

Técnicas de modelação de informação geográfica: uma síntese

Maribel Yasmina Santos

Universidade do Minho, Departamento de Sistemas de Informação, Guimarães, Portugal
maribel@dsi.uminho.pt

Luís Alfredo Amaral

Universidade do Minho, Departamento de Sistemas de Informação, Guimarães, Portugal
amaral@dsi.uminho.pt

Resumo

A modelação de dados, vista como o conjunto de actividades que conduz ao desenho de uma Base de Dados, é caracterizada por passar por três etapas. A primeira diz respeito ao *modelo conceptual de dados*, que determina os requisitos da base de dados na fase de análise de requisitos. A segunda etapa está associada ao *modelo lógico de dados* e corresponde à transformação do modelo conceptual em estruturas de dados consistentes com o Sistema Gestor de Bases de Dados seleccionado para a implementação. A terceira e última etapa prende-se com o *modelo físico de dados*, no qual se verifica a organização dos ficheiros e métodos de acesso aos dados, característicos do sistema gestor seleccionado.

A modelação de aplicações geográficas requer a utilização de técnicas específicas, que permitam armazenar e processar dados espaciais. Estes dados representam objectos cuja posição no espaço é relevante, uma vez que esta define relações topológicas entre os mesmos.

O desenho lógico de aplicações geográficas tem sido descuidado, em favor do desenho físico, permitindo a construção de aplicações sobre sistemas de ficheiros dedicados, nos quais os dados são armazenados em formatos próprios, cuja manutenção e portabilidade são difíceis de conseguir.

Neste artigo são apresentadas técnicas de modelação para informação geográfica. Ao nível conceptual, é apresentado o diagrama Entidades e Relacionamentos na modelação de informação geográfica. Ao nível lógico é apresentado o modelo georrelacional, uma extensão do modelo relacional, e ainda, uma abordagem formal à modelação de aplicações geográficas. Descreve-se, também, o *Unified Modeling Language*, uma linguagem que permite a modelação de dados tanto ao nível conceptual como lógico. Esta última, que não necessita de qualquer adaptação para utilização no domínio geográfico, é a linguagem utilizada pela *International Standard Organisation*, na definição das normas internacionais para *Informação Geográfica e Geomática*.

Palavras chave: modelação de dados, dados espaciais, sistema de informação geográfica.

1 Introdução

Um modelo de dados integra um conjunto de conceitos, utilizados para descrever a estrutura de uma Base de Dados (BD), assim como o conjunto de operações que podem ser executadas sobre a mesma [Navathe 1992]. Por estrutura entende-se tipos de dados, relacionamentos e restrições, que definem como a BD se encontra organizada. Um modelo de dados é utilizado para definir modelos de aplicação, que constituem a descrição de uma BD para um dado domínio de aplicação.

A modelação de dados é a actividade que "debruçando-se sobre a totalidade dos requisitos de informação de um sistema de informação, tenta encontrar um modelo que traduza a estrutura lógica dos dados que satisfaz esses requisitos" ([Pereira 1997] p. 16).

A modelação de dados, vista como o conjunto de actividades que conduz ao desenho da BD, passa por três etapas [Adam e Gangopadhyay 1997] [Navathe 1992]:

Desenho conceptual. Consiste na construção do *modelo conceptual de dados*, o qual reflecte a percepção que os utilizadores têm dos dados, sendo independente de qualquer implementação física. Uma das abordagens mais utilizadas na modelação conceptual de dados são os diagramas E-R (Entidades e Relacionamentos) [Chen 1976].

Desenho lógico. Corresponde à transformação do modelo conceptual no *modelo lógico de dados*, isto é, em estruturas de dados que são implementáveis no Sistema Gestor de Base de Dados (SGBD) seleccionado. O modelo relacional tem sido o modelo mais utilizado na modelação lógica de dados [Codd 1970].

Desenho físico. Passa pela definição dos detalhes físicos que serão considerados na implementação do modelo lógico. Permite definir os métodos de acesso aos dados e os detalhes associados à organização física dos ficheiros, e que são específicos do SGBD adoptado.

As limitações dos modelos de dados ditos convencionais, tais como o modelo relacional, começaram a evidenciar-se quando surgiram novas áreas de aplicação, tais como os *Sistemas de Informação Geográfica* (SIG), com diferentes necessidades ao nível do armazenamento e processamento dos dados. Algumas destas aplicações são construídas sobre sistemas de ficheiros dedicados [Pereira 1997], nos quais os dados são armazenados em formatos próprios, cuja manutenção e portabilidade são difíceis de conseguir. Hadzilacos e Tryfona [Hadzilacos e Tryfona 1996] defendem que as aplicações geográficas devem ser desenvolvidas recorrendo à modelação de dados, para o qual é possível estender (sempre que necessário) as metodologias habituais, por forma a estas suportarem as particularidades da informação espacial. Aplicações geográficas manuseiam objectos cuja posição no espaço é relevante. O facto de ocuparem determinada posição faz com que os objectos se encontrem relacionados uns com os outros, pelo que a modelação deverá permitir definir os relacionamentos topológicos existentes entre os mesmos.

A modelação de informação geográfica, ao nível do desenho lógico, deverá ser realizada recorrendo a ferramentas que permitam [Hadzilacos e Tryfona 1996]:

- Definir atributos espaciais, incluindo os seus tipos geométricos (ponto, linha, polígono);
- Organizar atributos espaciais por níveis;
- Definir atributos não espaciais e seus respectivos relacionamentos com os atributos espaciais;
- Especificar restrições topológicas ou de integridade dos dados espaciais;
- Etiquetar as versões das diversas entidades, introduzindo a componente *temporal* na modelação.

Este artigo, que apresenta diversas técnicas de modelação de informação geográfica, encontra-se organizado da seguinte forma. A secção 2 define os tipos de dados espaciais, apresentando as representações que podem ser utilizadas para os mesmos. A secção 3 evidencia como o modelo E-R pode ser utilizado na modelação conceptual de aplicações geográficas. Para a modelação lógica de dados, a secção 4 sintetiza o modelo geo-relacional, uma extensão do modelo relacional para dados geográficos. A secção 5 apresenta uma abordagem formal à modelação, que pode ser utilizada na especificação de aplicações geográficas. Este artigo prossegue, secção 6, com a descrição do *Unified Modeling Language* (UML), uma linguagem de modelação que não necessita de qualquer extensão para a modelação de dados espaciais. Na secção 7 conclui-se com uma breve síntese das diversas técnicas de modelação descritas neste documento.

