

# TDA - Um Método para Indexar Topologicamente Atributos Espaciais

Maurício R. Mediano, Marco A. Casanova, Marcelo Gattass  
Departamento de Informática  
Pontifícia Universidade Católica do Rio de Janeiro,  
Rua Marquês de São Vicente, 225, 22453-900 Rio de Janeiro, RJ, Brasil  
mediano,casanova,gattass@inf.puc-rio.br

## Resumo

Este trabalho introduz um método para indexar topologicamente um conjunto de atributos espaciais, chamado de *TDA - Topological Data Access Method*. Em determinadas circunstâncias, a representação topológica é ordens de grandeza mais compacta do que a representação geométrica exata dos atributos espaciais, oferecendo considerável ganho no processamento dos operadores topológicos. Este trabalho apresenta inicialmente o método de forma abstrata, incluindo o formalismo utilizado na definição da representação geométrica dos atributos espaciais e detalhes da estrutura topológica utilizada. Em seguida, discute questões ligadas ao armazenamento persistente das estruturas de dados subjacentes ao método.

## 1 Introdução

Sistemas de Informação Geográfica, ou simplesmente SIGs, são sistemas de informação construídos especialmente para armazenar, analisar e manipular objetos geográficos, ou seja, objetos que representam artefatos ou fenômenos em que a localização geográfica é uma característica inerente e indispensável para tratá-los [5]. SIGs tipicamente oferecem *linguagens de consulta espacial* que explicitamente incorporam *operadores espaciais* descrevendo relacionamentos (ou operações) entre objetos geográficos. Dentre a variedade de operadores espaciais, encontram-se os *operadores topológicos*, como *adjacente*, *intercepta*, etc..., que definem relacionamentos estritamente topológicos entre os objetos geográficos. Embora muito úteis para análise geográfica, estes operadores terão um custo bastante elevado se a representação geométrica da localização geográfica dos objetos for utilizada diretamente.

Este trabalho propõe então armazenar uma representação do relacionamento topológico entre os objetos geográficos e utilizar esta informação para otimizar o processamento dos operadores topológicos, em lugar de utilizar diretamente a representação geométrica da localização geográfica dos objetos. Em determinadas circunstâncias, esta representação adicional é ordens de grandeza mais compacta do que a representação geométrica exata, oferecendo considerável ganho no processamento das consultas. Esta representação foi batizada de *TDA - Topological Data Access Method*.

A quase totalidade dos esforços para otimizar o desempenho de consultas espaciais se concentraram em desenvolver estruturas de dados adequadas para armazenar e recuperar a

representação geométrica da localização geográfica dos objetos. Gaede [10] apresenta uma cronologia com a história de 54 métodos de acesso desta natureza. Já Mediano e outros [15] propõem um método de acesso para mapas baseado na estrutura de dados topológica *Winged-Edge*, inicialmente proposta por Baumgart [1]. Mediano e outros [16] propõem ainda um outro método de acesso para mapas com multiresolução geométrica e topológica baseado na estrutura de dados topológica HPS, inicialmente proposta por Celes e outros [9]. Saalfeld [18] apresenta métodos para ordenar entidades espaciais quanto a proximidade no espaço, chamados *tree-orderings*, que podem ser construídos a partir de estruturas de dados topológicas. Segundo Martha [11]: “A grande contribuição do uso de estruturas de dados topológicas é que consultas (por adjacência) ao banco de dados são executadas localmente e com algoritmos cuja complexidade é, no pior caso, linearmente proporcional ao número de elementos topológicos do resultado.”

O estudo de quais operadores topológicos devem ser adotados é um capítulo à parte. Egenhofer e Franzosa propõem [7] um *framework* para a definição de relações topológicas. A partir da matriz de 4 interseções  $M_{4IM}$  são geradas 16 combinações possíveis de interseções entre a fronteira ( $\partial$ ) e o interior ( $^{\circ}$ ) de dois atributos espaciais de dimensão dois,  $s_{o_i}$  e  $s_{o_j}$ . Egenhofer e Herring [8] ampliam a matriz do método de 4 interseções para 9 interseções. A matriz  $M_{9IM}$  do método de 9 interseções gera um total de 512 combinações de interseções entre a fronteira, o interior e o exterior ( $^-$ ) de dois atributos espaciais de dimensões zero, um ou dois. Egenhofer [6] prova que o método de 9 interseções é mais adequado para tratar atributos espaciais de dimensão um. Clementini e outros [4] propõem um conjunto de 8 operadores, chamado de *Calculus-based Method*, ou simplesmente CBM. No mesmo trabalho, os autores definem ainda outro método, chamado de método de 4 interseções estendido por dimensão, ou DE+4IM, onde é levada em consideração a dimensão dos resultados das interseções na matriz  $M_{4IM}$ . Por último, os autores provam que os operadores do método CBM são mais expressivos que os operadores do método DE+4IM. Clementini e Di Felice [2] definem um outro método, chamado de método de 9 interseções estendido por dimensão, ou DE+9IM, onde é levada em consideração a dimensão dos resultados das interseções na matriz  $M_{9IM}$ . Os autores provam ainda que o método CBM é equivalente ao método DE+9IM.

Este trabalho está organizado da seguinte forma. A seção 2 apresenta um resumo de conceitos utilizados ao longo do trabalho, incluindo o formalismo utilizado na definição da representação geométrica da localização geográfica dos objetos e dos operadores topológicos e detalhes da estrutura topológica utilizada. A seção 3 descreve o método TDA de indexação topológica. A seção 4 discute o armazenamento persistente das estruturas de dados subjacentes ao método. Finalmente, a seção 5 apresenta as conclusões e os trabalhos futuros.

Remetemos o leitor a Mediano [13] para uma definição detalhada do método TDA, incluindo uma análise do seu desempenho.

## 2 Preliminares

### 2.1 Conceitos Básicos

Um *atributo espacial* de um objeto geográfico  $o_i$  é qualquer atributo de  $o_i$  que contém uma definição da localização geográfica de  $o_i$ . Um objeto geográfico pode naturalmente ter mais de um atributo espacial. No contexto deste trabalho, estaremos interessados em atributos espaciais

cujos valores representem, no caso mais geral, conjuntos de *pontos complexos*, *linhas complexas* ou *áreas complexas*, definidos de acordo com a *point-set topology* de Clementini e Di Felice [3]. Chamamos a representação interna do valor de um atributo espacial de *representação geométrica* do atributo espacial.

