

# PUC

ISSN 0103-9741

Monografias em Ciência da Computação  
nº 04/07

## **Bancos de Dados Geográficos Inteligentes**

**Alexandre Gazola  
Antônio L. Furtado**

Departamento de Informática

**PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO DE JANEIRO  
RUA MARQUÊS DE SÃO VICENTE, 225 - CEP 22453-900  
RIO DE JANEIRO - BRASIL**

## **Bancos de Dados Geográficos Inteligentes**

Alexandre Gazola and Antônio L. Furtado

{agazola, furtado}@inf.puc-rio.br

**Abstract.** This paper presents an overview of the application of Intelligent Systems techniques to Geographic Information Systems (GIS) and their Geographic Databases. These techniques seem quite promising for the evolution of the GIS technology. In particular, this work brings reports and analyses of a few experiments with Ontologies, Rule-based systems and Data-mining in Geographic Information Systems.

**Keywords:** Geographic Information System, Geographic Databases, Intelligent Systems, Rules, Data-mining.

**Resumo.** Este artigo apresenta uma visão geral sobre a aplicação de técnicas de Sistemas Inteligentes a Sistemas de Informação Geográfica (SIG) e seus Bancos de Dados Geográficos. Essas técnicas se mostram bastante promissoras para a evolução da tecnologia de SIG. Em particular, este trabalho traz relatos e análises de alguns experimentos envolvendo Ontologias, Sistemas baseados em regras e Mineração de dados em Sistemas de Informação Geográfica.

**Palavras-chave:** Sistema de Informação Geográfica, Banco de Dados Geográficos, Sistemas Inteligentes, Ontologias, Regras, Mineração de Dados.

**Responsável por publicações:**

Rosane Teles Lins Castilho  
Assessoria de Biblioteca, Documentação e Informação  
PUC-Rio Departamento de Informática  
Rua Marquês de São Vicente, 225 - Gávea  
22453-900 Rio de Janeiro RJ Brasil  
Tel. +55 21 3114-1516 Fax: +55 21 3114-1530  
E-mail: [bib-di@inf.puc-rio.br](mailto:bib-di@inf.puc-rio.br)  
Web site: <http://bib-di.inf.puc-rio.br/techreports/>



## 1 Introdução

A tecnologia de Sistemas de Informação Geográfica (SIG) provê ferramentas de software bastante poderosas para realizar o tratamento de dados geográficos. A principal diferença de um SIG para um sistema de informação convencional é sua capacidade de armazenar tanto os atributos descritivos como as geometrias dos diferentes tipos de dados geográficos [CASANOVA et al., 2005].

Assim como os sistemas de informação convencionais dependem de um banco de dados, os SIG possuem como componente fundamental de sua arquitetura um banco de dados geográfico (BGD). No entanto, os dados geográficos manipulados pelos SIG são bastante complexos para representação e manipulação em bancos de dados comuns. Por isso, desenvolvimentos adicionais estão sendo empreendidos na tecnologia de SGBD-OR para permitir a adoção de bancos de dados geográficos.

Outro importante aspecto a ser considerado são as diferenças fundamentais entre os sistemas de bancos de dados tradicionais e os sistemas inteligentes. Um banco de dados tradicional construído sob o modelo relacional pode apenas representar fatos. Kannan and Geetha [1998] apontam características da área de gestão do conhecimento das quais bancos de dados tradicionais poderiam se beneficiar. São elas: facilidades para representação do conhecimento, raciocínio dedutivo, busca com backtracking, estruturas de controle para raciocínio indutivo, refinamento e validação de conhecimento e classificação automática de conhecimento. Bertino et. al. [2001] descrevem um banco de dados inteligente como um sistema caracterizado pela existência de um grande banco de dados com milhões de fatos persistentes associados com uma extensa base de regras que codifica o conhecimento do domínio em questão.

Os SIG possuem capacidades muito limitadas para manipulação de informações sobre análise espacial, bem como o tratamento das informações resultantes dessa análise. Além disso, existe uma grande necessidade de intercâmbio de informações geográficas presentes em SIGs, isto é, interoperabilidade semântica. Com isso, abordagens baseadas em sistemas inteligentes no domínio de SIGs se mostram bastante promissoras.

O objetivo deste artigo é apresentar um panorama sobre os bancos de dados geográficos e SIGs no contexto de sistemas inteligentes de bancos de dados. Questões relacionadas ao emprego de ontologias, sistemas baseados em regras e mineração de dados são discutidos no contexto de dados geográficos.

O restante deste trabalho está organizado da seguinte forma: a seção 2 resume algumas características básicas dos SIGs e seus bancos de dados geográficos. A seção 3 explora algumas abordagens inteligentes no âmbito de sistemas georreferenciados. Mais especificamente, a subseção 3.1 apresenta conceitos sobre representação e recuperação de conhecimento por meio de ontologias do ponto de vista de SIG, com ênfase em questões relacionadas à interoperabilidade semântica. A subseção 3.2 traz uma visão rápida do emprego de técnicas baseadas em regras para o processamento do conhecimento em bancos de dados geográficos. A subseção 3.3 relata uma abordagem de mineração de dados para descoberta e análise de informações não triviais e regras de associação em dados geográficos. Por fim, conclusões e considerações finais são apresentadas na seção 4.

## 2 Bancos de Dados Geográficos

Dados geográficos são aqueles que possuem uma dimensão espacial, ou uma localização, diretamente ligada ao mundo geográfico real como as imagens de satélites de sensoriamento remoto, os dados de inventários cadastrais, os dados ambientais coletados em campo e os modelos numéricos de terreno [Vinhas, 2006]. Bancos de dados geográficos (BDG) são coleções de dados georreferenciados, manipulados por Sistemas de Informação Geográficas (SIG). Os SIG são sistemas computacionais capazes de capturar, modelar, armazenar, recuperar, manipular, analisar e apresentar dados geográficos [Worboys and Duckham, 2004].

Considerando-se modelos formais para representação de dados geográficos, distinguem-se dois tipos principais de representação: o modelo de *geo-campos* e o modelo de *geo-objetos* [Goodchild, 1992]. O modelo de *geo-campos* enxerga o espaço geográfico como uma superfície contínua, sobre a qual variam os fenômenos a serem observados. Como exemplo, podem ser citados os mapas de vegetação e temperatura de uma determinada região geográfica. O modelo de *geo-objetos*, por sua vez, representa o espaço geográfico como uma coleção de entidades individualizadas distintas, onde cada entidade é definida por uma fronteira fechada.