2 Tipos de dados espaciais

A modelação de objectos espaciais passa pela sua abstracção recorrendo a pontos, linhas ou polígonos (Figura 1). Um **ponto** representa o aspecto geométrico de um objecto para o qual apenas a sua localização no espaço é relevante. Uma **linha** (constituída por um conjunto de segmentos de recta) permite a representação de objectos através dos quais é possível a movimentação no espaço. Um **polígono** é uma abstracção de um objecto cuja localização e extensão são relevantes.

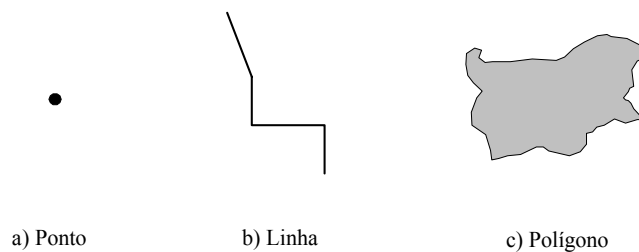


Figura 1 – Tipos de dados espaciais

Tipos de dados espaciais permitem representar entidades geométricas, pontos, linhas e polígonos, assim como os relacionamentos existentes entre as mesmas (l intersecta r). Estas entidades podem ser manipuladas verificando determinadas propriedades ($\text{área}(r) > 1000$), ou ainda executando operações específicas ($\text{intersecção}(l, r)$) sobre as mesmas [Güting 1994].

A integração de tipos de dados espaciais, no modelo de dados de um SGBD, permite a representação de objectos espaciais por entidades que possuem pelo menos um atributo do tipo espacial. Assim, são ampliados os tipos de dados normalmente disponíveis (INTEGER, REAL, STRING, ...) num SGBD, disponibilizando tipos de dados espaciais (POINT, LINE, REGION) [Shekhar, et al. 1997].

A **representação de dados espaciais** pode ser efectuada através de dois métodos: representação baseada em *células* ou baseada em *objectos*. Na representação baseada em células, o espaço geográfico é dividido num conjunto de células independentes que cobrem toda a região geográfica em análise. Os objectos geográficos encontram-se embebidos no espaço, sendo representados através do conteúdo das células. Esta representação matricial da informação é baseada em áreas, visto os objectos serem representados pela área que os constitui, e não pelo

seu limite [Adam e Gangopadhyay 1997]. O modelo matricial de resolução fixa (Figura 2) utiliza uma estrutura de dados denominada *raster*, que é caracterizada por ser constituída por um conjunto de células de tamanho fixo, as quais são utilizadas para descrever as entidades geográficas. Cada célula tem associados dois valores: a **posição**, que representa a sua localização, e o **valor** da área geográfica que representa. Este valor é uniforme ao longo de todas as células que constituem uma mesma entidade geográfica.

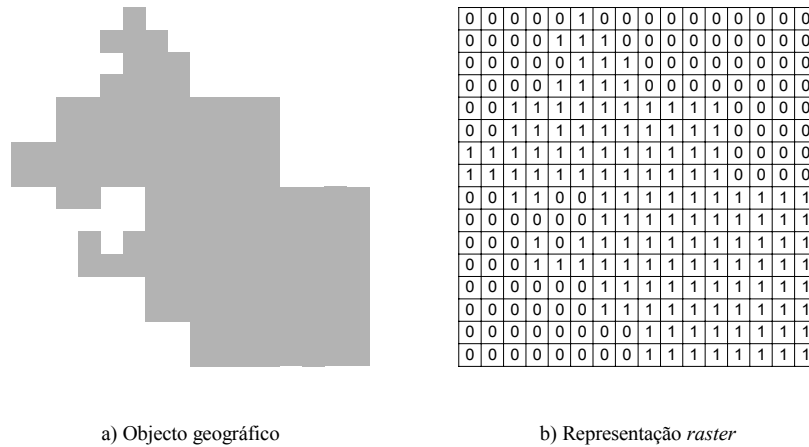
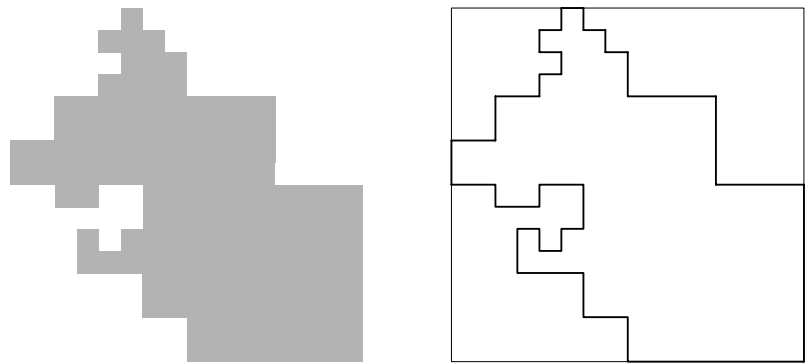


Figura 2 – Representação por células

Na representação baseada em objectos, modelo *vectorial*, os objectos são representados pelas entidades geográficas que os constituem. Pontos, linhas e polígonos são as três unidades geométricas fundamentais utilizadas na representação de objectos geográficos. Estas unidades podem ser combinadas por forma a representarem objectos *compostos* (agregação de várias unidades geométricas do mesmo tipo) ou objectos *complexos* (constituídos pela agregação de unidades geométricas de diferentes tipos) [Adam e Gangopadhyay 1997].

Um ponto é descrito por um único par de coordenadas, uma linha por uma sequência de pares de coordenadas que definem segmentos de recta, e um polígono, representado por uma área fechada, é descrito por uma sequência de pares de coordenadas, na qual o primeiro e último par de coordenadas, coincidem [Burrough 1986]. Este modelo constitui uma estrutura de armazenamento da informação eficiente, uma vez que apenas são representadas as coordenadas que definem um acontecimento. A Figura 3 apresenta a representação vectorial do objecto geográfico descrito na Figura 2 (na qual foi apresentado recorrendo a uma estrutura baseada em células).

A representação dos dados espaciais seguindo o modelo vectorial utiliza, comumente, o *modelo de dados topológico*, uma vez que este permite que os relacionamentos espaciais existentes entre entidades geográficas sejam expressos através da topologia. O elemento básico utilizado na representação é a *aresta*, a qual integra um conjunto de pontos, que iniciam e terminam num *nodo*. Um nodo representa um ponto de intersecção entre duas ou mais arestas. Uma *face* é representada recorrendo a uma cadeia de arestas, que delimitam a área representada. A topologia é neste modelo armazenada recorrendo a quatro tabelas (Figura 4), uma para cada tipo de elemento espacial (nodo, aresta ou face), e a quarta para as coordenadas das arestas [Aronoff 1989].

