

Aplicações de Ontologias na Recuperação de Informações Jurídicas na Web Semântica

Marcus de Melo Braga^{1,2}, Hélio Santiago Ramos Júnior¹, Tatianna de Faria Coelho¹

¹ Universidade Federal de Santa Catarina, Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento, Campus Universitário, Florianópolis, Santa Catarina, Brasil

² Universidade Federal de Alagoas, Instituto de Computação, Cidade Universitária, Maceió, Alagoas, Brasil

mbraga@fapeal.br, hramos@egc.ufsc.br, tannaco@hotmail.com

Resumo. A busca de informações na Web tem apresentado melhorias significativas com o advento das tecnologias introduzidas pela Web Semântica. Tais tecnologias possibilitam o acesso às informações nas diversas áreas do saber de modo mais preciso através da construção de mecanismos de busca que implementam ontologias visando a um tratamento mais inteligente das informações disponibilizadas na Internet. O presente estudo enfoca o uso de ontologias na recuperação de informações jurídicas na Internet, apresentando exemplos da sua aplicação.

Palavras-chave: Inteligência Artificial, Web Semântica, Ontologias, Engenharia do Conhecimento.

Applications of Ontologies in Juridical Informations Recovering on the Semantic Web

Abstract. The search for information on the Web has shown significant improvements with the introduction of the technologies used by semantic Web. These technologies allow a most accurated access to information in many domains through the development of searching engines that use ontologies for a more intelligent treatment of the available information on the Web. This study focus on the use of ontologies for recovering juridical information in the Internet, presenting some examples of its application.

Keywords: Artificial Intelligence, Semantic Web, Ontologies, Knowledge Engineering.

1 Introdução

Uma das principais inovações introduzidas na Internet nos últimos anos é o projeto de uma Web semântica, uma evolução do modelo atual. A idéia nasceu de uma proposta de Berners-Lee (2001) de dotar a rede de um contexto semântico que possibilite que informações atualmente disponíveis também possam ter significado para as máquinas, permitindo um tratamento automático deste conteúdo.

Trata-se de uma evolução significativa do estágio atual, onde as informações são compreendidas apenas por seres humanos, para um estágio em que as informações contidas na grande rede sejam representadas de forma que permita o seu tratamento por agentes de software, auxiliando de forma significativa a busca e o tratamento das informações.

Uma rede semântica não apenas tornará as buscas mais precisas como também agregará novas funcionalidades, possibilitando um avanço considerável nas suas aplicações em diversas áreas. Tais melhorias são relevantes, considerando o volume de informação que é disponibilizada em meio digital através da rede.

Para criar mecanismos de busca mais eficientes na Internet necessitamos mais do que simplesmente indexar palavras. Nos moldes atuais, tal indexação fornece muitas vezes informações irrelevantes para um pesquisador. Para aumentar a facilidade com que uma página seja encontrada na Internet, já existem atualmente profissionais que se especializam no modo de operação dos mecanismos de busca, visando otimizar a visibilidade de uma página na Internet.

A proposta semântica traz novas ferramentas que possibilitam indexar documentos de forma que suas propriedades possam também, ser reconhecidas pelos computadores. Tal iniciativa traz uma considerável ajuda para a tarefa colossal de indexar automaticamente milhões de páginas.

O projeto da Web semântica introduz novas formas de marcações de páginas da Internet que visam reduzir a sobrecarga dos motores de busca da Internet, auxiliando na classificação e obtenção de informações. Outras vantagens trazidas pela rede semântica são o aumento na acessibilidade e na sua usabilidade, possibilitando que documentos possam ser acessados através de celulares e outros dispositivos digitais com acesso à Internet.

2 Web Semântica: principais conceitos e arquitetura

De acordo com Hoeschl (2004), “a Web Semântica é uma teoria que pretende desenvolver uma linguagem universal para a Internet que permita a localização de endereços digitais pela compreensão dos conteúdos das páginas descritos internamente”.

Segundo Faria & Girardi (2006), “um dos desafios da Web Semântica é criar uma linguagem que seja capaz de expressar ao mesmo tempo o significado dos dados e definir regras para raciocinar sobre eles, de forma a deduzir novos dados e regras e, permitir que regras existentes em sistemas de conhecimento possam ser exportados para a Web”.

