Instituto de Matemática e Estatística
Universidade de São Paulo

Web Semântica

Paulo Augusto Rosa

MAC 5701: Tópicos em Ciência da Computação
Professora Orientadora: Leliane Nunes Barros

Dezembro / 2002

1. INTRODUÇÃO 3
2. WEB SEMÂNTICA 4
3. ARQUITETURA DA WEB SEMÂNTICA 6
CAMADA 1: UNICODE E URI 7
CAMADA 2: XML, NS E XML SCHEMA 8
CAMADA 3: RDF E RDF SCHEMA 10
CAMADA 4: ONTOLOGY VOCABULARY 11
CAMADA 5: LOGIC 11
CAMADA 6: PROOF 11
CAMADA 7: TRUST + DIGITAL SIGNATURE 11
4. RDF (RESOURCE DESCRIPTION FRAMEWORK) 13
4.1. MODELO BÁSICO DE RDF 13
4.2. EXEMPLOS DE DECLARAÇÃO 15
5. ONTOLOGIAS 17
6. FERRAMENTAS, SERVIÇOS E ÁREAS DE APLICAÇÃO 19
7. CONCLUSÃO 20
8. BIBLIOGRAFIA 22

Sumário

Este documento é o resultado dos estudos realizados durante o segundo semestre de 2002 para a disciplina de Tópicos em Ciência da Computação (MAC 5701). As áreas abordadas nesta monografia são Internet e Web Semântica. Primeiramente, é apresentado o conceito da Web Semântica, seguido do modelo de camadas no qual ela é estruturada. No decorrer do documento, cada camada é explicada detalhadamente, tendo a camada de RDF um maior detalhamento. Entre os temas que serão abordados estão: funcionamento da Web Semântica, a descrição das tecnologias necessárias para sua realização, o entendimento das principais linguagens de descrição de ontologias e a pesquisa dos possíveis serviços na Web Semântica.

1. Introdução

O crescimento extraordinário da Teia Mundial (*WWW - World Wide Web*) aumentou demasiadamente o número de documentos eletrônicos disponíveis. Atualmente, segundo o sítio do Google (http://www.google.com/), considerado a melhor ferramenta de busca da Internet, há cerca de 2,5 bilhões de páginas Web indexadas na Internet. Trata-se de uma quantidade extraordinária de documentos com os mais variados assuntos que estão distribuídos por todo o mundo de uma forma descentralizada e desestruturada. Apesar deste grande caos, a simplicidade com que foi idealizada a Web e a utilização de hipertexto tornaram a Web um grande sucesso.

Contudo, toda informação da Internet encontra-se difusa e desorganizada, afinal trata-se de um repositório mundial descentralizado de documentos eletrônicos. Desde agosto de 1994, com o surgimento do primeiro mecanismo de busca de documentos na Internet, o *Yahoo Directory Service* (http://www.yahoo.com/), motores de busca e programas de indexação de documentos vêm se aperfeiçoando no sentido de responder consultas da melhor forma possível para o usuário.

Apesar de todo este desenvolvimento das ferramentas de busca que atualmente se observa, é possível que já tenhamos alcançado o limite de poder dessas ferramentas, se considerarmos a forma atual de codificação de dados na Internet. As ferramentas tradicionais de busca atualmente são baseadas na indexação de documentos a partir do texto puro, privilegiando documentos através de marcações HTML (*tags)* associadas, quantidade de páginas referenciadas e quantidade de repetições da palavra sendo procurada. O principal problema desta abordagem é o fato que os agentes (*softwares*) de busca não conseguem entender o conteúdo das páginas, realizando consulta apenas por combinação de palavras-chave, pois os documentos Web foram criados de forma que

fossem entendidos apenas por humanos. Surge, então, a necessidade de se criar uma nova estrutura para a Web na qual seja possível escrever programas (*softwares*) capazes de entender o significado de documentos. Esta nova proposta de estrutura da Web é chamada de Web Semântica (*Semantic Web*) [1].

Esse documento tem como objetivo apresentar os conceitos envolvidos na Web Semântica bem como as tecnologias necessárias para o seu funcionamento. No decorrer deste documento, serão apresentadas as principais linguagens de representação do conhecimento para a Web Semântica, uma pequena lista das ferramentas existentes que dão suporte para o desenvolvimento desta tecnologia, algumas áreas de aplicação na qual ela pode ser utilizada e os principais projetos acadêmicos e comerciais atualmente em desenvolvimento.

Ao final da leitura deste texto, teremos algumas idéias de como responder às seguintes questões: O que a Web Semântica pode trazer de novo para a Internet? Quais os reais benefícios em utilizá-la? Quais as tecnologias necessárias para o funcionamento pleno da Web Semântica?

2. Web Semântica

A Web Semântica surgiu com a idéia de Tim Bernes-Lee, criador da Web e atual diretor do *W3C Consortiun*[1], na qual haveria uma extensão do conteúdo da Web atual de maneira a possibilitar o entendimento da informação por máquinas ("*Machine-understandable Information*") [1].

Em artigo publicado na revista *Scientific American* em 2001, Tim Bernes-Lee, Ora Lassila e James Hendler contam uma pequena história ilustrando o que seria a Web Semântica em sua plenitude. Resumidamente, dois irmãos, Pete e Lucy necessitam agendar sessões de fisioterapia para sua mãe, dividindo entre eles a tarefa de acompanhá-la. Para isto, através de um agente de software, utilizam a Web Semântica para encontrar algumas alternativas. Tendo conhecimento de informações a respeito de Pete, este agente percorre a Web em busca de informações relevantes e compatíveis com suas preferências. Em questão de minutos, o agente oferece uma lista de opções para que Pete escolha a que lhe for mais conveniente. Não satisfeito com a primeira lista, pois o término da sessão de fisioterapia coincidia com o horário de pico de trânsito, Pete pede ao agente que refaça a busca, agora acrescentando alguns dados para restringir as alternativas. Desta vez, o resultado oferecido é satisfatório e com algumas medidas simples, Pete muda parte de seus compromissos, obtém a aceitação do agente de Lucy através de certificados digitais e atualiza a agenda.