Geo-campos e geo-objetos são mapeados para estruturas de dados de duas naturezas: vetorial e matricial (ou raster). A Figura 1, extraída de Vinhas [2006], exemplifica estruturas matriciais e vetoriais para representar um conjunto de geo-objetos: duas construções e um lago próximo a uma estrada. A Figura 2 mostra estruturas vetoriais e matriciais para representar um *geo-campo*: a altimetria de uma região.

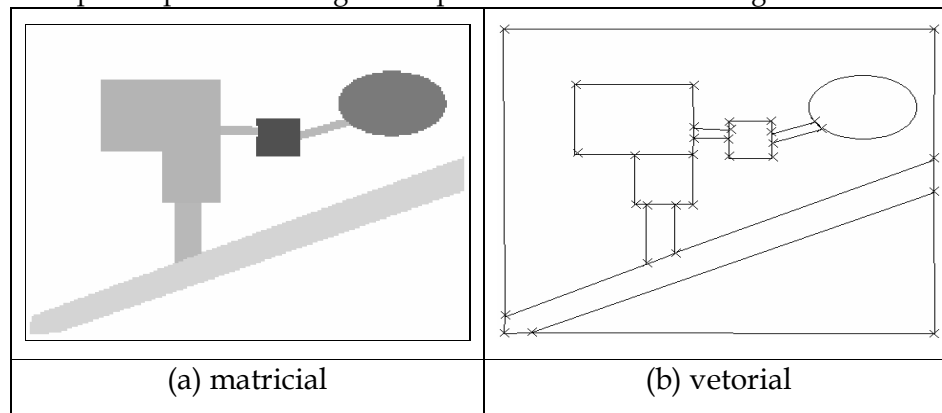


Figura 1. Representação de geo-objetos: (a) matricial e (b) vetorial. [Vinhas, 2006]

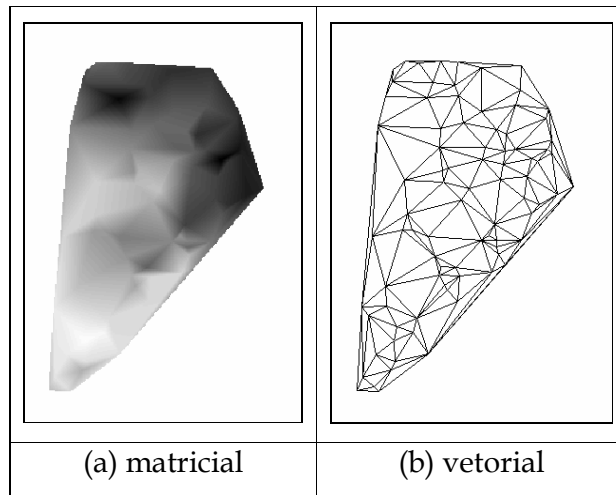


Figura 2. Representação de geo-campos: (a) matricial e (b) vetorial. [Vinhas, 2006]

Os SIGs freqüentemente se especializam quanto ao tratamento de uma ou de outra forma de representação possível. Mesmo assim, existem alguns SIG híbridos, que são sistemas capazes de tratar dados representados tanto sob a forma matricial quanto sob a forma vetorial.

Os bancos de dados geográficos utilizados pelo SIG possuem dados usualmente agrupados em duas componentes: a componente espacial (geográfica) e a componente convencional. A componente espacial refere-se ao valor de um atributo espacial de um objeto, ou seja, está diretamente relacionada à natureza geográfica do mesmo, podendo ser do tipo geo-objeto ou geo-campo, conforme explicado anteriormente. A componente convencional, ou descritiva, refere-se aos valores alfa-numéricos tradicionalmente manipulados pelos SGBDs relacionais comuns.

A componente espacial de um SIG representada sob a forma de geo-objetos normalmente se especializa nos seguintes tipos elementares da geometria plana:

- Ponto;
- Linha;
- Polígono

A componente espacial de um SIG representada sob a forma de geo-campos são usualmente classificadas nos seguintes tipos [Vinhas, 2006]:

- grades regulares (ou raster): o geo-campo é representado por um conjunto de células retangulares onde um único valor é atribuído a cada célula, representando o valor do geo-campo na extensão da célula;
- pontos amostrais: o geo-campo é representado apenas em localizações pontuais conhecidas;
- isolinhas: o geo-campo é representado por linhas ao longo das quais o valor do geo-campo é constante;
- subdivisões planares: o geo-campo é normalmente representado por um conjunto de áreas que não se interceptam e que recobrem todo o domínio do geo-campo.

- malhas triangulares (ou TINs): representam o geo-campo por um conjunto de triângulos que não se sobrepõem e cobrem totalmente a área do geo-campo.

Existem basicamente duas principais formas de integração entre os SIGs e os SGBDs, que são chamadas de *arquitetura dual* e *arquitetura integrada*, exibidas na Figura 3 [Casanova et. al., 2005].

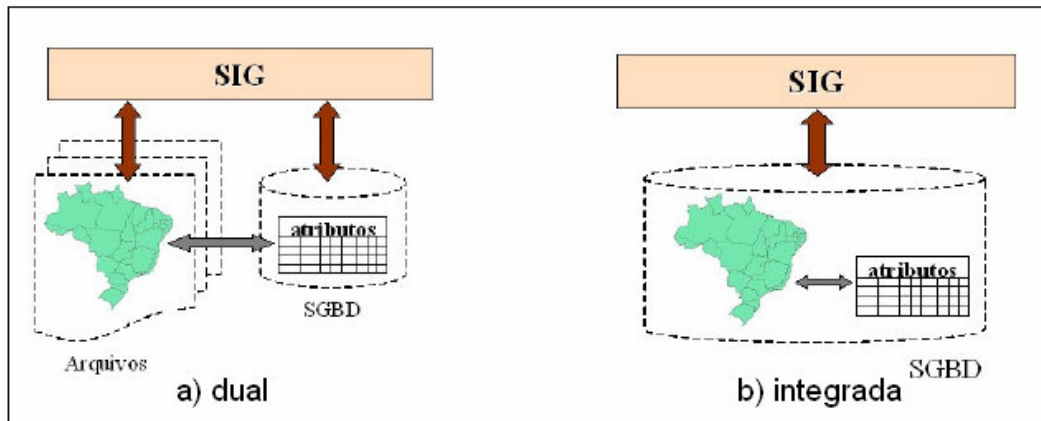


Figura 3. Integração entre SIGs e SGBDs: (a) Arquitetura Dual e (b) Arquitetura Integrada. [Casanova et. al., 2005]

A arquitetura dual armazena as componentes espaciais separadamente. A componente convencional é armazenada em um SGBD relacional e a componente espacial é armazenada em arquivos com formato proprietário. Essa arquitetura apresenta alguns problemas:

- Dificuldade para manipulação das componentes espaciais, principalmente em manter a integridade entre a componente espacial e a componente convencional;
- Dificuldade de interoperabilidade, pois cada sistema trabalha com arquivos com formato proprietário. Alguns formatos proprietários largamente utilizados no mercado incluem o Shapefile, o GeoTIFF e o GeoMedia. O problema da interoperabilidade de dados geográficos é discutido na próxima seção.