Para adicionar tais características ao atual modelo existente na Internet, foi concebida uma arquitetura de implantação que se baseia em camadas, como pode ser visto na Figura 1.

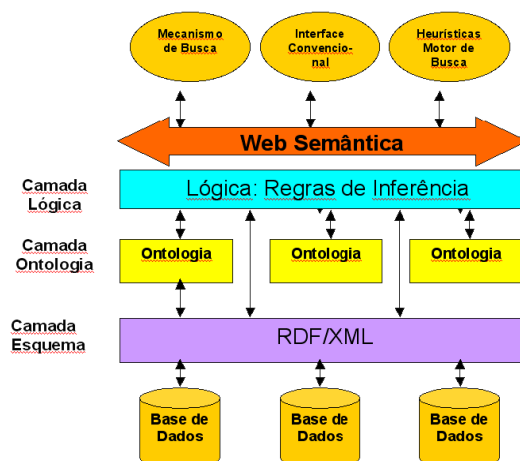


Fig. 1. Arquitetura em camadas da Web semântica (Afonso, 2001). caption.

A implantação da semântica da rede se dá através do uso intensivo de metadados nas três camadas, onde cada uma delas serve de suporte para a camada seguinte.

Na camada esquema, temos a junção das tecnologias do XML (eXtensible Markup Language) e RDF (Resources Definition Framework) que representam o primeiro passo na introdução de uma semântica na rede.

Na camada seguinte temos o uso de ontologias que possibilitam a definição de relacionamentos entre os dados em um domínio de uso compartilhado. Segundo Moura (2002), as ontologias têm um papel essencial neste modelo uma vez que possibilitam que as informações sejam acessadas com base em consultas que consideram a semântica do domínio, ultrapassando as dificuldades tradicionais encontradas numa linguagem tais como a sinonímia, homonímia e metonímia.

Finalmente na camada lógica são construídas as regras de inferência que poderão ser utilizadas por agentes de software que se relacionam visando obter informações de forma mais inteligente para seus usuários finais.

Outra forma de detalhar a arquitetura da Web Semântica pode ser vista na figura 2 abaixo.

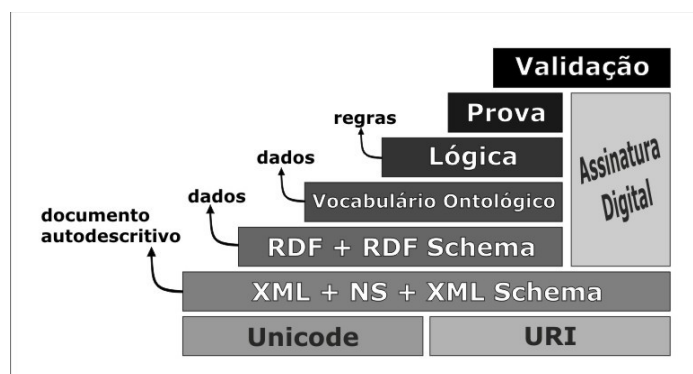


Fig. 2. Arquitetura em camadas da Web Semântica (W3C).

A primeira camada que serve de base para a Web Semântica repousa sobre o conjunto de caracteres Unicode e a URI (Uniform Resources Identifier) que é utilizada para referenciar os recursos na rede. Na segunda camada, a linguagem de marcação XML (eXtensible Markup Language) associada aos recursos de NS (NameSpaces) e XML Schema servem para descrever documentos que podem ser compreendidos por pessoas e software.

A partir deste ponto, todas as camadas seguintes estão estruturadas sobre a camada XML e tiram proveito dela. Assim, a estrutura seguinte (RDF + RDF Schema - Resource Description Framework) é uma camada de metadados que é construída sobre a de XML e que visa a interoperabilidade entre as aplicações, representando os dados sob a forma de triplas através de descrições de recursos, propriedades e valor.