Atualmente, toda a informação da Web esta codificada em HTML, uma linguagem de

[1] O World Wide Web Consortium (ou W3C Consortium), criado em outubro de 1994, tem como principal objetivo liderar o desenvolvimento de protocolos comuns para assegurar a interoperabilidade na Internet. Possui mais de 500 organizações membros que contribuem para o crescimento da Web.

marcação com o objetivo de formatar a visualização em navegadores. Assim, humanos processam informação e máquinas conseguem apenas consumir dados. O objetivo da Web Semântica, portanto, é transformar o conteúdo atual da Web num formato no qual não só humanos, mas também computadores entendam o significado dos dados (informação), de forma a permitir agentes inteligentes recuperar e manipular informação pertinente.

As palavras "dado" e "informação" são muito utilizadas neste documento e, portanto, há a necessidade de uma breve explicação da diferença entre ambas. A palavra "dado" refere-se a todo texto que não tem significado imbutido, ou seja, que não pode ser entendido. Já a palavra "informação" refere-se a todo o texto no qual o leitor é capaz de entender seu significado. Um exemplo simples esclarecerá a diferença um pouco melhor. Imagine uma pessoa que não entende a língua portuguesa recebesse um texto escrito em português. Para esta pessoa o texto é apenas um "dado", pois ela não conseguirá entender nada. Mas para outra pessoa que entenda a língua portuguesa, o texto torna-se uma informação pois é possível entender o que está escrito.

Utilizando-se desta nova proposta da Web, agentes de software poderiam realizar tarefas mais complicadas no conteúdo da Internet do que é possível realizar atualmente. Nas seções posteriores são apresentadas algumas aplicações já implementadas no contexto da Web Semântica e algumas áreas nas quais ela seria bem aproveitada.

Para resolver o problema de automatização dos processos da Web, a Web Semântica utilizará o conceito de ***metadado***. Este conceito deu origem ao projeto da Web Semântica do consórcio W3C que era chamado de Metadados. Metadados são dados que descrevem o conteúdo, a estrutura, a representação e o contexto de algum dado ou conjunto de dados. Embora o termo metadado seja relativamente novo, os conceitos básicos de metadados já vêm sendo utilizados há muito tempo. Vejamos, por exemplo, o catálogo de uma biblioteca, que tem servido como gerenciamento de coleção de informações sobre livros e para ferramentas de descoberta de recursos.

Nesse cenário, a Web Semântica proporciona o processamento por máquinas na semântica dos dados da Internet, ou seja da informação, possibilitando serviços inteligentes entenderem a informação e realizando um alto nível de transformação da informação, procura e recuperação. Assim, novos problemas poderão ser resolvidos.

3. Arquitetura da Web Semântica

Para que a realização da Web Semântica seja possível, uma série de tecnologias são necessárias, das quais algumas já são utilizadas e consideradas padrão enquanto outras ainda precisam ser desenvolvidas. Dentre essas teconologias, podemos colocar alguns pontos que precisam ficar bem definidos.

- A definição da arquitetura e da infra-estrutura da Web Semântica.
- O estabelecimento de linguagens formais para a representação do conhecimento no contexto da Web Semântica.
- A disponibilização da informação em linguagens baseadas em ontologias.
- A criação de serviços de inferência.
- O desenvolvimento de ferramentas de suporte para a Web Semântica.
- A criação de política de segurança e privacidade na Web Semântica.

Tim Berners-Lee, Ora Lassila e James Hendler definiram a arquitetura da Web Semântica em camadas de linguagens, como mostrado na figura 3.1.

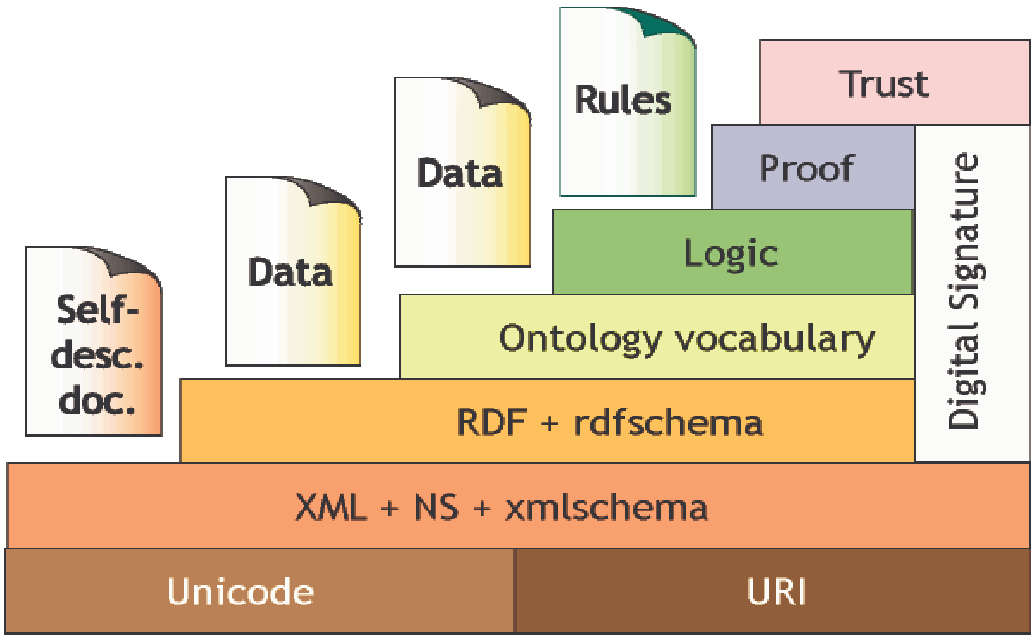


Figura 3.1: A arquitetura de linguagens da Web-Semântica.

