

GeoMap – Uma Ferramenta para Representação de Esquemas de Bancos de Dados Geográficos através de Ontologias

Danúbia Lima¹, Antonio Mendonça¹, Ana Carolina Salgado², Damires Souza¹

¹ Instituto Federal de Educação, Ciência e Tecnologia da Paraíba/IFPB, Brazil

² Centro de Informática/UFPE, Brazil

{danubialima@gmail.com, tony2415@gmail.com, acs@cin.ufpe.br, damires@ifpb.edu.br}

Resumo. *Um problema central em sistemas gerenciadores de dados em ambientes P2P (PDMSs) diz respeito à heterogeneidade dos esquemas de seus pontos. Como forma de minimizar tal problema, ontologias podem ser usadas como notação de representação conceitual padrão desses esquemas. Neste trabalho, estamos usando bancos de dados geográficos para instanciar um sistema PDMS. Em se tratando de dados geográficos, problemas específicos de representação e uso ocorrem. Neste sentido, nós desenvolvemos uma ferramenta, chamada GeoMap, que automaticamente extrai metadados de bancos de dados geográficos e constrói uma ontologia do ponto.*

Abstract. *One key issue in Peer Data Management Systems (PDMSs) is the heterogeneity of the peer schemas. To help matters, ontologies may be used as uniform conceptual representation of these schemas. In this work, we are working with geographic databases to be used in a PDMS. When dealing with geographic data, specific problems with representation and usage occur. In this light, we have developed a tool, named GeoMap, that automatically extract metadata from geographic databases and builds a peer ontology.*

1. Introdução

A necessidade crescente de disponibilização ágil das informações demanda a utilização de ambientes computacionais mais extensíveis que permitam o acesso transparente aos dados. Um desses ambientes envolve a utilização de uma rede *Peer-to-Peer* (P2P). No contexto de integração de dados estruturados e semi-estruturados, *PDMSs* (*Peer Data Management Systems*) ou sistemas de gerenciamento de dados em ambientes P2P vêm sendo alvo de pesquisa e desenvolvimento [Sung *et al.* 2005]. Um PDMS é um sistema de gerenciamento de dados que busca aliar os benefícios de sistemas P2P, como a ausência de controle central, com o poder semântico dos bancos de dados [Zhao 2006]. Deste modo, cada ponto pode tanto atuar como cliente realizando consultas como servidor provendo resultados às consultas formuladas. Entretanto, o gerenciamento de dados em ambientes P2P representa uma tarefa complexa devido ao grande número de pontos, sua natureza autônoma e à heterogeneidade de seus esquemas. O sistema *SPEED* (*Semantic PEER-to-Peer Data Management System*) é um exemplo de PDMS [Pires 2009, Souza 2009] baseado em semântica, onde pontos de dados (associados a fontes de dados) são agrupados de acordo com seu domínio em *clusters* semânticos. *Clusters* semânticos são agrupados em comunidades semânticas, onde um ponto

semântico atua como gerenciador da comunidade, oferecendo uma ontologia de domínio que pode ser usada como uma referência.

Para que um ponto possa participar do SPEED, ele deve publicar seu esquema de dados de modo que este possa ser visto pelos demais pontos da rede e, em contrapartida, o sistema provê mapeamentos deste ponto com os demais. Dentro deste contexto, estamos trabalhando com uma instanciiação do sistema SPEED através da utilização de pontos cujas fontes são bancos de dados geográficos. A heterogeneidade dos esquemas requer que o sistema possua uma capacidade diferenciada de modelagem, de maneira que o mesmo possa representar e reconciliar os esquemas das fontes, evitando a perda de semântica. Para isso, o SPEED usa uma notação ontológica como modelo de representação comum dos esquemas dos pontos. Neste sentido, este trabalho apresenta a ferramenta *GeoMap* que realiza o mapeamento do esquema de um banco de dados geográfico em uma representação de ontologia, codificada em OWL.

Este trabalho está organizado da seguinte forma: a Seção 2 introduz alguns conceitos e descreve a metodologia que fundamenta a ferramenta; a Seção 3 detalha os requisitos funcionais da *GeoMap*, apresentando um exemplo com o intuito de prover sua validação; finalmente, a Seção 4 tece algumas considerações e indica trabalhos futuros.