Na arquitetura integrada todos os dados são armazenados em um SGBD, tanto a componente espacial quanto a componente convencional. As principais vantagens dessa abordagem são a utilização dos recursos dos SGBDs, os quais já estão bem amadurecidos. Entre esses recursos podem ser citados gerência de transações, controle de integridade, concorrência e linguagens próprias de consulta [Casanova et. al., 2005].

A arquitetura integrada pode ser ainda subdividida em três outras, dependendo de como é armazenada a componente espacial dos objetos: (1) *baseada em campos longos*, (2) *em extensões espaciais* e (3) *combinada*. A primeira delas utiliza BLOBs para armazenar a componente espacial dos objetos, a segunda utiliza extensões espaciais oferecidas por alguns fabricantes de SGBDs (ex.: a extensão Oracle Spatial para o Oracle) para armazenar a componente espacial dos objetos; e a última forma combina as duas formas anteriores, armazenando geometrias vetoriais utilizando extensões espaciais e armazena geometrias matriciais em BLOBs. Um exemplo que utiliza a arquitetura integrada é a biblioteca TerraLib [Câmara et. al., 2000], a qual é uma biblioteca de classes e funções para o desenvolvimento de aplicativos geográficos.



### **3 Técnicas Inteligentes para Sistemas de Informação Geográfica**

A tecnologia de SIG precisa estar constantemente evoluindo para atender às necessidades de várias áreas como negócios, indústria, governo e educação. Como resultado desse processo, o nível de exigência das aplicações de SIG está se tornando bastante elevado, fazendo com que o desenvolvimento desses sistemas fique difícil de ser realizado apenas com uma linguagem de programação convencional. A nova geração de sistemas de SIG começa a empregar técnicas inteligentes, como ontologias, sistemas baseados em regras, mineração de dados, sistemas multi-agentes e outras tecnologias da inteligência artificial.

As próximas subseções apresentam uma visão geral de técnicas inteligentes baseadas em ontologias, sistemas baseados em regras e mineração de dados em sistemas georreferenciados.

#### **3.1 Interoperabilidade, Dados Geográficos e Ontologias**

##### **3.1.1 Interoperabilidade de Dados Geográficos**

A necessidade de troca de informações geográficas é notória [OpenGIS, 1996]. O grande volume de informações geográficas gerado por satélites e espalhado pela Internet torna bastante complexa a tarefa dos sistemas de informação geográfica, que estão bastante atrasados com relação à interoperabilidade. Heterogeneidade em SIG associada à complexidade e riqueza dos dados geográficos e a dificuldade de sua representação em sistemas de computação trazem problemas específicos para interoperabilidade em SIG [Fonseca and Egenhofer, 1999].

Algumas das primeiras tentativas de se obter interoperabilidade em SIG foram através da conversão direta do formato de um fabricante de software para outro. Outra variação foi o aparecimento de padrões intermediários. No Brasil, o formato IIG é utilizado para transferência de dados em alguns órgãos públicos [Davis, 1995].

Uma outra iniciativa proeminente para o intercâmbio de dados geográficos está sendo empreendida pelo consórcio OGC, o qual é formado por um amplo conjunto de companhias, agências governamentais e universidades com o propósito de criar e promover o desenvolvimento de tecnologias que facilitem a interoperabilidade entre sistemas que processam dados geográficos [Gardels, 1996; Lieberman, 2003]. O OGC publica seus resultados como especificações e padrões para intercâmbio de dados.

Apesar de existirem iniciativas como a do OGC, apenas o uso de padrões para resolver o problema da interoperabilidade de dados geográficos não é suficiente, visto que a heterogeneidade aparece naturalmente num universo livre de regulamentações oficiais [Elmsagarmid and Pu, 1990]. A utilização de tradutores semânticos é uma abordagem mais promissora do que as abordagens atuais baseadas em padrões [Bishr, 1997].

##### **3.1.2 Ontologias**

Uma ontologia é uma conceitualização formal, explícita e compartilhada de alguma área do conhecimento, também chamado de domínio de discurso [Uschold and Grüninger, 1996]. Em outras palavras, uma ontologia consiste de conceitos e relações, com propriedades e restrições descritas sob a forma de axiomas.

É útil distinguir a diferença entre ontologias e base de conhecimento. Uma ontologia modela os conceitos do domínio considerado, escolhendo termos adequados para descrever esse domínio apropriadamente. Em contraste, uma base de conhecimento usa os termos definidos na ontologia para descrever uma determinada realidade.

A figura 4, extraída de [Breitman, 2005], exibe um exemplo de uma ontologia elaborada para modelar o domínio de circuitos eletrônicos sendo comparada com uma base de conhecimento que utiliza essa ontologia.

