

Web Semântica: Uma Rede de Conceitos

Semantic Web: A Network of Concepts ONCEPTS

Professor Especialista José Maurício dos Santos Pinheiro

Artigo
Original

Original
Paper

Palavras-chaves:

Internet Semântica
Web 3.0
Rede de Conceitos
Ontologias
Metadados

Resumo

A popularização do uso dos computadores e da acessibilidade à Internet produziu uma mudança nos modos de uso da informação cuja extensão ainda não está completamente delineada. De um catálogo de informações universais e páginas estáticas, a Internet evoluiu para uma proposta de participação humana na criação do seu conteúdo, permitindo a maior interatividade dos usuários. Outra proposta, ainda mais revolucionária, considera a Internet uma Rede de Conceitos, onde as palavras fazem sentido às máquinas e deixam de existir apenas como letras transformadas em algoritmos. É a Internet 3.0, Web 3.0 ou ainda Web Semântica, que permite a recuperação da informação e organização do conhecimento por seres humanos e máquinas, e onde a pesquisa estabelece relações, ou seja, em vez de um grande catálogo de dados aleatórios, a Internet transforma-se em um guia universal e inteligente. O presente artigo apresenta a Web Semântica como um projeto que busca padronizar as formas de descrição do conteúdo e gerar verossimilhança na comparação dos dados para a recuperação das informações.

Abstract

The popularization of the use of computers and accessibility to the Internet produced a change in the ways of use of information whose extent is not yet fully delineated. From a universal catalog of information and static pages, the Internet has evolved into a proposal for human involvement in creating your content, allowing for greater interactivity for users. Another proposal, even more revolutionary, considers the Internet a network of concepts, where the words make sense to machines and no longer just letters turned into algorithms. It is the Internet 3.0, Web 3.0 or Semantic Web, which allows the retrieval of information and organization of knowledge by humans and machines, and where the research builds relationships, that is, instead of a large catalog of random data, the Internet becomes a universal and intelligent guide. This article presents the Semantic Web as a project that seeks ways to standardize the description of the content and generate likelihood of comparison of data for the recovery of information.

Key words:

Internet semantics
Web 3.0
Network Concepts
Ontologies
Metadata

Introdução

A partir da década de 1960, a disponibilidade de meios de armazenamento e tratamento automático da informação fez surgir os primeiros modelos e sistemas

de bases de dados. Os desenvolvimentos conceituais e o aumento da sofisticação dos sistemas oferecidos não pararam desde então, acompanhando o desenvolvimento tecnológico

e a ampliação das áreas de aplicação.

A massificação da disponibilização e do uso da informação na Internet obriga uma análise sobre a qualidade da informação disponível. A realidade é que, desde que surgiu oficialmente em 1991, a Internet vem acumulando informações desorganizadas de maneira exponencial. Recentemente, o advento da Web 2.0, que propõe o aproveitamento da inteligência coletiva no desenvolvimento de aplicativos para a Internet, trouxe à tona a necessidade da padronização dessa informação. De acordo com Berners e Hendler (2001), os computadores necessitam ter acesso a coleções estruturadas de informações e de conjuntos de regras de inferência que ajudem no processo de dedução automática para que seja administrado o raciocínio automatizado, ou seja, a representação do conhecimento.

Atualmente, as bases de dados suportam informações de todos os domínios de atividade em que seja necessária a persistência do conhecimento e uma parte substancial das aplicações incluem-nas de forma implícita ou explícita. O principal aspecto abordado na Web Semântica está em organizar o conhecimento armazenado na Internet por meio da compreensão da linguagem humana, pelas máquinas, na recuperação da informação, sem que o usuário precise dominar refinadas estratégias de linguagem ou conhecimento de máquinas de busca.

Objetivos

Na atual estrutura da Internet, grande parte da informação é disponibilizada de forma desorganizada e sem padronização. Apesar do resultado de uma pesquisa ser perfeitamente satisfatório para os seres humanos, na maioria das vezes, as máquinas, como os mecanismos de busca - por exemplo - não conseguem tirar o máximo proveito da informação disponível. Segundo Berners-Lee et al. (2001), “A Web Semântica é uma extensão da Web atual na qual a informação recebe um significado bem definido, permitindo que os computadores e as pessoas cooperem melhor.”

A Internet Semântica se desenvolve para que a informação possa ser compreendida pelas máquinas, na forma de agentes computacionais que são capazes de operar eficientemente sobre as informações, podendo entender seus significados, ou seja,

incorporar semântica às informações. Isso proporciona não somente que os usuários entendam as informações como também as máquinas possam usufruir dessa vantagem, pois há pretensão de fornecer estruturas e dar significado semântico ao conteúdo das páginas de internet, criando um ambiente onde os agentes de software e os usuários podem trabalhar de forma cooperativa.