A próxima estrutura corresponde à camada de ontologias (descrita mais detalhadamente no próximo capítulo). Ela define um vocabulário comum (ontológico) para que o conteúdo semântico possa ser compartilhado entre as aplicações. Neste nível, a linguagem padrão OWL (Ontology Web Language) usada para representar as ontologias possibilita agregar maior conteúdo semântico aos documentos utilizando recursos do RDF e do XML das camadas inferiores.

As últimas camadas, alicerçadas sobre a estrutura já existente na parte inferior da arquitetura, implementam a definição de regras e possibilitam a prova e a validação de inferências realizadas por agentes de software que fazem uso dos recursos de toda a estrutura existente para a representação do conhecimento.

3 Ontologias

Ontologias são um instrumento eficiente de auxílio à recuperação de informações na web e são extremamente relevantes para a criação de mecanismos de busca mais eficientes. Neste estudo daremos ênfase à sua aplicação na recuperação de

informações jurídicas, apresentando o seu conceito dentro do enfoque da engenharia e gestão do conhecimento.

Embora a palavra ontologia tenha a sua origem nas ciências filosóficas, este termo passou a ser utilizado por tecnólogos da informação, pesquisadores, engenheiros e gestores do conhecimento dentro de um novo contexto.

Neste sentido, "o termo 'ontologia' no campo da tecnologia faz referência à formulação de esquema conceitual, dentro de um certo domínio, com a finalidade de facilitar a comparação, classificação, organização e armazenamento dos textos analisados" (Hoeschl, 2004).

A recuperação de informações na web, no caso de se utilizar ontologias seria mais eficiente porque a pesquisa realizada pelos mecanismos de busca leva em consideração não apenas o termo ou expressão literal utilizada na pesquisa, mas permite que esta seja otimizada, apresentando informações relacionadas ao assunto pesquisado que não contenham necessariamente em seu teor as mesmas palavras que foram utilizadas na pesquisa.

De acordo com Staab & Maedche (2007), "a ontologia constitui a base para anotar na Web documentos da comunidade de aquisição do conhecimento com o objetivo de possibilitar o acesso inteligente a estes documentos e inferir o conhecimento implícito das regras e fatos declarados explicitamente na ontologia".

4 O Mecanismo de Busca OntoWeb®

Um exemplo de como o uso de ontologias pode contribuir para uma maior eficiência dos mecanismos de busca é o Ontoweb, que consiste em uma ferramenta inteligente de busca que utiliza ontologias para obter melhores resultados: "O Ontoweb é o primeiro buscador do mundo a trabalhar com engenharia de ontologias, e é fruto de um projeto de desenvolvimento científico, utilizando semânticas e estruturas valorativas para contextualizar as buscas e refinar os resultados. O uso das ontologias é um dos principais segredos da nova ferramenta (daí o nome Ontoweb), e é também um dos principais fatores responsáveis pela alta qualidade dos seus resultados" (Hoeschl, 2006).

O OntoWeb baseia-se em ontologias, na técnica de Inteligência Artificial denominada Raciocínio Baseado em Casos (RBC) e na extração de informações textuais através de técnicas tais como pesquisa contextual estruturada e representação do conhecimento contextualizado dinamicamente. Esta última, segundo Hoeschl (2004), trata-se de uma técnica que "prevê a construção de uma base de conhecimento, estruturada sobre um conjunto de expressões, utilizando diferentes referenciais, de forma a tratar com o máximo de personalização estas expressões. Cria-se, então, uma rede de conceitos relacionados por diversos tipos de conexão que representam cada um, um valor para o cálculo da similaridade entre os casos".

Na sua versão atual, o OntoWeb enfoca, especificamente, o cenário de Governo Eletrônico. Nesta especialidade, os autores afirmam ser a primeira ferramenta no âmbito mundial dirigida exclusivamente a este segmento, principalmente se considerarmos o uso de ontologias.

Uma das características interessantes do OntoWeb é a possibilidade de utilização de mais de 7 mil caracteres no campo de pesquisa, não se limitando apenas às palavras chave ou a expressões de busca reduzidas (IJURIS, 2007). Convém lembrar que, até recentemente, o mecanismo de busca padrão do Google aceitava apenas 256 palavras.

Outra característica interessante é a existência de um gráfico de linhas que possibilita visualizar a variação do assunto pesquisado ao longo do tempo, permitindo uma análise mais eficiente do material coletado.

