

A WEB SEMÂNTICA E SUA APLICABILIDADE NO MEIO ACADÊMICO

Graycielli Athanasio Mendes, Ana Carolina Xaves Ferreira, Rafael Couto dos S. Lima

mendes.grace@gmail.com, anacarolinaxaves@yahoo.com.br, rafaelcoutolima@yahoo.com.br

Associação Educacional Dom Bosco, FCEACDB – Resende, RJ, Brasil

RESUMO

Este artigo tem por finalidade abordar de forma simplificada conceitos básicos relacionados a Web Semântica, exemplificando através de suas principais ferramentas aplicações na estrutura de um domínio acadêmico.

Palavras-Chave: Web Semântica, Ontologias, Aplicações, Meio Acadêmico.

1. INTRODUÇÃO

Com o advento do hipertexto trazido pela Web, onde o objetivo maior era apresentar interfaces mais atraentes, o número de usuários cresceu de forma significativa e com eles a quantidade de informações indexadas.

O crescimento desordenado da Internet tem resultado na deterioração do desempenho da mesma, gerando dificuldades para localização, acesso, apresentação e manutenção da informação.

Atualmente, as ferramentas de busca são incapazes de alcançar resultados plenamente satisfatórios, trazendo juntamente com o que foi solicitado outras informações que não atendem as necessidades do usuário. Isso ocorre em decorrência da maneira como a Web foi implementada, não possuindo nenhuma estratégia satisfatória para a indexação dos documentos. Além disso, a recuperação das informações, baseia-se essencialmente em palavras-chave contidas no documento original, o que gera a dificuldade de determinar os contextos das informações e conseqüentemente a pertinência das mesmas em relação ao que se está buscando.

Diante disso, Tim Berners-Lee, autor do principal artigo sobre a Web Semântica propôs a evolução da Web atual, através da estruturação da informação de forma padronizada, de modo que tanto humanos quanto agentes automatizados pudessem manipular estas informações corretamente. Esta proposta abrange a atualização das tecnologias e linguagens utilizadas para o desenvolvimento de páginas Web focalizando além dos aspectos de exibição e apresentação dos dados, a descrição contextual da informação.

Diante da grande quantidade e diversidade de informações, a Internet tornou-se uma das ferramentas de pesquisa e trabalho mais comuns no meio acadêmico, sendo adotada em bibliotecas virtuais, cursos à distância, portais educacionais, pesquisas, na organização e até mesmo na estrutura de uma instituição de ensino superior. Por isso, veremos neste artigo exemplos de como o

projeto da Web Semântica pode se aplicar à organização de dados no domínio de aplicação acadêmico.

2. A WEB SEMÂNTICA

A Web Semântica é um novo conceito de web, na verdade, não se trata da substituição da atual, mas sim uma extensão e refinamento desta. Visa tornar a rede mais competente, fazendo com que o conteúdo disponível na Internet possa atender as solicitações dos usuários com mais eficiência, através da análise semântica dos documentos, consulta de ontologias e agentes inteligentes capazes de processar essas informações.

O pleno funcionamento dessa nova Web depende da standardização de tecnologias, linguagens e, principalmente, do uso de metadados descritivos. Dessa forma todos os usuários da web compartilharão de regras idênticas para o armazenamento de dados e descrição das informações armazenadas.

O principal desafio da Web Semântica é criar padrões de linguagem para descrição de dados, onde os mesmos sejam definidos e ligados de tal forma que seus significados sejam legíveis às pessoas e também operem como uma fonte de dados para que os computadores possam processá-los automaticamente.

3. A LINGUAGEM XML

Hoje em dia, a maior parte das páginas Web são construídas com o uso da linguagem HTML (HyperText Markup Language ou Linguagem de Marcação em HiperTexto). O HTML constitui um conjunto definido de tags que tratam somente da forma como os dados do documento serão exibidos. A linguagem HTML é muito fácil de ser aplicada e por essa razão se tornou tão popular entre usuários do mundo todo. Contudo, a ela possui muitas limitações, principalmente no que se trata da impossibilidade de atribuir significado ao conteúdo textual de documentos.

