

Módulo de Integração de Padrões de Análise com Ontologias de Domínio da ferramenta ArgoCASEGEO

Gabriel Silva Gonçalves, Jugurta Lisboa Filho, Evaldo de Oliveira da Silva, Guilherme Alves Lopes

Departamento de Informática – Universidade Federal de Viçosa (UFV)
36.570-000 – Viçosa – MG – Brasil

`gabrielsg@gmail.com, jugurta@ufv.br, evaldo.oliveira@gmail.com`

***Abstract.** Analysis patterns and domain ontologies are techniques that can be used in projects of information systems, in spite of presenting different levels of abstraction. With analysis patterns, the designers can reuse schemes previously validated, reducing costs and improving application's quality. Domain Ontologies are used to define the concepts and relationships within a specific domain. Together, these two techniques can help in the requirement elucidation and database design phases. This article describes the implementation of a module for integration and consulting (of analysis patterns and ontologies) for a CASE tool of geographic database modeling.*

***Resumo.** Padrões de análise e Ontologias de domínio são técnicas que, apesar de apresentarem níveis distintos de abstração, são ferramentas complementares no projeto de sistemas de informação. Com padrões de análise, os projetistas podem reutilizar esquemas já validados, reduzindo custos e melhorando a qualidade do sistema. Ontologias de domínio são usadas para definir conceitos e relacionamentos dentro de um domínio específico. Em conjunto, essas duas técnicas auxiliam na elucidação de requisitos da aplicação e na modelagem do banco de dados. Este artigo descreve a implementação de um módulo de integração e consulta (de padrões de análise e ontologias) para uma ferramenta CASE de modelagem de dados geográficos.*

1. Introdução

O projeto de aplicações de Sistema de Informação Geográfica (SIG) requer, inicialmente, a escolha das entidades a serem representadas e a descrição organizada destas por meio de conceitos.

Técnicas que facilitem o trabalho dos projetistas devem ser utilizadas por esses profissionais. O uso de Padrões de Análise [Fowler 1997], por exemplo, é uma técnica que permite a captura dos objetos de domínio e seus inter-relacionamentos, permitindo assim uma reutilização de componentes de software. Padrões de análise são idéias consideradas úteis dentro de certo contexto e que devem permanecer úteis em outros [Fowler 1997]. O conhecimento embutido nos padrões de análise, quando reutilizados, traz benefícios para a modelagem conceitual, aumentando a produtividade e qualidade da especificação de requisitos, reduzindo custos e facilitando a comunicação entre especialistas e projetistas [Silva et al 2008].

Ao fazer uso de Padrões de Análise para reutilizar esquemas de dados conceituais, surgem alguns problemas. Dentre os mais relevantes, tem-se a falta de formalismo, que dificulta o entendimento do significado dos conceitos, podendo gerar falsas interpretações; inconsistência, pois diferentes padrões podem ter estruturas diferentes; e redundância, consequência da inconsistência, pois esta complica o entendimento dos padrões e dificulta o reuso. Com isso, é necessário fazer uso de algum instrumento que permita melhorar a qualidade da reutilização. Hamza (2004) propõem que padrões sejam agrupados (catalogados) e reutilizados por meio de ontologias, ampliando o conhecimento sobre o domínio.

Ontologias são usadas para que o projetista conheça a natureza dos objetos daquele domínio a ser modelado. É um conjunto de conceitos dispostos de forma hierarquizada, que descrevem um domínio de conhecimento específico, podendo ser utilizado para criar uma base de conhecimento [Guarino 1997].

Este artigo apresenta uma proposta de integração entre ontologias e padrões de análise, que foram agrupados em um catálogo de padrões incorporado à ferramenta ArgoCASEGEO [Lisboa et al, 2004], que dá suporte à modelagem conceitual de bancos de dados geográficos, com base no modelo UML-GeoFrame [Lisboa e Iochpe 2008]. O artigo está estruturado da seguinte forma: A Seção 2 apresenta uma breve definição dos conceitos de padrões de análise e ontologias. A Seção 3 descreve a arquitetura da ferramenta ArgoCASEGEO, enquanto a Seção 4 descreve com mais detalhes o catálogo de padrões de análise dessa ferramenta. A Seção 5 aborda como a integração entre padrões de análise e ontologias é feita, seguido pelas considerações finais na Seção 6.

