



Tecnologias da Web Semântica no acesso e uso à informação aberta: estudo de caso com um blog de informação em saúde

Karina Sayuri Hagiwara*, Ivan Luiz Marques Ricarte

Resumo

A Web Semântica incorpora significado às informações da Web de modo que sejam processáveis por máquina. Neste projeto estudou-se e aplicou-se as tecnologias da Web Semântica, além de averiguar como estas ampliam as possibilidades de utilização e acesso às informações na Web, além de ter sido efetuado uma estratégia para tornar um blog convencional em uma fonte de dados abertos e conectados, a partir de publicações de um blog. Para este fim, foram realizados o desenvolvimento de um aplicativo de software em Java para captura de dados do blog e a incorporação das tecnologias da web semântica, a realização de um modelo RDF/RDFS e a construção de uma base para a mesma, disponibilização de um ponto de consulta SPARQL, avaliação das alterações da estrutura da informação no site e cadastro da aplicação em diretórios de aplicações de Web Semântica e de Dados Abertos.

Palavras-chave: Web Semântica, Dados Abertos Conectados, Informação em Saúde.

Introdução

A World Wide Web alterou a maneira como as pessoas compartilham conhecimento, diminuindo a barreira para publicar e acessar documentos (BIZER et al., 2011). Dessa maneira, há um grande número de conteúdo e utilizadores que a acessam, no entanto, apesar dos benefícios trazidos pela Web, o significado, o contexto e aplicabilidade dos dados encontrados nela não conseguem ser processados por máquinas. A solução para essa limitação está no conceito da Web Semântica, a idéia por trás deste conceito é tecer uma Web que não só vincula os documentos uns aos outros, mas também reconhece o significado da informação nesses documentos (FRAUENFELDER, 2001).

O objetivo deste trabalho é explorar a utilização das tecnologias da Web Semântica e avaliar como elas podem ampliar as possibilidades de acesso e utilização da informação disponível na Web. Foi realizado um estudo de caso a partir da informação disponibilizada em um blog sobre a saúde da criança e do adolescente, cujo título é Fale com o Dr. Risadinha. Como objetivos secundários, espera-se poder estabelecer diretrizes para que esse tipo de iniciativa possa ser reproduzido em outros contextos, bem como o desenvolvimento de ferramentas que apoiem esse processo.

Resultados e Discussão

Foi realizada uma análise das informações disponíveis no blog e também um estudo das tecnologias e plataformas da Web Semântica que serão utilizadas no projeto, tais como: RDF, RDFS, SPARQL, Apache Any23 e Apache Jena. Com base nisto, foi possível efetuar um modelo RDF/RDFS para capturar a informação apresentada nas publicações do blog, diferenciando dados genéricos e específicos, e conseqüentemente, foi dado início ao desenvolvimento de um aplicativo Java com a finalidade da criação dos dados semânticos para as publicações do blog.

O aplicativo de software em Java, inicialmente, obtém todos os dados do blog, através da utilização do Gson jar, uma biblioteca Java. Em seguida, de acordo o

modelo RDF, utiliza-se do Apache Jena para a criação dos dados semânticos para as publicações do blog. Jena é uma interface de programação de aplicativos Java que a HP Labs desenvolveu para tornar mais fácil o desenvolvimento de aplicativos que utilizam o modelo de informações e os linguagens da Web Semântica (MCBRIDE, 2002). Posteriormente, o aplicativo constrói uma base de dados RDF e, finalmente, é disponibilizado um ponto de consulta SPARQL para estes dados e o cadastro da aplicação em diretórios de aplicações de Web Semântica e de Dados Abertos.

Conclusões

Obteve-se a descrição de uma estratégia para tornar um blog convencional em uma fonte de dados abertos e conectados para a Web Semântica, contemplando aspectos genéricos e aspectos específicos do domínio da informação, além do desenvolvimento de uma ferramenta de software para apoiar esta estratégia e a disponibilização pública de um ponto de acesso de consulta SPARQL para as informações do blog e o cadastro da aplicação em diretórios de aplicações de Web Semântica e de Dados Abertos.

Além disso, com as informações obtidas, houve a avaliação das alterações da estrutura da informação no site e a avaliação das características de abertura e conexão dos dados disponibilizados, explorando conexão com outras fontes de dados da Web Semântica. Deve-se considerar que este projeto é apenas uma base e necessita de mais pesquisa para o amplo desenvolvimento desta área.

Agradecimentos

Ao CNPq, pela bolsa de IC, ao Orientador Prof. Dr. Ivan Luiz Marques Ricarte e a PRP Unicamp.

BIZER, Christian; HEATH, Tom; BERNERS-LEE, Tim. Linked data: The story so far. In: *Semantic services, interoperability and web applications: emerging concepts*. IGI Global, 2011. p. 205-227.

FRAUENFELDER, M. *A smarter Web*. 2001.

MCBRIDE, B. *Jena: A Semantic Web Toolkit*. IEEE Internet Computing. v.6 n.6, p.55-59. 2002.