Intuitivamente, quando projetamos no plano os valores de um conjunto de atributos espaciais referentes a uma dada região geográfica, quebrando as linhas nos cruzamentos, as regiões nas subregiões que se superpõem, etc... induzimos uma *subdivisão planar* daquela região, ou seja, um conjunto de pontos, linhas e áreas complexas que cobrem completamente a região e que só se tocam nas suas fronteiras.

O resto desta seção resume então as principais definições da *point-set topology*, define a representação geométrica adotada e discute questões associadas ao conceito de subdivisão planar.

## 2.2 Conceitos de Point-Set Topology

Seguindo Egenhofer e Franzosa [7], dado um ponto, linha ou área complexa  $T$ , usaremos a notação  $T^\circ$ ,  $\partial T$ ,  $T^-$ ,  $\bar{T}$  e  $\dim(T)$  para denotar o interior, fronteira, exterior, fechamento e dimensão de  $T$ , respectivamente

Um *ponto* é um *point-set* de dimensão zero (não vazio), que consiste de apenas um elemento. Um *ponto complexo*  $P$  é um *point-set* de dimensão zero (não vazio), que consiste de um número finito de elementos distintos  $P_1, \dots, P_n$ . A fronteira  $\partial P$  de um ponto complexo é vazia. Como consequência, o interior  $P^\circ$  de um ponto complexo  $P$  é igual à união de todos os elementos em  $P$ .

Uma *linha simples* é um *point-set* fechado  $L$  tal que  $\dim(L) = 1$ , definido como a imagem de um mapeamento contínuo  $f : [0, 1] \rightarrow \mathfrak{R}^2$ , tal que não admite auto-interseções. Os mapeamentos de 0 e 1 em  $f$  são os pontos finais que formam a fronteira de uma linha simples no plano. Removendo a restrição, temos a noção de *linha com auto-interseção*. Sejam  $f_1, f_2, \dots, f_n$  mapeamentos contínuos do intervalo  $[0, 1]$  no plano, chamamos de *linha complexa* qualquer *point-set* fechado  $L$  tal que  $\dim(L) = 1$ , definido como a união da imagem das funções  $f_1, f_2, \dots, f_n$ .

A noção de área complexa é definida a partir de noções de *conectividade* e *componente*. Uma *separação* de um *point-set*  $A \subset \mathfrak{R}^2$  é um par de conjuntos abertos disjuntos não vazios  $A_1$  e  $A_2$  cuja a união é  $A$ . O *point-set*  $A$  é *conectado* se não existirem separações em  $A$ , caso contrário,  $A$  é *desconectado* e  $A_1$  e  $A_2$  são chamados de *componentes* de uma separação de  $A$ .

Seja  $A$  um *point-set* tal que  $\dim(A) = 2$ . O *point-set*  $A$  é dito *regular fechado* se e somente se  $A = \bar{A}^\circ$ . Uma *área simples* é um *point-set* regular fechado (não vazio) de duas dimensões com interior conectado e o exterior conectado. Separações do exterior implicam que existe um *exterior externo* (conjunto sem fronteiras) e  $n > 0$  *exteriores internos* (conjuntos com fronteiras). O exterior externo é denotado por  $A_0^-$  e os exteriores internos por  $A_1^- \dots A_n^-$ . Sua união define o exterior  $A^-$  de  $A$ .

Uma *área com furos*  $A$  é um *point-set* regular fechado de duas dimensões (não vazio) com o interior conectado, tal que a interseção dos fechamentos de quaisquer dois exteriores diferentes é vazia ou igual a um conjunto finito de pontos.

Uma *área complexa*  $A$  é um *point-set* fechado de duas dimensões com componentes fechados  $A_1 \dots A_n$ , onde para todo  $i \in [1, n]$  cada componente  $A_i$  é uma área simples ou uma área

com furos; e para todo  $i, j \in [1, n]$  com  $i \neq j$ :

$$(A_i^\circ \cap A_j^\circ = \emptyset) \wedge ((\partial A_i \cap \partial A_j = \emptyset) \vee (\exists p_1, \dots, p_k \in \mathbb{R}^2)(\partial A_i \cap \partial A_j = \{p_1 \dots p_k\})). \quad (1)$$

### 2.3 A Representação Geométrica de Atributos Espaciais

Seja  $s_{o_i}$  um atributo espacial de um objeto geográfico  $O_i$ . Recorde que, no caso mais geral, o valor de  $s_{o_i}$  em um estado do banco de dados geográfico é um conjunto de *pontos complexos*, *linhas complexas* ou *áreas complexas*. Por abuso de notação, usaremos  $s_{o_i}^\circ$ ,  $\partial s_{o_i}$ ,  $s_{o_i}^-$ ,  $\overline{s_{o_i}}$  e  $\dim(s_{o_i})$  para denotar o interior, fronteira, exterior, fechamento e dimensão do valor de  $s_{o_i}$ , respectivamente.

Adotaremos neste trabalho uma particular representação geométrica para o valor de  $s_{o_i}$ , denotada por  $g(s_{o_i})$ , e definida indutivamente da seguinte forma:

- um ponto é representado por uma seqüência de coordenadas;
- uma linha poligonal é composta por um número  $n$  que indica quantos pontos a linha poligonal contém, seguido por uma seqüência de  $n$  pontos;
- uma linha poligonal fechada é representada por uma linha poligonal, onde o primeiro e o último pontos são iguais;
- um laço é representado por uma linha poligonal fechada;
- uma região é representada por um número  $n$  que indica quantos laços a região contém, seguido por uma seqüência de  $n$  laços; o primeiro laço é o laço externo, os demais são furos da região; a orientação do laço externo está sempre no sentido anti-horário e a orientação dos furos está sempre no sentido horário;
- quando  $\dim(s_{o_i}) = 0$ ,  $g(s_{o_i})$  é representada por um número que indica quantos pontos  $g(s_{o_i})$  contém, seguido por uma seqüência de pontos;
- quando  $\dim(s_{o_i}) = 1$ ,  $g(s_{o_i})$  é representada por um número que indica quantas linhas poligonais  $g(s_{o_i})$  contém, seguido por uma seqüência de linhas poligonais; e
- quando  $\dim(s_{o_i}) = 2$ ,  $g(s_{o_i})$  é representada por um número que indica quantas regiões  $g(s_{o_i})$  contém, seguido por uma seqüência de regiões.

