ESTUDO DE UM ROBÔ DE BUSCA BASEADO EM WEB SEMÂNTICA

Study of a search robot based on semantic Web

AREIAS NETO, Alexandre
Faculdade de Jaguariúna

PETROLI NETO, Silvio
Faculdade de Jaguariúna

Resumo: Esse artigo apresenta a estrutura que a Web Semântica vem propondo como um novo padrão para o desenvolvimento de páginas de Internet. Apresenta também, os conceitos básicos dessa nova estrutura, como o RDF (*Resource Description Framework*), que é recomendado para realizar a estruturação das informações; as Ontologias, que são utilizadas para demonstrar os relacionamentos entre os recursos; os benefícios que os sistemas de busca ganharão com a nova estrutura proposta e para finalizar apresenta uma aplicação que demonstra o funcionamento de um motor de busca semântica desenvolvido utilizando o Asp .Net e a linguagem C#, da Microsoft.

Palavras-chave: Web; Semântica; Busca.

Abstract: This article presents the structure of the Semantic Web proposed as a new standard for developing web pages. It also presents the basic concepts of this new structure, such a RDF (Resource Description Framework), which is recommended for making the allocation of information, Ontologies, which are used to demonstrate the relationships between resources, the benefits to use search engines and presents an application that demonstrates the operation of a semantic search engine that was developed using Microsoft ASP. Net and C# language.

Keywords: Web; Semantic; Search.

INTRODUÇÃO

Devido ao grande volume de informação que circula entre os computadores do mundo todo via Internet e da dificuldade de localizar uma informação na grande rede de computadores, pelo grande quantidade de resultados retornados pelos motores de busca, nos dias atuais surge a necessidade de desenvolvimento de novas técnicas e conceitos sobre esse assunto, visando tornar mais precisos os resultados apresentados pelos robôs de busca. Esse artigo vem apresentar o conceito da Web Semântica que está propondo uma nova estrutura para a criação de páginas da Internet com o intuito de realizar a estruturação das informações. Dessa forma, dividiu-se o artigo em duas partes:

Na primeira parte, são apresentadas as tecnologias que estão sendo recomendadas para o desenvolvimento dessa nova estrutura da Internet e os benefícios que a Web Semântica irá proporcionar aos sistemas de busca de informação. A segunda parte desse artigo apresenta uma aplicação, que baseada nas tecnologias que foram apresentadas na primeira parte do artigo, demonstra de forma prática, o funcionamento de um motor de busca diante da estruturação que está sendo proposta, tornando possível enxergar, de forma clara, a grande evolução que os sistemas de busca atuais irão sofrer, trazendo para o dia-a-dia do usuário mais agilidade e um melhor desempenho.

A Web Semântica

Com a evolução da Internet e das tecnologias de programação que atualmente permitem que os usuários que não tenham o conhecimento no setor de desenvolvimento consigam disponibilizar informações *online* com uma grande facilidade e rapidez, milhões de páginas estão disponíveis na grande rede de computadores contendo informações das mais diversas áreas [1]. Junto com essa evolução surgiram os sistemas de busca que utilizam-se de palavras chaves para localizar as informações desejadas pelos usuários da Internet com o intuito de ajudar e facilitar o acesso aos dados que sejam relevantes [2]. Porém, toda essa tecnologia está disposta de maneira que os computadores não conseguem interpretar o sentido dessas informações e, por

 ISSN 1679-8902

esse motivo, as buscas na Internet precisam ser bastante específicas para que não retornem resultados muito distantes do que, inicialmente, era a intenção do usuário. Diante desses fatos, surge o conceito da Web Semântica conforme é descrito por (Berners-Lee *apud* [1] pp. 133).

> A Web Semântica não é uma Web separada, mas uma extensão da atual. Nela, a informação é dada com um significado bem definido, permitindo melhor interação entre os computadores e as pessoas.

O principal objetivo da Web Semântica é estruturar as informações que estão disponíveis na Internet e tornar possível que os computadores consigam realizar a interpretação dos dados de maneira eficaz e não ambígua. Para isso ser possível, é necessário que as páginas possuam informações sobre seus recursos utilizando do conceito de metadados, ou seja, dados sobre os dados, para que seja feita a descrição das características e a forma de relacionamento desses recursos com os demais recursos existentes [1]. Portanto, a Web Semântica propõe uma nova estrutura que foi dividida em três camadas organizadas conforme é apresentado na Figura 1.