2. Padrões de Análise e Ontologias

Raramente são inventadas soluções realmente novas. Sempre que os problemas se assemelham com outros encontrados antes, há a reutilização de partes das soluções, ou seja, antigas soluções tornam-se padrões [Gamma et al 1994]. Na área de Computação isto ocorre, normalmente, sempre que um segundo sistema de informação é modelado e projetado, ou quando ocorrem novas alterações em uma aplicação já em produção.

Um padrão de análise é um mecanismo de reutilização que permite aos projetistas menos experientes reutilizarem o conhecimento de outros especialistas, durante a fase de análise. Em outras palavras, “um padrão apresenta a essência de uma solução para um problema recorrente, em um contexto específico” [Gamma et al, 1994].

A especificação mínima de um padrão deve conter os seguintes itens: Problema (e motivação), Contexto (e aplicabilidade), Forças e Solução. Outros itens também podem aparecer, como: Sinônimos, Estrutura, Variantes, Exemplos, Usos Conhecidos, Padrões Relacionados, etc. Esses itens devem ser apresentados de maneira formal, de tal forma que os padrões possam ser transmitidos para uma comunidade de pessoas.

Para exemplificar, a Figura 1 ilustra o padrão de análise Malha Viária Urbana. Esse padrão pode ser utilizado no desenvolvimento de diversas aplicações de SIG, em aplicações voltadas para o contexto de gestão urbana.

Segundo Guarino (1997), “Uma ontologia é uma especificação formal e explícita de um conhecimento compartilhado”. Uma ontologia pode ser usada para diversos propósitos, entre eles: conclusão e correção de informação insuficiente; fornecer um

conjunto fechado (controlado) de termos; eliminar ambigüidade do sentido das palavras; e para validação e verificação de consistência.

Em SIG, ontologias ajudam a entender como diferentes comunidades compartilham informações, auxiliam na descoberta de distorções presentes nos processos cognitivos de apreensão do mundo geográfico e ainda fornecem padrões para o desenvolvimento de SIG, que é o motivo que mais se enquadra no projeto de integração de ontologias e padrões de análise em questão [Fonseca et al 2000].

As ontologias podem ter algumas classificações, como: Ontologias de Aplicação, de Tarefas, de Alto Nível e de Domínio [Guarino 1997]. Este trabalho foca nas ontologias de domínio, as quais descrevem um vocabulário limitado a um domínio.