### 2.4 A Estrutura de Dados Topológica Half-Edge

Neste trabalho, adotaremos a estrutura topológica Half-Edge para representação de subdivisões planares. Comentaremos nesta seção apenas as referências, apresentadas na Figura 1, entre os descritores dos objetos que compõem a estrutura de dados topológica Half-Edge, referindo o leitor a Mäntylä [17] para maiores detalhes.

Cada face, vértice, aresta, laço ou semi-aresta da Half-Edge possui as seguintes referências:

- cada descritor de face possui uma referência para o seu laço externo e uma lista de laços, composta pelo laço externo e pelos furos da face (laços internos);

- cada descritor de aresta possui uma referência para a sua representação geométrica e duas referências para semi-arestas;
- cada descritor de vértice possui a sua representação geométrica (suas coordenadas) e uma referência para uma semi-aresta adjacente ao vértice;
- cada descritor de semi-aresta possui: duas referências que formam um elo de uma lista duplamente encadeada de semi-arestas, uma referência para o laço do qual ela faz parte, uma referência para a aresta que ela representa e uma referência para seu vértice inicial; e
- cada descritor de laço possui uma referência para a face da qual faz parte e uma referência para uma das semi-arestas que o compõem o laço.
- cada descritor de modelo possui três listas: de faces, de vértices e de arestas.

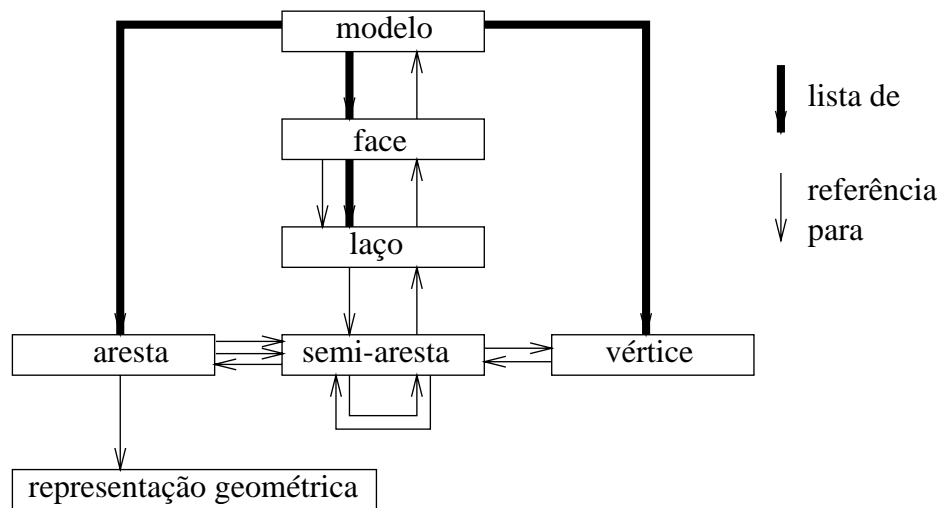


Figura 1: As referências entre os objetos da estrutura topológica Half-Edge.

### 3 O Método TDA

Apresentamos nesta seção os conceitos básicos do método TDA. Os detalhes de armazenamento, uma arquitetura em camadas para o método e um exemplo são deixados para a seção seguinte.

Seja  $O = \{o_1, o_2, \dots, o_n\}$  um conjunto de objetos geográficos e  $S_O = \{s_{o_1}, s_{o_2}, \dots, s_{o_n}\}$  um conjunto de atributos espaciais dos objetos em  $O$  (recorde que cada objeto pode ter mais de um atributo espacial), onde  $s_{o_i}$  é o atributo espacial escolhido de  $o_i$  e  $g(s_{o_i})$  é a representação geométrica de  $s_{o_i}$ .

A proposta do método TDA consiste em manter dinamicamente uma subdivisão planar armazenada em uma estrutura de dados topológica  $T_{S_O}$ , associada a  $S_O$ , tal que os valores dos atributos espaciais em  $S_O$  possam ser representados por conjuntos de faces, arestas e vértices

de  $T_{S_O}$  (chamados de *componentes* de  $T_{S_O}$ ). A estrutura  $T_{S_O}$  será atualizada dinamicamente a partir da representação geométrica de cada atributo espacial inserido ou removido em  $S_O$  de tal forma a manter esta propriedade.

Sejam  $F_{T_{S_O}} = \{f_1, \dots, f_k\}$ ,  $E_{T_{S_O}} = \{e_1, \dots, e_m\}$  e  $V_{T_{S_O}} = \{v_1, \dots, v_n\}$  os conjuntos (de teoria dos conjuntos), de faces, arestas e vértices de  $T_{S_O}$ .

Dado um atributo espacial  $s_{o_i}$  podemos definir os seguintes conjuntos:

- $F_{o_i}^\circ$ ,  $E_{o_i}^\circ$  e  $V_{o_i}^\circ$  contém as faces, arestas e vértices de  $T_{S_O}$  que representam o interior  $s_{o_i}^\circ$  de  $s_{o_i}$ :

$$F_{o_i}^\circ = \{c \in F_{T_{S_O}} \mid c^\circ \subseteq s_{o_i}^\circ\} \quad (2)$$

$$E_{o_i}^\circ = \{c \in E_{T_{S_O}} \mid c^\circ \subseteq s_{o_i}^\circ\} \quad (3)$$

$$V_{o_i}^\circ = \{c \in V_{T_{S_O}} \mid c^\circ \subseteq s_{o_i}^\circ\} \quad (4)$$

- $\partial F_{o_i}$ ,  $\partial E_{o_i}$  e  $\partial V_{o_i}$  contém as faces, arestas e vértices de  $T_{S_O}$  que representam a fronteira  $\partial s_{o_i}$  de  $s_{o_i}$ :

$$\partial F_{o_i} = \emptyset \quad (5)$$

$$\partial E_{o_i} = \{c \in E_{T_{S_O}} \mid c^\circ \subset \partial s_{o_i}\} \quad (6)$$

$$\partial V_{o_i} = \{c \in V_{T_{S_O}} \mid c^\circ \subseteq \partial s_{o_i}\} \quad (7)$$

