

Representando Esquemas de Fontes de Dados Geográficos através de Ontologias

Danúbia Almeida, Antonio Mendonça, Damires Souza

Instituto Federal de Educação, Ciência e Tecnologia da Paraíba - IFPB
Caixa Postal 15.064 – 91.501-970 – João Pessoa – PB – Brasil

{biaeccentric@gmail.com, tony2415@gmail.com, damires@ifpb.edu.br}

Abstract. *In the last few years, there has been a growing interest in approaches that define methodologies to automatically extract semantics from databases by using ontologies. When dealing with geographic data, specific problems with representation and usage may occur. In this light, mapping a geospatial database schema to an ontology becomes a hard task and an extension of existing methodologies is needed. This work aims to develop a tool able to support such need.*

Resumo. *Nos últimos anos, é crescente o interesse por abordagens que definam metodologias capazes de extrair semântica de bancos de dados através do uso de ontologias. Em se tratando de dados geográficos, problemas mais específicos de representação e utilização surgem. Neste cenário, mapear o esquema de um banco de dados geográficos para uma ontologia representa uma tarefa complexa, e a extensão de metodologias existentes é necessária. Este trabalho visa o desenvolvimento de uma ferramenta que supra essa necessidade.*

1. Introdução

A atual dinamicidade das organizações, com constantes evoluções nos modelos de negócios, e a necessidade de disponibilização ágil das informações demanda a utilização de arquiteturas de sistemas computacionais mais flexíveis e extensíveis. Uma destas soluções de arquitetura envolve a utilização de uma rede *Peer-to-Peer* (P2P). Entretanto, devido à necessidade de se lidar com dados estruturados, provenientes normalmente de fontes como bancos de dados, sistemas P2P vêm evoluindo para sistemas de gerenciamento de dados através dos chamados *Peer Data Management Systems* (PDMS) [Sung *et al.* 2005]. Um PDMS é um sistema de gerenciamento de dados que utiliza uma rede P2P e busca aliar benefícios como a ausência de autoridade central com o poder semântico dos bancos de dados [Zhao 2006].

O sistema SPEED (*Semantic Peer-to-Peer Data Management System*) é um exemplo de PDMS [Pires 2009]. É um sistema baseado em semântica onde pontos de dados (associados a fontes de dados) são agrupados de acordo com seu domínio em *clusters* semânticos. Dentro deste contexto, estamos trabalhando com uma instanciação do sistema SPEED através da utilização de fontes de dados geográficos. Para facilitar tarefas de integração, ontologias são utilizadas como forma de representação dos esquemas dos bancos de dados geográficos. Neste sentido, o presente trabalho visa o desenvolvimento de uma ferramenta que realize o mapeamento do esquema de um banco de dados geográfico em uma representação de ontologia. A seguir, introduzimos

alguns conceitos pertinentes ao escopo do trabalho na Seção 2, apresentamos seu estágio atual na Seção 3 e tecemos algumas considerações sobre o mesmo na Seção 4.

2. Fundamentação Teórica

A heterogeneidade das fontes de dados requer que o sistema PDMS possua uma capacidade diferenciada de modelagem, de maneira que o mesmo possa representar e integrar o conteúdo das diferentes fontes evitando a perda de semântica, normalmente, a partir de um modelo de representação comum. No sistema SPEED, utiliza-se ontologias como forma de representação dos esquemas das fontes. Uma ontologia é uma especificação explícita de uma conceitualização compartilhada onde objetos, conceitos, entidades e relacionamentos do mundo real são definidos dentro de um determinado domínio do conhecimento [Gruber 1995]. Como o formato OWL¹ (*Web Ontology Language*) tem sido indicado pelo W3C² como padrão, foi escolhido como modelo de representação para as ontologias dos esquemas das fontes.

Em se tratando de dados geográficos, problemas mais específicos de integração (de esquemas e de dados) surgem. Parte deles diz respeito aos conflitos estruturais e sintáticos que acontecem no momento em que correspondências devem ser identificadas entre elementos de esquemas de fontes distintas (processo denominado como *matching* de esquemas). Por exemplo, uma sede de município pode ter sido definida no esquema de uma determinada fonte A como um *ponto* e ter sido definida no esquema de uma fonte B como um *polígono*, o que implica em tipos de representação distintos.

3. Estágio Atual do Trabalho

No SPEED, para extrair a representação de esquemas de bancos de dados, utilizamos uma ferramenta que gera ontologias a partir de bancos de dados objeto-relacionais [Franco 2009]. Neste trabalho, como precisamos lidar com dados geográficos, a ferramenta está sendo estendida de modo a contemplar as características e requisitos desses tipos de dados. A ferramenta original já gera ontologias em formato owl para bancos Oracle³. Ela está sendo estendida também para permitir o uso de outros SGBDs (Sistemas de Gerenciamento de Banco de Dados), além do Oracle.

