (SHADBOLT; HALL; BERNERS-LEE, 2006), (FEIGENBAUM et al., 2007). Autores de ciências da vida acreditam que é possível terminarmos com uma web médica inteligente

(BOULOS; ROUDSARI; CARSON, 2002), o que pode se estender a web como um todo. Para isso, acreditamos que é necessário que os padrões já estabelecidos sejam realmente utilizados por uma grande quantidade de desenvolvedores. Também acreditamos que deve-se chegar a um consenso maior com relação a alguns pontos, como por exemplo a utilização de um padrão de representação de conhecimento e terminologias mais abrangentes, como o *openEHR*. Governos e o W3C vêm trabalhando em ambos os sentidos, e resultados já são observados fora do domínio da saúde. Inclusive comercialmente eles podem ser verificados, entretanto, os mais esperados ainda são os resultados que irão modificar nosso dia-a-dia, que ainda aparecem timidamente. Esperamos, daqui a alguns anos, ter uma web semântica o suficiente por trás da atual, que seja capaz de nos impressionar diariamente com a quantidade de serviços agregados e com a facilidade de encontrá-los e utilizá-los, como previsto por Tim Berners-Lee (BERNERS-LEE; HENDLER; LASSILA, 2001). E, para isso, acreditamos que a área da saúde tem papel crucial para que esse ponto seja alcançado, e esperamos que resultados cada vez mais impressionantes sejam apareçam através ela, como já vem acontecendo atualmente.

Este trabalho apresentou como principal contribuição o levantamento do estado da arte da Web Semântica aplicado ao domínio da área de cuidados com saúde. Verifica-se que algumas possibilidades levantadas no início do desenvolvimento da mesma já são alcançadas pela área em questão, que se beneficia amplamente com isso, pois antigos problemas são resolvidos, necessidades são supridas e novas possibilidades vislumbradas. Acreditamos que este trabalho pode servir como fonte básica de pesquisa para possíveis interessados no desenvolvimento de soluções em Web Semântica para a área de cuidados com saúde, pois apresenta casos de sucesso nos campos científico, governamental e privado. Por fim, cremos que a Web Semântica é um grande recurso que pode ser utilizado para melhorar não somente o campo de estudo deste trabalho, mas a maneira como a Web é utilizada. Conseqüentemente, atividades sociais, como foi visto no campo apresentado, podem ser otimizadas como o atendimento a população, otimização e simplificação de processos realizados por humanos, entre outras atividades, possibilitando novos usos revolucionários da Web.

Acreditamos que este trabalho também possa ser útil como fonte de trabalhos futuros relacionados a Web Semântica, principalmente nos relacionados a pesquisa de ontologias, área que possui a maior quantidade de pesquisas atualmente no ramo da Web Semântica. Também acreditamos que este trabalho possa ser utilizado fora do escopo da área da saúde, apesar de abordar mais profundamente atividades desenvolvidas por essa área de conhecimento, as estratégias aqui estudadas podem ser utilizadas em outras área de conhecimento.

Referências Bibliográficas

- AXELSSON, J. et al. *XHTMLTM 2.0*. [S.l.], 2006. Disponível em: <<http://www.w3.org/TR/xhtml2/>>.
- BECHHOFFER, S. et al. *OWL Web Ontology Language Reference*. [S.l.], 2004. Disponível em: <<http://www.w3.org/TR/owl-ref/>>.
- BEGED-DOV, G. et al. *RDF Site Summary (RSS) 1.0*. [S.l.], 2000. Disponível em: <<http://web.resource.org/rss/1.0/spec>>.
- BERNERS-LEE, T. *Tim Berners-Lee's proposal*. Internet. Disponível em: <<http://info.cern.ch/Proposal.html>>.
- BERNERS-LEE, T. *Information Management: A Proposal*. [S.l.], 1989. Disponível em: <<http://www.w3.org/History/1989/proposal%20-.html>>.
- BERNERS-LEE, T.; CONNOLLY, D. W. *Hypertext Markup Language - 2.0*. [S.l.], 1995. Disponível em: <<http://www.ietf.org/rfc/rfc1866.txt>>.
- BERNERS-LEE, T.; FIELDING, R. T.; MASINTER, L. *Uniform Resource Identifier (URI): Generic Syntax*. [S.l.], 2005. Disponível em: <<http://tools.ietf.org/html/rfc3986>>.
- BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. The semantic web. *Scientific American*, 2001.
- BOULOS, M. N. K.; ROUDSARI, A. V.; CARSON, E. R. Towards a semantic medical web: Healthcybermap's tool for building an rdf metadata base of health information resources based on the qualified dublin core metadata set. *Med Sci Monit*, v. 8, p. 124–136, 2002.
- BRAY, T. et al. *Extensible Markup Language (XML) 1.0 (Fourth Edition)*. [S.l.], 2006. Disponível em: <<http://www.w3.org/TR/2006/REC-xml-20060816/>>.
- BREITMAN, K. K. *Web Semântica: A internet do futuro*. [S.l.]: LTC - Livros Técnicos e Científicos Editora S.A., 2005.
- BRICKLEY, D.; GUHA, R.; MCBRIDE, B. *RDF Vocabulary Description Language 1.0: RDF Schema*. [S.l.], 2004. Disponível em: <<http://www.w3.org/TR/rdf-schema/>>.
- CLARK, K. G. *RDF Data Access Use Cases and Requirements*. [S.l.], 2005. Disponível em: <<http://www.w3.org/TR/rdf-dawg-uc>>.
- COMMUNICATIONS, C. W. *Tim Berners-Lee's original World Wide Web browser*. Internet. Disponível em: <<http://info.cern.ch/NextBrowser.html>>.
- COMMUNICATIONS, C. W. *The website of the world's first-ever web server*. Internet. Disponível em: <<http://info.cern.ch/>>.