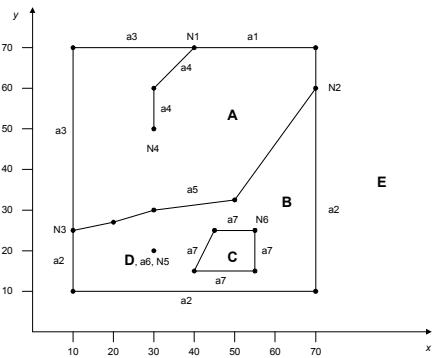


a) Objecto geográfico

b) Representação *vectorial*

Figura 3 – Representação *vectorial*

Na tabela correspondente às faces (Figura 4), armazenam-se as arestas que as delimitam, seguindo o sentido dos ponteiros do relógio. Sempre que uma face contiver outros elementos, a identificação dos mesmos é precedida de um 0. Na tabela de nodos identificam-se as arestas a que cada nodo está associado, enquanto que na tabela de arestas estabelece-se a relação existente entre estas e os nodos e as faces.



Faces		Nodos		Arestas					Coordenadas das Arestas			
Face	Arestas	Nodo	Arestas	Aresta	Nodo inicial	Nodo final	Face esquerda	Face direita	Aresta	(x, y) inicial	(x, y) intermédios	(x, y) final
A	a1, a5, a3	N1	a1, a3, a4	a1	N1	N2	E	A	a1	(40,70)	(70,70)	(70,60)
B	a2, a5, 0, a6, 0, a7	N2	a1, a2, a5	a2	N2	N3	E	B	a2	(70,60)	(70,10), (10,10)	(10,25)
C	a7	N3	a2, a3, a5	a3	N3	N1	E	A	a3	(10,25)	(10,70)	(40,70)
D	a6	N4	a4	a4	N4	N1	A	A	a4	(40,70)	(30,60)	(30,50)
E	área exterior do mapa	N5	a6	a5	N3	N2	A	B	a5	(10,25)	(20,27), (30,30), (50,32)	(70,60)
		N6	a7	a6	N5	N5	B	B	a6	(30,20)	(30,20)	(30,20)
				a7	N6	N6	B	C	a7	(55,25)	(55,15), (40,15), (45,25)	(55,25)

Figura 4 – Modelo de dados *topológico*

3 O diagrama E-R na modelação de dados geográficos

O diagrama E-R como técnica de modelação de dados geográficos [Laurini e Thompson 1992] permite representar objectos espaciais como *entidades*, cujas propriedades estruturais e topológicas com outros objectos espaciais, são representadas através de *relacionamentos* entre entidades.

No caso de uma representação vectorial da informação, os objectos espaciais são descritos recorrendo a agrupamentos de objectos do tipo ponto, linha ou polígono. Cada objecto espacial, representado como uma entidade, é relacionado com os objectos que o constituem, através de relacionamentos estruturais de diferentes cardinalidades.

A Figura 5 apresenta o diagrama E-R correspondente ao modelo de dados topológico, apresentado na secção anterior.

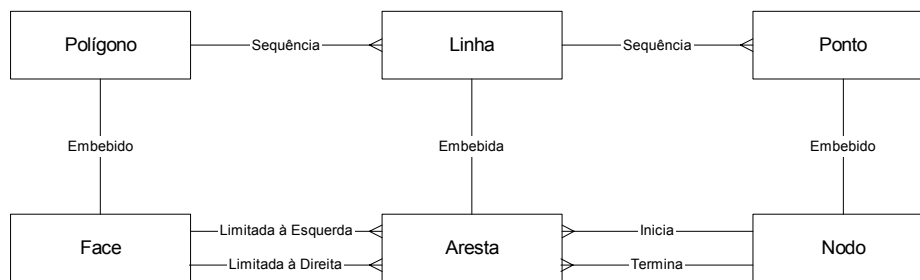


Figura 5 – Diagrama E-R na modelação de dados espaciais

4 Extensão do modelo relacional para dados geográficos: o modelo geo-relacional

O modelo **geo-relacional** estende o modelo relacional por forma a este permitir modelar dados geográficos. A incorporação no modelo de componentes necessários ao domínio geográfico, tais como *níveis*, *relações*, *níveis virtuais*, *classes de objectos* e *restrições de integridade*, permitem a construção de modelos lógicos de dados (esquemas lógicos), assim como vistas (esquemas externos) sobre os mesmos [Hadzilacos e Tryfona 1996]. Os esquemas lógicos são construídos recorrendo a componentes do tipo *níveis* e *relações*, enquanto que os esquemas externos utilizam *níveis virtuais*, *classes de objectos* e *restrições de integridade*.

O modelo geo-relacional (Figura 6) permite que a informação seja armazenada em diversas tabelas, relacionadas através de atributos comuns. Estas ligações permitem que a pesquisa de informação possa ser iniciada nos dados espaciais ou nos dados não espaciais. Diferentes conjuntos de atributos são armazenados em diferentes tabelas, através das quais é possível, verificando as relações existentes entre as mesmas, seleccionar todos os atributos associados a determinada entidade espacial [Shepherd 1991].

Segundo Shepherd [Shepherd 1991], a maioria dos SIG do tipo vectorial adoptam o modelo geo-relacional na conceptualização das ligações existentes entre dados espaciais e dados não espaciais. Esta aproximação passa por armazenar separadamente dados espaciais e dados não espaciais, estando o SGBD utilizado encarregado de assegurar a ligação entre os dois conjuntos de dados.