Basicamente, estas camadas podem ser agrupadas em três subcamadas. A primeira, relacionada à base sintática da Web Semântica, é formada por: Unicode, URI e XML, NS e XML Schema. Nesta camada, Unicode e URI são responsáveis pela identificação de recursos e XML, NS e XML Schema são responsáveis pela estruturação dos elementos. A segunda, relacionada à semântica, é formada por: RDF, RDF Schema e Ontology Vocabulary. E, a terceira camada, relacionada à lógica, é formada por Logic, Proof, Trust e Digital Signature.

Este diagrama de camadas mostra-nos a estrutura de linguagens da Web Semântica. Desta maneira, cada linguagem é construída baseada nas linguagens definidas em níveis inferiores, sendo os níveis de URI e Unicode os mais fundamentais, pois tratam da representação de caracteres e recursos. A camada XML também é importante pois define a sintaxe de todas as linguagens das camadas superiores. Os níveis mais altos do modelo de camadas servem para dar um significado maior para a informação que vai além do que RDF pode proporcionar.

A seguir, cada uma das camadas será explicada.

Camada 1: Unicode e URI

Unicode[2] é uma linguagem que define uma forma padrão para a representação de caracteres. Unicode proporciona uma forma única para a representação de um caracter não importando a plataforma, o programa nem a linguagem que está sendo utilizada.

A utilização de Unicode na Web Semântica proporciona a capacidade de troca de símbolos de maneira universal, requisito fundamental para o sucesso desta nova proposta de representação de informação na Internet.

Basicamente, **URI**[3] (*Uniform Resource Identifier*) estabelece uma forma padrão para a identificação de recursos. Dentre os diversos recursos que podem ser identificados, pode-se citar: documentos, imagens, arquivos, serviços e páginas HTML.

Através da utilização de URI faz-se a referência para recursos representados na Web Semântica.

No contexto da Internet, o conceito de URI já é bem utilizado. Na Web é utilizado um tipo de URI chamado URL. Através da URL é possível endereçar documentos utilizando protocolos específicos da Internet como *http* e *ftp*. Um exemplo de URL é a página principal da USP, *http://www.usp.br/*.

---

[2] http://www.unicode.org/

[3] http://www.w3c.org/Addressing/

Camada 2: XML, NS e XML Schema

XML[4] (*Extensible Markup Language*) é uma linguagem de marcação muito simples e flexível derivada da linguagem SGML[5] (*Standard Generalized Markup Language*), projetada para uso específico da Web. SGML é um padrão internacional de processamento de texto. XML proporciona o padrão para a representação das estruturas, na qual todas as linguagens, para expressar significado da Web semântica, serão baseadas. Como contribuição à estrutura da Web Semântica, XML faz com que seja possível a interoperabilidade sintática entre agentes.

O consórcio W3C aponta alguns dos benefícios que a linguagem XML ou oferece:

- linguagem independente de fornecedor e de fácil compreensão;
- redução de custos de treinamento e desenvolvimento devido ao formato simples;
- linguagem legível por humanos e por agentes de software;
- aumento da confiança devido à automatização dos agentes de software;
- trata-se da base de apoio da Web Semântica na qual proporciona interoperabilidade;
- liberdade na criação das estruturas por usuários;
- permite a reutilização de dados sem nenhuma ligação com ferramentas proprietárias ou formatos não documentados.

XML é uma linguagem de marcação, utilizando marcações (*tags*) de início e fim para a definição da informação. Cada marcação, que é delimitada pelos caracteres '<' e '/>', possui um elemento (*element*) e pode ou não ter atributos (*attribute*).

Um exemplo de documento xml é apresentado abaixo:

```
<person id=1>
    <name>Paulo Rosa</name>
    <email>par@ime.usp.br</email>
<person>
```

No exemplo, *'person'*, *'name'*, e *'email'* são elementos e *'id'* é um atributo.

A linguagem **XML Schema**[6] permite a definição e a descrição de estruturas e de conteúdos de documentos XML. Através dessa linguagem, define-se o formato válido de

[4] http://www.w3.org/XML/

[5] site do SGML

[6] http://www.w3.org/XML/Schema/

um documento XML, incluindo quais os elementos e atributos são permitidos ou não, quais são as suas localizações, o número de ocorrências de cada elemento e outras características. Ou seja, proporciona mecanismos para definição de gramáticas para correção de documentos XML.

Torna-se importante notar a diferença entre documento XML válido e um documento XML bem formado. Um documento XML bem formado é aquele no qual está sintaticamente correto na visão de XML. Entretanto, um documento XML válido é aquele que segue as especificações do seu documento XML Schema correspondente.

Desta maneira, os documentos XML Schema acabam sendo razoavelmente grandes, pois a definição de uma gramática tende a ser complicada. A seguir, um exemplo simples mostra a utilização de um documento XML Schema.