- CONSORTIUM, W. W. W. *Semantic Web Layer Cake*. Internet. Disponível em: <<http://www.w3.org/2007/03/layerCake.png>>.
- DATASUS. *Cartão Nacional de Saúde*. Internet. Disponível em: <<http://dtr2001.saude.gov.br/cartao/>>.
- DATASUS. *PRC - Padronização de Registros Clínicos*. Internet. Disponível em: <<http://www.datasus.gov.br/prc/datasus.htm>>.
- DAWSON, F.; HOWES, T. *vCard MIME Directory Profile*. [S.l.], 1998. Disponível em: <<http://www.ietf.org/rfc/rfc2426.txt>>.
- DOGAC, A. et al. Artemis: Deploying semantically enriched web services in the healthcare domain. *Science Direct Information Systems*, v. 31, p. 321–339, 2006.
- DOLIN, R. H. et al. *HL7 Clinical Document Architecture, Release 2*. [S.l.], 2008. Disponível em: <<http://www.w3.org/2001/sw/hcls/>>.
- FEIGENBAUM, L. et al. The semantic web in action. *Scientific American*, December 2007.
- FENSEL, D. et al. [S.l.]: Massachusetts Institute of Technology, 2003.
- FERRACIOLI, F.; CAMARGO-BRUNETTO, M. A. de O. A utilização das recomendações do w3c no sacarweb. *Anais do XVI EAIC*, 2007.
- FIELDING, R. et al. *Hypertext Transfer Protocol – HTTP/1.1*. [S.l.], 1999. Disponível em: <<http://tools.ietf.org/html/rfc2616>>.
- FIELDING, R. T. *Architectural Styles and the Design of Network-based Software Architectures*. Tese (Doutorado) — University of California, 2000. Disponível em: <<http://www.ics.uci.edu/~fielding/pubs/dissertation/top.htm>>.
- FOUNDATION openEHR. *openEHR Primer*. Internet. Disponível em: <http://www.openehr.org/shared-resources/openehr_primer.html>.
- FREITAS, F. L. G. de. *Ontologias e a Web Semântica*. Dissertação (Mestrado) — Universidade Católica de Santos - UniSantos, 2003.
- GLOBE, C. The semantic web: an evolution for a revolution. *Computer Networks*, v. 42, p. 551–556, 2003.
- GROMOV, G. R. *History of Internet and WWW: The Roads and Crossroads of Internet History*. Internet. Disponível em: <<http://www.netvalley.com/intval1.html>>.
- HAWKE, S. *Rule Interchange Format Working Group Charter*. [S.l.], 2005. Disponível em: <<http://www.w3.org/2005/rules/wg/charter.html>>.
- HEFLIN, J. *OWL Web Ontology Language: Use Cases and Requirements*. [S.l.], 2004. Disponível em: <<http://www.w3.org/TR/webont-req>>.
- HERMAN, I. *About W3C - Technology*. [S.l.], 2008. Disponível em: <<http://www.w3.org/Consortium/technology>>.

