

Instituto de Matemática e Estatística  
Universidade de São Paulo

## **Ontologias & Web Semântica**

Paulo Augusto Rosa

Plano de Estudos

MAC 5701: Tópicos em Ciência da Computação

Prof. Orientadora: Leliane Nunes Barros

Setembro/2002

# Sumário

Este documento é um plano de estudos para a disciplina de Tópicos em Ciência da Computação (MAC 5701), no qual apresenta as atividades de pesquisa que serão desenvolvidas durante o segundo semestre de 2002. Os estudos estão direcionados para as áreas de Inteligência Artificial e Web Semântica. Entre os temas que serão abordados estão: funcionamento da Web Semântica, entendimento das principais linguagens de descrição de ontologias e pesquisa das possibilidades de serviços para a 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/>), considerada a melhor ferramenta de busca da Internet, há cerca de 2,5 bilhões de páginas Web indexadas na Internet.

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 destas ferramentas, dado o formato atual de dados da Internet. As ferramentas tradicionais de busca atualmente são baseadas na indexação de documentos a partir do texto puro, privilegiando documentos através de *tags* associadas, quantidade de *links* apontados e quantidade de repetições da palavra sendo procurada. O principal problema desta abordagem é o fato que os agentes (softwares) de busca de documentos não conseguem entender o conteúdo das páginas, 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 (agentes de software) capazes de entender o significado de documentos. Esta nova estrutura é hoje chamada de Web Semântica (*Semantic Web*) [1]. A Web Semântica é baseada em extensões da linguagem HTML que permitam colocar informação sobre semântica de páginas Web. Segundo, Tim Berners-Lee, criador da Web e um dos líderes do projeto de Web Semântica, ela pode ser definida como "*Machine-understandable Information*".

Neste contexto, portanto, será direcionado o tema da pesquisa. Pretende-se entender o funcionamento da Web Semântica, obtendo o conhecimento das tecnologias envolvidas

para a sua realização. A motivação principal será a avaliação dos serviços que poderão ser obtidos com a utilização da Web Semântica.

## 2. Web Semântica

A pesquisa sobre Web Semântica será direcionada para o entendimento de seu funcionamento e dos fundamentos necessários para sua concretização. Pretende-se estudar a infra-estrutura necessária para que a informação seja bem definida e entendida por agentes de software, enfatizando os aspectos relacionados à área de Inteligência Artificial.

Dentre os fundamentos necessários para o funcionamento da Web Semântica, serão estudados:

- As linguagens e as gramáticas para definir as informações da Web Semântica. A principal linguagem, que já é base para diversas outras, é a RDF (Resource Description Framework) utilizada pelo consórcio W3C;
- As principais propostas de ontologias, necessárias para interpretação do significado das informações;

Como objetivo principal, será realizada uma pesquisa nos possíveis serviços que podem ser obtidos com a Web Semântica.

## 3. Ontologias

Parte da tarefa de prover informação com significado bem definido é criar uma linguagem de ontologia para descrever e representar objetos na Web. Basicamente, uma ontologia é uma especificação formal de uma conceitualização [6]. 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 conhecidas da área de Inteligência Artificial são: KIF, Loom e PowerLoom, e Frame-Logic, todas baseadas em lógica de primeira ordem.

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, seguindo o padrão XML, para descrição de ontologias para a Web Semântica.

Algumas das principais linguagens para descrever ontologias são:

- SHOE (Simple HTML Ontology Extension)  
(<http://www.cs.umd.edu/projects/plus/SHOE/>);
- OML (Ontology Markup Language)

(<http://www.ontologos.org/OML/OML.html>);

- RDF (Resource Description Framework) (<http://www.w3.org/RDF/>);
- OIL (Ontology Inference Layer) (<http://www.ontoknowledge.org/oil/>);
- DAML+OIL (<http://www.daml.org/2001/03/daml+oil-index>).

Será realizada uma pesquisa sobre estas principais linguagens observando suas características, vantagens e desvantagens. Também serão, também, a existência de ferramentas para a criação de ontologias e para a criação de páginas HTML com ontologias. Será pesquisado como resolver o problema que ocorre quando duas ontologias diferentes descrevem o mesmo conhecimento, ou seja, como pode ser possível integrá-las.

Com o surgimento 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. 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 inteligentes e de uma base de conhecimento. Torna-se viável a organização de conteúdos semânticos na Teia Mundial.

A busca de informações na Web não é o único serviço que pode obter benefícios com o desenvolvimento da Web Semântica. Sendo este um dos pontos que se pretende pesquisar dentro desse plano de estudos: quais são os novos tipos de serviços que se tornam possíveis graças a Web Semântica? Somente para citar alguns, temos: ferramentas avançadas de busca de documentos, agentes inteligentes móveis (*softbots*) para busca de *Web Services*, comércio eletrônico inteligente, entre outros.

## 4. Bibliografia

[1] Ian Horrocks and Sergio Tessaris. Querying the Semantic Web: a Formal Approach. University of Manchester.

[2] Peter F. Patel-Schneider and Dieter Fensel. Layering the Semantic Web: Problems and Directions. Bell Labs Research.

[3] Tim Berners-Lee, James Hendler and Ora Lassila. The Semantic Web. Scientific American: Feature Article, May 2001.

<http://www.scientificamerican.com/2001/0501issue/0501berners-lee.html>

[4] Tim Berners-Lee. Semantic Web Road map. September, 1998.

<http://www.w3.org/DesignIssues/Semantic.html>

[5] Abraham Bernstein, Mark Klein. Towards High-Precision Service Retrieval.

[6] Tom Gruber. What is an Ontology? <http://www.-ksl.stanford.edu/kst/what-is-na-ontology.html>