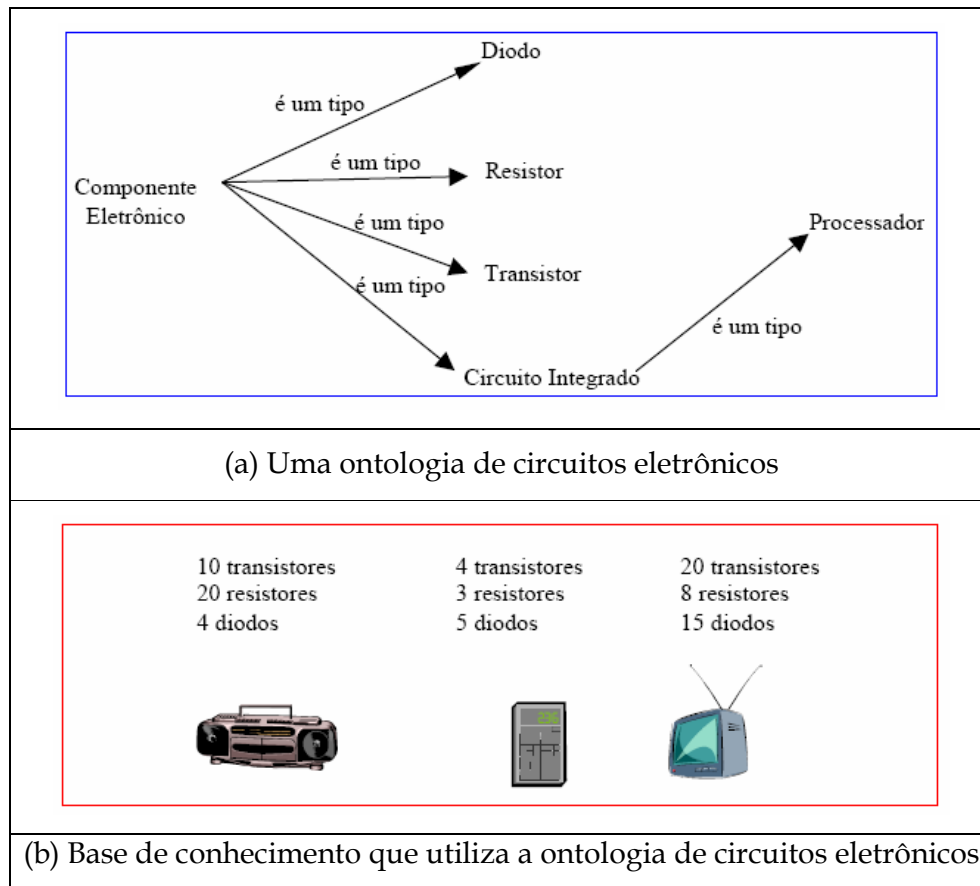


Figura 4. Comparação entre ontologia e base de conhecimento [Breitman, 2005]

Existem vários benefícios advindos da utilização de ontologias em sistemas de computação. A seguir, são enumerados alguns desses benefícios [Breitman, 2005]:

- Ontologias fornecem um vocabulário comum e formal para a representação do conhecimento, ou seja, interpretações ambíguas são evitadas;
- Ontologias permitem o compartilhamento do conhecimento. Pessoas que desenvolvam ontologias para um determinado domínio de aplicação podem disponibilizar suas ontologias para que estas sejam reaproveitadas por outras pessoas que estejam desenvolvendo aplicações no mesmo domínio.
- Em muitos casos, é possível estender o uso de uma ontologia genérica de maneira a adequá-la a um domínio específico. Isso traz mais flexibilidade para a construção de novas ontologias.

Existem várias classificações distintas de ontologias, fornecidas por diversos autores. Maedche [2002] classifica as ontologias quanto ao nível de generalidade necessária. Sua classificação compreende os seguintes tipos:

- Ontologias de alto-nível – Modelam conceitos bastante gerais tais como tempo, espaço, etc. Esses conceitos são geralmente independentes de um domínio ou problema particular.
- Ontologias de domínio – Modelam o vocabulário relacionado a um domínio genérico, por meio da especialização de conceitos apresentados nas ontologias de alto-nível. Alguns exemplos de ontologias que podem ser citados são ontologias de veículos, ontologias de objetos geográficos, ontologias de documentos, etc.
- Ontologias de tarefa – Modelam um vocabulário relacionado a uma tarefa ou atividade específica, por meio da especialização de conceitos apresentados nas ontologias de domínio.
- Ontologias de aplicação – São as ontologias mais específicas por serem utilizadas dentro das aplicações.

Por fim, existem diversas formas para representação de ontologias, sendo a forma de triplas RDF a mais utilizada, por ser uma recomendação do W3C. Em RDF, as ontologias são representadas em arquivos XML seguindo um modelo de dados composto de dois objetos básicos:

- Um recurso qualquer identificado por uma URI;
- Uma propriedade que define uma relação binária entre recursos

A seguir, é apresentada uma análise comparativa de alguns estudos realizados empregando técnicas baseada em ontologias para permitir a interoperabilidade semântica de sistemas de informação geográficas diversos.

### **3.1.3 SIG Baseados em Ontologias**

Aplicações de ontologias em SIG focam em problemas práticos de definição de um vocabulário comum para descrever o domínio geo-espacial que pode facilitar a interoperabilidade e diminuir problemas de integração de dados [Agarwal, 2005 and Fonseca et. al., 2002]. Na Web, já se fala da *Geospatial Semantic Web* [Egenhofer, 2002], que é uma tentativa de trazer semântica aos serviços de busca e integração de dados geo-espaciais.

Perry and Hakimpour [2006] descrevem uma abordagem baseada em ontologias para a análise de três aspectos bastante comuns no tratamento de dados geográficos: tema, espaço e tempo. A figura 5 mostra a ontologia apresentada nesse trabalho.

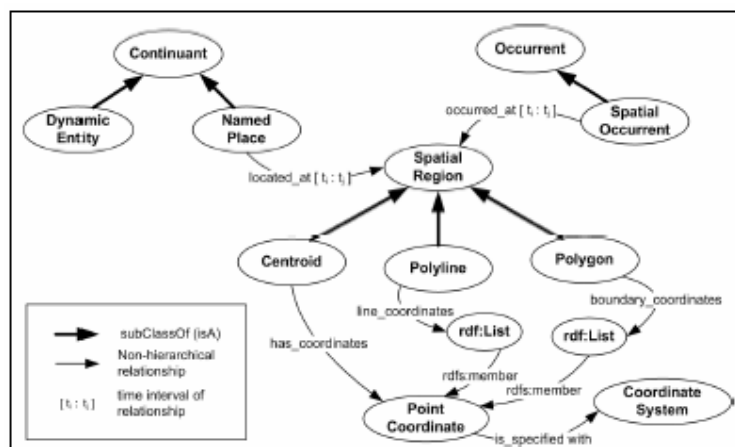


Figura 5. Ontologia para a modelagem de tema, espaço e tempo [Perry and Hakimpour, 2006]

Um tema diz respeito a um conjunto de objetos geográficos que compartilham algumas características em comum. Por exemplo, um tema de hidrografia do norte do país englobaria todos os objetos geográficos relacionados à informação de hidrografia e que tivessem a propriedade de estar no norte do país. Espaço está relacionado à posição relativa de um objeto na superfície da terra. Finalmente, tempo está diretamente associado a um período no qual propriedades de um determinado objeto podem ter sido alteradas.

A ontologia da figura 5 detalha e explicita os relacionamentos espaciais entre os vários tipos de dados geográficos. Além disso, os autores da ontologia construíram um conjunto de operadores de consulta simbólicos que levam em consideração esses relacionamentos espaço-temporais definidos pela ontologia. Com isso, objetiva-se construir um framework para um sistema de informação de busca e análise de relacionamentos espaciais. Pretende-se como próximo passo do trabalho utilizar essa modelagem desenvolvida para criar operações concretas de mais alto nível para os usuários utilizando o Oracle Spatial.