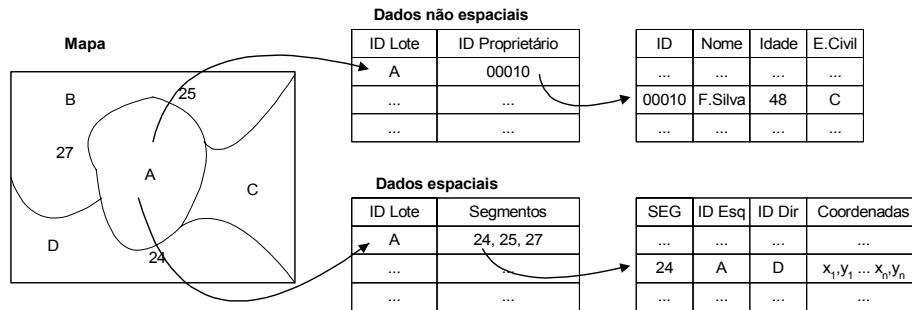


Figura 6 – Modelo geo-relacional

No modelo geo-relacional, um **nível** é utilizado para descrever propriedades geográficas, que definem associações entre o espaço geográfico e um dado conjunto de atributos. Cada nível é caracterizado por representar um tipo de entidade geométrica (ponto, linha, polígono), declarado através das palavras reservadas `GEOMETRIC TYPE`, devendo ser identificado por um número, e opcionalmente, por um identificador.

```

DEFINE LAYER n <identificador> <TEMPORAL>
ATTR (at1, Domain, <UNIQUE>), ..., (atn, Domain, <UNIQUE>)
GEOMETRIC TYPE Geometric_type
<POSITIONING sistema_coordenadas>
<CONSTRAIN restrição >

Domain = {INT, REAL, STRING}
Geometric_type = {POINT, LINE, REGION}

```

As **relações** são utilizadas para modelar entidades não geográficas. A chave de uma relação resulta da concatenação dos atributos caracterizados com a palavra reservada `KEY`.

```

DEFINE RELATION nome_relação
ATTR (at1, Domain, <KEY>), ..., (atn, Domain, <KEY>)
<TIME_POINT atj>

```

Os **níveis virtuais** são utilizados para descrever informação que é obtida (calculada) a partir da informação espacial definida nos diversos níveis do esquema lógico. Entre as operações que podem ser realizadas encontra-se a *sobreposição*, *intersecção*, *união*, etc., permitindo que novos níveis de informação sejam gerados a partir dos níveis existentes. O modelo geo-relacional disponibiliza quatro tipos de operações:

- *derivação de atributos*, que permite adicionar novos atributos a um nível, sem alterar as características geométricas do mesmo (construção de um novo atributo, baseado num ou mais atributos já existentes no nível);

```

DEFINE V_LAYER n <nome> <TEMPORAL>
AS COMPUTE ATTRS (m, novo_at1 = f1(at1, ..., atn), ..., novo_atk = fk(at1, ..., atn))

```

- *computação geométrica*, permite construir um novo nível, caracterizado por representar entidades com determinada característica geométrica, que foi calculada a partir das entidades geográficas existentes no nível que lhe deu origem;

```

DEFINE V_LAYER n <nome> <TEMPORAL>
AS COMPUTE GEOMETRIC (m, novo_at, f(at1, ..., atn))

```

- *sobreposição* de dois níveis, permitindo criar um novo nível que é o resultado da intersecção das entidades geométricas contidas nos níveis que lhe dão origem;

```

DEFINE V_LAYER n <nome> <TEMPORAL>
AS OVERLAY (m, k)

```

- *agregação*, que permite integrar entidades geométricas adjacentes, que apresentem o mesmo valor para determinado atributo.

```

DEFINE V_LAYER n <nome> <TEMPORAL>
AS RECLASS OF (m, at1, ..., atn)

```

As aplicações geográficas requerem frequentemente a integração de atributos pertencentes a diversos níveis e relações. **Classes de objectos** permitem a agregação de entidades geométricas de diferentes tipos, mas que estão inseridas no mesmo espaço geográfico. A definição de uma classe de objectos pode ser efectuada: i) agregando todas as entidades geométricas que se encontram nos níveis especificados em `ON LAYERS`; ii) limitando as entidades geométricas às discriminadas em `GEOMETRIC TYPES`; ou iii) restringindo os objectos intervenientes aqueles que possuem como atributos o conjunto mencionado em `WITH ATTRIBUTES`.

As opções `ON LAYERS` e `GEOMETRIC TYPES` poderão ser omitidas se a classe de objectos que está a ser definida constitui um subtipo, `SUBTYPE OF`, de outra. Neste caso é assumido que a classe a criar possui o mesmo nível e tipo geométrico.

```

DEFINE OBJECT CLASS nome_classe_objecto
<GEOMETRIC TYPE Geometric_type>
<SUBTYPE OF classe>
<ON LAYERS id_nível1, ..., id_nívelk>
<WITH ATTRIBUTES at1, ..., atn>
<CONSTRAIN restrição>

```

Uma **restrição**, explícita através de predicados em lógica de 1ª ordem, indica o domínio de valores que as entidades da BD podem tomar. Ao nível dos predicados, estes podem ser construídos recorrendo a operadores aritméticos, operadores lógicos, funções espaciais (distância, área, ...) e operadores topológicos.

```

DEFINE CONSTRAINT restrição
ON {nome_classe_objecto | id_nível}
AMONG {classe_obj1, ..., classe_objk | id_nível1, ..., id_nívelk}
AS condição

```

De seguida é apresentado um pequeno exemplo, que visa evidenciar a utilização do modelo geo-relacional na definição das estruturas de dados e das restrições necessárias à execução de determinada questão.

Exemplo de utilização do modelo geo-relacional

Formulada a questão "Determinar todos os distritos que possuem pelo menos três escolas", as estruturas de dados e as restrições necessárias à sua satisfação são:

```
DEFINE LAYER 1 Distrito
ATTR (ID_distrito, INT, UNIQUE), (nome_distrito, CHAR(20))
GEOMETRIC TYPE REGION
POSITIONING UTM /* UTM é um sistema de coordenadas */

DEFINE LAYER 2 Escola
ATTR (ID_escola, INT, UNIQUE), (nome_escola, CHAR(20))
GEOMETRIC TYPE POINT
POSITIONING UTM

DEFINE OBJECT CLASS Distritos_com_diversas_escolas
GEOMETRIC TYPE REGION, POINT
ON LAYERS Distrito, Escola
CONSTRAINT Diversas_Escolas

DEFINE CONSTRAINT Diversas_Escolas
ON Distrito
AMONG Distrito, Escola
AS ( $d \in \text{Distrito} \wedge \exists_{x,y,z} (x \in \text{Escola} \wedge y \in \text{Escola} \wedge z \in \text{Escola} \wedge \text{inside}(x, d) \wedge \text{inside}(y, d) \wedge \text{inside}(z, d) \wedge x \neq y \wedge x \neq z \wedge y \neq z)$ )
```

5 A especificação formal no domínio geográfico

Roman [Roman 1990] propõe uma abordagem formal na modelação lógica de aplicações geográficas, na qual a utilização de predicados em lógica de 1ª ordem permite a especificação dos requisitos de informação. Estes são compatíveis com as especificidades do PROLOG, permitindo a inferência lógica de informação. Predicados em lógica de 2ª ordem são utilizados para definir as regras que possibilitam a incorporação no sistema, de mecanismos de raciocínio espacial e temporal.