O XML (eXtensible Markup Language ou Linguagem Extensível de Formatação) é uma linguagem que surgiu para suprir as carências do HTML. Com o XML é possível definir ou nomear Tags livremente e com isso, tratá-las como objetos, atribuindo um significado conhecido pela aplicação na qual ela é definida. Além disso, ela permite a manipulação e a transferência de dados pela Internet ou intranet de forma prática e abrangente possibilitando que qualquer tipo de aplicação independente de sua linguagem, sistema operacional e plataforma possa manuseá-los inteligentemente, ou seja, o XML é a linguagem mais indicada para a Web Semântica no tocante a descrever o conteúdo semântico e os significados contextuais dos documentos.

A Figura 3.1 apresenta parte da bibliografia de uma disciplina marcada em XML. Este exemplo ressalta que a linguagem XML é capaz de definir tags arbitrárias, tais como: <autor>, <local>, <editora>, entre outros. Atuando de acordo com a necessidade da aplicação.

```
<BIBLIOGRAFIA>
<BIB_BAS> Bibliografia básica
<LVRO>
<AUTOR> BERNSTEIN, P. A. </AUTOR>
<TITULO>Concurrency control and recovery in database systems</TITULO>
<LOCAL> Reading</LOCAL>
<EDITORIA> Addison-Wesley</EDITORIA>
<ANO_PUBLIC> 1987</ANO_PUBLIC>
<NRO_PAGINAS> 370 </NRO_PAGINAS>
</LVRO>
<LVRO>
<AUTOR> DATE, C. J. </AUTOR>
<TITULO> Introdução a sistemas de bancos de dados </TITULO>
<NRO_EDICAO> 2 </NRO_EDICAO>
<LOCAL> Rio de Janeiro </LOCAL>
<EDITORIA> Campus </EDITORIA>
<ANO_PUBLIC>1993</ANO_PUBLIC>
<NRO_PAGINAS>513</NRO_PAGINAS>
</LVRO>
. . .
```

Fig. 3.1 – Parte de uma bibliografia marcada em XML

Devido às características funcionais do padrão XML, o mesmo é oficialmente recomendado pela W3C (World Wide Web Consortium) como padrão de troca de dados na web. Contudo ainda existe um empecilho a ser transposto: a falta de padrões na descrição de dados, assunto que trataremos a seguir.

4. A LINGUAGEM RDF

O RDF (Resource Description Framework ou Sistema de Descrição de Recursos) é um padrão de definição e manipulação de metadados para ser embutido na codificação XML.

O que muitas pessoas confundem é que embora o XML permita adicionar estruturas arbitrárias em seus documentos, não trata do significado (semântica) destas estruturas, ou seja, as tags XML permitirão um claro entendimento para o leitor humano, no entanto, para os computadores não haverá diferença, já que essas tags tratam somente da sintaxe. Sendo assim, lançamos mão do RDF, a fim de adicionar semântica aos documentos.

O RDF vale-se de estruturas ou afirmações denominadas triplas, que são compostas de sujeito, predicado e objeto. Uma tripla RDF também pode ser chamada de declaração ou afirmação.

Para compreender melhor a estrutura RDF, é necessário que os seguintes conceitos sejam esclarecidos:

Resource (Recurso): constitui tudo o que é descrito através de expressões RDF. Um recurso pode ser uma página *Web*, parte de uma página, entre outros. Os recursos são sempre nomeados através de URIs (*Uniform Resource Identifier*) que permitem a introdução de identificadores para qualquer entidade.

Property (Propriedade): É um atributo ou característica que descreve o recurso. Uma propriedade representa também o relacionamento entre recursos.

Statement (Expressão/Afirmação): Corresponde à associação de um recurso específico, uma propriedade e um valor desta propriedade para este recurso. Uma declaração é dividida em sujeito, predicado e objeto. O objeto desta declaração pode ser um recurso ou um literal.