- $F_{o_i}^-$ ,  $E_{o_i}^-$  e  $V_{o_i}^-$  contém as faces, arestas e vértices de  $T_{S_O}$  que representam o exterior  $s_{o_i}^-$  de  $s_{o_i}$ :

$$F_{o_i}^- = \{c \in F_{T_{S_O}} \mid c^\circ \subseteq s_{o_i}^-\} \quad (8)$$

$$E_{o_i}^- = \{c \in E_{T_{S_O}} \mid c^\circ \subset s_{o_i}^-\} \quad (9)$$

$$V_{o_i}^- = \{c \in V_{T_{S_O}} \mid c^\circ \subset s_{o_i}^-\} \quad (10)$$

Note que  $F_{o_i}^\circ$ ,  $E_{o_i}^\circ$ ,  $V_{o_i}^\circ$ ,  $\partial F_{o_i}$ ,  $\partial E_{o_i}$ ,  $\partial V_{o_i}$ ,  $F_{o_i}^-$ ,  $E_{o_i}^-$  e  $V_{o_i}^-$  não são *point-sets*, são conjuntos de teoria dos conjuntos; por exemplo, não faz sentido afirmar que  $F_{o_i}^\circ$  é o interior de  $F_{o_i}$ , pois a definição de  $F_{o_i}$  a partir de  $F_{o_i}^\circ$  não é óbvia e não faz sentido; os operadores  $\cap$  e  $\cup$  quando aplicados a esses conjuntos são os operadores de interseção e união de teoria dos conjuntos, respectivamente. Em contra-partida, a representação geométrica de faces, vértices e arestas e a representação geométrica de atributos espaciais representam *point-sets*; os operadores  $\cap$  e  $\cup$  aplicados a *point-sets* são os operadores de interseção e união de *point-set topology*, respectivamente. O leitor poderá distinguir se os operadores  $\cap$ ,  $\cup$ ,  $\subset$  etc, são relativos à teoria dos conjuntos ou à *point-set topology* em função dos operandos envolvidos.

O método TDA mantém, para cada atributo espacial  $s_{o_i}$ , uma representação  $t_{o_i}$  chamada de *representação topológica* de  $s_{o_i}$ . A representação topológica  $t_{o_i}$  é formada por componentes de  $T_{S_O}$  tal que:

$$t_{o_i} = F_{o_i}^\circ \cup E_{o_i}^\circ \cup V_{o_i}^\circ \cup \partial F_{o_i} \cup \partial E_{o_i} \cup \partial V_{o_i}. \quad (11)$$

Vale notar ainda as seguintes propriedades existentes entre componentes de  $T_{S_O}$  e um atributo espacial  $s_{o_i} \in S_O$ :

$$F_{T_{S_O}} = F_{o_i}^\circ \cup \partial F_{o_i} \cup F_{o_i}^-, \quad (12)$$

$$E_{T_{S_O}} = E_{o_i}^\circ \cup \partial E_{o_i} \cup E_{o_i}^- \quad (13)$$

$$V_{T_{S_O}} = V_{o_i}^\circ \cup \partial V_{o_i} \cup V_{o_i}^-. \quad (14)$$

Dada a dimensão do atributo espacial, alguns dos nove conjuntos definidos em (2), (3), (4), (5), (6), (7), (8), (9) e (10) são vazios, pois  $s_{o_i}^\circ$  contém apenas componentes de  $T_{S_O}$  de uma dimensão menor ou igual à dimensão de  $s_{o_i}$  e  $\partial s_{o_i}$  contém apenas componentes de  $T_{S_O}$  de dimensão menor que a dimensão de  $s_{o_i}$  ( $s_{o_i}^-$  contém componentes de  $T_{S_O}$  de qualquer dimensão).

	$F_{o_i}^\circ$	$E_{o_i}^\circ$	$V_{o_i}^\circ$	$\partial F_{o_i}$	$\partial E_{o_i}$	$\partial V_{o_i}$	$F_{o_i}^-$	$E_{o_i}^-$	$V_{o_i}^-$
$dim(s_{o_i}) = 0$	$\emptyset$	$\emptyset$		$\emptyset$	$\emptyset$	$\emptyset$			
$dim(s_{o_i}) = 1$	$\emptyset$			$\emptyset$	$\emptyset$				
$dim(s_{o_i}) = 2$				$\emptyset$					

Tabela 1: Conjuntos de componentes vazios em função da dimensão do atributo espacial.

A partir da Tabela 1 podemos simplificar (11), (12), (13) e (14) em função da dimensão de  $s_{o_i}$ :

- quando  $dim(s_{o_i}) = 2$  temos que:

$$t_{o_i} = F_{o_i}^\circ \cup E_{o_i}^\circ \cup V_{o_i}^\circ \cup \partial E_{o_i} \cup \partial V_{o_i}, \quad (15)$$

$$F_{T_{S_O}} = F_{o_i}^\circ \cup F_{o_i}^-, \quad (16)$$

$$E_{T_{S_O}} = E_{o_i}^\circ \cup \partial E_{o_i} \cup E_{o_i}^- e \quad (17)$$

$$V_{T_{S_O}} = V_{o_i}^\circ \cup \partial V_{o_i} \cup V_{o_i}^-; \quad (18)$$

- quando  $dim(s_{o_i}) = 1$  temos que:

$$t_{o_i} = E_{o_i}^\circ \cup V_{o_i}^\circ \cup \partial V_{o_i}, \quad (19)$$

$$F_{T_{S_O}} = F_{o_i}^-, \quad (20)$$

$$E_{T_{S_O}} = E_{o_i}^\circ \cup E_{o_i}^- e \quad (21)$$

$$V_{T_{S_O}} = V_{o_i}^\circ \cup \partial V_{o_i} \cup V_{o_i}^-; e \quad (22)$$

- quando  $dim(s_{o_i}) = 0$  temos que:

$$t_{o_i} = V_{o_i}^\circ, \quad (23)$$

$$F_{T_{S_O}} = F_{o_i}^-, \quad (24)$$

$$E_{T_{S_O}} = E_{o_i}^- e \quad (25)$$

$$V_{T_{S_O}} = V_{o_i}^\circ \cup V_{o_i}^-. \quad (26)$$

Em (11) temos os seis subconjuntos de componentes de  $T_{S_O}$  com os quais podemos compor  $t_{o_i}$ . Ao restringir (11) para cada uma das dimensões 2, 1 e 0, temos (15), (19) e (23), respectivamente. Vamos propor duas formas de armazenar a representação topológica  $t_{o_i}$  de um atributo espacial  $s_{o_i}$ . A primeira forma é através da *representação topológica completa*, ou simplesmente CTR:

$$CTR(s_{o_i}) = \begin{cases} (F_{o_i}^\circ, E_{o_i}^\circ, V_{o_i}^\circ, \partial E_{o_i}, \partial V_{o_i}) & \text{se } dim(s_{o_i}) = 2, \\ (E_{o_i}^\circ, V_{o_i}^\circ, \partial V_{o_i}) & \text{se } dim(s_{o_i}) = 1, \\ (V_{o_i}^\circ) & \text{se } dim(s_{o_i}) = 0. \end{cases} \quad (27)$$

Os conjuntos  $F_{o_i}^-$ ,  $E_{o_i}^-$  e  $V_{o_i}^-$  não fazem parte de  $t_{o_i}$  em nenhuma das três dimensões por serem proporcionais ao número de componentes de  $T_{S_O}$ , como pode ser visto em (16), (17), (18), (20), (21), (22), (24), (25) e (26).