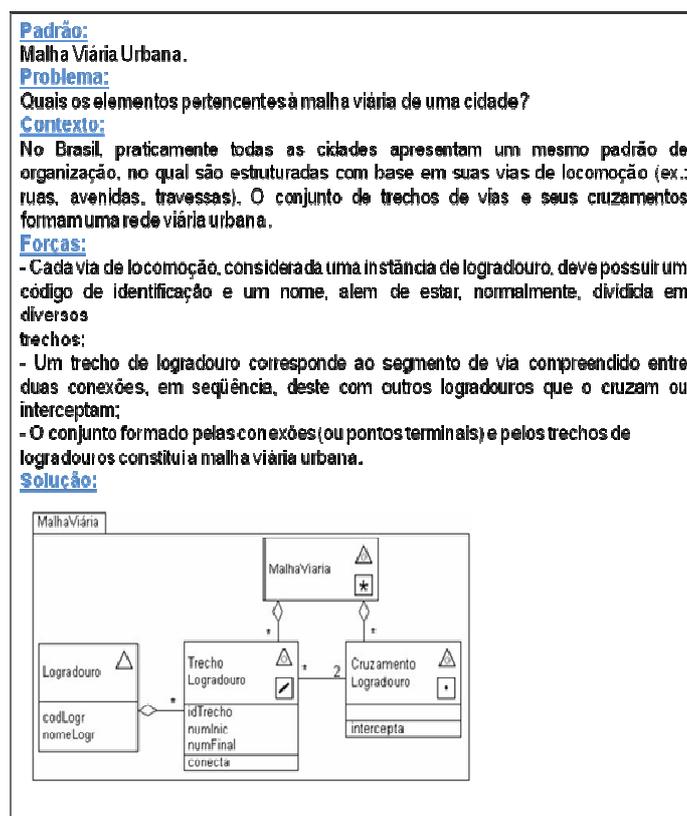


Figura 1. Padrão Malha Viária Urbana

3. Ferramenta ArgoCASEGEO

A ferramenta CASE ArgoCASEGEO [Lisboa et al 2004] tem como principal função dar suporte à modelagem de bancos de dados geográficos, com base no modelo UML-GeoFrame. A ferramenta, que é uma extensão do software ArgoUML [TIGRIS 2005], está sendo desenvolvida no Departamento de Informática da Universidade Federal de Viçosa, tem código fonte aberto e é escrita em linguagem Java.

Os esquemas de dados criados usando a ArgoCASEGEO são armazenados em formato XMI, uma sintaxe própria para armazenamento de esquemas conceituais de dados em documentos XML. A arquitetura da ArgoCASEGEO dispõe de cinco módulos (Figura 2).

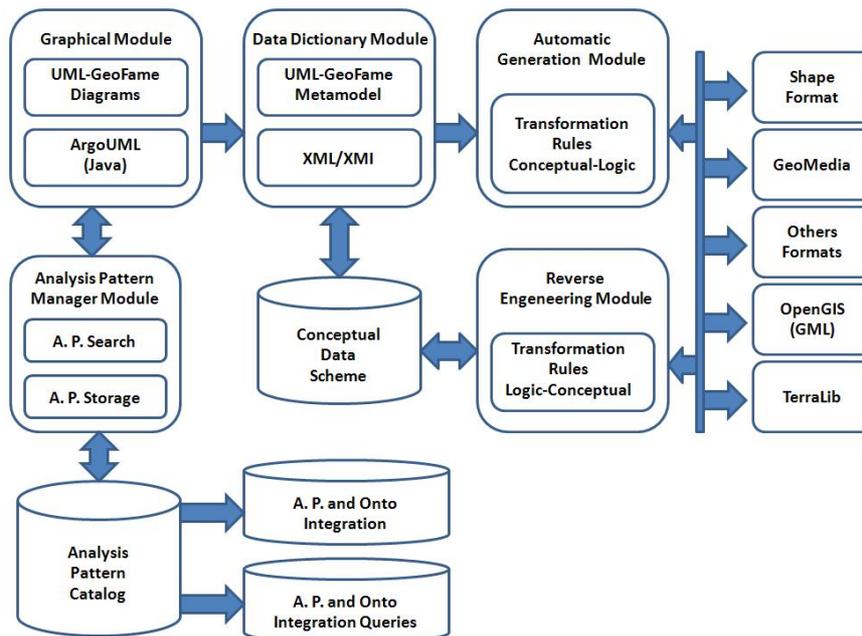


Figura 2. Arquitetura da Ferramenta ArgoCASEGEO (versão 2.0)

O Módulo Gráfico permite o desenho do esquema conceitual, provendo um conjunto de construtores do modelo UML-GeoFrame. O Módulo Dicionário de Dados armazena a descrição dos elementos do diagrama criado pelo projetista. O Módulo de Geração Automática permite a geração de esquemas lógicos usados em SIG comerciais, a partir do esquema conceitual armazenado no dicionário de dados. O Módulo de Engenharia Reversa (ainda não implementado) permitirá ao projetista obter os esquemas conceituais a partir de aplicações SIG existentes. O Catálogo de Padrões de Análise e seu gerenciador oferecem apoio ao projetista do banco de dados a encontrar soluções já usadas e testadas anteriormente, em outras aplicações similares, melhorando a qualidade da versão final do banco de dados. Por fim, foram desenvolvidos dois novos submódulos do Catálogo de Padrões de Análise: o módulo de integração de padrões de análise e o módulo de consultas de integrações, que estão abordados na Seção 5.