Inicialmente, um documento XML:

```
<?xml version="1.0" encoding="UTF-8" ?>
    <xsd:schema xmlns:xsd=http://www.w3.org/2001/XMLSchema>
    <xsd:element name="vehicles">
        <xsd:complexType>
            <xsd:sequence>
                <xsd: element name="nickname" type="xsd:string" maxOccurs="unbounded"/>
            </xsd:sequence>
        </xsd:complexType>
    </xsd:element>
</vehicles>
```

Agora, o documento XML correspondente:

```
<?xml version="1.0" encoding="UTF-8" ?>
<vehicles xmlns:xsi=http://www.w3.org/2001/XMLSchema-instance
        xsi:noNamespaceSchemaLocation=http://lucas.ucs.ed.ac.uk/xml-schema/smlns/simple.xsd>
    <nickname>Bog Hopper</nickname>
    <nickname>Wee Beastie</nickname>

</vehicles>
```

O consórcio W3C define **XML NameSpace**[7] da seguinte maneira: Um XML Namespace é uma coleção de nomes, identificados por uma referência URI [RFC 2396], que é utilizada em documentos XML, como tipos de elementos e nomes de atributos.

Trata-se de um método para qualificar nomes de elementos e atributos usados em documentos XML, através da associação de referências URI. Através deste mecanismo

[7] http://www.w3.org/TR/REC-xml-names/

de espaço de nomes, é possível a combinação de documentos com utilização de vocabulário compartilhado. Através do mecanismo de espaço de nomes definido em XML, é possível compartilhar e reutilizar a definição de outros esquemas XML sem que haja problemas de colisão de nomes.

No documento XML Schema apresentado anteriormente, o elemento *xsd:schema* possui um atributo denominado *xmlns* que é justamente o exemplo da utilização do XML NameSpace, e possui o valor *xsd=http://www.w3c.org/2001/XMLSchema*. Através deste atributo faz-se referência a outras definições pela utilização de um nome. Ou seja, tudo que é definido no documento *http://www.w3c.org/2001/XMLSchema* pode ser utilizado com referência ao nome *xsd*. Por exemplo, o elemento *complexType* é definido neste documento e portanto referenciado como *xsd:complexType*.

Camada 3: RDF e RDF Schema

Desenvolvido pelo consórcio W3C, **RDF**[8] **(Resource Description Framework)** é uma linguagem para representação de informação na Web. Trata-se de uma infra-estrutura que fornece a habilidade para codificação, troca e reutilização de metadados. RDF define um modelo de dados para descrição de semântica de dados para o entendimento pelo computador.

É o fundamento para o processamento de metadados (informação sobre informação). Basicamente, *metadado* pode ser definido como "dado estruturado sobre dado".

Atualmente, o desenvolvimento desta linguagem faz parte das atividades do grupo de Web Semântica do consórcio W3C, cujo trabalho visa obter uma linguagem que troque informação processável por computador através da Web. Através da utilização de primitivas de modelagem pré-definidas, RDF proporciona a interoperabilidade semântica, isto é, permite que as aplicações troquem informações de maneira que possam ser entendidas por computador. De acordo com o modelo de camadas apresentado, RDF utiliza a sintaxe XML como forma de troca e processamento de metadados.

Como seu próprio nome diz, RDF proporciona um modelo para descrição de recursos. Um recurso é qualquer objeto que é identificado por uma URI. Recursos possuem propriedades, que são atributos ou características do objeto, e são identificadas por tipo de propriedade (*property-type)*. Cada tipo de propriedade possui valores (*values)* e expressa o relacionamento com algum outro recurso. Valores podem assumir valores primitivos como textos e números ou podem referenciar outros recursos.

Na seção 4, RDF será apresentado com um pouco mais de detalhe.

Similarmente à linguagem XML Schema, **RDF Schema**[9] é uma linguagem que define a

[8] http://www.w3.org/TR/rdf-schema/

[9] url de rdf schema

estrutura válida para dos documentos RDF.

RDF e RDF Schema são recomendações do consórcio W3C que definem o padrão para a representação de metadados. São a base de todas as linguagens para expressar semântica da Web Semântica, devido à adoção pelo consórcio W3C.

Camada 4: Ontology Vocabulary

A camada de Ontologia refere-se às linguagens para representação do conhecimento, igualmente ao que a camada de RDF propõe-se a realizar. No entanto, as linguagens de Ontologia aparecem com um poder maior de representação de recursos e relacionamento entre eles do que a do modelo de dados RDF proporciona.

Linguagens de Ontologias originaram-se de duas subáreas de Inteligência Artificial chamadas de Engenharia do Conhecimento e sistemas baseados em conhecimento.

Mais adiante será dado um enfoque maior na definição de Ontologias e nas linguagens disponíveis atualmente.

Camada 5: Logic

A camada de Lógica proporciona a definição de semântica em linguagem formal habilitando a execução de serviços inteligentes. É composta principalmente por regras de inferência, com as quais os agentes poderão se utilizar para relacionar e processar informação.

Camada 6: Proof

De posse das regras de inferência da camada imediatamente inferior a esta (camada de prova), os agentes podem ter mais poder para raciocinar sobre conceitos e relacioná-los na camada de ontologia.

Esta é a camada na qual pode-se obter explicações (provas) sobre as respostas dadas por agentes que consomem alguma informação com o objetivo de verificar se a dedução foi correta.

Camada 7: Trust + Digital Signature

A camada de confiança (*Trust*) conjuntamente com a camada de assinatura digital (*digital signature*) proporciona mecanismos para prevenção de inconsistências na Web Semântica. Através de aplicações criadas neste nível, é possível criar agentes que saibam dizer, identificar e validar algum tipo de informação.

Trata-se de outra característica importante da Web Semântica e muito importante no ambiente da Internet, na qual blocos de dados encriptados podem ser utilizados para garantir a autenticidade das fontes e a confiabilidade da informação que os agentes

consultam.