Note que, para armazenar  $t_{o_i}$ , podemos ainda omitir alguns dos subconjuntos da CTR. Em linhas gerais, quanto menor o número de componentes de  $T_{S_O}$  em  $t_{o_i}$ , menor o espaço ocupado em memória secundária pelas representações topológicas (o que é aconselhável em técnicas de indexação). Porém veremos que o armazenamento da CTR significa um ganho de eficiência no cálculo de operadores. O critério básico utilizado para escolher os subconjuntos de componentes de  $T_{S_O}$  que serão armazenados por  $t_{o_i}$  é o seguinte: os subconjuntos escolhidos devem ser suficientes para que não seja necessário recorrer à representação geométrica  $g(s_{o_i})$  de  $s_{o_i}$  ou à representação geométrica das arestas de  $T_{S_O}$  para calcular operadores topológicos.

A segunda forma de armazenar a representação topológica  $t_{o_i}$  de um atributo espacial  $s_{o_i}$  é através da *representação topológica reduzida*, ou apenas RTR:

$$RTR(s_{o_i}) = \begin{cases} (F_{o_i}^\circ, \partial E_{o_i}, \partial V_{o_i}) & \text{se } \dim(s_{o_i}) = 2, \\ (E_{o_i}^\circ, V_{o_i}^\circ, \partial V_{o_i}) & \text{se } \dim(s_{o_i}) = 1, \\ (V_{o_i}^\circ) & \text{se } \dim(s_{o_i}) = 0. \end{cases} \quad (28)$$

Em conclusão, podemos dizer que a representação topológica resumida é mais compacta que a completa. Porém, ao se utilizar a representação topológica resumida torna-se necessário navegar na estrutura de dados topológica para executar o cálculo de alguns operadores topológicos.

Para provar a abrangência do método TDA, Mediano [13] mostra que a matriz  $M_{DE+9IM}$ , definida como [3]:

$$M_{DE+9IM} = \begin{pmatrix} \dim(\partial s_{o_j} \cap \partial s_{o_i}) & \dim(\partial s_{o_j} \cap s_{o_i}^\circ) & \dim(\partial s_{o_j} \cap s_{o_i}^-) \\ \dim(s_{o_j}^\circ \cap \partial s_{o_i}) & \dim(s_{o_j}^\circ \cap s_{o_i}^\circ) & \dim(s_{o_j}^\circ \cap s_{o_i}^-) \\ \dim(s_{o_j}^- \cap \partial s_{o_i}) & \dim(s_{o_j}^- \cap s_{o_i}^\circ) & \dim(s_{o_j}^- \cap s_{o_i}^-) \end{pmatrix}.$$

que é a mais completa de todas as matrizes citadas na introdução, pode ser facilmente calculada a partir das representações topológicas introduzidas pelo método TDA.

## 4 Armazenamento

Para armazenar o conjunto  $S_O$ , indexado topologicamente de acordo com o método TDA, é preciso armazenar as representações topológicas dos atributos espaciais em  $S_O$  e uma estrutura topológica  $T_{S_O}$  associada a  $S_O$ . Vamos então descrever como serão armazenados  $T_{S_O}$  e  $S_O$ , nas Seções 4.2 e 4.3, exemplificando estes conceitos na Seção 4.4. Adiantamos que todos os descritores das estruturas de dados, utilizadas para armazenar  $S_O$  e  $T_{S_O}$ , possuem tamanho fixo e são alocados de forma persistente. Utilizamos o subsistema OPS de armazenamento, proposto por Mediano [12], para gerenciar a alocação e liberação de memória em meio persistente.

### 4.1 A Arquitetura em Camadas do Método TDA

A Figura 2 mostra a arquitetura em camadas do método TDA. Note que o descritor do atributo espacial no sistema de gerência de bancos de dados geográficos (SGBDG) referencia o descritor

do atributo espacial no índice topológico e vice-versa. Como consequência, o SGBDG é capaz de otimizar operadores topológicos através do acesso a representação topológica do atributo espacial e o método TDA é capaz de acessar a representação geométrica do atributo espacial armazenada no SGBDG.

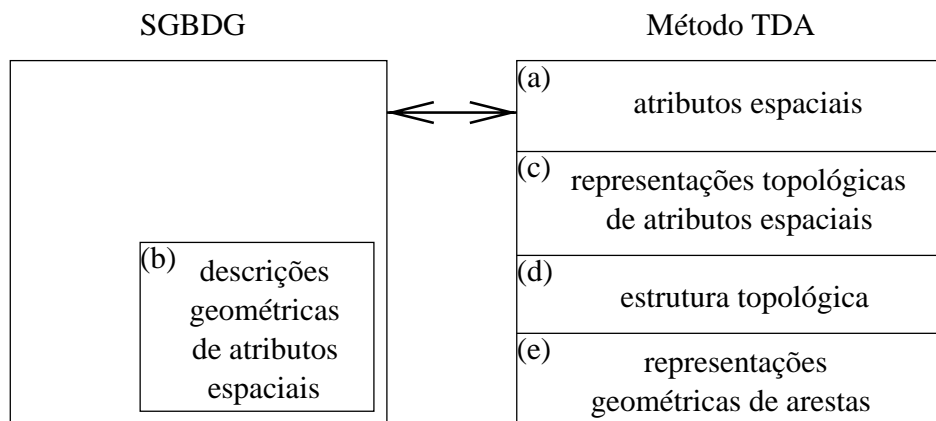


Figura 2: A arquitetura em camadas do método TDA.