Para compreender melhor como funciona a estrutura de uma declaração RDF aplicada a organização de uma estrutura Acadêmica, tomemos como exemplo a seguinte afirmação: “Ana Carolina é estudante de Sistemas de Informações”, onde o sujeito é “Ana Carolina”, “é estudante de” é o predicado, e “Sistemas de Informações” é o objeto da declaração. O endereço correspondente a Ana Carolina no servidor da faculdade é <http://projetos.aedb.br/si2004/si20457073>. Vejamos o exemplo na tabela 5.1.

Tabela 4.1 - Modelo de declaração RDF

Recurso (Sujeito)	Ana Carolina http://projetos.aedb.br/si2004/si20457073
Propriedade (Predicado)	Estudante
Valor (Objeto)	Sistemas de informação

Veja, agora, a figura 5.1 que representa o grafo da seguinte declaração: Graycielli é estudante de Sistemas de Informações, cursando a disciplina de Tópicos Avançados em informática I, cujo professor é Eduardo Barrére.

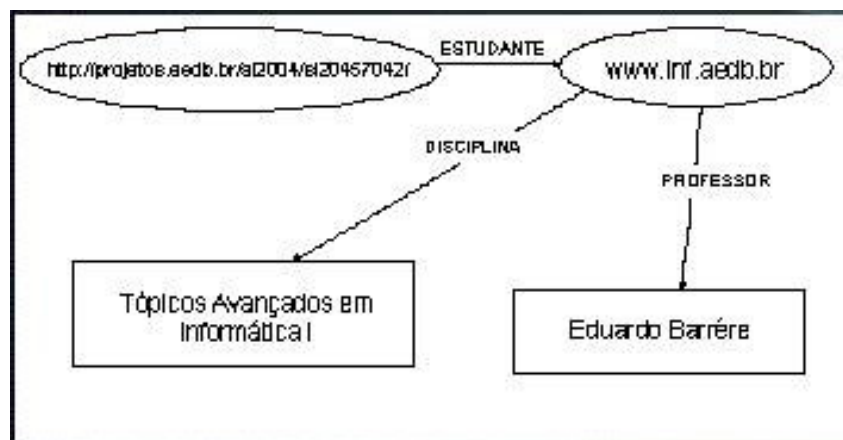


Figura 5.1 – Grafo da declaração

A figura 5.1 descreve o recurso “curso de Sistemas de Informação”, representado pelo site do mesmo; disciplina e professor são propriedades de “sistemas de informação”. Assim, concluímos que um recurso é um objeto a ser descrito, não se tratando necessariamente do sujeito de uma declaração.

```
<?xml version="1.0"?>
<rdf:RDF
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns="http://www.inf.aedb.br/">
<rdf:Description about="projetos.aedb.be/si2004/si20457042/">
<telefone>99-9999-9999</telefone>
<email>mendes.grace@gmail.com</email>
<estudante>
<rdf:Description ID="www.inf.aedb.br">
<disciplina>Tópicos Avançados I</disciplina>
<professor>Eduardo Barrére</professor>
</rdf:Description>
</estudante>
</rdf:Description>
</rdf:RDF>
```

Fig. 3.3 – Sintaxe RDF

Os grafos e declarações são apenas meios de representar as idéias que pretendemos relacionar através do RDF, afinal, para que as máquinas processem tais informações é necessário que estas idéias estejam codificadas, como está exemplificado na figura 3.3.

Descrevemos os recursos através do elemento *Description* juntamente com uma lista de propriedades que se aplicam a esse recurso. As propriedades do primeiro recurso que estamos descrevendo são telefone e e-mail. Aqui, o elemento *estudante* indica a relação entre os recursos. O segundo elemento *Description* descreve o recurso Sistemas de Informação e suas propriedades disciplina e professor.

5. O PADRÃO DUBLIN CORE

RDF é a base para processamento de metadados. Os metadados em documentos *Web* têm por finalidade definir características dos dados que descrevem, a forma com que serão utilizados, exibidos, ou mesmo seu significado em um contexto. Especificamente, no contexto de RDF, os metadados são utilizados para definir recursos da *Web*.