Relembrando da história de Tim Berners-Lee, apresentada anteriormente, esta camada torna-se presente no momento em que os agentes de Pete e Lucy trocam informações bem como na comunicação do agente de Pete com as clínicas para obtenção de informações. É muito importante para o agente de Pete que as informações sejam autênticas e que seus dados, que podem ser confidenciais, trafeguem na Internet com segurança.

Com esta visão do modelo de camadas, podemos entender que para o funcionamento da Web Semântica, computadores devem ter acesso a coleções de informações estruturadas e conjuntos de regras de inferências que podem ser utilizadas para conduzir o raciocínio automatizado.

4. RDF (Resource Description Framework)

RDF é uma linguagem de infra-estrutura que permite a codificação, a troca e a reutilização de metadados. É a base para a representação e o processamento de metadados. As linguagens RDF e RDF Schema compõem a camada mais importante da Web Semântica, sendo recomendações do consórcio W3C, formam a base de todas as propostas de linguagens para descrição de ontologias na Web semântica.

RDF é uma aplicação de XML que impõe restrições estruturais para fornecer métodos precisos (sem ambigüidade) para expressar semântica no ambiente da Internet. Também fornece um meio para publicar vocabulário para humanos lerem (*human-readable*) e para serem processados pelo computador (*machine-processable*) projetado para encorajar a reutilização e extensão de semântica de metadados entre diferentes comunidades de informação.

As restrições estruturais de RDF impostas para garantir a consistência da codificação e troca (compartilhamento) de metadados padronizados permitem o compartilhamento de pacotes distintos de metadados, definidos por diferentes comunidades de descrição de recursos.

Com relação à sintaxe, XML é apenas uma das possíveis e a que se enquadra melhor no ambiente da Internet. Através da utilização de XML e conjuntamente com as linguagens padrões de representação de recursos (URI) e do conjunto de caracteres universais (Unicode) o transporte e o armazenamento de informações na Web ficam facilitados.

O principal objetivo da linguagem RDF é definir um mecanismo para descrever recursos que não façam absolutamente nenhuma suposição sobre um domínio particular de aplicação, nem definam a semântica de qualquer domínio de aplicação. Soma-se a isto a característica que RDF proporciona de maximizar a interoperabilidade de servidores e clientes Web desenvolvidos de maneira independente.

Para a definição de metadados, RDF utiliza um sistema de classes, chamado de *RDF Schema*, muito parecido com sistemas de modelagem de entidade e relacionamento e programação orientada a objetos. Através do compartilhamento de esquemas (*RDF Schema*), obtém-se o suporte a reutilização de definições de metadados.

4.1. Modelo básico de RDF

O desenvolvimento da linguagem RDF vem sendo influenciado por diversas comunidades. Algumas delas são: as comunidades de padronização da Web; a comunidade de documentos estruturados (SGML e XML); a comunidade de programação orientada a objetos; a comunidade de banco de dados; a comunidade de representação de conhecimento.

Como aspecto importante do RDF para a comunidade de representação do conhecimento

é o fato que a linguagem não especifica nenhum mecanismo de raciocínio. O mecanismo de raciocínio deveria ser construído no topo desta linguagem.

Formalmente, utiliza-se de um modelo básico de dados para sua definição e consiste de três tipos de objetos, todos identificados por URI's: **Recursos**, (*resources)*, **Propriedades** (*properties)* e **Declarações** (*statements)*.

As propriedades RDF podem ser encaradas como atributos de recursos, isto é, pares de atributo-valor. As propriedades RDF também representam relações entre recursos lembrando um diagrama de entidade e relacionamento. Na terminologia de projeto orientado a objeto, recursos correspondem a objetos e propriedades correspondem a instâncias de variáveis.

O modelo de dados de RDF é uma maneira livre de sintaxes para a representação de expressões, o que permite descrever uma relação de equivalência.

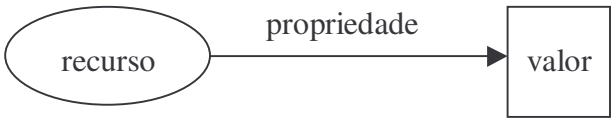
Relação de Equivalência: Duas expressões RDF são equivalentes se e somente se suas representações de modelos de dados são as mesmas.

Essa definição de equivalência permite alguma variação sintática em expressões sem alterar o significado delas. A representação do modelo de dados pode ser feita através de triplas, grafos ou XML.

Todos os objetos descritos por expressões RDF são chamados de **Recursos**. Um recurso pode ser: uma página Web inteira, como por exemplo, o documento HTML http://www.w3c.org/Overview.html; um elemento HTML ou XML específico dentro de um documento; uma coleção completa de páginas; um objeto que não é acessível diretamente pela Web. Recursos são sempre nomeados por URI's e opcionalmente adicionados de âncora.

Uma **Propriedade** é um aspecto, característica, atributo ou relação específica usada para descrever um recurso. Cada propriedade possui: um significado específico, define seus valores permitidos, suas relações com outras propriedades e os tipos de recursos que ela pode descrever.

Uma **Declaração** RDF é definida em três partes: um recurso específico, um nome de propriedade e o valor desta propriedade associada ao recurso. Essas partes individuais de uma declaração são chamadas, respectivamente, de sujeito, predicado e objeto. Na figura abaixo é apresentado o modelo de declaração.



O objeto de uma declaração, ou seja, o valor de uma propriedade, pode ser um outro recurso ou pode ser um literal, isto é, um recurso especificado por uma URI, uma simples

string ou um tipo de dado primitivo definido em XML.