## 4.2 O Armazenamento de $T_{S_o}$

A partir deste ponto, consideraremos que a estrutura topológica  $T_{S_o}$  é uma estrutura do tipo Half-Edge, apresentada na Seção 2.4. Para adaptar  $T_{S_o}$  em memória secundária, a primeira modificação a ser feita é substituir as listas de faces, arestas, vértices e laços por métodos de acesso adequados. Esta modificação é baseada na constatação de que, durante o acesso à essas listas, são feitas inserções e remoções de objetos espaciais e consultas do tipo: “Recupere os componentes na lista de faces (vértices, arestas ou laços) que interceptam uma determinada região ou que contém um ponto.” Sendo assim, é mais adequado, ainda que em memória principal, utilizar estruturas de dados que agrupam objetos espaciais próximos entre si. Dado que estas listas são armazenadas em memória secundária, escolhemos utilizar R-trees. O descritor de uma R-tree contém o número de níveis da R-tree e uma referência para a raiz da R-tree.

A representação geométrica do vértice é armazenada junto com o descritor do vértice. A representação geométrica de cada aresta deve ser armazenada através de um método de acesso para dados vetoriais que seja capaz de otimizar algoritmos geométricos utilizados pela estrutura topológica e de armazenar, de forma eficiente, a representação geométrica da aresta em memória secundária. A SV-tree (V-tree estática) proposta por Mediano e outros [14] é um método de acesso adequado a estes requisitos. A SV-tree possui quase 100% de ocupação em seus nós e é capaz de recuperar em tempo  $O(N)$  os  $N$  pontos que descrevem a linha poligonal armazenada. O descritor de uma SV-tree contém o número de níveis da SV-tree e uma referência para a raiz da SV-tree.

Remetemos o leitor a trabalhos similares, propostos por Mediano e outros [15, 16], que abordam em maior detalhe a adaptação das estruturas topológicas Winged-Edge e HPS em memória secundária, respectivamente.

### 4.3 O Armazenamento de $S_O$

A representação topológica  $t_{o_i}$  é armazenada através de B-trees, onde cada B-tree armazena um dos subconjuntos  $F_{o_i}^\circ, E_{o_i}^\circ, V_{o_i}^\circ, \partial E_{o_i}$  e  $\partial V_{o_i}$ . A representação  $t_{o_i}$  possui três B-trees quando  $\dim(s_{o_i}) = 2$  ou  $\dim(s_{o_i}) = 1$  e apenas uma B-tree quando  $\dim(s_{o_i}) = 0$ . As B-trees armazenam referências para os componentes de  $T_{S_O}$ . Para armazenar a CTR, seriam necessárias cinco B-trees quando  $\dim(s_{o_i}) = 2$ .

Armazenaremos uma R-tree associada ao conjunto  $S_O$  de atributos espaciais, indexado topologicamente através do método TDA, onde cada atributo espacial no conjunto  $S_O$  conterá uma entrada na R-tree de  $S_O$ . Cada entrada na R-tree de  $S_O$ , relativa à um atributo espacial  $s_{o_i} \in S_O$ , é composta pelo retângulo envolvente de  $g(s_{o_i})$  e uma referência para o descritor de  $s_{o_i}$ . O descritor do atributo espacial  $s_{o_i}$  contém um número que representa  $\dim(s_{o_i})$ , uma referência para o  $g(s_{o_i})$  no SGBDG e uma referência para o descritor de  $t_{o_i}$ . O descritor da representação topológica  $t_{o_i}$  contém, quando  $\dim(s_{o_i}) = 2$  ou  $\dim(s_{o_i}) = 1$ , três descritores de B-trees e, quando  $\dim(s_{o_i}) = 0$ , um descritor de B-tree. Cada descritor de B-tree contém o número de níveis da B-tree e uma referência para a raiz da B-tree.

### 4.4 Exemplos

A Figura 3 detalha a arquitetura em camadas do método TDA através do exemplo de um atributo espacial de dimensão 1, cuja representação topológica é composta por dois vértices ( $\partial V_{o_i}$ ) e uma aresta ( $E_{o_i}^\circ$ ). Subdividimos as 3 camadas do método TDA e a camada do SGBDG que manipula a representação geométrica do atributo espacial, da seguinte forma:

- a camada (a) dos atributos espaciais do método TDA é composta pelas camadas: (a1) que contém a R-tree que referencia os descritores dos atributos espaciais indexados topologicamente e (a2) que contém os descritores de atributos espaciais do método TDA;
- a camada (b) das representações geométricas de atributos espaciais armazenadas no SGBDG é composta pelas camadas: (b1) que contém os descritores de atributos espaciais no SGBDG e (b2) que contém as representações geométricas de atributos espaciais;
- a camada (c) das representações topológicas do método TDA é composta pelas camadas: (c1) que contém os descritores de representações topológicas e (c2) que contém as B-trees que armazenam as referências para os componentes da estrutura topológica que compõem as representações topológicas;
- a camada (d) da estrutura topológica do método TDA é composta pelas camadas: (d1) que contém as três R-trees de faces, vértices e arestas; (d2) que contém os descritores de faces; (d3) que contém as R-trees de laços (uma R-tree para cada face) e (d4) que contém os descritores de vértices, arestas, laços e semi-arestas; e
- a camada (e) das representações geométricas das arestas da estrutura topológica é composta pelas SV-trees que armazenam as representações geométricas das arestas da estrutura topológica.

O exemplo a seguir ilustra dois atributos espaciais  $s_{o_j}$  e  $s_{o_k}$  armazenados em um conjunto  $S_O$  indexado topologicamente pelo método TDA. A Figura 4 mostra  $s_{o_j}$  e  $s_{o_k}$ . Cada atributo é

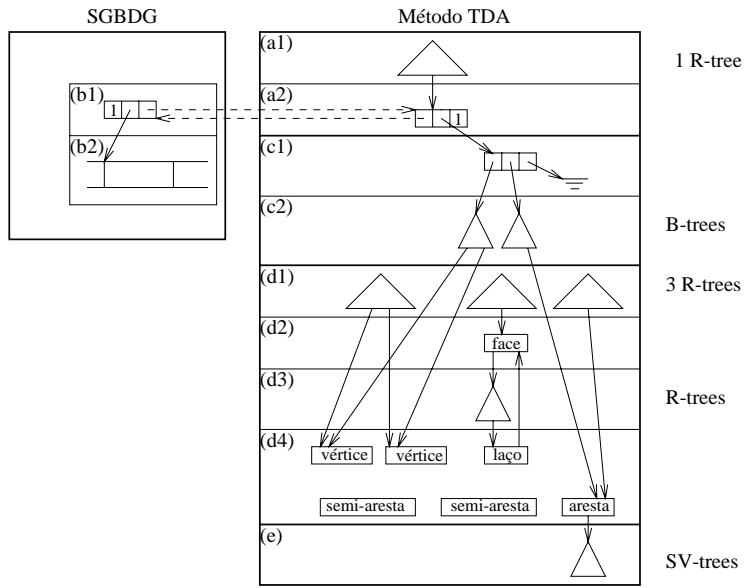


Figura 3: Um exemplo de um atributo espacial de dimensão um na arquitetura em camadas do método TDA.

formado por duas regiões separadas, ambos os atributos possuem um furo em uma das regiões. As Figuras 5, 6 e 7 mostram  $T_{S_O}$  (29),  $t_{o_j}$  (30) e  $t_{o_k}$  (31), respectivamente.