Esta abordagem assume que o mundo real pode ser modelado por uma colecção de *objectos*, que permitem retratar as diferentes perspectivas que os utilizadores têm das entidades geográficas. Estes objectos são formalmente definidos recorrendo a predicados em lógica de 1ª ordem, os quais constituem **conhecimento básico** (*basic facts*), representando os factos assumidos como verdadeiros, ou **conhecimento virtual** (*virtual facts*), que é construído recorrendo a outros factos (básicos ou virtuais).

Ao nível semântico, predicados em lógica de 2ª ordem são utilizados na definição do conhecimento espacial e temporal a incorporar no sistema de raciocínio. O **conhecimento semântico** permite a definição do instante temporal e do espaço geográfico em que determinado facto é verdadeiro. A consistência semântica entre os factos é conseguida recorrendo à especificação de restrições, que dependem do contexto em que a informação é utilizada. Esta

aproximação permite a construção de diversos modelos sobre os dados, os quais retratam diferentes perspectivas dos mesmos.

Meta-factos e meta-restrições podem ser definidos em lógica de 2ª ordem, permitindo a especificação de conhecimento que não está limitado a determinado contexto (utilizador ou aplicação), mas que constitui os mecanismos básicos de raciocínio. Estes podem ainda ser agrupados em **meta-modelos**, que possibilitam o encapsulamento do conjunto de regras aplicáveis a determinado domínio de aplicação.

A especificação de um modelo de dados, utilizando a abordagem formal proposta por Roman [Roman 1990], passa então pela definição de:

Conhecimento básico. Um facto pode assumir a forma de uma propriedade considerada verdadeira para um dado objecto, ou de uma propriedade verificada entre objectos, definida recorrendo a uma relação:

```
estrada(x1) .  
estrada(x2) .  
intersecta(x1, x2) .
```

Restrições. Uma restrição tem como objectivo validar o conhecimento básico existente no sistema, detectando eventuais inconsistências e alertando o utilizador no caso destas ocorrerem:

$$(\forall_{x,y,z}) : (\text{concelho_de}(X,Y) \wedge \text{concelho_de}(X,Z) \wedge (Y \neq Z) \\ \Rightarrow \text{ERRO}(\text{pertence_a_dois_distritos}, X)) .$$

Conhecimento semântico. O conhecimento semântico permite a utilização de valores (quantificação) na qualificação das propriedades dos objectos. Estes valores não podem ser tratados como objectos. A sua especificação passa pela extensão dos predicados, por forma a estes integrarem o valor do atributo e o respectivo objecto. Por exemplo, o conhecimento semântico associado a: "*A temperatura média de Braga é de 20 graus*", pode ser especificado através de:

```
temperatura_média(20)(Braga) .
```

Modelos. O conceito de modelo tem como objectivo permitir retratar diferentes perspectivas dos dados, ou alterações no domínio de aplicação, que requerem a re-interpretação dos mesmos. Os modelos permitem indicar o contexto em que os factos são válidos. Esta especificação é conseguida acrescentando o nome do modelo, seguido de um qualificador ('), antes da definição do facto: `celsius'temperatura_congelação(0)(X)`. Este facto indica que na escala Celsius, a temperatura de congelação do objecto X é de 0°.

Meta-factos (regras de inferência). Os meta-factos permitem especificar as regras de inferência que permitem o raciocínio num domínio semântico específico. Integram conhecimento, sob a forma de regras, que é independente de predicados particulares. A independência é conseguida recorrendo a predicados em lógica de 2ª ordem, que permitem a utilização de variáveis quantificadas universalmente, válidas para determinado conjunto de predicados. O meta-facto "*um facto não explícito como sendo verdadeiro é assumido como sendo falso*", pode ser especificado através de:

$$(\forall_{M,P,X}) : (M'P(X) \Rightarrow M'P(\text{true})(X)) ,$$

$$(\forall_{M,P,X}) : (\text{MODELO}(M) \wedge \text{PREDICADO}(P) \wedge \text{OBJECTO}(X) \wedge \text{not}(M'P(\text{true})(X)) \Rightarrow M'P(\text{false})(X))$$

Meta-restrições. As meta-restrições têm o mesmo objectivo das restrições, mas possuem a vantagem de poder utilizar na sua definição meta-factos já especificados. A meta-restrição "*nenhum facto pode ser simultaneamente verdadeiro e falso*", é conseguida através do predicado:

$$(\forall_{M,P,X}) : (M'P(\text{true})(X) \wedge M'P(\text{false})(X) \Rightarrow M'ERROR(\text{contradição}, P, X))$$

Meta-modelos. Os meta-factos, e as meta-restrições, podem ser agrupados num conjunto denominado de **meta-regras**, que quando integrado com o conhecimento semântico específico de um dado domínio de aplicação, constitui um meta-modelo. Os meta-modelos permitem que diferentes regras de inferência sejam utilizadas sobre os mesmos dados, aumentando a produtividade na avaliação de vários modelos.

Utilizando a abordagem formal apresentada, o exemplo descrito na secção anterior pode ser especificado através de:

```

distrito(Aveiro).
distrito(Braga).
distrito(Coimbra).

...

escola(Alberto Sampaio).
escola(Carlos Amarante).
escola(Da Maria).

...

@p escola_de(Braga) (Alberto Sampaio).
@p escola_de(Braga) (Carlos Amarante).
@p escola_de(Braga) (Da Maria).

...

(∀D,X,Y,Z,P) : (@P escola_de(X) (D) ∧ @P escola_de(Y) (D) ∧
                @P escola_de(Z) (D) ∧ distrito(D) ∧ escola(X) ∧
                escola(Y) ∧ escola(Z) ∧ X≠Y ∧ X≠Z ∧ Y≠Z) .

```

Neste exemplo, o operador “@” é utilizado para qualificar factos, permitindo especificar que determinada propriedade é verdadeira num ponto específico do espaço. O facto @_p q(y) (x) retrata que a propriedade q(y) do objecto x é válida na posição p.