O processo de geração das ontologias de representação inclui etapas como extração dos metadados, identificação de entidades espaciais e não espaciais, identificação de propriedades de tipos simples, identificação e classificação de relacionamentos espaciais e não espaciais, entre outros. Para tal, regras vêm sendo especificadas e testadas. Estas regras constituem a base para a implementação da ferramenta.

A idéia geral desse processo e do trabalho é mostrada na Figura 1 que representa o esquema da captura dos dados de uma base geográfica, onde tais dados são lidos de um SGBD pela ferramenta geradora de ontologias. Após esta etapa e, de acordo com as regras de mapeamento, é gerado o arquivo OWL – a ontologia geográfica de representação do esquema do banco.

¹ <http://www.w3.org/TR/owl-features/>

² <http://www.w3.org/>

³ <http://www.oracle.com/br/products/database/index.html>

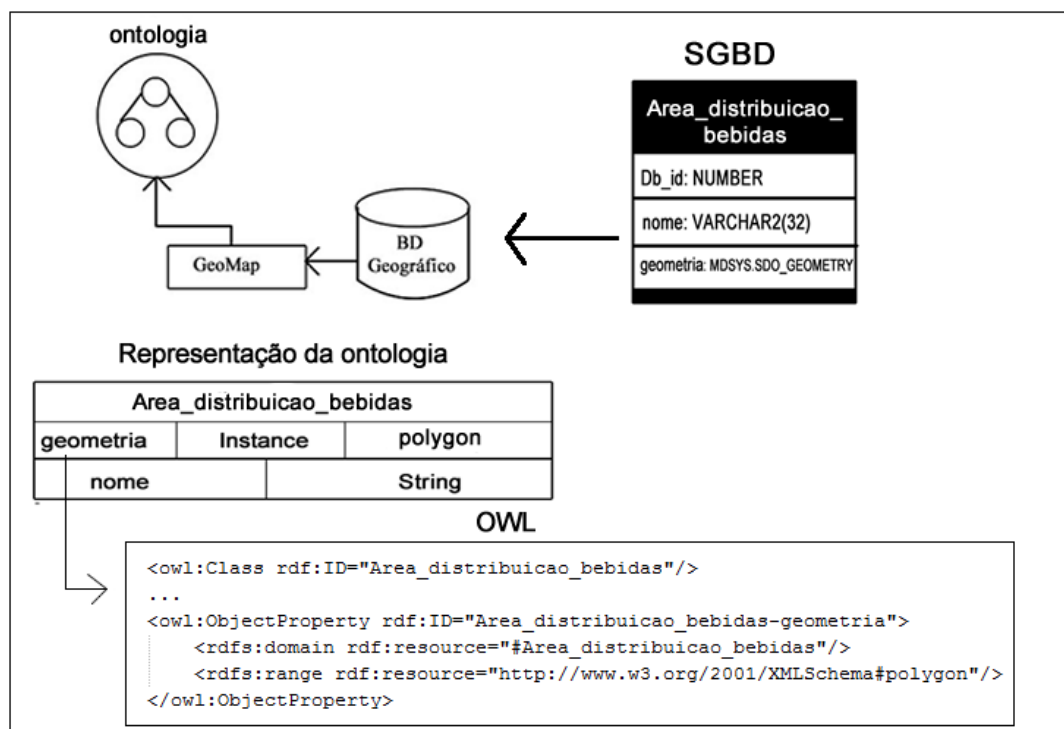


Figura 1: Esquema da geração de uma ontologia geográfica.

4. Considerações

Este trabalho encontra-se em andamento, as regras de mapeamento estão sendo testadas através de estudos de casos. A partir daí, a ferramenta será completamente estendida de forma a suportar estas regras.

Referências

- Franco, F. (2009) "Uma Ferramenta de Transformação de Esquemas Objeto-Relacionais para Ontologias". Trabalho de Graduação, Centro de Informática, Universidade Federal de Pernambuco (UFPE), Recife, PE, Brasil.
- Gruber T. (1995) "Toward principles for the design of ontologies used for knowledge sharing". International Journal Human-Computer Studies Vol. 43, Issues 5-6, November 1995, p.907-928.
- Pires C.E.S. (2009) "Ontology-Based Clustering in a Peer Data Management System". PhD thesis, Center for Informatics, UFPE, 2009.
- Souza D. (2009) "Using Semantics to Enhance Query Reformulation in Dynamic Distributed Environments". PhD Thesis, Federal University of Pernambuco (UFPE), Recife, PE, Brazil.
- Sung L. G. A., Ahmed N., Blanco R., Li H., Soliman M. A., Hadaller D. (2005) "A Survey of Data Management in Peer-to-Peer Systems". School of Computer Science, University of Waterloo.
- Zhao J. (2006) "Schema Mediation and Query Processing in Peer Data Management Systems". Master Thesis, University Of British Columbia, October, 2006.