4. O Catálogo de Padrões de Análise

Com um catálogo de padrões de análise na ArgoCASEGEO [Sodré 2005], o projetista passa a dispor de soluções já utilizadas em aplicações de SIG similares, permitindo uma melhora na qualidade do banco de dados e uma redução dos custos de projeto.

O módulo do catálogo de padrões de análise permite: criar, abrir, editar e utilizar um padrão, assim como realizar pesquisas dentro do catálogo. Qualquer diagrama conceitual de dados criado pode ser salvo como um padrão de análise. O projetista deve, então, fornecer a descrição do padrão de análise, e adicioná-lo ao catálogo de padrões de análise da ferramenta. A Figura 3 ilustra o ambiente de trabalho da ArgoCASEGEO com o módulo de catalogação de padrões de análise.

Na tela da criação do padrão, é possível visualizar a solução para o problema, modelada através de um diagrama de classes, e preencher os campos que descrevem o padrão. Os campos *Name*, *Problem*, *Context* e *Forces* constituem as informações

mínimas que um padrão deve possuir para ser catalogado, e estão em destaque por serem de preenchimento obrigatório.

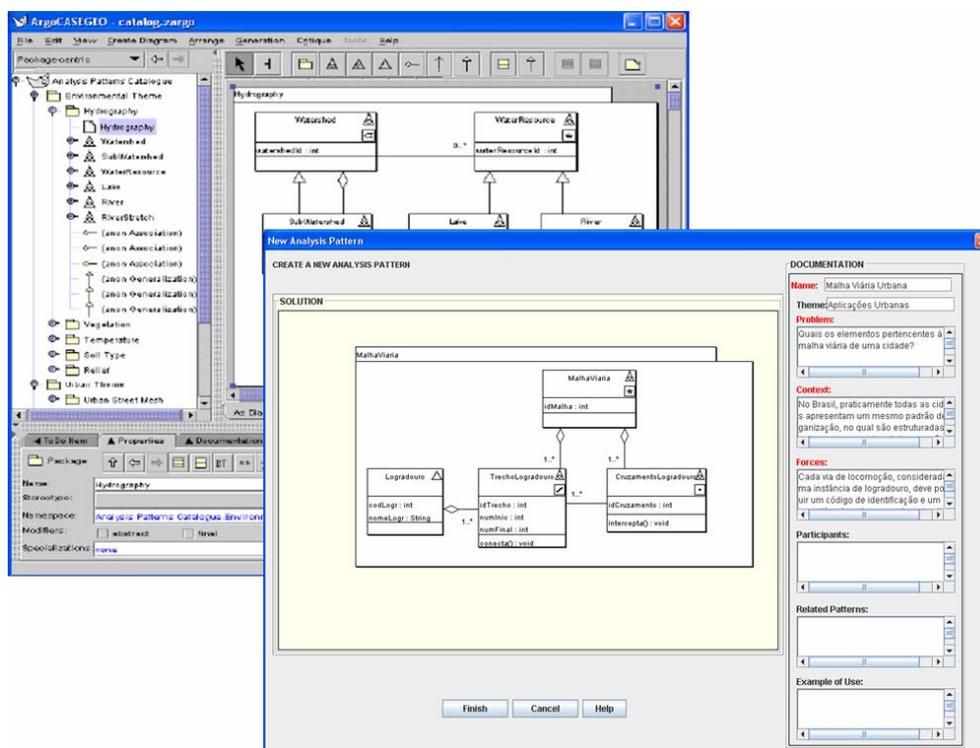


Figura 3. Ferramenta ArgoCASEGEO e Criação de um Padrão de Análise

Após criar o padrão, a ferramenta o armazena em um diretório específico do sistema. Os arquivos contendo os padrões são armazenados numa pasta, dentro da pasta do catálogo, juntamente com sua documentação, que é armazenada em formato XML, com o mesmo nome do padrão.