2. Abordagem e Metodologia para a *GeoMap*

A heterogeneidade de esquemas é um dos grandes desafios para o compartilhamento de dados em ambientes distribuídos como os *PDMSs*. Como forma de minimizar essas diferenças, ontologias têm sido empregadas como modelo de representação, criando uma camada conceitual entre os serviços do sistema e os dados das fontes. Existem várias ferramentas que lidam com o mapeamento de um esquema de banco de dados para uma notação de ontologia, entretanto, a maioria delas está focada em bancos de dados relacionais. Um exemplo desse tipo de ferramenta é a *RDBToOnto* [Cerbah 2008]. Uma ferramenta para extração de esquema de um banco de dados objeto-relacional foi implementada em [Franco 2009]. Em se tratando de fontes de dados geográficos, alguns trabalhos têm sido desenvolvidos com o objetivo de prover alinhamentos entre ontologias geográficas [Cruz *et al.* 2004] ou merge de ontologias [Bucella *et al.* 2008]. O trabalho de Baglioni e seu grupo [Baglioni *et al.* 2007] cria um método para, de forma semi-automática, construir uma ontologia para um banco de dados geográfico. Comparando estes trabalhos com o nosso, o mais próximo é o último citado, entretanto, no nosso trabalho, a geração da ontologia é realizada de forma automática. Para isso, usamos uma ontologia de domínio como referência de termos para a criação de construtores específicos (do domínio geográfico) da ontologia de representação.

Um aspecto fundamental no nosso trabalho é a manipulação de metadados geográficos. Dados geográficos, quando representados através do modelo vetorial, são tratados como objetos individualizáveis e armazenados como pontos, linhas ou polígonos, dependendo da escala de sua obtenção. Dessa forma, a heterogeneidade das fontes é ainda maior: dados podem ter múltiplas representações (um mesmo dado pode ser representado como um ponto numa fonte e como um polígono em outra fonte), dados podem ter resoluções e sistemas de coordenadas diferentes e propriedades temporais associadas (múltiplas versões). Neste sentido, as heterogeneidades sintática,

semântica e de formato dos dados geográficos devem ser consideradas na criação de uma ontologia de representação. A ferramenta deve estar habilitada para identificar diferenças entre modelos de SGBDs geográficos (por exemplo, o Oracle armazena dados geográficos diferentemente do PostGIS), assim como diferenças de formato para os objetos existentes (se é ponto ou polígono, por exemplo).

Considerando OWL como modelo de codificação comum, o SPEED minimiza os problemas de heterogeneidade através da transformação dos esquemas das fontes de dados em ontologias. Mas, o conjunto de construtores pré-definido em OWL não inclui especificações para a descrição de conceitos e propriedades geográficas. Diante disso, foi necessária a utilização de uma ontologia de domínio do contexto geográfico que pudesse ser usada como referência de termos. Esta ontologia foi construída, por não ter sido encontrada uma ontologia padrão que atendesse às nossas necessidades. Nela, conceitos como *Entity*, *SpatialEntity* (entidade com componente espacial), *Geometry* (ponto, linha ou polígono), *SpatialRelationship* (por exemplo, topológicos), entre outros, foram criados, assim como os relacionamentos entre eles. Um fragmento da ontologia é mostrado na Figura 1.