O autor de uma página pode criar a tag <AUTOR> em XML para descrever o fato de que aquela página específica foi criada por ele, entretanto, outro sujeito poderia utilizar uma tag <CRIADOR> pelo mesmo ensejo, o que provocaria uma ambigüidade e impossibilitaria a coerência na recuperação de informações. Por essa razão, a W3C e as comunidades de usuários têm trabalhado em conjunto para determinar os padrões necessários para o compartilhamento de informações na *Web*. Neste contexto, foi criado o padrão Dublin Core (Dublin Metadata Core Element Set), com a finalidade de standardizar as nomenclaturas utilizadas na marcação semântica de documentos.

O padrão Dublin Core é composto de 15 elementos:

TITLE - o nome dado ao recurso, ou título;
CREATOR - a pessoa ou organização responsável pelo conteúdo;
SUBJECT - o assunto, ou tópico coberto pelo documento;
DESCRIPTION - descrição do conteúdo;
PUBLISHER - o responsável por tornar o recurso ou documento disponível;
CONTRIBUTORS - aqueles que contribuíram para o conteúdo;
DATE - data em que o recurso foi tornado disponível;
TYPE - uma categoria preestabelecida para o conteúdo;
FORMAT - o formato no qual o recurso se apresenta;
IDENTIFIER - identificador numérico para o conteúdo, tal como uma URL;
SOURCE - fonte de onde foi originado o conteúdo;
LANGUAGE - a linguagem em que está escrito;
RELATION - como o conteúdo se relaciona com outros recursos, como, por exemplo, se é um capítulo em um livro;
COVERAGE - onde o recurso está fisicamente localizado;
RIGHTS - um ponteiro ou *link* para uma nota de *copyright*;

Como exemplo, considere a seguinte declaração: “O trabalho Web Semântica foi feito por Graycielli”. Explicitamente está descrito que o mesmo (trabalho) se encontra no endereço correspondente a matrícula de Graycielli no servidor.

```
<?xml version="1.0"?>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:DC="http://purl.org/dc/elements/1.0/">
  <rdf:Description about="projetos.aedb.be/si2004/si20457042/web semantica.doc">
    <DC:Creator>Graycielli</DC:Creator>
    <DC:Title>Web Semântica</DC:Title>
    <DC:Date>31-05-2006</DC:Date>
  </rdf:Description>
</rdf:RDF>
```

Figura 5.1 – Código RDF com Tags DC.

Neste exemplo foram utilizadas as propriedades creator, title e date do padrão Dublin Core, repare que aparecem sempre precedidas pelo prefixo DC.

4. ONTOLOGIAS

Após a marcação e estruturação semântica do conteúdo das páginas, é necessário criar meios de viabilizar o processamento dos dados. Esse processamento

deverá promover o intercâmbio de informações entre humanos e agentes informatizados. Para isso, as ontologias possibilitarão a compreensão de termos de forma consensual entre si. Além disso, servirão de auxílio ao computador para deduzir significados de informações inseridas nos documentos.

A princípio o termo ontologia era mais popular no campo filosófico, mais especificamente na área que trata do Existencialismo e questões metafísicas, na verdade, na área de Tecnologia da Informação este termo começou a ser utilizado a partir da década de 90 no ramo da Inteligência Artificial.

Podemos empregar ontologias para descrever um domínio de aplicação específico. De maneira geral, uma ontologia comum define o vocabulário com o qual as pesquisas e afirmações são trocadas entre os agentes, registrando os significados dos termos e as relações entre eles, organizando-os em uma classificação e contendo as visões primitivas de modelagem da informação.

Para exemplificar o uso de ontologias de acordo com o que se propõe este artigo faremos uso da DAML+OIL, desenvolvida pelo USAmerican/European Committee, um projeto da DARPA (Defense Advanced Research Project Agency).

A sintaxe utilizada para a implementação é a RDF.