- HERMAN, I. *W3C Semantic Web Activity*. [S.l.], 2008. Disponível em: <<http://www.w3.org/2001/sw/>>.
- HICKSON, I.; HYATT, D. *HTML 5*. [S.l.], 2008. Disponível em: <<http://www.w3.org/TR/html5/>>.
- HOURS, A. L.; JACOBS, I. *HTML 4.01 Specification*. [S.l.], 1999. Disponível em: <<http://www.w3.org/TR/html401/>>.
- HULSE, N. C. et al. Kat: A flexible xml-based knowledge authoring environment. *Journal of American Medical Informatic Association*, v. 12, p. 418–430, 2005.
- HYVONEN, E.; VILJANEN, K.; SUOMINEN, O. Healthfinland - finnish health information on the semantic web. *Proceedings of the 6th International Semantic Web Conference (ISWC 2007)*, November 2007.
- IANNELLA, R. *Representing vCard Objects in RDF/XML*. [S.l.], 2001. Disponível em: <<http://www.w3.org/TR/vcard-rdf/>>.
- INITIATIVE, D. C. M. *Dublin Core Metadata Element Set, Version 1.1*. [S.l.], 2008. Disponível em: <<http://dublincore.org/documents/dces/>>.
- JACOBS, I. *World Wide Web Consortium Process Document*. [S.l.], 2005. Disponível em: <<http://www.w3.org/2005/10/Process-20051014/>>.
- JACOBS, I. *About W3C*. [S.l.], 2008. Disponível em: <<http://www.w3.org/Consortium/Overview.html>>.
- JACOBS, I. *About W3C: History*. [S.l.], 2008. Disponível em: <<http://www.w3.org/Consortium/history>>.
- JONGE, A. de. *HTML V5 and XHTML V2: Two competing standards coexist*. [S.l.], 2007. Disponível em: <www.ibm.com/developerworks/xml/library/x-xml5xhtml2.html>.
- KOIVUNEN, M.-R.; MILLER, E. *W3C Semantic Web Activity*. [S.l.], 2001. Disponível em: <<http://www.w3.org/2001/12/semweb-fin/w3csw>>.
- LEAO, B. de F. et al. The brazilian national health informatics strategy. *Studies in Health Technology and Informatics*, v. 84, p. 38–42, 2001.
- LINDEN, H. van der et al. Proper revisited. *International Journal of Medic Informatics*, v. 74, p. 235–244, 2004.
- MANOLA, F.; MILLER, E.; MCBRIDE, B. *RDF Primer*. [S.l.], 2004. Disponível em: <<http://www.w3.org/TR/rdf-primer/>>.
- MCCARRON, S. et al. *XHTMLTM 1.1 - Module-based XHTML - Second Edition*. [S.l.], 2007. Disponível em: <<http://www.w3.org/TR/xhtml11/>>.
- MCDONALD, C. J. et al. Open source software in medical informatics - why, how and what. *International Journal of Medic Informatics*, v. 69, p. 175–184, 2003.
- MCQUEEN, C. M. S.; THOMPSON, H. *XML Schema*. [S.l.], 2008. Disponível em: <<http://www.w3.org/XML/Schema>>.

NORMALISATION, C. E. de. *CEN/TC 215 Executive Summary*. Internet. Disponível em: <<http://www.cen.eu/CENORM/Sectors/TechnicalCommitteesWorkshops/CENTechnicalCommittees/CENTechnicalCommittees.asp>>.

O'REILLY, T. *What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software*. [S.l.], 2005. Disponível em: <<http://www.oreillynet.com/pub/a-oreilly%20-tim/news/2005/09/30/what-is-web-20.html>>.

ORGANIZATION, H. L. S. *What's HL7*. Internet. Disponível em: <<http://www.hl7.org>>.

ORGANIZATION, I. S. *ISO/TC 215 Business Plan Draft 2*. Internet. Disponível em: <http://www.iso.org/iso/iso_technical_committee?commid=54960>.

ORGANIZATION, W. H. *International Classification of Diseases (ICD)*. Internet. Disponível em: <<http://www.who.int/classifications/icd/en/>>.

PASSIN, T. B. *Explorer's Guide to the Semantic Web*. [S.l.]: Manning Publications Co., 2004.

PEMBERTON, S. et al. *XHTMLTM 1.0 The Extensible HyperText Markup Language (Second Edition)*. [S.l.], 2002. Disponível em: <<http://www.w3.org/TR/xhtml1/>>.

PETRY, K.; LOPES, P. M. A.; WANGENHEIM, A. von. Padrões para a interoperabilidade na saúde. *Anais do CBIS 2006*, 2006.

PRUD'HOMMEAUX, E. *Semantic Web Health Care and Life Sciences (HCLS) Interest Group*. [S.l.], 2008. Disponível em: <<http://www.w3.org/2001/sw/hcls/>>.

PRUD'HOMMEAUX, E.; SEABORNE, A. *SPARQL Query Language for RDF*. [S.l.], 2008. Disponível em: <<http://www.w3.org/TR/rdf-sparql-query/>>.

SAMWALD, M. et al. Experiences with the conversion of senselab databases to rdf/owl. *W3C Interest Group Note*, June 2008. Disponível em: <<http://www.w3.org/TR/hcls-senselab/>>.

SAÚDE, S. de Gestão de Investimentos Ministério da. *O cartão nacional de saúde - Instrumento para um novo modelo de atenção*. [S.l.].

SCHOROEDER, M.; NEUMANN, E. Semantic web for life sciences. *Web Semantic: Science, Services and Agents on the World Wide Web*, v. 4, p. 167, 2006.

SHADBOLT, N.; HALL, W.; BERNERS-LEE, T. The semantic web revisited. *IEEE Intelligent Systems*, June 2006.

SILVA, A. F. da; COSTA, E. de B.; JATOBÁ, A. Web semântica e gestão de conteúdos: Um estudo de caso em um departamento acadêmico. 2007.

SUPLEMENTAR, M. da Saúde Agência Nacional de S. *Padrão TISS - Troca de Informações em Saúde Suplementar*. [S.l.], 2007.

SWICK, R. *Metadata Activity Statement*. [S.l.], 2002. Disponível em: <<http://www.w3.org/Metadata/Activity.html>>.