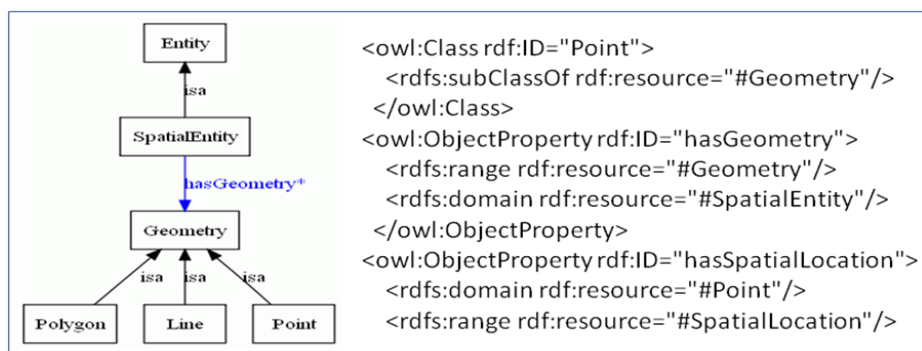


Figura 1. Fragmento da Ontologia de Referência

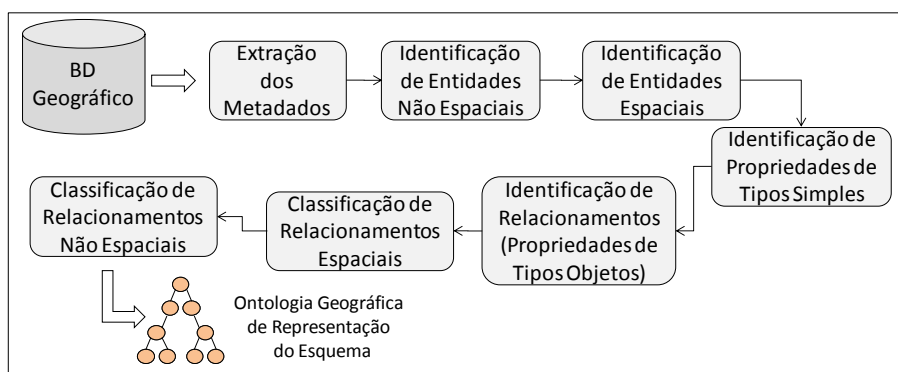


Figura 2. Processo de Geração das Ontologias

Fundamentado nesses aspectos, o processo de geração da ontologia de representação inclui as etapas mostradas na Figura 2. Durante este processo, são aplicadas regras para transformar os metadados do esquema origem do banco de dados geográfico em elementos correspondentes numa ontologia destino. Essas regras dizem respeito à: (i) extração dos metadados do banco de dados; (ii) identificação das entidades não espaciais (tabelas que não possuem colunas geográficas); (iii) identificação das entidades espaciais (tabelas que representam um conceito geográfico -

possuem coluna que representa sua geometria); (iv) identificação de propriedades de tipo simples (por exemplo, *varchar* e *number*); (v) identificação de tipos de relacionamentos (propriedades de tipos de objetos), com a especificação de *domain* e *range*; (vi) classificação desses relacionamentos como espaciais ou não espaciais. Com base nessas etapas, um arquivo OWL é gerado contendo a ontologia do ponto. Na seção seguinte, apresentamos a ferramenta, onde descrevemos com mais detalhes seus requisitos funcionais.

3. A Ferramenta *GeoMap*

Os requisitos funcionais da *GeoMap* foram especificados através de um diagrama de casos de uso (Figura 3). Existem dois atores no sistema. O primeiro é a própria ferramenta *GeoMap* que inicializa todo processo de geração da ontologia através da conexão com o banco de dados. Este, por sua vez, caracteriza o segundo ator. Neste desenvolvimento, por enquanto, as opções de geração são iniciadas através de um “usuário”, ou seja, de forma manual. Entretanto, esta ferramenta será inserida no SPEED como um serviço a ser chamado sempre que um ponto solicitar sua entrada no sistema.