As ontologias são compostas pelas descrições de objetos e é necessário que se descreva o tipo de objeto que se está relacionando. Para demonstrar a utilidade de ontologias na organização do meio acadêmico, representaremos o domínio inerente ao curso de *Sistemas de Informação*.

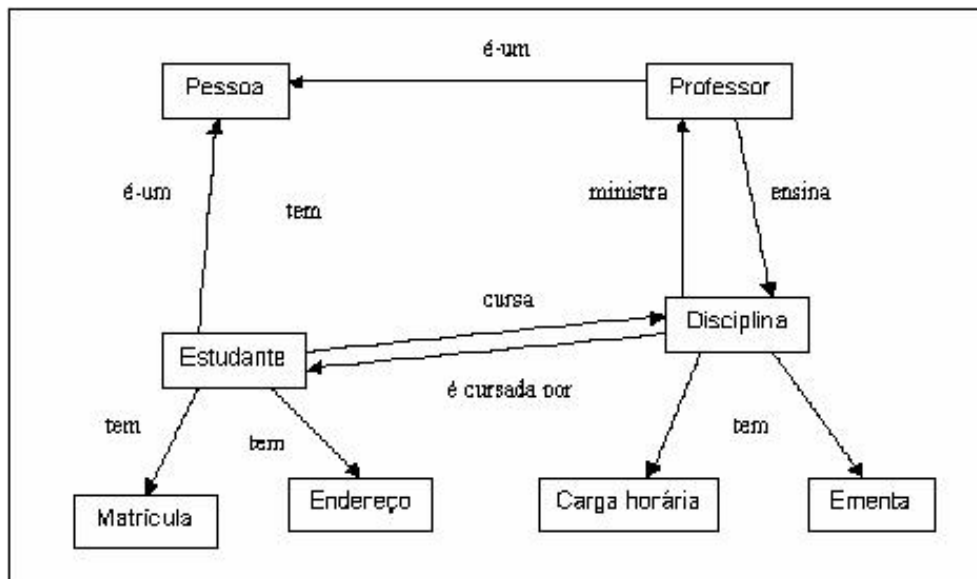


Fig. 4.1 – Ontologia de um Meio Acadêmico

Mais especificamente pretendemos representar as pessoas envolvidas nesse curso, portanto, podemos criar uma classe geral denominada *Pessoas*, através da tag Class:

```
<daml:Class rdf:ID="Pessoas">
```

Declaramos a existência de uma classe denominada *Pessoas*, porém, até este momento nenhuma informação foi fornecida a respeito da mesma. Para identificar e fazer comentários sobre esta classe serão utilizadas as marcas label e comment.

```
<rdfs:label>Pessoas</rdfs:label>
  <rdfs:comment>
    Esta classe descreve as pessoas envolvidas no Curso de Sistemas de Informação.
  </rdfs:comment>
```

Para encerrar a descrição usa-se:

```
</daml:Class>
```

Para que cada agente esteja representado coerentemente descreveremos os tipos de *Pessoas* que interagem no domínio de aplicação, tais como *estudante*, *professor* e *coordenador*, dessa forma, usa-se:

```
<daml:Class rdf:ID="Professor">
  <rdfs:subClassOf rdf:resource= "#Pessoas"/>
</daml:Class>

<daml:Class rdf:ID="Estudante">
  <rdfs:subClassOf rdf:resource= "#Pessoas"/>
</daml:Class>

<daml:Class rdf:ID="Coordenador">
  <rdfs:subClassOf rdf:resource= "#Pessoas"/>
</daml:Class>
```

A marca *subClassOf* afirma que o sujeito *Professor* e o sujeito *Estudante* são subclasses de seu objeto, o recurso identificado por *#Pessoas*. Para afirmar que *Pessoas* não podem ser *Estudante* e *Professor* ou *Estudante* e *Coordenador* simultaneamente, usam-se as marcas *about* e *disjointWith*:

```
<daml:Class rdf:about= "#Professor">
  <daml:disjointWith rdf:resource= "#Estudante"/>
</daml:Class>
<daml:Class rdf:about= "#Coordenador">
  <daml:disjointWith rdf:resource= "#Estudante"/>
</daml:Class>
```