Ontologias

O termo ontologia vem sendo discutido e aplicado em diversas áreas do conhecimento científico. Originado na filosofia, significa “categoria que pode ser usada para classificar alguma coisa” (ALMEIDA e BAX, 2003, p.7). Segundo Gruber (1993), ontologia representa a definição explícita de conceitos e suas relações, propriedades e restrições expressas formalmente, ou seja, a ontologia permite especificar uma visão abstrata e simplificada de um assunto que se deseja representar. Em termos de utilização, a ontologia serve de vocabulário para a troca de informações entre aplicações. A adoção de uma ontologia é um compromisso de consistência com os seus conceitos e a garantia de que uma aplicação, dentro dos limites das suas funcionalidades, irá interrogar outras aplicações e responder aos pedidos destas, de acordo com o vocabulário comum.

Para os Sistemas de Informação, a ontologia compreende um conjunto de conceitos e linguagens que podem ser usadas para descrever alguma área do conhecimento ou construir uma representação para esse conhecimento. O uso de linguagens permite escrever afirmações que captam o significado dos dados numa aplicação. Se cada aplicação pode criar os seus próprios conceitos, e se pretende que outras aplicações possam manipulá-los de uma forma sensível ao significado, então, é preciso que o significado desses conceitos seja exposto de uma forma que se possa manipular automaticamente. Esse é o papel das ontologias na Web Semântica.

As regras da Web Semântica são especificadas através das ontologias, as quais permitem representar explicitamente a semântica dos dados, tornando possível elaborar uma estrutura que complementa o processamento das máquinas e melhorar qualitativamente o nível de serviços na Internet.

Segundo Hendler (2001): “a Web Semântica pode ser considerada como a composição de um grande número de pequenos componentes ontológicos que apontam entre si”.

O uso de ontologias que relacionem dados semânticos permite a interação entre diferentes dados. Por exemplo, se for dito que “Fulano vai ao Seminário de Virtualização” e que “O Seminário de Virtualização será dia 10 de maio, na Universidade XYZ”, uma aplicação especializada poderá responder uma questão do tipo “Onde está o Fulano?”, retornando que “dia 10 de maio, Fulano estará na Universidade XYZ”.

Metadados

Metadados referem-se à estrutura descritiva da informação sobre outro dado, o qual é usado para ajudar na identificação, descrição, localização e gerenciamento de recursos da Internet. A finalidade principal é documentar e organizar, de forma estruturada, os dados, com objetivo de minimizar duplicação de esforços e facilitar a manutenção dos dados.

Berners-Lee et al. (2001) afirma que “o desenvolvimento da Web Semântica, se dará a partir do uso intensivo de metadados, visando fornecer acesso automatizado à informação com base no processamento de dados e heurísticas feito por máquinas”. Seguindo essa filosofia, agentes automatizados podem realizar tarefas para os usuários, usando metadados semânticos. Assim, é possível criar vocabulários comuns para metadados (ontologias) e conexões entre vocabulários que permitem que os autores de documentos saibam como estruturá-los para que os agentes automatizados possam usar as informações de acordo com o contexto. Por exemplo, o autor no sentido de “autor da página” não será confundido com o “autor do livro” que está sendo revisado na página.

Moura (2002) afirma:

A utilização das arquiteturas de metadados é importante para as aplicações que trocam dados estruturados na Web, pois fornecem suporte à codificação, transporte e a interoperabilidade semântica, sintática e estrutural a uma variedade de metadados desenvolvidos de forma independente.

Uma das aplicações dos metadados está na formação de dicionários de informações. Nesses dicionários são descritas informações que dizem respeito à origem dos dados, do formato dos dados, do fluxo dos dados, das regras de transformação, entre outros aspectos.

Componentes da Web Semântica

As linguagens usadas na Internet foram desenhadas com o objetivo de serem de acesso fácil para os seres humanos. Entretanto, o uso intenso da Web fez surgir também o conceito de “domínio de serviços”, ou seja, aplicações que se destinam a realizar tarefas em nome do usuário, explorando a informação disponível e fornecendo um resultado em um tempo extremamente baixo.