4.2. Exemplos de declaração

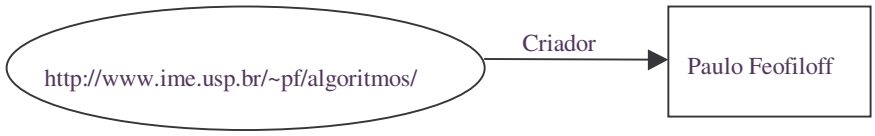
Considere a seguinte sentença:

Paulo Feofiloff é o criador do recurso http://www.ime.usp.br/~pf/algoritmos/.

De acordo com o modelo de dados de RDF, temos:

Sujeito (recurso)	Http://www.ime.usp.br/~pf/algoritmos/
Predicado (propriedade)	Criador
Objeto (literal)	"Paulo Feofiloff"

E graficamente:



Na representação XML, o documento fica:

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
         xmlns:s="http://description.org/schema/">
    <rdf:Description about="http://www.ime.usp.org/~pf/algoritmos">
        <s:Creator>Paulo Feofiloff</s:Creator>
    </rdf:Description>
</rdf:RDF>
```

O documento RDF acima possui vários elementos, com propósitos e significados diferentes. Primeiramente, um recurso RDF é identificado pelas marcações *<rdf:RDF>* e *</rdf:RDF>*. Neste documento, o elemento *rdf:RDF* possui dois atributos para a identificação de nomes. Um deles é referido como ***rdf*** para os conceitos definidos da própria linguagem RDF e o outro como ***s*** para referência a conceitos definidos na URI especificada.

Desta maneira, no documento o sujeito da declaração é especificado pela marcação *rdf:Description*, o predicado é definido como *s:Creator* e o objeto é o valor deste elemento.

Contudo, geralmente há a necessidade de referirmos a coleção de recursos. Para mostrar um exemplo um pouco mais complexo, vamos mostrar o conceito de *Containers* de RDF, que é utilizado para conter lista de recursos ou literais.

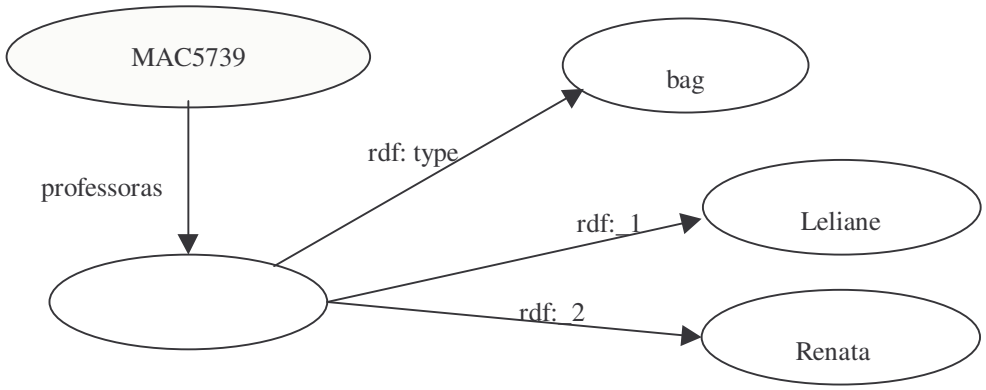
Considere agora o seguinte exemplo:

As professoras do curso de IA são Leliane e Renata.

De acordo com o modelo de dados de RDF, temos:

Sujeito (recurso)	http://sistemas.usp.br/fenixweb/fexDisciplina?sgldis=MAC5739
Predicado (propriedade)	Professoras
Objeto (recurso)	Leliane e Renata

E graficamente:



Na representação XML, o documento fica:

```
<rdf:RDF>
        <rdf:Description about="http://sistemas.usp.br/fenixweb/fexDisciplina?sgldis=MAC5739">
                <s:professores>
                  <rdf:Bag>
                        <rdf:li resource="http://www.ime.usp.br/~leliane/">
                        <rdf:li resource="http://www.ime.usp.br/~renata/">
                  </rdf:Bag>
                </s:professores>
        </rdf:Description>
</rdf:RDF>
```

Neste documento RDF, não são apresentadas as marcações de XML Name Space. O que há de novidade em relação ao exemplo anterior é a utilização elemento *Bag* definido no modelo de RDF Schema.

Portanto, no exemplo, o recurso "professores" é utilizado como um objeto do tipo Bag contendo recursos, no caso, as professoras da matéria.

5. Ontologias

Parte da tarefa de prover informação com significado bem definido é criar uma linguagem de ontologia para descrever e representar objetos na Web. Na literatura, encontramos diversas definições de ontologias, mas basicamente, uma ontologia é uma especificação formal de uma conceitualização[6], que proporciona um entendimento comum de um domínio que pode ser comunicado através de pessoas e sistemas. Ontologias são conjuntos de asserções que definem as relações entre conceitos e estabelecem regras lógicas de raciocínio sobre eles. Para a representação de uma conceitualização é necessária a utilização de uma linguagem para representação de conhecimento. Alguns exemplos deste tipo de linguagem já conhecidos da área de Inteligência Artificial são: KIF, Loom e PowerLoom, e Frame-Logic, todos baseados em lógica de primeira ordem.

Ontologias definem uma semântica formal da informação permitindo processamento pelo computador. Permite a ligação entre o entendimento do mundo real e o conteúdo processado por computador através de terminologias consensuais. A iniciativa sobre ontologias do W3C Consortium tem como objetivo explorar linguagens que estendam a capacidade de XML e RDF de expressarem semântica.

Ontologias permitirão a definição de estrutura semântica dos documentos criando novas possibilidades, comparado com os atuais sistemas de gerenciamento de documentos.

Segundo James Hendler, uma Ontologia é uma coleção de informações com uma especificação formal e consensual de conceitos, definindo as relações entre os conceitos e que providencia um compartilhamento e um entendimento comum de um domínio que pode ser comunicado entre pessoas e sistemas de aplicações.

Contudo, alguns problemas podem ocorrer com a utilização de ontologias: no compartilhamento e no entendimento de ontologias diferentes. Perguntas que vão além do escopo deste documento.

No ambiente da Internet, entretanto é fundamental a utilização de uma sintaxe padrão. Assim, XML apresenta-se como padrão para criação de linguagens de representação de conhecimento. Desta maneira, diversas linguagens surgiram, utilizando a sintaxe padrão XML, para descrição de ontologias para a Web Semântica.

Algumas das principais linguagens para descrever ontologias são:

- SHOE[10] (*Simple HTML Ontology Extension*);
- DAML[11] (*DARPA Agent Markup Language*);

[10] http://www.cs.umd.edu/projects/plus/SHOE/

- OIL[12] (*Ontology Inference Layer*);
- DAML+OIL[13]
- OWL[14] (*Ontology Markup Language*);

SHOE é uma linguagem para representação do conhecimento baseada na linguagem HTML, na qual são adicionadas novas marcações (*tags*) necessárias para expressar semântica nos dados das páginas. SHOE permite a utilização de regras de inferência para ser definida na especificação da ontologia. Criada em meados de agosto de 1999 e desenvolvida pelo Departamento de Ciência da Computação da Universidade de Maryland, o projeto foi descontinuado, pois a equipe foi mudou-se para o projeto da linguagem DAML.

O desenvolvimento da linguagem **DAML** foi iniciado em agosto de 2000 pelo programa do grupo DARPA, cujo principal objetivo seria desenvolver uma linguagem e ferramentas para viabilizar a Web Semântica. A linguagem DAML é baseada nos padrões W3C XML e RDFe em predicados de lógica de primeira ordem. Trata-se de um projeto no qual já foi desenvolvida uma série de ferramentas para suporte da linguagem.

A linguagem de ontologia **OIL** é desenvolvida pelo projeto On-to-knowloegde[15]. Trata-se de uma linguagem com um poder de expressividade para representação de conhecimento e suporte a inferências muito maior que a linguagem RDF proporciona.

A linguagem **DAML+OIL** foi criada em dezembro de 2000 a partir da fusão das linguagens DAML e OIL, com o objetivo de obter o que havia de melhor em cada linguagem. A última versão desta linguagem foi lançada em março de 2001 e proporciona um rico conjunto de construções para criação de ontologias e marcações de informações. Como todas as outras linguagens, também é baseada em padrões W3C, XML, RDF e RDF Schema.

Seguindo o desenvolvimento das linguagens, surge a linguagem **OWL**, que é derivada da linguagem DAML+OIL e atualmente tem a possibilidade de tornar-se uma recomendação do W3C Consortium. São poucas as diferenças entre a linguagem DAML+OIL. Dentre elas, podemos citar: remoção de algumas construções não utilizadas e a habilidade para diretamente afirmar que propriedades podem ser simétricas.

---

[11] http://www.daml.org/

[12] http://www.ontoknowledge.org/oil/

[13] http://www.daml.org/2001/03/daml+oil-index

[14] http://www.w3.org/2001/sw/WebOnt/

[15] http://www.ontoknowledge.org/

6. Ferramentas, Serviços e Áreas de Aplicação

Com realização da Web Semântica e sabendo que as máquinas terão a capacidade de entender o conteúdo das páginas HTML, um típico serviço que logo surge a nossa mente é a busca de informações. Com esta nova tecnologia, uma nova geração de buscadores proporcionará serviços mais avançados de recuperação de informação possibilitando respostas mais precisas, através da utilização de agentes de software e de uma base de conhecimento. Torna-se viável a organização de conteúdos semânticos na Teia Mundial.

A idealização da Web Semântica de Tim Bernes-Lee iniciou a criação de uma infra-estrutura de linguagens e aplicativos necessários para o seu desenvolvimento. Algumas questões surgem referentes à utilização da Web Semântica. Quais tipos de serviços poderão se beneficiar com essa nova estrutura? Os serviços realmente serão melhores do que atualmente já são providos?