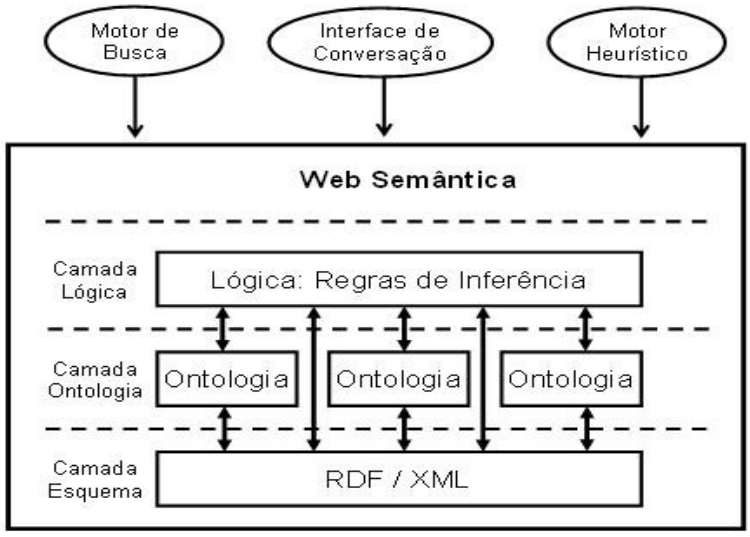

Figura 1: Arquitetura da Web Semântica [2]

Na camada esquema é feita a estruturação e a disposição dos dados de uma maneira que os sistemas web possam interpretá-los. Para realizar esse processo a W3C (Word Wide Web Consortium), recomenda o uso do padrão

RDF (Resource Description Framework). Já a segunda camada, que ganhou o nome de Ontologia, é responsável por definir o relacionamento entre os recursos e pode utilizar o padrão OIL (Ontology Inference Layer) para ser representada. A terceira e última camada, que recebeu o nome de camada lógica, é responsável por realizar a inferência dos dados. [2]

O RDF

Utilizado na camada de esquema, o RDF, é um padrão para descrição de metadados recomendado pela W3C, que utiliza do padrão XML (eXtensible Markup Language) para realizar a estruturação dos dados das páginas web. O padrão RDF permite representar o significado das informações no formato de triplas onde existe: um recurso, um atributo e um valor, conforme é ilustrado na Figura 2.

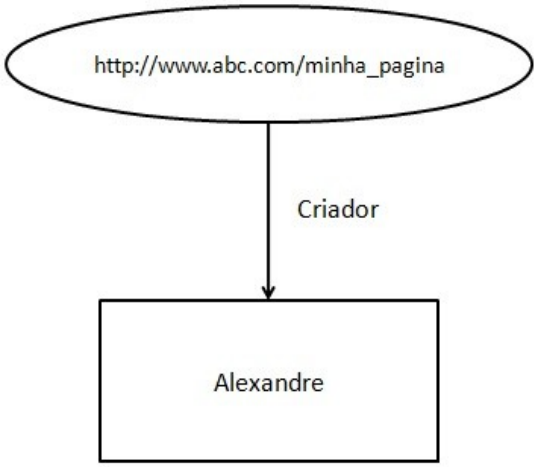

Figura 2: Grafo representando a tripla do RDF [2]

A demonstração feita na Figura 2 pode ser lida como sendo o recurso “http://www.abc.com/minha_pagina” que possui um atributo “criador” e que tem o valor Alexandre, ou seja, podemos afirmar dessa forma, que a página “http://www.abc.com/minha_pagina” foi criada pelo Alexandre [2]. Esse grafo pode ser escrito no padrão RDF conforme é mostrado no Bloco de Código 1.

ISSN 1679-8902

```
<?xml version=”1.0”?>
<rdf:RDF
  xmlns:rdf=”http://www.w3.org/1999/02/22-rdf-syntax-ns#”
  xmlns:x=”http://description.org/schema/”>

       <rdf:Description=”http://WWW.abc.com/minha_pagina”>
             <s:Creator> Alexandre </s:Creator>
       </rdf:Description>
</rdf:RDF>
```

Bloco de Código **1: Descrição da Figura 2 em RDF**

Ontologias

As ontologias são utilizadas na segunda camada da estrutura da Web Semântica e seu principal objetivo é definir como os dados se relacionam. Pode ser comparada com o modelo de entidades e relacionamentos utilizado com grande frequência no setor de planejamento de software. Segundo (SemanticWeb.org *apud* [1]), segue a definição de Ontologia:

> Uma ontologia é uma especificação de uma conceituação. É designada com o propósito de habilitar o compartilhamento e reuso de conhecimentos, de forma a criar ‘compromissos ontológicos’, ou definições necessárias à criação de um vocabulário comum.

