UMA FERRAMENTA DE APOIO A DETERMINAÇÃO DE
EQUIVALÊNCIAS SEMÂNTICAS ENTRE ESQUEMAS GML
UTILIZANDO ONTOLOGIAS OWL

LEONARDO ROMAN DA ROSA

Trabalho de conclusão de curso apresentado como parte
dos requisitos para obtenção do grau de Bacharel em
Sistemas de Informação.

Florianópolis - SC

2006 / 2
LEONARDO ROMAN DA ROSA

UMA FERRAMENTA DE APOIO A DETERMINAÇÃO DE EQUIVALÊNCIAS SEMÂNTICAS ENTRE ESQUEMAS GML UTILIZANDO ONTOLOGIAS OWL

Trabalho de conclusão de curso apresentado como parte dos requisitos para obtenção do grau de Bacharel em Sistemas de Informação.

Orientador: Ronaldo dos Santos Mello, Dr.

Banca examinadora

Lia Caetano Bastos, Dr.

Marília Abrahão Amaral, Dr.
Dedicatórias
Agradecimentos
Sumário

Agradecimentos........................................................................................................... 4
Sumário .......................................................................................................................... 5
Lista de Figuras............................................................................................................. 6
Resumo ........................................................................................................................... 7
Palavras-chave............................................................................................................... 7
Introdução..................................................................................................................... 8
SIGs e a GML............................................................................................................... 9
Interoperabilidade entre Esquemas GML ................................................................. 9
O Uso da Ontologia na Integração ............................................................................ 10
Trabalhos Relacionados............................................................................................. 13
O Método para a Interoperabilidade entre GMLs ..................................................... 15
A Ferramenta de Interoperabilidade Semântica ....................................................... 16
Trabalhos Futuros ....................................................................................................... 17
Referências Bibliográficas ......................................................................................... 19
Anexos ......................................................................................................................... 21
Lista de Figuras

Ilustração 1- Funcionamento do método proposto (Frozza 2006) ..................... 15
**Resumo**

Um dos problemas existentes na Web é a identificação semântica adequada do conteúdo de um site. O uso da XML permite resolver ambigüidades no significado desses conteúdos. Em sistemas de informações geográficas existe um problema similar relacionado à integração semântica entre esquemas GML.

Sendo um padrão aberto, o GML é muito utilizado para armazenamento de informações geográficas codificadas em XML. Enquanto a GML facilita o intercâmbio de informações geográficas, ela não resolve por si só a questão da integração semântica de esquemas heterogêneos entre diferentes sistemas de informações geográficas. O que queremos, portanto, é fazer uma análise entre dois esquemas relacionados a um mesmo domínio e encontrar seus elementos semanticamente equivalentes. Para facilitar essa análise, será utilizada uma ontologia OWL com a representação dos conceitos do domínio da aplicação.

Neste trabalho temos como propósito final a criação da ferramenta com o objetivo de demonstrar na prática essa integração de dados geográficos.

**Palavras-chave**

XML, Extensible Markup Language, GML, Geography Markup Language, OWL, Web Ontology Language, Ontologias, SIG, GIS, Integração Semântica
**Introdução**

A maior parte da informação na Web é atualmente apresentada na linguagem HTML (Hypertext Markup Language), que foi desenvolvida para possibilitar o acesso a informações através dos navegadores Web. Enquanto a linguagem HTML permite visualizar essas informações, ela não tem muita capacidade de descrevê-las. Essa capacidade de descrição da informação facilita a interoperabilidade entre diferentes aplicações.

A Geography Markup Language (GML) é um padrão aberto para o transporte e armazenamento de informações geográficas em arquivos codificados na linguagem XML. A integração entre dois esquemas GML que tratam sobre as mesmas informações geográficas tem grande importância para órgãos públicos e privados que utilizem informações geográficas de diferentes fontes.

Nesse trabalho usaremos especificamente esquemas GML referentes ao domínio de cadastro urbano, muito utilizado no planejamento da infra-estrutura das cidades.
SIGs e a GML

Um Sistema de Informações Geográficas ou SIG, é definido como um sistema computacional capaz de coletar, armazenar, gerenciar, manipular, recuperar, transformar e visualizar dados geo-referenciados, com habilidade para integrar grandes quantidades de informação e com capacidade de utilizar ferramentas analíticas para explorar esses dados [SILVA 03].

A linguagem GML, de Geography Markup Language, é uma gramática XML para a modelagem, transporte e armazenamento de informações geográficas [GML 04]. A GML inclui uma variedade de objetos para descrição geográfica, tais como sistemas de coordenadas, geometria, topologia, unidades de tempo e medida.

Por ser um padrão aberto, a GML tem sido muito utilizada como meio de armazenamento de dados geográficos de diferentes SIGs, da mesma forma como a XML é utilizada por uma grande variedade de aplicativos e sistemas das áreas mais diversas como formato para armazenamento de dados e informações.

Interoperabilidade entre Esquemas GML

Informações geográficas são abstrações de fenômenos do mundo real [GML 04]. Elas são representações digitais de alguma parte do mundo real. Como tal, estão sujeitas à cultura (regionalismos), ao domínio dos programadores sobre o tema, entre outros.

Diferenças a nível semântico entre esquemas GML dizem respeito aos aspectos relacionados à compreensão e utilização de seus dados, em
diferentes aplicações, diferentes modelos de dados e diferentes interpretações desses dados.

O nível semântico da integração abrange a questão da unificação da nomenclatura utilizada para representar os fenômenos da realidade a serem modelados e os relacionamentos entre eles. Neste sentido, é necessário criar uma estrutura de organização do conhecimento, tal como um vocabulário controlado, uma taxonomia, um thesaurus ou uma ontologia [FROZZA 06].