Uma propriedade ou uma relação binária conecta dois itens. Em DAML+OIL as propriedades relacionam objetos com outros objetos, ou objetos com valores de

tipo de dados. No exemplo seguinte, usa-se a marca *ObjectProperty* para definir a propriedade *temMatricula* para *Estudante*.

```
<daml:ObjectProperty rdf:ID= "temMatricula">
  <rdfs:domain rdf:resource= "#Estudante"/>
  <rdfs:range rdf:resource= "Estudante"/>
</daml:ObjectProperty>
```

Neste caso a propriedade *temMatricula* é aplicada somente a *Estudante*.

5. LÓGICA

É através da camada lógica que são possíveis os relacionamentos de informação e as inferências (dedução) de conhecimento da Web Semântica.

As regras de inferência possibilitam aos agentes inteligentes (programas) o poder de raciocínio sobre os termos e seus significados e de raciocinar a respeito dos relacionamentos entre os conceitos (ontologia).

Divide-se em três níveis: Lógica (propriamente dita), Proof (prova) e Trust (Confiança).

Em sua parte Lógica (lógica de descrição), o agente poderá deduzir informações. Por exemplo: O Histórico Escolar de um aluno mostra que este já cursou o Ensino Superior. O sistema (baseado em Web Semântica) afirma que, ao ter cursado Ensino Superior, está apto a fazer pós-graduação MBA e, caso tenha os requisitos necessários (tais como artigos publicados no campo almejado), seria classificado também como apto a fazer mestrado.

Se puséssemos esta regra em forma de algoritmo, seria:

```
INÍCIO
"SE CURSOU(ALUNO, ENSINO SUPERIOR) = VERDADEIRO ENTÃO
  MBA = VERDADEIRO;
  SE REQUISITOS(ALUNO, MESTRADO) = VERDADEIRO ENTÃO
    MESTRADO = VERDADEIRO;
SENÃO
  "O ALUNO AINDA PRECISA CURSAR O ENSINO SUPERIOR PARA
  FAZER UMA MBA E/OU PÓS-GRADUAÇÃO";
FIM.
```

No nível Proof, é possível verificar a consistência dos dados acessíveis na Web Semântica, pois, ao utilizar-se a lógica definida no nível inferior, pode-se verificar a existência de dados contraditórios e apontá-los. Como exemplo podemos citar uma universidade que utiliza um sistema padronizado através da Web Semântica com o qual verifica se o aluno está apto a fazer sua inscrição para o vestibular. Para isto, o aluno envia seu Histórico Escolar para o sistema analisar. Este, por sua vez, utiliza a lógica que

“Se o aluno cursou o Ensino Médio, está apto a fazer sua inscrição para o vestibular, senão não poderá”, provando assim que o aluno pode (ou não) inscrever-se.

Por fim, o nível Trust baseia-se na assinatura digital, visando garantir a procedência de um documento ou definição, o que é vital para decidir se uma informação é confiável ou não. As assinaturas digitais fornecem provas de que uma certa pessoa escreveu um documento ou afirmação. Pode-se assinar digitalmente qualquer conteúdo virtual (e.g. trabalhos acadêmicos, dissertações e teses), de forma que, durante uma pesquisa poderá ser comprovada a procedência e integridade daquele conteúdo; desta forma, o usuário pode decidir se a informação é confiável ou não. Pode-se, ainda, criar um agente que trabalhe somente com dados confiáveis na Web Semântica para que, além de verificar a consistência dos dados, possa ser programado para confiar ou não em uma fonte, através da assinatura digital.

6. WEB SEMÂNTICA: VANTAGENS PARA A EDUCAÇÃO BASEADA EM WEB

No contexto da Educação Baseada em Web (EBW), ainda não existem meios automatizados para compartilhamento e reutilização de objetos de aprendizagem entre aplicações. Mesmo porque, grande parte dos sistemas, atualmente, utilizam linguagens e vocabulários distintos para representação de tais informações (assim como as demais aplicações de outras áreas), além disso, se preocupam mais com a área de gestão da educação, deixando as tarefas rotineiras dos professores, como planejamento de aulas e busca de materiais para ilustração das mesmas em segundo plano.