Para demonstrar de forma conceitual a representação de uma ontologia, vamos imaginar um cenário de um *site* de comércio eletrônico onde são apresentados diversos tipos de produtos organizados por categorias. A Figura 3 exemplifica de como ficaria a estrutura dessa loja em forma de ontologia.

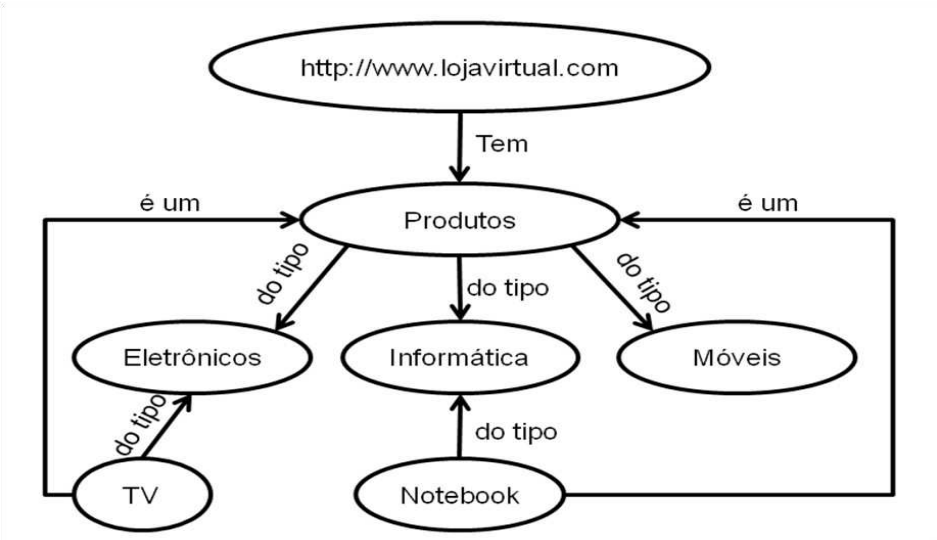

Figura 3: Exemplo de Ontologia

Diversas linguagens estão sendo estudadas para que possam representar o conceito de uma Ontologia e ser aplicada na segunda camada da Web Semântica, contudo, a W3C reuniu um grupo de estudos com o objetivo de desenvolver uma linguagem compatível com as necessidades da Web Semântica e o resultado desse estudo foi à criação da linguagem OWL (Ontology Web Language), cujo principal objetivo é realizar a representação de conceitos e relacionamentos. [4]

Aplicações

Com esse novo padrão de estruturação da Internet, existirá a possibilidade dos motores de busca se tornarem muito poderosos, eficazes e precisos. Eles serão capazes, por exemplo, de realizem a comparação de preços entre sites de comércio eletrônico, realizando o cálculo de frete e impostos do produto escolhido de forma automática e retornar a melhor opção de acordo com as necessidades especificadas pelo usuário, que poderá, nesse instante, realizar a compra do produto. Serão capazes também de realizar o agendamento de consultas médicas de maneira automática, fazendo a

ISSN 1679-8902

comparação entre a agenda do paciente com a agenda dos médicos, conseguindo, através dessa comparação, identificar a melhor opção para o usuário, ou seja, com a Web Semântica a informação irá circular de forma inteligente, o que vai gerar um ganho de desempenho nas rotinas do dia-a-dia dos usuários da Internet.

ESTUDO DE CASO

Com base nas informações obtidas durante o desenvolvimento desse trabalho, foi desenvolvida uma pequena aplicação, que simula o ambiente de uma loja virtual.

O objetivo é demonstrar o funcionamento de um robô de busca simples que utiliza somente a primeira camada da Web Semântica, sendo possível visualizar como a informação é estruturada e a forma com que o robô de busca encontra essa informação. A aplicação foi desenvolvida utilizando a linguagem C#, através do ASP.Net da Microsoft, por ser uma linguagem bastante utilizada para o desenvolvimento Web, com padrões bem definidos de controles de acesso e segurança.