6 O UML na modelação de dados geográficos

O UML é uma linguagem de modelação, que pode ser utilizada para visualizar, especificar, construir e documentar os componentes de uma aplicação. O UML utiliza três tipos de componentes básicos na modelação: **entidades** (*things*), **relacionamentos** (*relationships*) e **diagramas** (*diagrams*). Entidades representam abstrações de objectos do mundo real, agrupadas dentro de diagramas, e relacionadas através de associações definidas pelo utilizador.

As entidades podem ser classificadas em quatro grandes grupos: **estrutura**, **comportamento**, **agrupamento** e **anotação**. Estes quatro conjuntos de entidades constituem os blocos OO (*Object-Oriented*) básicos utilizados em UML [Booch, et al. 1999]. A notação gráfica, associada a cada um dos elementos que constituem estes quatro grupos, é apresentada na Figura 7. De seguida são descritas as principais características de cada um destes elementos.

Entidades de estrutura. Constituem os elementos estáticos de um modelo UML, podendo os mesmos possuir uma existência conceptual, lógica ou física. Existem sete entidades estruturais básicas. As cinco primeiras podem ser utilizadas para representar entidades conceptuais ou entidades lógicas, enquanto que as duas últimas são utilizadas para descrever entidades físicas. As sete entidades são:

- i. a *classe*, que corresponde a uma descrição de um conjunto de objectos com semântica comum, os quais partilham atributos, operações e relacionamentos;
- ii. a *interface*, utilizada na definição de um conjunto de operações que especificam uma tarefa de uma *classe* ou de um *componente*;
- iii. a *colaboração*, que define uma *interacção*, e constitui um agrupamento de papéis ou outros elementos que trabalham em conjunto para disponibilizar determinado comportamento cooperativo;
- iv. o *caso de uso*, permite a descrição de acções que o sistema executa para atingir determinado resultado;
- v. a *classe activa*, que permite descrever classes que integram um ou mais processos;
- vi. o *componente*, que representa uma parte física do sistema, implementando a realização física de um conjunto de *interfaces*;
- vii. o *nodo*, que representa um elemento físico que tem existência durante a execução da aplicação, sendo normalmente um recurso computacional.

Entidades de comportamento. Constituem as partes dinâmicas dos modelos UML. Existem dois tipos básicos:

- i. a *interacção*, que integra um conjunto de mensagens trocadas entre objectos, inseridas em determinado contexto e para atingir determinado fim;
- ii. o *estado*, que permite especificar a sequência de estados que um objecto ou interacção atravessa, na resposta a determinados eventos.

Entidades de agrupamento. Constituem as partes organizacionais dos modelos UML. Basicamente podem ser consideradas caixas, nas quais um modelo pode ser decomposto. O tipo básico é o *módulo*, que permite organizar, conceptualmente, elementos em grupos.

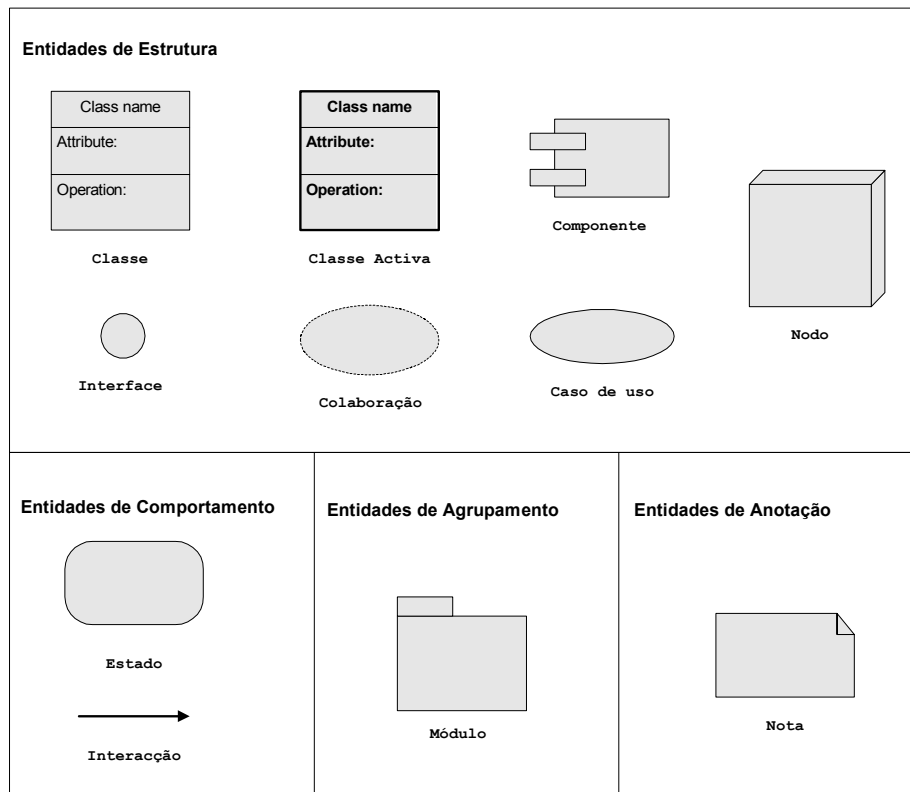


Figura 7 – Entidades básicas do UML

Entidades de anotação. Permitem explicar partes de modelos UML, através da descrição de qualquer elemento do modelo. O elemento básico deste tipo de entidades é a nota.

Ao nível dos relacionamentos, a linguagem UML disponibiliza quatro tipos básicos: dependência, associação, generalização e realização (Figura 8). Basicamente:

- Uma dependência constitui um relacionamento semântico entre duas entidades, no qual qualquer alteração de uma das entidades (entidade independente), pode afectar a semântica da outra entidade (entidade dependente).
- Uma associação constitui um relacionamento estrutural, que descreve um conjunto de ligações entre objectos. A agregação constitui um caso especial de associação, representando uma relação estrutural entre o *todo* e as suas *partes*.
- Uma generalização constitui um relacionamento de generalização/especialização, no qual os objectos da entidade *especializada* partilham a estrutura e o comportamento da entidade *generalizada*.
- Uma realização descreve relacionamentos semânticos entre classificadores, no qual um classificador especifica as tarefas que outro classificador desempenha. Este tipo de relacionamento pode existir entre interfaces e as classes ou componentes que as implementam; ou entre casos de uso e as colaborações que os implementam.