Como pôde ser visto, Perry and Hakimpour [2006] se preocupam fundamentalmente com a modelagem dos dados geográficos sob a tríade tema, espaço e tempo ao considerar o uso de ontologias. Nolan and Simon [2001], por sua vez, consideram não apenas a modelagem dos dados geográficos em ontologias, mas de todo o núcleo de um SIG. Ou seja, três componentes são considerados: os *dados* utilizados no processamento, as *consultas* e os *algoritmos* que processam os dados. A figura 6 ilustra essa visão.

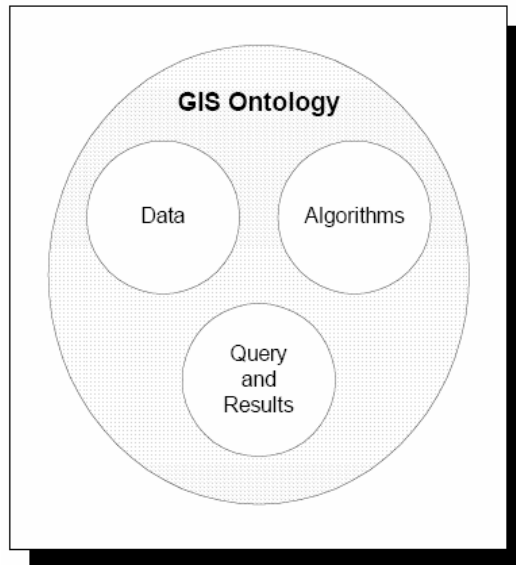


Figura 6. Uma ontologia para um SIG . [Nolan and Simon, 2001]

Devido ao fato de existirem muitos sistemas legados com dados disponíveis em vários formatos, com novos tipos de dados constantemente surgindo, é bastante difícil se chegar a um padrão comum, apesar de existirem muitos esforços nesse sentido. Por exemplo, a linguagem GML proposta pelo OGC é uma linguagem baseada em XML para a representação de dados geográficos [Cox et. al., 2001]. Nolan and Simon [2001] sugerem uma ontologia que decompõe dados e algoritmos em suas componentes descritivas, independentes dos formatos dos dados ou linguagem de programação. A abordagem consiste em construir essa ontologia e representá-la sob a forma de XML usando o modelo de dados RDF e permitir que ela seja utilizada por agentes de software encarregados de processar consultas definidas pelo usuário. Esse sistema foi implementado na linguagem Java, usando pacotes comerciais para a construção dos agentes, Jini como middleware e o modelo RDF codificado em XML para representar tanto a ontologia do núcleo do SIG quanto os conceitos pertencentes aos agentes. Perguntas do tipo “qual é a rota ótima entre a cidade A e a cidade B?” ou “Qual é a colheita de trigo esperada para este ano no estado X?” puderam ser tratadas satisfatoriamente pelo sistema.

Pode-se dizer que um sistema de informação é *baseado* em ontologias quando uma ontologia explícita tem uma função na construção e utilização do sistema. Esse é o caso do trabalho apresentado em Fonseca and Egenhofer [1999], o qual descreve um sistema de SIG que trata as ontologias como estruturas dinâmicas, orientadas a objetos, que o usuário pode folhear. Existe um servidor de ontologias que contém um catálogo padrão e está disponível para o usuário pesquisar e navegar. A idéia, então, é criar classes a partir de um mapeamento orientado a objetos de ontologias-padrão e extrair os dados geográficos (objetos) de múltiplas fontes, integrando-os por meio dessas ontologias. Essa abordagem provê um grande nível de interoperabilidade e permite integração parcial de informação quando a integração total não é possível.

Outro sistema de informação geográfica baseado em ontologias é o sistema *OntoCarta* [Venancio et. al., 2003], desenvolvido no laboratório LIS da Unicamp. Esse sistema permite a navegação em mapas dirigida por ontologias. Por exemplo, quando um usuário efetua uma operação de aproximação em um mapa (*zoom in*), freqüentemente ele perde a noção de contexto. Com uma ontologia de suporte, facilmente tem-se uma

relação explícita entre todos os objetos e conceitos, sendo possível a localização de um objeto em foco numa base de conhecimento.

Esta seção apresentou um relato de alguns experimentos bem sucedidos na utilização de ontologias no tratamento de dados geográficos, com destaque para a interoperabilidade de dados geográficos. Ontologias também têm se mostrado uma técnica promissora para a extração automática de dados geográficos [Buscaldi et. al., 2006].

A abordagem de ontologias pode ser aprimorada utilizando-se sistemas de regras para governar o fluxo de execução dos cenários comuns independentes de ações e decisões específicas. A seção seguinte apresenta algumas características dos sistemas baseados em regras e sua utilidade em sistemas de informação geográfica.

## **3.2 Sistemas Baseados em Regras e SIG**

### **3.2.1 Sistemas Baseados em Regras**

Um dos principais objetivos da pesquisa em Inteligência Artificial tem sido a aplicação de mecanismos gerais de inteligência para resolver problemas novos específicos [Riesbeck and Schank, 1989]. Um desses mecanismos consiste numa abordagem baseada em regras para se resolver problemas específicos de domínios específicos. Sistemas que trabalham assim são chamados de sistemas especialistas. Um princípio básico que os desenvolvedores de sistemas especialistas assumem é que o conhecimento de um especialista pode ser traduzido num conjunto de regras.

Comportamento baseado em regras parte do pressuposto que a mente humana consiste de um mecanismo de raciocínio que pode aplicar uma coleção de regras gerais para resolver qualquer problema [Lanter, 1992]. Ou seja, a resolução de problemas gerais é possibilitada através de um mecanismo de raciocínio que é independente de conhecimento específico do domínio.

Sistemas baseados em regras aplicam mecanismos de inferência gerais em conhecimento expresso sob regras de produção. O objetivo é aplicar conhecimento de forma que um nível alto de performance de conhecimento possa ser reproduzido.

Alguns dos benefícios identificados em se representar conhecimento sob a forma de regras de produção são [Lanter, 1992]:

- modularidade;
- uniformidade;
- naturalidade

Modularidade torna possível estruturar o conhecimento do domínio sendo estudado como uma série de regras de produção distintas coletadas conjuntamente para uma aplicação dentro de um determinado domínio. Uniformidade é evidenciada no formato padrão SE-ENTÃO que as regras de produção geralmente seguem. Naturalidade é obtida pelo fato de que os humanos que são especialistas em um assunto geralmente o expressam por meio de construções SE-ENTÃO.

### 3.2.2 SIG Baseados em Regras

Sistemas baseados em regras (ou especialistas) podem ser empregados com bastante eficácia em várias questões relacionadas ao processamento geral de dados geográficos, incluindo interpretação, planejamento, monitoramento, controle, e infra-estrutura.