Paralelo ao código fonte dessa aplicação foi construído um arquivo RDF, que é parcialmente apresentado no Bloco de Código 2, onde, todos os produtos e suas características ficam armazenados de forma estruturada. Tal arquivo será utilizado para carregar as informações dos produtos na página da aplicação e será utilizado pelo robô de busca da aplicação. É importante salientar que o arquivo RDF e o arquivo HTML (HyperText Markup Language) compõem uma mesma página. Enquanto o HTML descreve os dados, o RDF fornece o significado semântico aos mesmos.

```
<?xml version="1.0"?>
<rdf:RDF
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:produto="http://localhost/produtos#">
<rdf:Description rdf:about="dvd">
  <produto:id>1</produto:id>
```

```
  <produto:produto>dvd</produto:produto>
  <produto:titulo>Velozes e Furiosos 4</produto:titulo>
  <produto:audio1>Português</produto:audio1>
  <produto:audio2>Espanhol</produto:audio2>
  <produto:audio3>Inglês</produto:audio3>
  <produto:legenda1>Português</produto:legenda1>
  <produto:legenda2>Espanhol</produto:legenda2>
  <produto:legenda3>Inglês</produto:legenda3>
  <produto:fornecedor>Universal</produto:fornecedor>
  <produto:preco>29.90</produto:preco>
  <produto:img>img/dvd/vf4.gif</produto:img>
    ... </rdf:Description>
</rdf:RDF>
```

Bloco de Código 2: Parte do arquivo RDF desenvolvido na aplicação

Na Figura 4 é exibida a página principal da aplicação, onde são apresentados os produtos com seus respectivos títulos e preços, divididos por categorias e também o sistema de busca da aplicação.

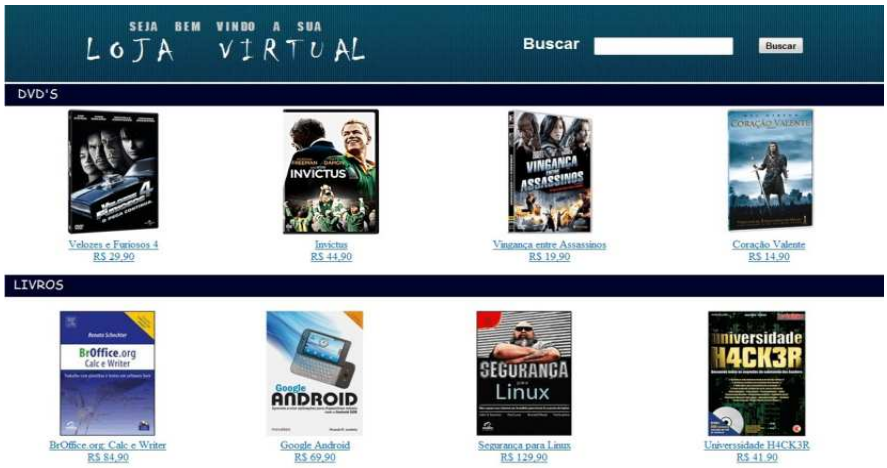

Figura 4: Página principal da aplicação

O sistema de busca da aplicação é onde podemos visualizar, de forma prática, a vantagem de se ter as informações estruturadas de forma que os computadores consigam realizar a interpretação dos dados e os conceitos que foram apresentados nesse artigo. Quando o sistema de busca é acionado, o arquivo RDF é percorrido para verificar se a informação procurada está relacionada com alguma propriedade de algum produto que esteja disponível. Caso essa condição seja verdadeira a aplicação retorna um link para que o

ISSN 1679-8902

usuário possa ter acesso ao produto encontrado. A forma como esse motor de busca percorre o arquivo RDF pode ser vista no Bloco de Código 3.

```
protected void btn_buscar_Click(object sender, EventArgs e)
{

  bool acho = false;
  menu.InnerHtml = "Resultado da Busca";
  produtos.InnerHtml = "";
  String buscando = texto.Text;
  produtos.InnerHtml = produtos.InnerHtml +
  "Elementos encontrados com a propriedade: " + buscando + " <br />";

  XmlDocument doc = new XmlDocument();
  doc.Load("e:\\produtos.xml");

  XmlElement Elem = doc.DocumentElement;

  for (int i = 0; i < Elem.ChildNodes.Count; i++)
  {

   for (int a = 0; a < Elem.ChildNodes.Item(i).ChildNodes.Count; a++)
   {
                    if(Elem.ChildNodes.Item(i).ChildNodes.Item(a).InnerText.ToString
        ().Contains(buscando))
      {
         acho = true;
                String id =
                Elem.ChildNodes.Item(i).ChildNodes.Item(0).InnerText +"";
                String tipo =  Elem.ChildNodes.Item(i).ChildNodes.Item(1).InnerText +
                "";
                String nome =
                Elem.ChildNodes.Item(i).ChildNodes.Item(2).InnerText+"";

         produtos.InnerHtml = produtos.InnerHtml + "<br />" +
                "<a href=detalhes.aspx?id=" + id + "&tpo=" + tipo +
                "> Categoria: " + tipo+ " - Título: " + nome + "</a>";
      }}}
  if (acho == false)
  {produtos.InnerHtml = "Nenhum produto encontrado"; }
    produtos.InnerHtml = produtos.InnerHtml + "<br /><br />
        <a href=home.aspx>Home Page</a>";
}
```