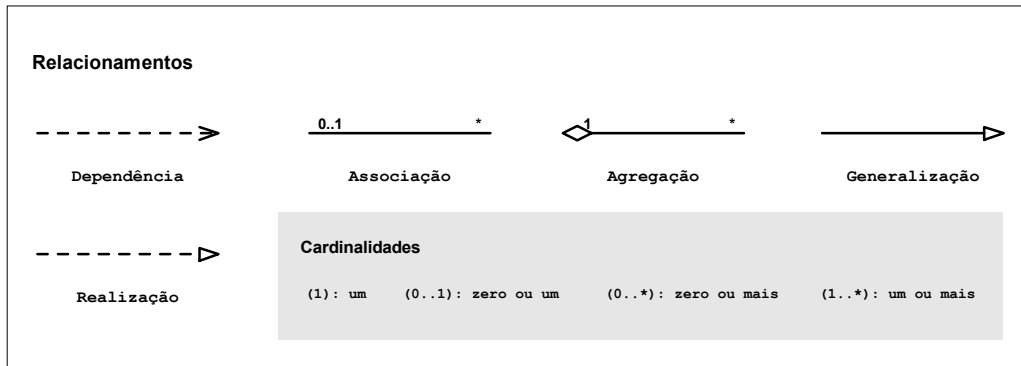


Figura 8 – Relacionamentos existentes em UML

Os diagramas são utilizados para agrupar um conjunto de elementos. Diferentes diagramas permitem visualizar um sistema de diferentes perspectivas. Um mesmo elemento pode estar representado em todos os diagramas, apenas em alguns, ou mesmo em nenhum, sendo esta uma situação excepcional. O UML permite a construção de nove tipos de diagramas:

Diagrama de classes. Um diagrama de classes evidencia um conjunto de classes, interfaces, colaborações e seus respectivos relacionamentos. Permitem desenhar uma visão estática do sistema.

Diagrama de objectos. Um diagrama de objectos agrega um conjunto de objectos e seus relacionamentos. Permitem representar perspectivas estáticas das instâncias, cujas entidades estão representadas num dado diagrama de classes.

Diagrama de caso de uso. Um diagrama de caso de uso agrupa um conjunto de entidades do tipo caso de uso, actores (um tipo particular de classe) e seus respectivos relacionamentos. Estes diagramas são particularmente importantes na organização e modelação do modo de funcionamento de um sistema, permitindo retratar uma visão (perspectiva de utilização) estática do mesmo.

Diagrama de interacção. Os diagramas de sequência e os diagramas de colaboração constituem os dois tipos de diagramas de interacção utilizados em UML. Estes diagramas permitem modelar os aspectos dinâmicos do sistema em termos dos objectos e suas relações, com base nas mensagens trocadas entre os objectos. Um **diagrama de sequência** é um diagrama de interacção que evidencia a sequência temporal existente entre as mensagens. Um **diagrama de colaboração** é um diagrama de interacção que retrata a organização estrutural dos objectos que recebem e enviam mensagens.

Diagrama de estados. Estes diagramas integram um conjunto de estados, transições, eventos e actividades. Representam características dinâmicas, e são normalmente utilizados na modelação do comportamento de uma interface, classe ou colaboração, evidenciando a ordem de eventos que caracterizam o comportamento de um objecto.

Diagrama de actividades. Constituem um caso particular dos diagramas de estados, permitindo representar o fluxo de actividades dentro de um sistema.

Diagrama de componentes. Permitem retratar a organização e dependência existente entre um conjunto de componentes. São utilizados na modelação de uma vista estática da implementação dum sistema.

Diagrama de desenvolvimento. Permite retratar a configuração dos nodos de processamento e dos componentes ligados aos mesmos. Estes diagramas estão relacionados com o diagrama de componentes, uma vez que um nodo agrega normalmente um ou mais componentes.

O UML tem sido amplamente utilizado na área da informação geográfica, sendo a linguagem de modelação utilizada pelo ISO TC 211, a comissão técnica da *International Standard Organisation* (ISO) com responsabilidades na produção de normas para a **Informação Geográfica e Geomática** [ISO/TC-211 1999]. Esta comissão foi constituída em Abril de 1994 e é presidida pela Noruega, englobando 27 países observadores e 29 países colaboradores, sendo este último o estatuto de Portugal.

A Figura 9 apresenta um exemplo de utilização do UML na modelação de informação geográfica, o qual retrata a caracterização dos aspectos geométricos e topológicos da informação geográfica, recorrendo ao esquema espacial G1 que integra as pré-normas europeias produzidas pela comissão técnica TC 287 do Comité Europeu de Normalização (CEN) [CEN/TC-287 1996]. Esta comissão técnica foi constituída em Outubro de 1991, e integrou 22 países, entre os quais Portugal. Apesar do CEN TC 287 não ter utilizado o UML na especificação das pré-normas europeias, dois dos esquemas que integram as directivas do CEN, o esquema espacial G1 e o esquema de identificadores geográficos, foram adaptados por Santos [Santos 2001] para a notação UML.

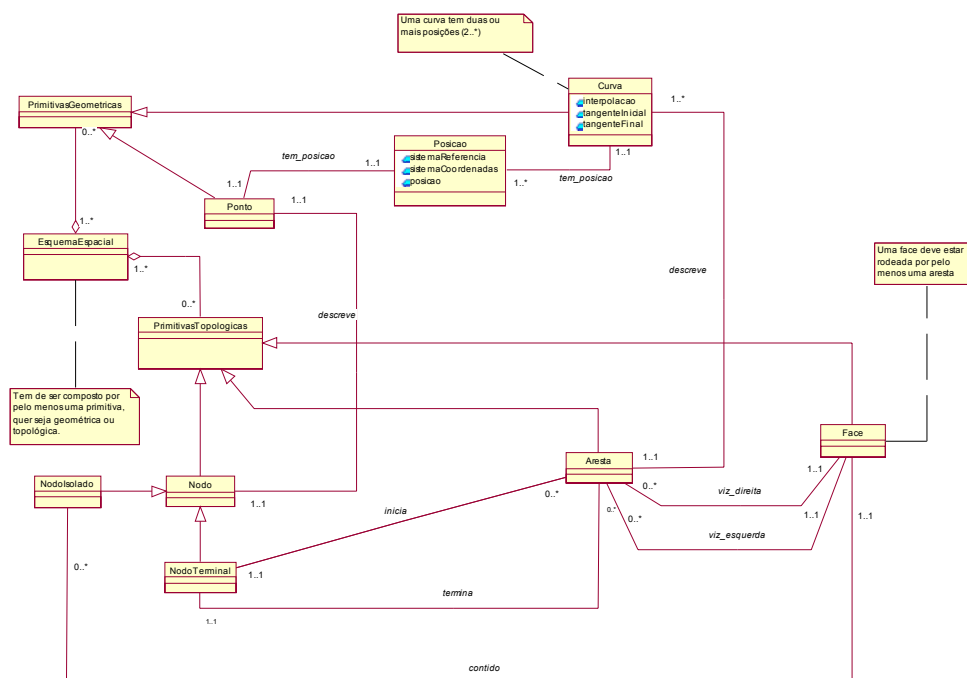


Figura 9 – Esquema espacial G1: primitivas geométricas e topológicas

O esquema espacial G1 permite caracterizar mapas de topologia planar [CEN/TC-287 1996]. Deve ser utilizado sempre que o domínio geográfico em causa for constituído por mapas planos com faces que cobrem toda a superfície em análise, não existindo qualquer sobreposição entre as mesmas. Este esquema possui duas ramificações: uma geométrica e uma topológica. As primitivas topológicas permitem descrever a conectividade existente entre as entidades