Kalogirou [2002] apresenta uma abordagem de combinação das tecnologias de sistemas especialistas com sistemas de informação geográfica para auxiliar na implementação de um modelo de avaliação de terrenos. As regras consideradas modelam o cultivo geral e cinco culturas específicas (trigo, cevada, milho, algodão e beterraba). O sistema especialista foi construído para ajudar na avaliação de terrenos e permitir alteração em suas regras baseada em diferentes performances observadas em áreas locais.

O sistema Geolineus [Lanter, 1991] provê assistência inteligente por meio de seu motor de inferência baseado em regras. Ele torna possível o uso de representação de conhecimento como memória de trabalho ao aplicar conjuntos de regras de propagação cartográfica e restrições de integridade semântica. A figura 7 mostra o esquema desse sistema auxiliando uma aplicação de SIG.

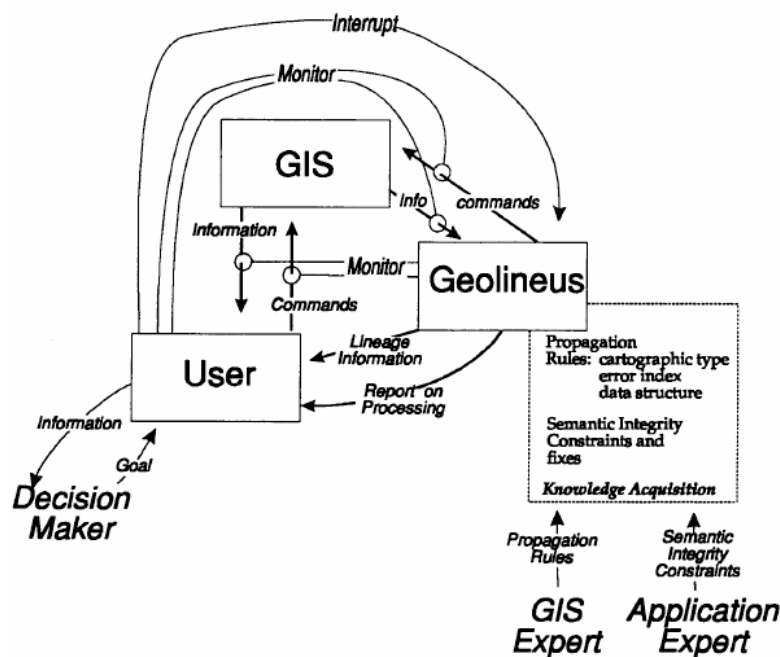


Figura 7. Funcionamento do sistema Geolineus [Lanter, 1991]

Regras de propagação de tipos cartográficos são adquiridas de maneira interativa pelo sistema e armazenadas em sua base de regras. O resultado é um conjunto de regras aplicadas para realizar a inferência automática de tipos cartográficos de dados derivados [Lanter, 1992]. Essas regras são armazenadas juntamente com meta-dados e consultadas pelo usuário durante uma consulta de banco de dados, construções de mapa e seleção de funções de SIG.

Pullar [1997] utiliza a abordagem de regras sob outro aspecto: programação de SIG por usuários finais inexperientes. Vários problemas que são endereçados pelos SIG requerem algum tipo de programação por parte do usuário do sistema. Dado que a maioria dos usuários de SIG são profissionais ligados às ciências e engenharias da terra, torna-se importante permitir que a interface de programação dos sistemas seja tão fácil de ser utilizada quanto possível. Pullar [1997] revisa diferentes paradigmas de programação e conclui que uma abordagem que combine os paradigmas orientados a re-

gras e orientados a objetos permite que os usuários formulem consultas relativamente complexas para o tratamento por SIG. Para chegar a essa conclusão, foi implementado um protótipo de interface para consultas para operar com um sistema de SIG comercial. Com isso, os conceitos pesquisados puderam ser verificados, apesar de o sistema ainda não ter sido estendido a uma versão completamente funcional. Planeja-se distribuir uma versão robusta da aplicação que possa executar uma vasta gama de consultas espaciais.

Outro sistema baseado em regras é descrito em Haj-Yehia and Peled, 2004. Nesse sistema, a preocupação é em facilitar a atualização de dados geográficos. Levando-se em consideração que os esforços de produção e manutenção de dados geográficos são uma grande preocupação ao se lidar com SIG, ter um processo de revisão e atualização desses dados é de grande valia. O sistema desenvolvido leva em consideração objetos geográficos e seus relacionamentos topológicos (ao lado, dentro, fora, etc), definindo regras que incluem uma descrição única para cada tipo de objeto.

Sistemas baseados em regras precisam de conhecimento para a codificação de regras válidas e gerais. A técnica de mineração de dados é bastante útil para se extrair informação não trivial de grandes bases de dados que pode ser utilizada na elaboração de novas regras. A seção seguinte introduz o tópico de mineração de dados e sua aplicação a dados geográficos.

### **3.3 Mineração de dados e Dados Geográficos**

#### **3.3.1 Mineração de dados**

Mineração de dados (*data mining*) é definida como o processo de realizar buscas automáticas em grandes volumes de dados, procurando por padrões e regras de associação. De maneira mais formal, mineração de dados é a extração não trivial de informações implícitas, previamente desconhecidas e potencialmente úteis de grandes bases de dados [Frankley et. al., 1992].

Mineração de dados é usado hoje em dia por empresas com um grande foco no consumidor, ou seja, organizações ligadas à área de varejo, comunicações, finanças e marketing. Ela possibilita que essas empresas descubram relacionamentos entre fatores internos como preço, posicionamento do produto, habilidades de equipe; e fatores externos como indicadores econômicos, competição e demografia dos clientes.

O software que realiza a atividade de mineração de dados analisa os relacionamentos e padrões em dados de transações armazenados oriundos de consultas dos usuários finais. O processo de mineração de dados consiste de três estágios [Edelstein, 1999]: 1 - exploração inicial, 2 - construção de um modelo ou identificação de padrões e 3 - distribuição (*deployment*).

O estágio de exploração geralmente se inicia com a preparação dos dados, que pode envolver a limpeza dos dados (*data cleaning*), transformações dos dados e seleção de subconjuntos de registros. Depois disso, dependendo da natureza do problema analítico, o primeiro estágio do processo de mineração pode envolver desde uma simples opção de preditores simples e diretos até um modelo de regressão formal, que elabore análises exploratórias utilizando um grande variedade de métodos gráficos e estatísticos.

O segundo estágio, chamado de estágio de construção de um modelo, envolve a consideração de vários modelos e a escolha do melhor deles baseada em sua perfor-



mance preditiva. Existem várias técnicas para se atingir esse objetivo [Weiss and Indurkha, 1997], como *mineração de dados preditiva*, *boosting* e *meta-aprendizagem*.