Bloco de Código 3: Função que realiza a busca no arquivo RDF

Basicamente a função de busca carrega o conteúdo do arquivo RDF para a variável chamada de “doc”**,** armazenando essas informações em forma

de uma estrutura de árvore, sendo que a cada tag que descreve um produto como um todo vira um nó e cada nó possui como seus filhos as propriedades que descrevem aquele produto. Após o carregamento do arquivo ser realizado, através do comando “for” o motor de busca vai percorrer todas as propriedades de cada um dos produtos e quando a informação que está sendo procurada for igual a uma propriedade de um produto, o sistema de busca retorna o nó pai daquela propriedade. É, criado então, um link na página da aplicação, tornando possível que o usuário possa visualizar todas as informações referente aquele produto. A apresentação dos resultados do sistema de busca é apresentada na Figura 5, tomando como informação a ser buscada a palavra “Velozes”.

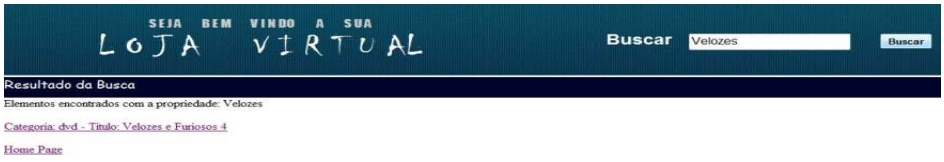

Figura 5: Exemplo de busca

Clicando sobre o link gerado, o usuário irá visualizar a tela que é exibida na Figura 6, com todos os detalhes do produto selecionado, onde ele terá a opção de retornar para a página anterior.



Figura 6: Exibição dos detalhes do produto

ISSN 1679-8902

DISCUSSÕES

A grande dificuldade no desenvolvimento de todo esse trabalho foi encontrar informações precisas sobre a Web Semântica, pois ainda é um projeto que se encontra em fase de estudos e testes.

No desenvolvimento da aplicação, o mais complicado foi realizar a busca no arquivo RDF, pois não foi encontrado nenhum exemplo de como realizar essa busca. Sendo que o RDF é uma extensão do padrão XML, foram testados diversos métodos de busca em arquivo XML até que a busca no arquivo RDF fosse favorável.

É preciso deixar claro que, na Web Semântica, a página que é apresentada pelo navegador está totalmente estruturada dentro do arquivo RDF, ou seja, toda informação - seja um título, uma imagem ou qualquer outro objeto - está descrito no RDF.

CONSIDERAÇÕES FINAIS

A Web Semântica poderá se tornar um marco histórico na evolução da Internet, onde, as informações passarão a circular de forma inteligente e os computadores serão capazes de interpretar as informações das páginas web. Com isso, os motores que realizam a busca da informação se tornarão muito mais poderosos, trazendo para o dia-a-dia dos usuários da Internet uma melhor agilidade e precisão na busca da informação.

Entretanto, esse processo deverá ocorrer de forma lenta, pois seria preciso a criação de novas IDE’s (Interface Development Environment) para o desenvolvimento de páginas semânticas. Também é necessário realizar transformações nas redes sociais, onde um grande volume de informação é disponibilizada por usuários que não possuem nenhum tipo de conhecimento em desenvolvimento. Seria necessário que essas informações fossem estruturadas de forma automática e transparente a esses usuários e, por fim,

havia necessidade ainda de uma grande transformação dos motores de busca para se adaptarem ao novo padrão de Internet.

REFERÊNCIAS

[1] SOUZA, R. R.; ALVARENGA, L. **A Web Semântica e suas contribuições para a ciência da informação**. Ci. Inf., 2004. Disponível em <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-19652004000100016&lng=pt&nrm=iso>. Acessado em 15 mai. 2010.

[2] MORAIS, E. F.; SOARES, M. B. **Web Semântica para Máquinas de Busca.** Universidade Federal de Minas Gerais, 2003. Disponível em <http://homepages.dcc.ufmg.br/~nivio/cursos/pa03/seminarios/seminario7/seminario7.pdf>. Acessado em 28 mai. 2010.

[3] MENDES, G.; XAVES, A. C.; TEIXEIRA, J.; GOMES, F. **Web Semântica**. Associação Educacional Dom Bosco, 2006. Disponível em http://www.aedb.br/seacIV/SI/Poster/websemantica.pdf. Acessado em 06 jun. 2010.

[4] BREITMAN, K. **Web Semântica: A Internet do Futuro.** LTC, 2006.

 ISSN 1679-8902