Unioeste - Universidade Estadual do Oeste do Paraná
CENTRO DE CIÊNCIAS EXATAS E TECNOLÓGICAS
Colegiado de Informática
Curso de Bacharelado em Informática

Banco de Dados Geográfico

Marcio Machry e Rafael Orssatto

CASCAVEL
2005

Marcio Machry e Rafael Orssatto

Banco de Dados Geográficos

Monografia apresentada como requisito parcial para a disciplina de banco de dados II do curso de Bacharelado em Informática, do Centro de Ciências Exatas e Tecnológicas da Universidade Estadual do Oeste do Paraná - Campus de Cascavel

Docente: Carlos José Maria Olguin

CASCAVEL
2005

Lista de Figuras

2.1 Exemplo de Uso de Pontos 6
2.2 Exemplo de Uso de Linhas 6
2.3 Exemplo de Uso de Poligono 7

3.1 Modelos de dados baseado na visão de campos 10
3.2 Processo de Construir Topologia 13

5.1 Processamento de Consultas Espaciais[5] 26
5.2 Seleção por índice[5] 28
5.3 Filtro por árvores-R[5] 29

Lista de Tabelas

2.1 Tipos básicos de objetos espaciais 4

Sumário

Lista de Figuras **iii**

Lista de Tabelas **iv**

Sumário **v**

1 Introdução **1**

2 Sistema de Informação Geográfico e Elementos Geográficos **2**

2.1 Sistema de Informação Geográfico 2

2.2 Conceitos Básicos em Banco de Dados Geográfico 3

2.2.1 Identidade 3

2.2.2 Atributo 4

2.3 Objetos Espaciais 5

2.3.1 Ponto 5

2.3.2 Linha 6

2.3.3 Polígono 7

2.3.4 Representação de Superfícies Contínuas 7

3 Modelagem e Armazenamento de Dados Geográficos **9**

3.1 Modelagem de dados geográficos 9

3.2 Armazenando topologia em banco de dados 11

3.3 Armazenamento de Dados Geográfico 13

3.3.1 Armazenamento de Representações Matriciais 14

3.3.2 Armazenamento de Indexação de Representações Vetorias 14

4 Manipulação de dados geográficos **18**

4.1 Operações sobre os dados geográficos 18

4.1.1 Operações sobre Geo-Objetos 18

4.1.2 Operações sobre Geo-Campos 20
4.1.3 Operações Mistas 21
4.2 Recuperação de dados geográficos 21
4.2.1 Um exemplo de linguagem: LEGAL 22
4.3 Visualização de dados geográficos 23
5 Processamento e Computação em Banco de Dados Geográficos 24
5.1 Processamento de Consultas 24
5.1.1 Fase 1: Filtragem 24
5.1.2 Fase 2: Refinamento 25
5.1.3 Fase 3: Pós-Processamento 25
5.2 Computação de operações básicas 27
5.2.1 Computação de seleções espaciais 27
5.3 Exemplo de otimização de consulta 30
5.4 Gerência da transações 31
6 Aplicações e Sistemas Gerenciadores de Banco de Dados Geográficos 32
6.1 O Sistema Spring 32
6.2 O Projeto Sagre 32
6.3 Exemplo de Banco de Dados Geográficos Livres 33
Referências Bibliográficas 35

Capítulo 1

Introdução

Neste trabalho será feita uma breve abordagem sobre os banco de dados geográficos e os sistemas de informações geográficos - SIG. Abordar-se-á a respeito dos elementos a serem armazenados no banco de dados, o porquê que banco de dados convencionais não são suficientes. Também será dada uma abordagem de como os dados são armazenados, sua importância, as formas de otimização de consulta.

E por fim são apresentadas algumas informações de uso e a respeito de sistemas de gerenciamento de banco de dados.

Capítulo 2

Sistema de Informação Geográfico e Elementos Geográficos

2.1 Sistema de Informação Geográfico

Um Sistema de Informação Geográfica (SIG ou GIS - Geographic Information System) é um sistema de hardware, software, informação espacial e procedimentos computacionais, que permite e facilita a análise, gestão ou representação do espaço e dos fenômenos que nele ocorrem.

Existem vários modelos de dados aplicáveis em SIG. Por exemplo, o SIG pode funcionar como uma base de dados com informação geográfica (dados alfanuméricos) que se encontra associada por um identificador comum aos objetos gráficos de um mapa digital. Desta forma, assinalando um objeto pode-se saber o valor dos seus atributos, e inversamente, selecionando um registro da base de dados é possível saber a sua localização e apontá-la em um mapa.