5. Padrões de Análise e Ontologias e Domínio como Mecanismos de Reuso

Conforme descrito anteriormente, o uso de padrões de análise contribui para a reutilização de modelos conceituais já consolidados, agilizando o processo de levantamento de requisitos. As ontologias, por sua vez, são úteis para melhorar a comunicação, evitando desacordos com relação a definições e restrições dentro de certo domínio, facilitando a captura, representação e reuso do conhecimento.

Padrões de análise, no entanto, ainda apresentam alguns problemas, como falta de formalismo, inconsistência e redundância [Hamza2004]. Por sua vez, ontologias de domínio podem estar integradas a padrões de análise, visando melhorar a qualidade de reuso desta classe de padrões. Desta forma, os projetistas de sistemas de informação poderiam, através dessa integração, consultar os conceitos de um domínio e recuperar os padrões relacionados, permitindo a reutilização do conhecimento de forma mais ampla.

5.1 Metodologia PatternOnto

Para resolver os problemas decorrentes do uso isolado de padrões de análise, Silva et al (2008) propõe um método que permite integrar um conjunto de padrões a uma ontologia

de domínio, com base em uma revisão dos padrões usando a técnica VERONTO [Villela 2004]. O método compreende três fases:

- Adequação: adequar os padrões por meio de meta-propriedades ontológicas, visando obter uma única estrutura para o padrão, definindo o seu domínio de reutilização por meio da técnica VERONTO;
- Catalogação: catalogar os padrões para serem recuperados por meio de um repositório compartilhado por projetistas;
- Associação: analisar a estrutura do padrão de análise e da ontologia, relacionando as classes do padrão e da ontologia que tenham o mesmo significado dentro do domínio;

5.2 Módulo de integração

A terceira fase da metodologia diz respeito à associação entre as classes do padrão de análise e os conceitos descritos em uma ontologia. Para fazer a associação, foi desenvolvido um módulo para a ferramenta ArgoCASEGEO, permitindo ao projetista fazer a integração desejada de forma amigável, como ilustrado na Figura 4.

Através da interface do módulo, o projetista deve definir os detalhes da integração. Primeiramente, ele informa o diretório em que se encontra o catálogo dos padrões de análise. O sistema então gera uma lista com os padrões encontrados na pasta do catálogo. Após a escolha de um dos padrões listados (em formato XMI), deve informar qual ontologia (em formato OWL) deseja integrar ao padrão escolhido.

Definidos os dois elementos fundamentais da integração, uma árvore é montada com a hierarquia da ontologia (*ontology classes*) e outra com a organização das classes do padrão (*pattern classes*). Além das árvores, o projetista dispõe de um campo (*description*) que exibe a descrição do elemento selecionado na ontologia.

Finalmente, para realizar a integração deve-se selecionar a classe do padrão e a entidade da ontologia que compartilham o mesmo significado dentro do domínio em questão. Assim que forem selecionadas as classes a serem integradas, uma tabela (parte inferior da janela) vai sendo construída, dispondo das classes do padrão na coluna da esquerda, e as respectivas classes da ontologia na coluna da direita.

Após a confirmação será gerado um arquivo de integração (em formato XML), que poderá ser utilizado em futuras consultas, recuperando os padrões relacionados com a ontologia desejada.

5.3 Módulo de consultas

Durante o processo de modelagem de um novo banco de dados, usando a ferramenta ArgoCASEGEO, o projetista pode consultar o catálogo de padrões de análise visando reutilizar soluções já validadas em outras aplicações. Conceitos que podem não fazer parte do vocabulário diário do projetista podem, então, serem revisados a partir de suas definições pertencentes à ontologia, tudo dentro de um único ambiente.

Para isto, foi desenvolvido um módulo para consultar os padrões integrados a uma ontologia. A interface do módulo pode ser visto na Figura 4. O usuário deve preencher alguns campos da interface para poder realizar a consulta. Primeiramente, deve selecionar o diretório em que se encontram os arquivos de integração. Em seguida, define a ontologia integrada aos padrões de análise.