Uma aplicação tem habitualmente um domínio cujos conceitos são captados tanto na informação que é manipulada como nas funcionalidades oferecidas ao usuário da aplicação, ou seja, a semântica está implicitamente captada na aplicação. Quando há necessidade de comunicação entre aplicações, deve-se estabelecer uma linguagem própria, que mais uma vez tem implícitos os conceitos do domínio. Estas linguagens, normalmente chamadas de “protocolos”, são faladas pelas aplicações quando precisam trocar informações entre si.

O desenvolvimento da Web Semântica apresenta diversas tecnologias desenvolvidas e em desenvolvimento. As mais básicas estão ao nível da uniformização da sintaxe usada para representar dados elementares. Este primeiro nível garante um formato aberto e a possibilidade de criar anotações arbitrárias, ou seja, completa liberdade em termos de criação de conceitos.

Nesse contexto, a Web Semântica compreende linguagens conforme descrito a seguir:

- **XML (eXtensible Markup Language)** - provê uma sintaxe de estruturação de documentos, mas não impõe regras semânticas nesses documentos;
- **XML Schema** - linguagem para restringir a estrutura de documentos XML;
- **RDF (Resource Description Framework)** - modelo de dados para se referir a objetos (resources) e como eles são relacionados. Um

modelo baseado em RDF pode ser representado usando uma sintaxe XML;

- **RDF Schema** - vocabulário para descrever propriedades e classes de recursos RDF, juntamente com semânticas para criar hierarquias para essas propriedades e classes;
- **OWL (Ontology Web Language)** - adiciona vocabulários para descrever classes e propriedades: relacionamentos entre classes, cardinalidade, igualdade, maior variedade de tipos de propriedades, características de propriedades e classes numeradas. Pode incluir descrições de classes e suas respectivas propriedades e seus relacionamentos.

Esses componentes são utilizados para aumentar a usabilidade e funcionalidade da Internet e de seus recursos relacionados. Conforme descrito em Afonso (2001), uma arquitetura para a Web Semântica é composta por três camadas, conforme mostrado na Figura 1:

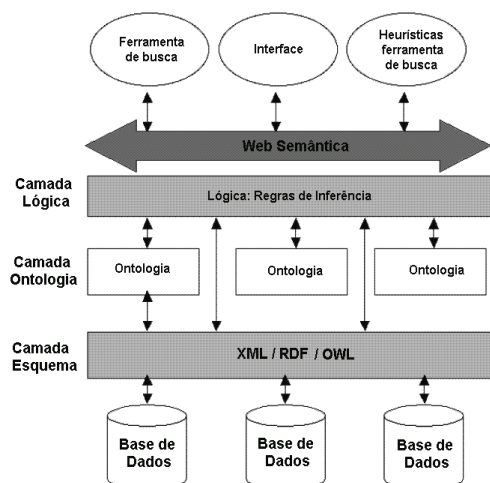


Figura 1 - Estrutura para Web Semântica

- **Camada Esquema** - responsável por estruturar os dados e definir seu significado para que possa elaborar um raciocínio lógico;
- **Camada Ontologia** - responsável por definir relações entre os dados. Nesse nível se dá o entendimento comum e compartilhado de um domínio;
- **Camada Lógica** - responsável por definir mecanismos para fazer inferência sobre os dados. Composta por um conjunto de regras de inferência que os agentes poderão utilizar para relacionar e processar informações.

Desafios

O crescimento da Internet e a popularização do seu uso geraram um grande

volume de informação. Em consequência, as máquinas de busca não conseguem filtrar o que é realmente relevante para uma dada pesquisa. No modelo atual, os documentos são estruturados, considerando principalmente a apresentação e usando um tipo de marcação, muitas vezes, gerado por ferramentas automatizadas que, na maior parte do tempo, não observam a sintaxe ou a semântica da informação. Por exemplo, em uma ferramenta de busca tradicional, os resultados obtidos são baseados nas palavras-chaves digitadas, de acordo com o número de vezes que elas se repetem na rede. Esse resultado não considera o sentido ou a relação entre elas.

O desafio da Internet Semântica é construir uma linguagem que faça sentido para a máquina, em que o maior volume possível de informação seja disponibilizado de modo lógico e coerente para o usuário, ou seja, a pesquisa estabelece relações. Para evitar uma sobrecarga nos mecanismos de busca e para facilitar a classificação e obtenção de informações, a proposta da Web Semântica é o uso de novos vocabulários de marcação como RDF e OWL, por exemplo, e novos usos para vocabulários já existentes.

A modularização proporcionada pelos vocabulários da Web semântica facilita a criação de mecanismos de busca e aplicações específicas para cada tipo de dado, sendo descrito por uma semântica. A ideia é que a ferramenta de busca entenda o sentido, por meio de inteligência artificial e da programação da página, ou seja, o uso da semântica e do contexto para se realizar buscas na Internet, bem como a captura e utilização do perfil de interesse para se filtrar informação.