A busca de informações na Web não é o único serviço que pode obter benefícios com o desenvolvimento da Web Semântica. Há diversas áreas de aplicação na qual a utilização da Web Semântica poderia trazer benefícios na qualidade dos serviços prestados. Na área de gerenciamento do conhecimento, vejamos como exemplo as grandes empresas que precisam manipular milhares e até milhões de páginas. Estas possuem um grande problema quando é necessário procurar um documento. Com a utilização da Web Semântica este trabalho poderia ser reduzido em muito. Na área de comércio eletrônico, haveria um benefício muito grande na troca de informações entre parceiros.

Abaixo é listado alguns exemplos de serviços que utilizam a estrutura da Web Semântica.

Somente para citar alguns, temos: ferramentas avançadas de busca de documentos, agentes inteligentes móveis (*softbots*) para busca de W*eb Services*, comércio eletrônico inteligente, entre outros.

- ***ItTalks***[16]: Sistema para notificação automática de palestras sobre tecnologia da informação;
- ***DAML-S**:* Linguagem de ontologia derivada da linguagem DAML específica para a utilização em Web Services[17].
- ***Information Retrieval***: O programa DARPA, que desenvolve a linguagem DAML, desenvolveu ferramentas para busca em documentos escritos na linguaem DAML.
- ***Browsing Schedules***[8]: Agentes criados para agendamento de conferências. Um agente chamado RETSINA Calendar que pode navegar na Web Semântica para obter e raciocinar sobre datas de eventos e seus detalhes. Outro agente, chamado Scheduler

[16] http://ittalks.org/jsp/Controller.jsp

[17] http://www.w3.org/2002/ws/

Browser, que organiza detalhes de conferências como localização, data, hotéis e publica em formato HTML e com informação adicional permitindo a ação de outros agentes.

- ***Protege***[18]: Um editor de ontologias. O projeto DARPA também tem editores de ontologias.
- ***DAML Viewer, DAML Validator, DAML Ontology Cheker***[19]: Ferramentas criadas pelo projeto DARPA para, respectivamente, visualizar, validar e verificar ontologias.

[18] http://protege.stanford.edu/

[19] http://www.daml.org/tools/

7. Conclusão

Um dos principais problemas da Web atual refere-se ao fato que o conteúdo codificado nas páginas HTML representam informação para humanos e apenas dados para máquinas, ou seja, os dados da Web são apresentados em linguagem natural. Somando-se a esse problema, há disponível na Internet uma quantidade de informação muito grande, mas que não pode ser tratada automaticamente de forma eficiente.

A Web Semântica surge para resolver este problema e outros, bem como trazer uma abertura muito grande para a criação de aplicações novas cuja ação de agentes seja realizada de maneira mais eficiente e inteligente. Juntamente, a "marcação" da informação através de uma linguagem semântica (ex. DAML, SHOE) os agentes proporcionarão, por exemplo, resultados de busca de informação mais eficientes e precisos.

Sendo um pouco mais realista, o computador não "entenderá" a informação como os humanos, mas poderá manipular dados com muito mais eficiência de maneira mais útil e significativa.

O verdadeiro poder da Web Semântica será alcançado quando for criado um número de ferramentas suficientes para coleta de informação na Web, para o processamento de informação e para a troca de resultados entre agentes. A quantidade de agentes aumentará exponencialmente quanto mais conteúdo processável por computador e serviços automatizados estiverem disponíveis.

Agora, podemos tentar responder algumas questões feitas no início do texto. Dentre elas, o que a Web Semântica pode trazer de novo para a Internet? O que se pode dizer é que a qualidade dos serviços poderia ser aumentada sensivelmente, o que pode ser visto nos exemplos de aplicações apresentados. Quais as tecnologias necessárias para o funcionamento da Web Semântica? Basicamente, as tecnologias envolvidas na Web Semântica seriam aquelas apresentadas no modelo de camadas.

Contudo alguns problemas ainda ficam sem solução. Qual será a linguagem de ontologia a ser padronizada pelo W3C Consortium? Como o conteúdo atual da Web será modificado?

8. Bibliografia

[1] IAN HORROCKS and SERGIO TESSARIS. *Querying the Semantic Web: a Formal Approach*. ISWC 2002, LNCS 2342, pp 177-191, 2002.
http://www.cs.man.ac.uk/~horrocks/Publications/download/2002/iswc2002.pdf

[2] PETER F. PATEL-SCHNEIDER and DIETER FENSEL. *Layering the Semantic Web: Problems and Directions.* ISWC 2002, LNCS 2342, pp. 16-29,2002.
http://www.cs.vu.nl/~dieter/ftp/paper/layering.pdf

[3] TIM BERNERS-LEE, JAMES HENDLER and ORA LASSILA. *The Semantic Web.* Scientific American: Feature Article, May 2001.
http://www.scientificamerican.com/2001/0501lissue/0501berners-lee.html

[4] TIM BERNERS-LEE. *Semantic Web Road map*. September, 1998.
http://www.w3.org/DesignIssues/Semantic.html

[5] D. FENSEL et. al. *Semantic Web Application Areas.*
http://www.cs.vu.nl/~mcaklein/papers/NLDB02.pdf

[6] TOM GRUBBER. What is an Ontology?
http://www.-ksl.stanford.edu/kst/what-is-an-ontology.html

[7] R. SCOTT COST et. al. *ITTalks: A case student in how the semantic Web helps.* Em Int. Semantic Web Working Symposium, Pp. 477-494, 2001.
http://www.semanticweb.org/SWWWS/program/full/paper41.pdf

[8] TERRY R. PAYNE, RAHUL SINGH AND KATIA SYCARA. *Browsing Schedules - An Agent-Based Approach to Navigating the Semantic Web.* ISWC 2002, LNCS 2342, pp. 469-473, 2002.
http://www.daml.ri.cmu.edu/ont/iswc2002/iswc02-Rcal.html

[9] ALEXANDER MAEDCHE and STEFFEN STAAB. *Applying Semantic Web Technologies for Tourism Information Systems.*
http://ww.ontoprise.de/com/download/Applying_SemanticWeb_Technologies_forTourism_Information_Systems.pdf

[10] ERIC MILLER. *An Introduction to the Resource Description Framework.* D-Lib Magazine, May 1998.
http://ww.dlib.org/dlib/may98/miller/05miller.html

[11] MARUT BURANARACH. *The Foundation for Semantic Interoperability on the World Wide Web*. University of Pittsburgh, School of Information Sciences, November 8, 2001.

[12] W3C Web Ontology (WebOnt) Working Group (OWL).
http://www.w3.org/2001/sw/WebOnt/

[13] GANGMIN LI, VICTORIA UREN, et. al. *ClaiMaker: Weaving a Semantic Web of Research Papers.* ISWC 2002, LNCS 2342, pp. 436--442, 2002.
http://kmi.open.ac.uk/projects/scholonto/docs/ClaiMaker-ISWC2002.pdf

[14] 1st International Semantic Web Conference. http://iswc2002.semanticweb.org/

[15] The Semantic Web Community Portal. http://www.semanticweb.org/