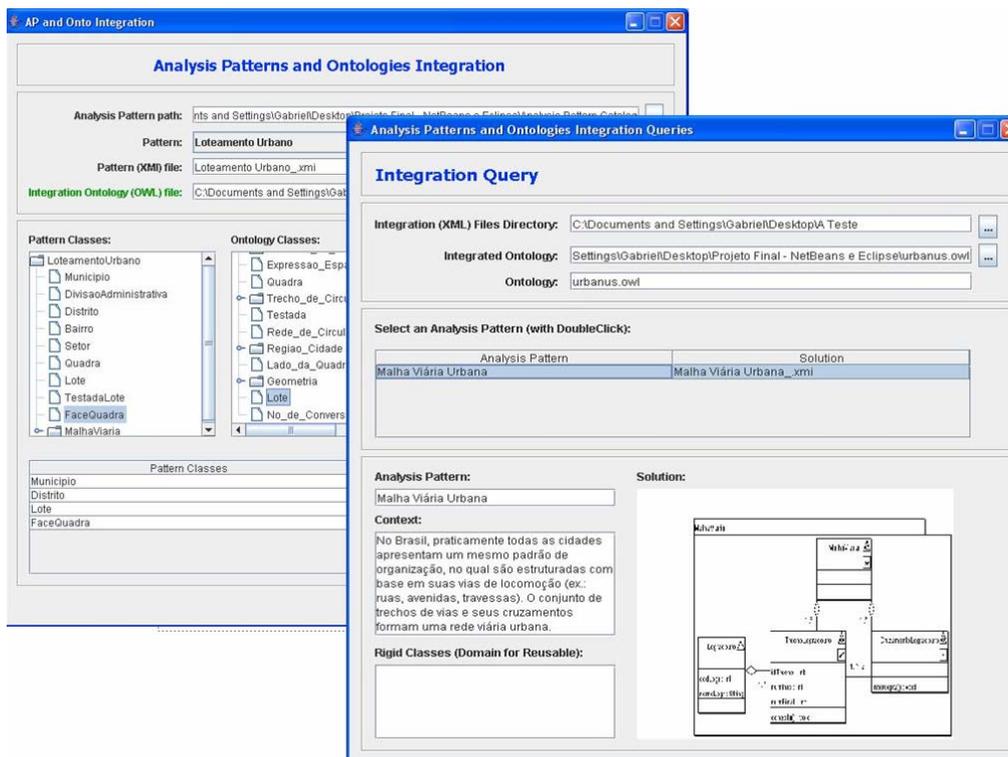


Figura 4. Módulos de Integração e de Consulta

Com isso, o programa se encarrega de fazer uma leitura em todos os arquivos de integração que estejam dentro da pasta escolhida e exibir, em uma tabela, todos os padrões que foram integrados com aquela ontologia. Ao selecionar um desses padrões, são exibidos dados como: nome do padrão de análise, contexto ao qual o padrão se aplica, a solução apresentada como diagrama de classes e as suas classes rígidas.

Para a implementação do módulo de integração de padrões de análise com ontologias de domínio foi utilizado o *framework* Jena, escrito em Java e desenvolvido pela HP (Hewlett-Packard). Jena é destinado ao desenvolvimento de aplicações para a Web semântica. A primeira versão da API Jena possui métodos para fazer a manipulação de ontologias em DAML+OIL, enquanto que na segunda versão (utilizada nesse projeto) foram desenvolvidos métodos que permitem a manipulação em ontologias RDFS e OWL. Para manipular ontologias OWL, a segunda versão da API Jena oferece um pacote específico denominado API Jena 2 Ontology. A API dispõe, nesse pacote, de duas classes específicas para realizar essa tarefa: a *OntClass* e a *Object Property*. Apesar de o *framework* ter várias capacidades, ele foi utilizado apenas para fazer a leitura do arquivo da ontologia.