A opção do uso de uma ontologia deve-se exatamente pela maior facilidade de comparação entre dois esquemas GML se utilizarmos como modelo de comparação uma estrutura padronizada.

O Uso da Ontologia na Integração

Em computação, uma ontologia é um modelo de dados que representa um domínio. É utilizada para abstrair sobre os objetos desse domínio e as relações entre eles [WIKI 06-1].

Segundo Duarte 2000, uma ontologia é uma descrição de conceitos e relações que existem em um domínio de interesse. Basicamente, uma ontologia consiste desses conceitos e relações, e suas definições, propriedades e restrições, descritas na forma de axiomas. Ontologias são úteis para apoiar a especificação e implementação de qualquer sistema de computação complexo. Uma ontologia pode ser desenvolvida para vários fins, mas, de uma forma geral, os seguintes propósitos são atingidos:

• Ajuda as pessoas a compreender melhor certa área de conhecimento: no desenvolvimento de uma ontologia, as pessoas envolvidas no processo se vêem diante de um desafio: explicar
seu entendimento sobre o domínio em questão, o que as faz refletir e melhorar sua compreensão sobre esse domínio;

- Ajuda as pessoas a atingir um consenso em seu entendimento sobre uma área de conhecimento: geralmente, em uma determinada área de conhecimento, diferentes especialistas têm entendimento diferenciado sobre os conceitos envolvidos, o que gera problemas na comunicação. Ao se construir uma ontologia, essas diferenças são explicitadas e busca-se um consenso sobre seu significado e sua importância;

- Ajuda outras pessoas a compreender certa área de conhecimento: uma vez que haja uma ontologia sobre uma determinada área de conhecimento desenvolvida, uma pessoa que deseje aprender mais sobre essa área não precisa se reportar sempre a um especialista. Ela pode estudar a ontologia e aprender sobre o domínio em questão, absorvendo um conhecimento geral e de consenso.

A linguagem de ontologia escolhida para a integração dos esquemas GML foi a OWL. A OWL (Web Ontology Language) é uma linguagem de representação criada justamente para que a informação contida em documentos possa ser processada por aplicações, ao invés de somente ser apresentada para visualização [OWL 04]. A OWL pode ser utilizada para representar o significado de termos em vocabulários e o relacionamento entre esses termos.

Essa linguagem também possui um suporte muito melhor a interpretação de conteúdo Web do que outras linguagens com abordagem semelhante como
Trabalhos Relacionados

Em Dorneles 2005 é descrita uma solução de abordagem para consultas que utilizem argumentos de busca booleanos. A proposta é a adoção de uma função de similaridade em substituição aos operadores booleanos. Com o uso de funções de similaridade, o sistema pode gerar um escore de semelhança entre cada objeto na base e o argumento de consulta.

Hess 2004 propõe uma solução de conflitos entre esquemas conceituais de bancos de dados geográficos, suportada por uma arquitetura de software, com duas etapas: uma sintática que deve converter os esquemas para o formato GML e a outra semântica que utiliza uma ontologia para a integração semântica. Esse modelo de resolução de conflitos entre esquemas conceituais de BDGs é proposto para ser universal, de forma a aceitar esquemas conceituais de diferentes domínios geográficos. Para solucionar os conflitos entre esquemas conceituais foi desenvolvida uma metodologia, suportada por uma arquitetura de software, que divide a fase de pré-processamento em duas etapas, uma sintática e uma semântica. A fase sintática visa converter os esquemas em um formato canônico, a Geographic Markup Language (GML). Para cada um dos diferentes modelos de dados um conjunto de regras foi desenvolvido e um **wrapper** implementado. Para suportar a etapa semântica da integração uma ontologia é utilizada para integrar semanticamente os esquemas conceituais dos diferentes projetos.

Uma abordagem para a interoperabilidade na representação de dados geográficos é feita por Silva 2003. Seu estudo tem como objetivo analisar e validar a hipótese da viabilidade da integração entre duas soluções para representação de dados geográficos: GeoBR/INPE e GML 3.0/OGC. Esses
modelos de representação foram escolhidos em virtude de seu uso no INPE, para fazer um estudo de caso. O tratamento do tempo na representação da informação geográfica foi abordado e explorado dentro do contexto do caso estudado, utilizando os recursos da especificação GML 3.0 e estendida ao formato GeoBR. Para a validação do trabalho, ferramentas de conversão e visualização foram desenvolvidas e adaptadas, em particular adaptações no conversor TerraTranslator, no aplicativo geográfico para Internet denominado SpringWeb, e um demonstrativo do uso de SVG para dados animados.
O Método para a Interoperabilidade entre GMLs

Um método para a determinação de equivalências semânticas entre esquemas GML é proposta por Frozza 2006, especificamente para o domínio do cadastro urbano. Em seu estudo sobre a interoperabilidade de SIGs desse domínio, é feito todo o embasamento teórico necessário para a criação de uma ferramenta de integração semântica de esquemas GML através de ontologias OWL. Sua proposta tem como objetivo a determinação de equivalências através de métodos semi-automáticos e prevê a aplicação de seus resultados em softwares que possibilitem a troca de informações entre bases de dados geográficos sobre uma determinada cidade.

Abaixo, a ilustração 1 fornece uma visão geral do funcionamento do método.

![Ilustração 1 - Funcionamento do método proposto [Frozza 06]](image-url)
A Ferramenta de Interoperabilidade Semântica

.....

....
Trabalhos Futuros
Referências Bibliográficas


Anexos

-código - fonte