O estágio final, a distribuição, consiste em utilizar o modelo selecionado como melhor modelo no estágio anterior e aplicá-lo a novos dados para gerar previsões ou estimativas da renda esperada.

A maioria das ferramentas de mineração de dados disponibiliza pacotes que geralmente oferecem várias das funcionalidades seguintes:

- Análise de dados exploratória;
- Displays para visualização de mapas ligados e grafos;
- Métodos estatísticos variados;
- Modelos de regressão linear e logística;
- Classificação;
- Árvores de decisão e regressão;
- Redes neurais;
- Raciocínio baseado em memória;
- Descoberta de seqüência.

### 3.3.2 Mineração de dados Geográficos

Sistemas de informação geográfica geram uma vasta quantidade de dados num ambiente computacional bastante rico. A cobertura e o volume do conjunto de dados geográficos digitais é bastante extensa e está em contínuo crescimento [Miller, 2001]. Sendo assim, técnicas e ferramentas analíticas mais avançadas como mineração de dados podem ser bastante úteis para a descoberta de conhecimento escondido em bancos de dados geográficos.

Existem várias ferramentas para realizar mineração de dados em bancos de dados relacionais; por exemplo, o *Intelligent Data Miner* do DB2/IBM, *MineSet* da SiliconGraphics e o *Clementine* da SPSS.. No entanto, os modelos de análise usados em mineração de dados convencionais precisam ser modificados ou adaptados para levar em consideração os diferentes estilos e necessidades de análise e modelagem relevantes para o mundo dos SIG. Existem várias características que devem ser levados em consideração ao analisar dados geográficos [MILLER and HAN, 2001], tais como:

- Incerteza dos dados aliada ao fato de os erros serem frequentemente estruturados de forma espacial;
- Relacionamentos frequentemente estão relacionados a uma posição geográfica específica, em vez de serem globais;
- O tempo geralmente interage com o espaço;
- Alto-nível de multi-variedade nos dados, associada à redundância;
- O elemento localização é importante.

Segundo Miller [2001], padrões como classes, associações, regras e tendência possuem expressões espaciais bem definidas, tendo, assim, aplicações em técnicas para mineração de dados em bancos geográficos. A seguir, são descritas algumas técnicas que utilizam esses padrões.

- Classificação espacial – Essa técnica mapeia objetos geográficos em categorias com significados bem definidos, considerando relacionamentos de distância, direção ou conectividade e morfologia desses objetos. Koperski [1998] utiliza buffers espaciais para classificar objetos baseados em similaridade de atributos. A abordagem de Ester et. al. [2001] generaliza essa idéia para um algoritmo de aprendizado para classificação espacial considerando relacionamentos espaciais definidos como caminhos entre os objetos numa vizinhança conhecida.
- Associação espacial – Essa técnica define regras de associação espacial que definem contêm predicados espaciais na parte precedente ou antecedente de uma cláusula. Koperki [1998] formulou uma técnica de busca top-down numa árvore que explora o conhecimento implícito na forma de uma hierarquia de conceitos geográficos.
- Predição espacial – Malerba et. al. [2001] utiliza algoritmos de aprendizagem indutiva para extrair informações de mapas topográficos de propósito geral, como os produzidos por organizações nacionais ligadas à cartografia. Nesse método, uma heurística de busca constrói predicados locais baseados em objetos espaciais, conhecimento implícito, conceitos de alto-nível previamente definidos e critérios de performance.

Mineração de dados geográficos é um tópico de estudo da área de Descoberta do Conhecimento (*Knowledge Discovery*) que ainda está começando. A pesquisa nessa área tende a crescer, visto que cada vez mais cientistas estão dando especial atenção aos dados geográficos e, com isso, novas técnicas para explorar esses dados vem sendo desenvolvidas. Além disso, o crescimento do poder das plataformas computacionais contribui bastante para viabilizar a aplicação de novos métodos.

## 4 CONCLUSÃO

Bancos de Dados Geográficos são constituídos por uma coleção de dados georreferenciados, também chamados de dados geográficos. Sistemas de Informação Geográfica permitem visualizar, processar e analisar dados geográficos. A tecnologia de SIG tem sido utilizada em muitas situações de tomada de decisões de natureza crítica e, com isso, há uma demanda crescente por resultados precisos e significativos.

No entanto, os SIG “tradicionais” são bastante restritos quanto a sua capacidade para realizar análises espaciais significativas. Além disso, existe o problema da falta de interoperabilidade entre os vários sistemas devido à proliferação de padrões proprietários sem nenhuma regulamentação.

Este trabalho apresentou um relato de alguns experimentos com técnicas de sistemas inteligentes aplicadas no domínio dos SIG e seus bancos de dados geográficos com o objetivo de amenizar os problemas apontados acima.

Sistemas inteligentes são sistemas que possuem facilidades para representação do conhecimento, raciocínio dedutivo, busca com backtracking, refinamento e validação de conhecimento e classificação automática de conhecimento. Esses sistemas frequentemente utilizam um banco de dados inteligente, que pode ser descrito como um banco com um grande volume de dados e regras que codificam o conhecimento do domínio considerado.

Abordagens baseadas em sistemas inteligentes no domínio de SIGs se mostram bastante promissoras. Em especial, foram apresentadas técnicas baseadas em ontologias, regras e mineração de dados. Ontologias se mostraram bastante úteis para facilitar a representação do conhecimento e interoperabilidade de dados geográficos. A abordagem baseada em regras se mostrou adequada para tratar questões ligadas ao processamento de dados geográficos, incluindo interpretação, planejamento, monitoramento, controle e infra-estrutura. Por fim, mostrou-se como a análise e a descoberta de conhecimento não trivial em bases de dados geográficas vêm sendo endereçadas por técnicas de mineração de dados geográficos.

## Referências Bibliográficas

AGARWAL, P. Ontological considerations in GIScience. *Int'l Journal of Geographical Information Science*, 19, 5, p. 501-536, 2005.

BERTINO, E., Catania, B., and Zarri, G.P. **Intelligent Database Systems**. London, Addison-Wesley and ACM Press, 2001.

BISHR, Y. *Semantic Aspect of Interoperable GIS*. Ph.D., Wageningen Agricultural University and ITC, 1997.

BREITMAN, K. **Web Semântica: A Internet do Futuro**. Brasil, LTC, 212P, 2005.

BUSCALDI, D; ROSSO, P; SANCHIS, E. Wordnet as a geographical information resource. In *Proceedings of the 3rd Global WordNet Association (GWA06)*, 2006.