Todo projeto utilizado pela ferramenta ArgoCASEGEO é composto por um arquivo compactado que contém: a estrutura lógica do esquema conceitual (arquivo XML/XMI), responsável pela descrição das classes, atributos, relacionamentos etc; e a estrutura gráfica do esquema conceitual (arquivo XML/PGML), responsável pelas informações de localização dos elementos, figuras, tamanho dos elementos, etc.

6. Considerações Finais

Este artigo apresentou o módulo de integração e consulta a padrões de análise e ontologias de domínio, da ferramenta ArgoCASEGEO. A abordagem de padrões de

análise apresenta grande potencial para melhorar a qualidade das aplicações, reduzindo o tempo e custos das etapas de análise de requisitos e modelagem conceitual do banco de dados. Um padrão de análise não necessita apresentar uma solução original. Pelo contrário, padrões devem documentar soluções já testadas e validadas, pois são soluções para problemas recorrentes.

Integrar padrões a ontologias pode facilitar a comunicação e melhorar a reutilização e organização do conhecimento, contribuindo para uma melhor modelagem conceitual do sistema de informação e melhorando a qualidade do reuso dos padrões. A incorporação de um catálogo de padrões de análise a uma ferramenta CASE pode proporcionar ganho de produtividade, registrando definições do contexto aos quais os padrões estão inseridos e permitindo seu compartilhamento entre outros projetistas. A ferramenta ArgoCASEGEO está disponível para *download* no sítio <http://www.dpi.ufv.br/projetos/argocasegeo>.

Agradecimentos

Projeto parcialmente financiado com recursos da FAPEMIG e do CNPq/MCT/CT-Info.

Referências Bibliográficas

- Fonseca, F., Egenhofer, M. and Borges, K. A. V. (2000) “Ontologias e Interoperabilidade Semântica entre SIGs”. In: Workshop Brasileiro em Geoinformática. São Paulo.
- Fowler, M. (1997) Analysis Patterns: Reusable Object Models. Menlo Park, CA: Addison Wesley Longman.
- Gamma, E. et al. (1994) Design Patterns: Elements of Elements of Reusable Object-Oriented Software. Addison Wesley.
- Gruber, T. R. (1995) “Toward Principles for the Design of Ontologies Used for Knowledge Sharing”. Int. Journal of Human-Computer Studies, v. 43, p. 907-928.
- Guarino, N. (1997). “Semantic Matching Formal Ontological Distinctions for Information Organizatios, Extraction and Integration”. International Summer School, SCIE-97, Frascati, Italy.
- Hamza, Haitam (2004) “Improving Analysis Patterns Reuse: An Ontological Approach”. Ontologies as Software Engineering Artifacts Workshop, OOPSLA’04.
- Lisboa Filho, J., Sodr , V.F., Daltio, J., Rodrigues, M.F., Vilela, V. (2004) A CASE tool for geographic database design supporting analysis patterns. Proc. Conceptual Modeling for GIS, ER2004. LNCS 3289.
- Lisboa F., J. e Iochpe, C. (2008) “Modeling with a UML profile”. In: Shashi Shekhar and Hui Xiong. Encyclopedia of GIS. Germany: Springer-Verlag.
- Silva, E. O., Lisboa F., J., Oliveira, A. P., Gonalves, G. S. (2008) “Improving analysis patterns in the geographic domain using ontological meta-properties”. 10th Int. Conf. on Enterprise Information Systems. Barcelona, Spain
- Sodr , V. F., Lisboa F., J., Vilela, V., Andrade, M. V. A. (2005) “Improving productivity and quality of GIS databases design using an analysis pattern catalog”. In: Asia-Pacific Conf. on Conceptual Modelling, Newcastle, Australia.
- TIGRIS: ArgoUML Project Home (2005), <http://argouml.tigris.org/>, Agosto.
- Villela, M. L. B., Oliveira, A. P. e Braga, J. L. (2004) “Modelagem ontol gica no apoio   modelagem conceitual”. XVIII Simp sio Brasileiro de Engenharia de Software, Bras lia.