É importante ressaltar que a semântica não está somente relacionada ao conteúdo de um recurso, mas também à forma de como este se relaciona com os demais recursos na internet. Para que essa estrutura funcione, os computadores devem ter acesso às coleções de informações estruturadas e a um conjunto de regras de inferência que podem ser utilizadas para levar ao raciocínio automatizado. Portanto, é essencial que os recursos disponibilizados sejam expressivos, para que as máquinas ou agentes sejam capazes de processar e entender o real significado do dado, intermediando as necessidades de cada usuário e as fontes de informações disponíveis.

Conclusão

A Web Semântica é uma proposta de evolução da Internet atual, uma rede em que o significado da informação está acessível às pessoas e máquinas. Nela, busca-se incluir mecanismos que capturem o significado das páginas, criando um ambiente onde os computadores possam processar e relacionar conteúdos provenientes de várias fontes. Trata-se de um conceito que visa facilitar a obtenção, classificação e organização de informações, de forma estruturada, de modo que os dados estejam disponíveis tanto para as pessoas quanto para as máquinas.

A Web Semântica busca estruturar e dar semântica aos dados, facilitando a construção de aplicações, visando diminuir ou eliminar as dificuldades relacionadas com o acesso à informação. Entretanto, a tarefa de construir tal estrutura requer o desenvolvimento de novas soluções tecnológicas para normalizar as representações dos significados semânticos e da capacidade de elaborar modelos adequados para a representação da informação. O objetivo é explicitar o significado da informação para as máquinas, por meio do estabelecimento de linguagens para os conceitos usados. Essa abordagem, centrada na tecnologia, valoriza a informação estruturada, elegendo como objetos os itens de informação elementares e os seus relacionamentos, com ênfase no tratamento automático do seu significado semântico.

Bibliografia

AFONSO, M. M. R. Semantic web. Porto: Universidade do Porto, 2001. Disponível em: <<http://www.fe.up.pt/~mgi00014/ari/SW.doc>>. Acesso em dez. 2008.

ALMEIDA, M. B.; BAX, M. P. Uma visão geral sobre ontologias: pesquisa sobre definições, tipos, aplicações, métodos de avaliação e de construção. Ciência da Informação, Brasília, v.32, n.3, p.7-20, 2003.

BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. The Semantic Web. Scientific American, maio 2001. Disponível em: <<http://www.sciam.com/article.cfm?id=the-semantic-web>> Acesso em dez. 2008.

DZIEKANIAK, G. V.; KIRINUS, J. B. Web

Semântica. Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação. Universidade Federal de Santa Catarina, n.18, p. 20-39. Florianópolis, 2004.

GRUBER, T. R. What is an Ontology? Disponível em: <<http://www.ksl.stanford.edu/kst/what-is-an-ontology.html>>. Acesso em dez 2008.

HENDLER, J. Agents and the semantic web. IEEE Intelligents Systems, mar./abr. 2001.

MOURA, A. M. C. A web semântica: fundamentos e tecnologias. Rio de Janeiro: IME, 2002. Disponível em: <<http://www.ipanema.ime.eb.br/~anamoura/publicacoes.html>>. Acesso em: dez. 2008.

W3C - Semantic Web. Disponível em <<http://www.w3.org/2001/sw/>>. Acesso em nov. 2008.

W3C - Web-Ontology (WebOnt) Working Group. Disponível em <<http://www.w3.org/2001/sw/WebOnt/>>. Acesso em nov. 2008.

Endereço para Correspondência:

Professor Especialista José Maurício dos Santos Pinheiro (Curso Tecnológico de Redes de Computadores – UniFOA) – jm.pinheiro@uol.com.br

Centro Universitário de Volta Redonda
Campus Três Poços
Av. Paulo Erlei Alves Abrantes, nº 1325,
Três Poços - Volta Redonda / RJ
CEP: 27240-560

Informações bibliográficas:

Conforme a NBR 6023:2002 da Associação Brasileira de Normas Técnicas (ABNT), este texto científico publicado em periódico eletrônico deve ser citado da seguinte forma: PINHEIRO, José Maurício dos Santos. "Web Semântica: Uma Rede de Conceitos", **Cadernos UniFOA**, Volta Redonda, ano IV, n. 9, abril. 2009. Disponível em: <http://www.unifoa.edu.br/portal_pesq/caderno/edicao/09/23.pdf>