Atualmente, existem dificuldades em relação à busca de materiais coerentes. Quando é encontrado algum material útil, ainda há dificuldade em reutilizá-lo, devido às adaptações que precisam ser feitas.

A Web Semântica é cogitada entre meios de pesquisa e universidades para dar novos rumos a EBW, através das suas tecnologias. Uma solução possível é desenvolver aplicações educacionais nas quais os objetos de aprendizagem sejam baseados em ontologias. Esse processo de conversão de tecnologias resultaria no ganho de eficiência na busca, recuperação e adaptação de materiais didáticos.

7. CONCLUSÃO

A Web Semântica é um projeto inovador desenvolvido para acessar de maneira “inteligente” dados heterogêneos, propondo uma evolução baseada na eficiência de resultados. Entretanto, este não é um projeto simples: organizar e estruturar todos os dados analisando-os semanticamente não é uma tarefa fácil, visto que os códigos têm que ser compreendidos tanto por seres humanos quanto por máquinas. Todavia, essa tecnologia pode ser aplicada em setores específicos (o que a torna um pouco menos densa), tais como a medicina, na tomada de decisões médicas, na geografia e na educação, através de bibliotecas virtuais, ensino a distância, desenvolvimento de pesquisas, entre outros. Vale a pena ressaltar que o problema do contexto de semântica na web é a falta de exemplos práticos que demonstrem como implementá-la. Considerando o universo

acadêmico e sua influência no avanço da tecnologia e do conhecimento, é importante buscar desenvolver exemplos reais de acordo com as necessidades vigentes.

8. REFERÊNCIAS BIBLIOGRÁFICAS

BERNERS-LEE, Tim. HENDLER, James. LASSILA, Ora. **The Semantic Web**. Disponível em: <<http://www.sciam.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21>> Acesso em: 24 fev. 2006.

BERNERS-LEE, Tim. **Semantic Web Road Map**. Disponível em: <<http://www.w3c.org>> Acesso em: 12 mar. 2006.

CIUFFO, Leandro Neumann. **Linguagens e Ferramentas para a Web Semântica**. Disponível: <<http://www.areaweb.com.br/websemantica/index.xml>> Acesso em: 15 mar. 2006.

CHAHOU, Juliana Jabra. **RELATÓRIO DE ESTUDO: Web Semântica**. Disponível: <<http://www.ime.usp.br/~yoshi/2003i/mac5701/Relatorios/Juliana.relatorio.pdf>>. Acessado em: 09. mar.2006

DZIEKANIAK, Gisele Vasconcelos; KIRINUS, Josiane Boeira. **WEB SEMÂNTICA**. Disponível: <http://www.encontrosbibli.ufsc.br/Edicao_18/2_Web_Semantica.pdf>. Acessado em: 08 mai. 2006.

PARREIRA, Fernando Silva. **Web Semântica: a solução para o caos da Internet**. Disponível: <<http://www.fernando.parreiras.nom.br/content/view/36/42/>>. Acessado em: 08 mai. 2006.

MORAIS, Erikson Freitas de. SOARES, Marcelo Borghetti. **Web Semântica para Máquinas de Busca**. Disponível em: <<http://www2.dcc.ufmg.br/~nivio/cursos/pa03/seminarios/seminario7/seminario7.pdf>> Acesso em: 12 mar. 2006.

SILVA, Geysa Cristina da; LIMA, Tarcísio de Souza. **RDF e RDFS na Infra-Estrutura de Suporte à Web Semântica**. Disponível em: <<http://www.revolucao.etc.br/archives/introducao-a-semantica-web/>> Acesso em: 12 mai. 2006.

SOUZA, Renato Rocha. **A Web Semântica e suas contribuições para a ciência da informação**. 2003.