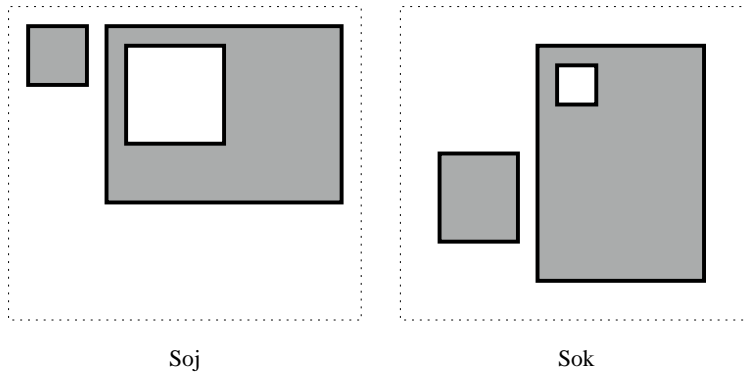


Figura 4: Os atributos espaciais  $s_{o_j}$  e  $s_{o_k}$ .

$$T_{S_O} = \begin{cases} F_{T_{S_O}} = \{f_0, f_1, f_2, f_3, f_4, f_5, f_6, f_7, f_8, f_9\} \\ E_{T_{S_O}} = \{e_1, e_2, e_3, e_4, e_5, e_6, e_7, e_8, e_9, e_{10}, e_{11}, e_{12}, e_{13}, e_{14}, e_{15}, e_{16}, e_{17}, e_{18}\} \\ V_{T_{S_O}} = \{v_1, v_2, v_3, v_4, v_5, v_6, v_7, v_8, v_9, v_{10}, v_{11}, v_{12}\} \end{cases} \quad (29)$$

$$t_{o_j} = \begin{cases} F_{o_j}^\circ = \{f_1, f_2, f_5, f_8\} \\ \partial E_{o_j} = \{e_1, e_2, e_3, e_4, e_5, e_6, e_7, e_8, e_9, e_{10}\} \\ \partial V_{o_j} = \{v_1, v_2, v_3, v_5, v_6, v_7, v_8, v_9, v_{10}, v_{12}\} \end{cases} \quad (30)$$

$$t_{o_k} = \begin{cases} F_{o_k}^\circ = \{f_4, f_5, f_7, f_8, f_9\} \\ \partial E_{o_k} = \{e_9, e_{11}, e_{12}, e_{13}, e_{14}, e_{15}, e_{16}, e_{17}, e_{18}\} \\ \partial V_{o_k} = \{v_4, v_5, v_6, v_7, v_8, v_9, v_{10}, v_{11}, v_{12}\} \end{cases} \quad (31)$$

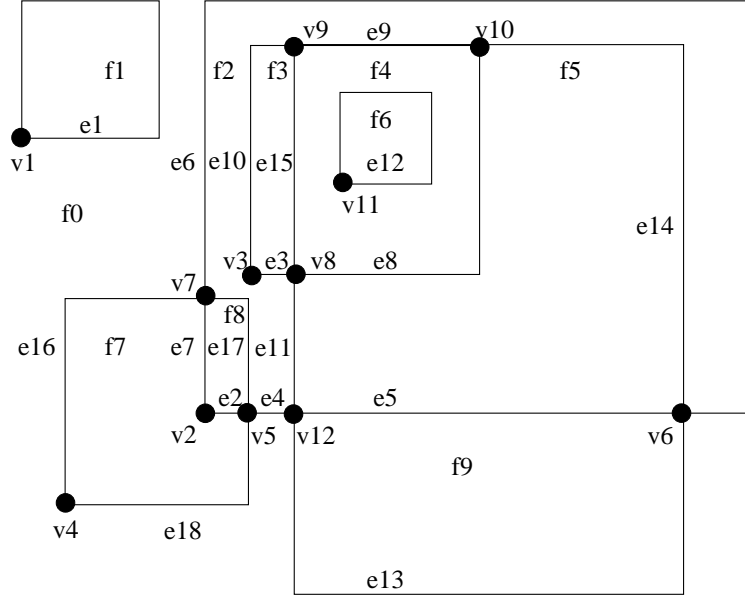


Figura 5: A estrutura topológica  $T_{S_O}$  após a inserção de  $s_{o_j}$  e  $s_{o_k}$ .

## 5 Conclusões

Introduzimos neste trabalho um método capaz de indexar topologicamente atributos espaciais complexos. Apesar do método TDA ser extenso, fizemos questão de enfatizar ao longo do texto que ele é fundamentado em formalismos tais como *point-set topology* e estruturas de dados topológicas, conceitos utilizados em modelagem geométrica.

Apresentamos dois tipos de representações topológicas de atributos espaciais, batizadas de RTR e CTR, onde a primeira é mais compacta do que a segunda. Mediano [13] contém resultados indicando que a complexidade dos operadores topológicos da matriz  $M_{DE+9IM}$  para a CTR é menor que a complexidade dos mesmos operadores para a RTR e que quanto maior for a sobreposição entre os atributos espaciais indexados, pior será o desempenho dos algoritmos do método TDA, pois o tamanho das representações topológicas é diretamente proporcional à sobreposição entre os atributos espaciais. Ou seja, quanto menor for o tamanho das