5 Aplicações na Recuperação de Informações Jurídicas

Uma das principais vantagens do uso de ontologias é que estas, uma vez bem elaboradas, podem contribuir para recuperar informações de forma mais eficiente contidas em um banco de dados em virtude do sistema de busca poder realizar a pesquisa por sinônimos e também por relações semânticas entre as palavras.

Hoeschl (2004) aponta para um outro aspecto positivo do uso das ontologias, aplicadas no âmbito do direito, que está vinculado à possibilidade de contribuir para a inclusão digital de toda a sociedade "na medida em que é capaz de relacionar e reconhecer tanto os rebuscados termos jurídicos quanto a linguagem leiga (...). Isso porque os diferentes usuários têm maior e mais eficaz acesso aos documentos jurídicos com a aplicação de ontologias, podendo melhor compreendê-los e aproveitá-los".

No domínio do Direito, há experiências de aplicação de ontologias e sistemas de conhecimento jurídico, principalmente no que concerne às técnicas de Raciocínio Baseado em Casos (RBC), que apresentaram resultados significativos para a sociedade, como, por exemplo, o projeto Portal do Consumidor Catarinense.

O Portal do Consumidor Catarinense criou um portal de informações para o consumidor na internet, disponibilizando um software denominado Aletheia que utiliza técnicas de inteligência artificial em seus recursos permitindo que o cidadão possa, através do programa, tirar suas dúvidas referentes a uma situação qualquer envolvendo relação de consumo, obtendo como resposta o trecho correspondente do Código de Defesa do Consumidor (Garcia, 2003).

Este portal teve a finalidade de funcionar como um canal aberto para a prestação de serviços de apoio ao consumidor em ambiente virtual, promovendo a difusão dos direitos do consumidor brasileiro em âmbito global. Para tanto, foi desenvolvido o software Aletheia, que permite obter uma resposta para uma determinada questão jurídica envolvendo relação de consumo.

Acerca do software Aletheia, explica Garcia (2004) que se trata de "um sistema que utiliza técnicas de Inteligência Artificial (IA) - como o Raciocínio Baseado em Casos (RBC) e a Pesquisa Contextual Estruturada (PCE) - para a recuperação de informações contidas no Código de Proteção e Defesa do Consumidor (CDC)".

A Pesquisa Contextual Estruturada é uma metodologia que resulta da combinação de técnicas de Recuperação de Informação e de Inteligência Artificial (IA) com o conhecimento jurídico do especialista, enquanto que o Raciocínio Baseado em Casos

é uma técnica bastante utilizada no direito porque utiliza praticamente o mesmo raciocínio dos juristas na solução de um problema, pois busca resolver um problema atual buscando soluções utilizadas em casos anteriores.

Portanto, o sucesso da aplicação das técnicas de RBC na área do Direito se deve ao fato delas utilizarem a mesma forma de raciocínio humano para encontrar a solução de um problema.

Um outro exemplo significativo de sua aplicação é o Sistema Olimpo que resultou na criação de um sistema inteligente para aplicação nas resoluções do Conselho de Segurança da ONU, recuperando as resoluções mais semelhantes ao texto oferecido como entrada (Hoeschl, 2001).

No atual estágio da Web Semântica, existe a possibilidade de se obter uma maior eficácia na recuperação de informações jurídicas e na construção de um sistema de conhecimento jurídico a partir do uso de ontologias juntamente com as técnicas de RBC.

Entretanto é oportuno salientar que no que concerne à construção de um sistema de conhecimento jurídico de apoio à decisão judicial quanto a aspectos relacionados ao direito penal, mais especificamente quanto à tipificação de crimes, deve-se ter um cuidado especial em sua elaboração e verificar se os resultados são juridicamente válidos, tendo em vista que o ordenamento jurídico brasileiro não admite a analogia para forçar a consumação de um delito cuja conduta não esteja expressamente prevista em lei.

Com relação à eficiência dos mecanismos de busca na sociedade da informação, as ontologias podem contribuir para que estes sejam mais eficazes, permitindo que a pesquisa seja realizada dentro de um contexto de forma que a ferramenta automatizada seja capaz de entender o que está sendo pesquisado e, assim, apresentar um resultado com maior precisão e qualidade.