geográficas, propriedades que permanecem invariantes a transformações do espaço, como sejam mudanças de escala ou de sistema de referência quantitativo. As primitivas geométricas permitem a descrição, total ou parcial, de um objecto recorrendo a coordenadas ou funções matemáticas.

7 Conclusões

Dados geográficos são dados respeitantes a fenómenos associados implícita ou explicitamente a uma localização relativa à superfície da Terra, sendo representações, tratáveis automaticamente, de informação geográfica. Um conjunto de dados geográficos inclui entre os dados que o constituem, pelo menos um aspecto espacial, que permite a definição das características geométricas e topológicas associadas ao mesmo.

A modelação de aplicações geográficas requer a utilização de técnicas específicas, que permitam armazenar e processar dados espaciais. Estes dados representam objectos cuja posição no espaço é relevante, definindo relações topológicas entre os mesmos.

A modelação de aplicações geográficas pode ser realizada estendendo metodologias convencionais, ou utilizando técnicas OO, que tradicionalmente não requerem qualquer adaptação para utilização no domínio geográfico.

Este artigo apresentou diversas técnicas de modelação, que podem ser utilizadas na definição do modelo conceptual de dados, ou ainda, na estruturação do modelo lógico de dados. Ao nível conceptual apresentou-se o modelo E-R, que permite a modelação de informação geográfica. Nesta abordagem, objectos espaciais são representados como *entidades*, cujas propriedades estruturais e topológicas são representadas através de *relacionamentos* entre entidades.

O modelo geo-relacional permite a definição de modelos lógicos de dados, assim como vistas sobre os mesmos. Estes modelos são construídos recorrendo a *níveis*, *relações*, *níveis virtuais*, *classes de objectos* e *restrições de integridade*, permitindo que a informação esteja armazenada em diversas tabelas, relacionadas através de atributos comuns. As ligações estabelecidas permitem que a pesquisa de informação possa ser iniciada nos dados espaciais ou nos dados não espaciais.

A utilização de paradigmas lógicos na programação requer a utilização de abordagens formais na especificação dos requisitos de informação, pelo que também foi apresentada uma abordagem formal na modelação lógica de aplicações geográficas. Na abordagem apresentada, predicados em lógica de 1ª ordem permitem a definição de *conhecimento básico* e *conhecimento virtual*. Predicados em lógica de 2ª ordem são utilizados na definição das regras que possibilitam a incorporação no sistema, de mecanismos de raciocínio espacial e temporal.

O UML pode ser utilizado para visualizar, especificar, construir e documentar os componentes de uma aplicação. Não necessita de qualquer adaptação ou extensão para a sua utilização no domínio geográfico. Esta linguagem permite a modelação de dados tanto ao nível conceptual como ao nível lógico. Constitui, pela sua ampla utilização, uma opção muito apropriada à modelação de informação geográfica, sendo a linguagem adoptada pela comissão técnica da ISO encarregue da produção de normas para o domínio da *Informação Geográfica e Geomática*.

8 Referências

- Adam, N. R., e A. Gangopadhyay, *Database Issues in Geographic Information Systems*, Kluwer Academic Publishers, Massachusetts, 1997.
- Aronoff, S., *Geographic Information Systems: a management perspective*, WDL Publications, Ottawa, 1989.

- Booch, G., J. Rumbaugh, e I. Jacobson, *The Unified Modeling Language User Guide*, Addison Wesley Longman, Inc., 1999.
- Burrough, P. A., *Principles of Geographical Information Systems for Land Resources Assessment*, Oxford University Press, Oxford, 1986.
- CEN/TC-287, *Geographic Information: Data Description, Spatial Schema*, European Committee for Standardisation, prENV 12160, 1996.
- Chen, P. P., "The Entity-Relationship Model: Toward a Unified View of Data", *ACM Transactions on Database Systems*, 1, 1 (1976), 9-35.
- Codd, E. F., "A Relational Model for Large Shared Data Banks", *Communications of the ACM*, 13, 6 (1970), 377-387.
- Güting, R., "An Introduction to Spatial Database System", *The VLDB Journal*, 3, 4 (1994), 357-400.
- Hadzilacos, T., e N. Tryfona, "Logical Data Modelling for Geographical Applications", *International Journal of Geographical Information Systems*, 10, 2 (1996), 179-203.
- ISO/TC-211, *Geographic Information - Overview*, International Standard Organisation, N723, 1999.
- Laurini, R., e D. Thompson, *Fundamentals of Spatial Information Systems*, Academic Press, San Diego, 1992.
- Navathe, S. B., "Evolution of Data Modeling for Databases", *Communications of the ACM*, 35, 9 (1992), 112-123.
- Pereira, J. L., *Tecnologia de Bases de Dados*, FCA - Editora de Informática, Lisboa, 1997.
- Roman, G.-C., "Formal specification of geographic data processing requirements", *IEEE Transactions on Knowledge and Data Engineering*, 2, 4 (1990), 370-380.
- Santos, M. Y., *PADRÃO: Um Sistema de Descoberta de Conhecimento em Bases de Dados Geo-referenciadas*, Tese de Doutoramento, Universidade do Minho, 2001.
- Shekhar, S., M. Coyle, B. Goyal, D.-R. Liu, e S. Sarkar, "Data Models in Geographic Information Systems", *Communications of the ACM*, 10, 4 (1997), 103-111.
- Shepherd, I., *Information Integration and GIS*, in Maguire, D. J., M. F. Goodchild, e D. W. Rhind (Eds.), *Geographical Information Systems: Principles and Applications*, Longman Scientific and Technical, Harlow, 1991.