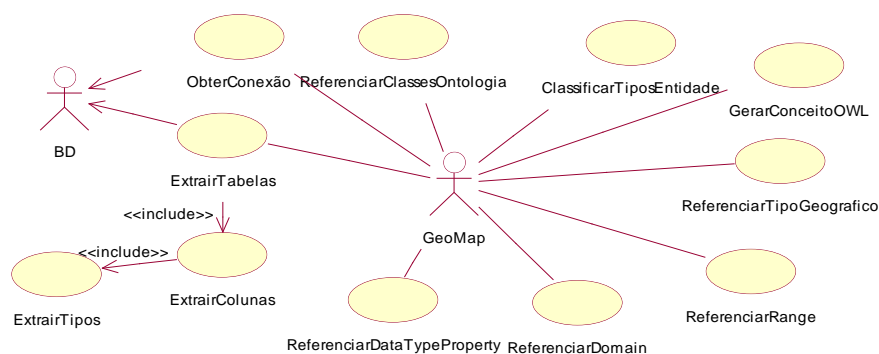


Figura 3. Diagrama de Casos de Uso da *GeoMap*

Após a conexão com o banco de dados, a *GeoMap* recupera as tabelas e os tipos existentes no esquema. Através dessa recuperação, é possível extrair todas as colunas, identificando o que é atributo simples e o que é atributo que representa uma geometria. Tabelas que possuem colunas com representação geográfica são classificadas como espaciais, enquanto que as demais são classificadas como não espaciais. Em seguida, a ferramenta identifica o tipo de geometria das tabelas (entidades) espaciais – se é ponto, linha ou polígono. Com base no levantamento de todas as entidades e suas propriedades (simples ou de objetos), a ferramenta faz uso da ontologia de domínio como referência de termos para a criação dos construtores específicos do contexto geográfico. Esta ontologia é necessária porque o conjunto de construtores pré-definido de OWL não contém tipos específicos do contexto geográfico (por exemplo, *point* ou *line*). Assim, ao criar a representação das entidades e fazer menção aos tipos geográficos, usamos a referência da ontologia de domínio. Ao referenciar o *domain*, identificamos o domínio das propriedades. Quando referenciamos o *range* para uma coluna geográfica especificamos os tipos *line*, *polygon* ou *point*, presentes na ontologia de referência. Após estas etapas, é realizada uma associação dessas informações, montando-se a estrutura do arquivo OWL, com a ontologia geográfica do esquema do banco.

A ferramenta *GeoMap* foi implementada na linguagem JAVA, como uma extensão de uma ferramenta para geração de ontologias a partir de bancos de dados objeto-relacionais [Franco 2009]. Para esta versão, a *GeoMap* utilizou o SGBD Oracle 10g para persistência do banco de dados geográfico. A manipulação da ontologia foi realizada através da API Protege-OWL [Protege 2011] e do framework JENA [Jena 2011]. A seguir, a ferramenta é apresentada através de um exemplo de utilização.