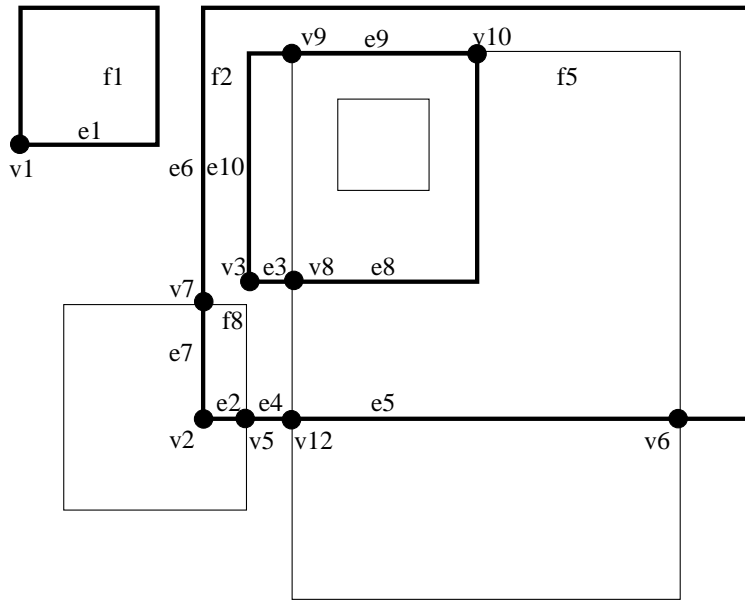


Figura 6:  $t_{o_j}$  após a inserção de  $s_{o_j}$  e  $s_{o_k}$ .

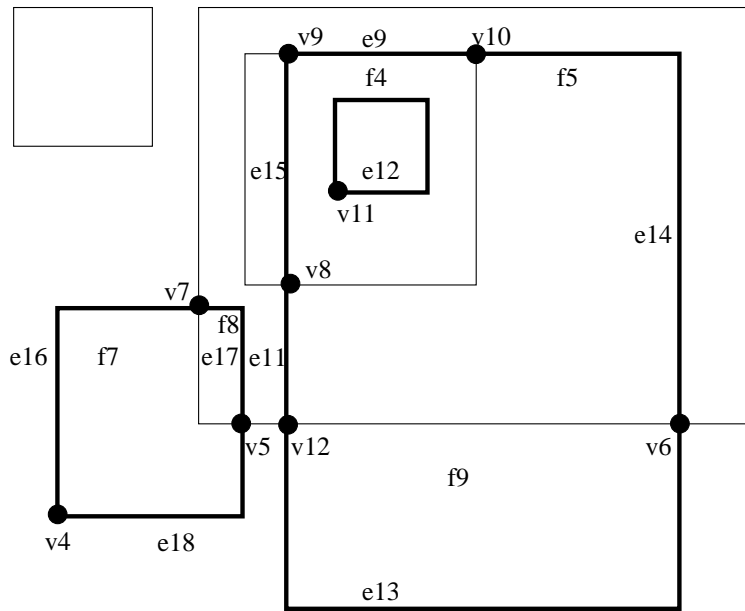


Figura 7:  $t_{o_k}$  after the insertion of  $s_{o_j}$  e  $s_{o_k}$ .

representações topológicas quando comparadas às representações geométricas, melhor será o desempenho do método TDA.

Quanto a trabalhos futuros, sugerimos investigar, entre outras técnicas de agrupamento e *cache* para que o método TDA, técnicas especiais de controle de concorrência para permitir bom desempenho e concorrência durante acessos simultâneos ao índice.

## Referências

- [1] B. Baumgart. Winged-edge polyedron representation. In *AFIPS Proc.*, number 44 in 1, pages 589–596, 1975.
- [2] E. Clementini and P. Di Felice. A Comparison of Methods for Representing Topological Relationships. *Information Sciences*, 3(3):149–178, 1995.
- [3] E. Clementini and P. Di Felice. A Model for Representing Topological Relationships Between Complex Geometric Features in Spatial Databases. *Information Sciences*, 90(1-4):121–136, 1996.
- [4] E. Clementini, P. Di Felice, and P. van Oosteron. A Small Set of Formal Topological Relationships Suitable for End-User Interaction. In *Proc. SSD*, pages 277–295, 1993.
- [5] G. Câmara, M. Casanova, A. Hemerly, G. Magalhães, and C. Medeiros. *Anatomia de Sistemas de Informação Geográfica*. 10<sup>a</sup> Escola de Computação, 1996.
- [6] M. Egenhofer. Definitions of Line-Line Relations for Geographic Databases. *Data Engineering*, 16(6):40–45, 1993.
- [7] M. Egenhofer and R. Franzosa. Point-set topological spatial relations. *International Journal of Geographical Information Systems*, 5(2):161–174, 1991.
- [8] M. Egenhofer and J. Herring. Categorizing Binary Topological Relations Between Regions, Lines, and Points in Geographic Databases. Technical report, University of Maine - NCGIA, 1990.
- [9] W. Celes Filho, L. Figueiredo, M. Gattass, and P. Cezar Carvalho. A topological data structure for hierarquical planar subdivisions. Technical Report CS-95-53, Department of Computer Science, University of Waterloo, 1995.
- [10] V. Gaede and O. Günther. Multidimensional Access Methods. *ACM Computing Surveys*, 30(2):170–231, June 1998.
- [11] L. Martha. *Topological and geometrical modelling approach to numerical discretization and arbitrary fracture simulation in three-dimensions*. PhD thesis, Cornell University, August 1989.
- [12] M. Mediano. OPS: Um subsistema extensível e configurável de armazenamento de objetos persistentes. Relatório Técnico MCC37/97, Departamento de Informática, Pontifícia Universidade Católica, Rio de Janeiro, 1997.

- [13] M. Mediano. *TDS: Um Método de Indexação Topológica para Bancos de Dados Geográficos*. Tese de Doutorado, Departamento de Informática, Pontifícia Universidade Católica, Abril 1999.
- [14] M. Mediano, M. Casanova, and M. Dreux. V-tree - A Storage Method for Long Vector Data. In *Proceedings of the 20th VLDB Conference*, September 1994.
- [15] M. Mediano, M. Casanova, and M. Gattass. Map-tree: Um método de acesso para mapas longos. In *11º Simpósio Brasileiro de Bancos de Dados*, Setembro 1996.
- [16] M. Mediano, M. Gattass, and M. Casanova. HPS-tree: Um método de acesso para armazenar mapas longos com multi-resolução geométrica e topológica. In *IX Simpósio Brasileiro de Computação Gráfica e Processamento de Imagens*, pages 219–226, Outubro 1996.
- [17] M. Mäntylä. *Solid Modelling*. Computer Science Press, 1988.
- [18] A. Saalfeld. Sorting Spatial Data for Sampling and Other Geographic Applications. *GeoInformatica*, 2(2):37–57, 1998.