6 Considerações Finais

O aprimoramento da Web com o advento dos recursos de semântica é extremamente significativo. Isto pode ser comprovado pelo volume de recursos investidos em diversos grupos de pesquisa que estão atualmente engajados no desenvolvimento e na implantação de uma web Semântica.

Os principais benefícios de uma Web semântica terão uma maior efetividade a partir do momento em que uma massa crítica seja atingida. Segundo Berners-Lee, em uma entrevista para a Technology Review, comenta que “a previsão é que atingiremos uma massa crítica onde tudo começa a ser interligado de forma inimaginável e em grande escala. Daí, o incentivo para adicionar mais conteúdo aumenta exponencialmente, assim como o valor do que já foi agregado” (Fraunfelder, 2004).

A integração de ontologias à Web poderá contribuir significativamente para a criação de ambientes, sistemas e arquiteturas mais sólidas e, também, para o desenvolvimento e a consolidação de várias atividades na Web, tais como: mecanismos de buscas mais precisos, como o Ontoweb, e ainda prover a interoperabilidade entre serviços e aplicações, além do entrelaçamento semântico entre os conteúdos das páginas.

Referências Bibliográficas

1. Afonso, M. M. R. "Semantic Web". Porto: Universidade do Porto, 2001. Disponível em: <<http://paginas.fe.up.pt>>. Acesso em: 11 mar. 2007.
2. Berners-Lee, T; Hendler, J.; Lassila, O.: "The Semantic Web". Scientific American, maio 2001.
3. Faria, C. G.; Girardi, R. "Uma Análise da Web Semântica e suas Implicações no acesso a Informação." Disponível em: <<http://maae.deinf.ufma.br>>. Acesso em: 11 mar 2007.
4. Fraunfelder, Mark. "Sir Tim Berners-Lee: He Created the Web. Now He's Working on Internet 2.0. Technology Review, October 2004.
5. Garcia T. H. B. et al. "Portal do consumidor catarinense: um modelo de atendimento ao cidadão no âmbito do governo eletrônico". In: Anais do II Ciberética - Simpósio Internacional de Propriedade Intelectual, Informação e Ética. Florianópolis, 2003.
6. Garcia, T. H. B. et al. "Sistemas inteligentes na era do governo eletrônico: uma aplicação para enquadramento legal no âmbito do direito do consumidor". In: Rover, A. J. (org.). Direito e Informática. Barueri (SP): Manole, 2004. 479-491.
7. Hoeschl, H. C. "Sistema Olimpo: tecnologia da informação jurídica para o Conselho de Segurança da ONU". Florianópolis, 2001. 133 f. Tese (Doutorado) - Universidade Federal de Santa Catarina.
8. Hoeschl, H. C. et al. "Ontologias e sistemas de conhecimento jurídico: uma abordagem sobre entorpecentes". In: I Conferência Sul-Americana de Ciência e Tecnologia Aplicada ao Governo Eletrônico. Florianópolis: Ijuris, 2004. 161-168.
9. Hoeschl, H. C. "Ontoweb: A nova era das ferramentas de busca". In: Revista Consultor Jurídico. Publicado em 1º de fevereiro de 2006. Disponível em: <<http://conjur.estadao.com.br/static/text/41493,1>>. Acesso em: 14 mai. 2007.
10. Ijuris. "Ontoweb: Ferramenta Informacional de Governo Eeletrônico", Instituto de Governo Eletrônico, Inteligência Jurídica e Sistemas. Disponível em <<http://www.ijuris.org>>. Acesso em: 14 mai. 2007.
11. Moura, A. M. "A Web Semântica: Fundamentos, Tecnologias e Tendências". Rio de Janeiro, IME, 2002. Disponível em: <<http://ipanema.ime.eb.br>>. Acesso em 11 mar. 2007.
12. Staab, S.; Maedche, A. "Knowledge Portals – Ontologies at Work". Disponível em: <<http://www.aifb.uni-karlsruhe.de>>. Acesso em: 14 mai. 2007.