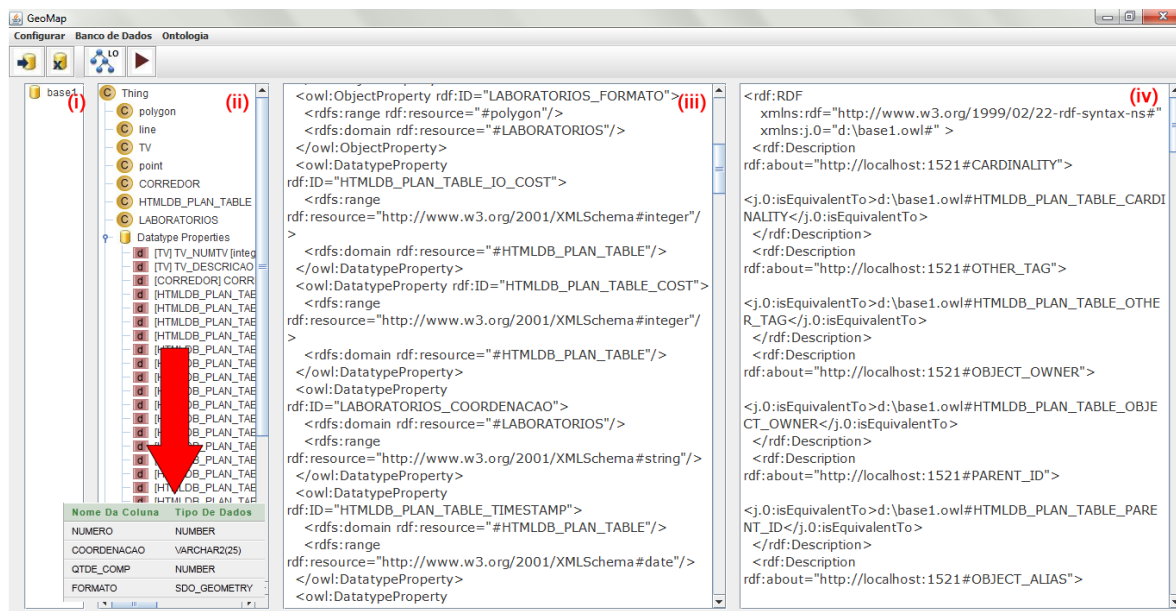


Figura 4. Interface da *GeoMap* e Ampliação do Esquema do BD

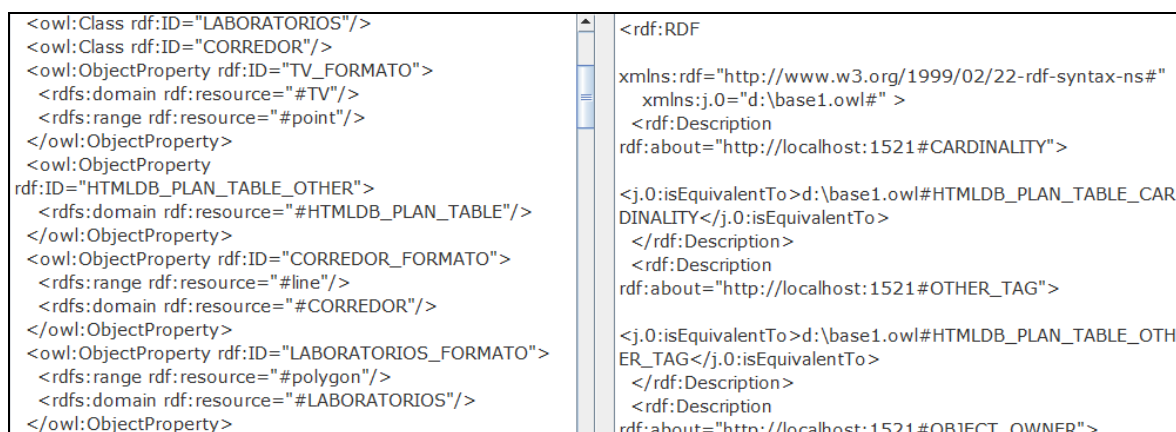


Figura 5. *Snapshot* com Fragmento da Ontologia do Ponto e do Conjunto de Correspondências

Na Figura 4, a interface da *GeoMap* é apresentada, com a indicação de suas principais áreas: (i) área com o nome do banco; (ii) área com a estrutura do esquema do banco; (iii) área com a ontologia do ponto gerada; e (iv) área com o arquivo de correspondências obtido. Neste exemplo, estamos utilizando um banco de dados geográfico que armazena laboratórios do IFPB. Assim, uma tabela “Laboratório” é representada como polígono (*polygon*) no banco (através do tipo *MDSYS.SDO_GEOMETRY*). A estrutura do esquema origem do banco é ampliada na Figura 4. Seguindo o processo explicado na Seção 2, a ferramenta gera a ontologia para

este esquema, como mostrado na interface da ferramenta (Figura 4). A *GeoMap* também gera um arquivo de correspondências que identificam equivalências entre componentes obtidos na ontologia do ponto e os elementos existentes no esquema do banco. Para facilitar a visualização, apresentamos um fragmento da ontologia do ponto gerada, bem como do arquivo de correspondências obtido na Figura 5.

4. Considerações e Trabalhos Futuros

Neste trabalho, foi apresentada a ferramenta *GeoMap*, que realiza a extração de metadados de bancos de dados geográficos e os representa em termos de uma ontologia. Para tal, etapas de identificação e classificação de entidades e propriedades geográficas e não geográficas são realizadas. Uma ontologia de referência está sendo usada como forma de suprir a semântica de relacionamentos e tipos geográficos ausente do conjunto de conceitos existentes no modelo OWL. No momento, esta versão gera ontologias para bancos de dados Oracle. Como trabalho futuro, ela será estendida para também extrair dados de outros SGBDs, como o PostGIS e de outras fontes como a GML.

Referências

- Baglioni M., Masserotti M., Renso C., Spinsanti L. (2007) “Building Geospatial Ontologies from Geographical Databases”. *GeoS 2007*: 195-209.
- Buccella A., Perez L. and Cechich A. (2008) “GeoMergeP: Supporting an Ontological Approach to Geographic Information Integration”. In *SCCC '08: Proceedings of the International Conference of the Chilean Computer Science Society*, pp. 52–61, 2008.
- Cerbah, F. (2008) “Learning Highly Structured Semantic Repositories from Relational Databases – The RDBtoOnto Tool”. In: *5th European Semantic Web Conference (ESWC'08)*, pages 777-781, Tenerife, Spain (2008).
- Cruz, I.F., Sunna, W., Chaudhry, A. (2004) “Semi-automatic ontology alignment for geospatial data integration”. In: Egenhofer, M.J., Freksa, C., Miller, H.J. (eds.) *GIScience 2004*. LNCS, vol. 3234, pp. 51–66. Springer, Heidelberg (2004).
- 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.
- Jena (2011). Jena - A Semantic Web Framework for Java. <http://jena.sourceforge.net/>
- OWL API (2011) – API da Protégé. <http://protege.stanford.edu/plugins/owl/api/>
- 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.