CÂMARA, G.; SOUZA, R. C. M.; PEDROSA, B.; VINHAS, L.; MONTEIRO, A. M. V.; PAIVA, J. A.; CARVALHO, M. T.; GATTASS, M. TerraLib: technology in support of GIS innovation. In: *Simpósio Brasileiro em Geoinformática, 2, 2000, São Paulo, SP. Proceedings...*

CASANOVA, M. et al. **Bancos de Dados Geográficos**. Brasil, MundoGeo, 2005. 506 p.

COX, S.; CUTHBERT, A.; LAKE, R.; MARTELL, R. *Geography Markup Language (GML) 2.0*. Disponível em: <http://opengis.net/gml/01-029/GML2.html>, 2001.

DAVIS JR., C. A. *Intercâmbio de informações geográficas: proposta preliminar, Relatório Técnico*, Prodabel, 1995.

EDELSTEIN, H. *Introduction to data mining and knowledge discovery (3rd ed)*. Potomac, MD: Two Crows Corp, 1999.

EGENHOFER, M. J. *Toward the Semantic Geospatial Web*. In *Proceedings of the Tenth ACM Int'l Symposium on Advances in Geographic Information Systems (ACM-GIS '02)*, McLean, VA, 2002.

ELMAGARMID, A. K.; PU, C. Guest editors' Introduction to the Special Issue on Heterogeneous Databases. *ACM Computing Surveys* 22(3), p. 175-178, 1990.

ESTER, M.; KRIEGEL, H. P.; SANDER, J. Algorithms and applications for spatial data mining. In: H. J. Miller and J. Han (eds.) *Geographic Data Mining and Knowledge Discovery*, London: Taylor and Francis, 160-187, 2001.

FONSECA, F. and EGENHOFER, M. *Sistemas de Informação Geográficos Baseados em Ontologias*, *Informática Pública*, 1, p. 47-65, 1999.

- FONSECA, F. T., EGENHOFER, M. J., AGOURIS, P. and CAMARA, G. Using Ontologies for Integrated Geographic Information Systems. *Transactions in GIS*, 6, 3, p. 231-257, 2002.
- FRAWLEY, W.; PIATETSKY-SHAPIRO, G. ; MATHEUS, C. Knowledge Discovery in Databases: An Overview. *AI Magazine*, pp. 213-228, 1992.
- GARDELS, K. The Open GIS Approach to Distributed Geodata and Geoprocessing. In: *Third International Conference/Workshop on Integrating GIS and Environmental Modeling*. Santa Fe, NM, USA, p. 21-25, 1996.
- GOODCHILD, M. F. Geographical data modeling. *Computers and Geosciences*, v. 18, n. 4, p. 401-408, 1992.
- HAJ-YEHIA, B.; PELED, A. Rule-based System for Updating Spatial Database. *ISPRS - XXth Congress*, Istanbul, Turkey, 2004.
- KALOGIROU, S. Expert systems and GIS: an application of land suitability evaluation. *Computers, Environment and Urban Systems*, v. 26, n. 2, p. 89-112, 2002.
- KANNAN, A. and GEETHA, T.V. Temporal Reasoning with Intelligent Databases. *Vivek -Bombay, India*, v. 11, n. 4, p. 11 - 20, 1998.
- KOPERSKI, K.; HAN, J.; STEFANOVIC N. An efficient two-step method for classification of spatial data, *Proceedings of the Spatial Data Handling Conference*, Vancouver, Canada, 1998.
- LANTER, D. P. Design Of A Lineage-Based Meta-Data Base For GIS. *Cartography and Geographic Information Systems*, 18(4):255-261, 1991.
- LANTER, D. P. Intelligent Assistants for Filling Critical Gaps in GIS. *Technical Publication 92-4*, National Center for Geographic Information and Analysis, Santa Barbara, CA, 1992.
- LIEBERMAN, J., PEHLE, T., DEAN, M. Semantic Evolution of Geospatial Web Services. In: *Proc. Workshop on Frameworks for Semantics in Web Services*. DERI, Innsbruck, Austria, 2005.
- MAEDCHE, A. *Ontology Learning for the Semantic Web*. Kluwer Academic Publishers, 2002.
- MALERBA, D.; ESPOSITO, F.; LANZA, A.; LISI, F. A. Machine learning for information extraction from topographic maps. In: H. J. Miller and J. Han (eds.) *Geographic Data Mining and Knowledge Discovery*, London: Taylor and Francis, in press, 2001.
- MILLER, H. J. and HAN, J. Geographic data mining and knowledge discovery: An overview. In: H. J. Miller and J. Han (eds.) *Geographic Data Mining and Knowledge Discovery*, London: Taylor and Francis, in press, 2001.
- NOLAN, J. J. and SIMON, R., et al. Developing an Ontology and ACL in an Agent-based GIS. *Ontologies in Agent Systems (0AS2001)*, Montreal, Canada, p. 99-101, 2001.
- PERRY, M; HAKIMPOUR, F; SHETH, A. Analyzing Theme, Space, and Time: An Ontology-based Approach. In: *14th International Symposium on Advances in Geographic Information Systems*, ACM-GIS'06, Arlington, Va, USA, 2006.
- PULLAR, D. Rule-Based Modelling in GIS. *Conference of GeoComputation '97 & SIRC '97*, 2, Otago, New Zealand, 1997.
- University of Otago, New Zealand, 26-29 August 1997

RIESBECK, C. K. and SCHANK, R. C. **Inside Case-based Reasoning**. Lawrence Erlbaum Associates, Northvale, NJ: Erlbaum., 1989. 448p.

USCHOLD, M. and GRÜNINGER, M. Ontologies: Principles, Methods and Applications. *Knowledge Engineering Review*, 11(2): 93-155, 1996.

The OpenGIS® Guide - Introduction to Interoperable Geoprocessing and the OpenGIS Specification. Open GIS Consortium, Inc, MA, 1998.

VENANCIO, L. R.; FILETO, R; MEDEIROS, C. B. Aplicando Ontologias de Objetos Geográficos para Facilitar Navegação em GIS. Simposio Brasileiro de Geoinformatica (GEOINFO), São José dos Campos, 2003.

VINHAS, Lúbia. **Um Subsistema Extensível para o Armazenamento de Geo-Campos em Bancos de Dados Geográficos**. 2006. 114 f. Tese (Doutorado) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos - SP, 2006.

WEISS, S. M.; INDURKHIA, N. Predictive data mining: A practical guide. New York: Morgan-Kaufman, 1997.

WORBOYS, M.; DUCKHAM, M. **GIS A Computing Perspective Second Edition**. Boca Raton, Florida, USA: CRC Press, 2004. 426 p.