O Sistema de Informação Geográfica separa a informação em diferentes camadas temáticas e armazena-as independentemente, permitindo trabalhar com elas de modo rápido e simples, permitindo ao operador ou utilizador a possibilidade de relacionar a informação existente através da posição e topologia dos objetos, com o fim de gerar nova informação.

Os modelos mais comuns em SIG são o modelo *raster* ou matricial e o modelo vetorial. O modelo de SIG matricial centra-se nas propriedades do espaço, compartimentando-o em células regulares (habitualmente quadradas, mas podendo ser retangulares, triangulares ou hexagonais). Cada célula representa um único valor. Quanto maior for a dimensão de cada célula (resolução) menor é a precisão ou detalhe na representação do espaço geográfico. No caso do modelo de SIG vetorial, o foco das representações centra-se na precisão da localização dos elementos no

espaço. Para modelar digitalmente as entidades do mundo real utilizam-se três formas espaciais: o ponto, a linha e o polígono.

Os campos de aplicação dos Sistemas de Informação Geográfica, por serem muito versáteis, são muito vastos, podendo-se utilizar na maioria das atividades com uma componente espacial, da cartografia a estudos de impacto ambiental. A profunda revolução que provocaram as novas tecnologias afetou decisivamente a evolução da análise espacial.

2.2 Conceitos Básicos em Banco de Dados Geográfico

Um Banco de Dados Geográfico é um componente de um Sistema de Informação Geográfico, é definido como uma coleção de dados referenciados espacialmente, que funciona como um modelo da realidade. Um banco de dados é um modelo da realidade por representar um conjunto selecionado de fenômenos da realidade, que podem estar associados a diferentes períodos de tempo (passado, presente ou futuro). As seguintes definições fazem parte de um trabalho de padronização de termos proposto pelo US National Digital Cartografic Standart :

2.2.1 Identidade

Elementos da realidade modelados em um banco de dados geográfico têm duas identidades: o elemento na realidade, denominado entidade e o elemento representado no banco de dados, denominado objeto. Uma terceira identidade usada em aplicações cartográficas é o símbolo usado para representar entidades/objetos como uma feição no mapa.

Entidade

É qualquer fenômeno geográfico da natureza, ou resultante da ação direta do homem, que é de interesse para o domínio específico da aplicação.

Objeto

É a representação digital de uma entidade, ou parte dela. A representação digital varia de acordo com a escala utilizada (ex.: um aeroporto pode ser representado por um ponto ou uma área, dependendo da escala em uso).

Tipo de Entidade

É a descrição de um agrupamento de entidades similares, que podem ser representadas por objetos armazenados de maneira uniforme (ex: o conjunto das estradas de uma região). Fornece uma estrutura conceitual para a descrição dos fenômenos.

Tipo de Objeto Espacial

Cada tipo de entidade em um Banco de Dados Espacial é representado de acordo com um tipo de objeto espacial apropriado. Na tabela 2.1 são mostrados os tipos básicos de objetos.

Dimensão	Tipo	Descrição
0D	Ponto	Um objeto com posição no espaço, mas sem comprimento
1D	Linha	Tem comprimento. Composto por 2 ou mais objetos 0D.
2D	Área	Objeto com comprimento e largura. Limitado por pelo menos 3 objetos 0D.
3D	Volume	Objeto com comprimento, largura e altura.

Tabela 2.1: Tipos básicos de objetos espaciais

Classe de Objeto

Descreve um conjunto de objetos que representa um conjunto de entidades. Por exemplo, o conjunto de pontos que representam um conjunto postes de uma rede elétrica ou o conjunto de polígonos representando lotes urbanos.

2.2.2 Atributo

Descreve as características das entidades, normalmente de forma não-espacial. Exemplos são o nome da cidade, diâmetro de um duto, etc.

Valor de Atributo

Valor quantitativo ou qualitativo associado ao atributo. (ex.: nome da cidade = 'Recife', diâmetro do duto = $1\ \frac{1}{2}$").

Camada (*layer*)

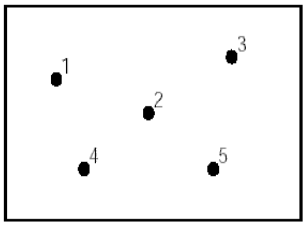
Os objetos espaciais em um BD Geográfico podem ser agrupados e dispostos em camadas. Normalmente, uma camada contém um único tipo de entidade ou um grupo de entidades conceitualmente relacionadas a um tema (ex.: uma camada pode representar somente as rodovias de uma região, ou pode representar também as ferrovias).

2.3 Objetos Espaciais

Os objetos espaciais são as representações das entidades do mundo real, armazenadas no BD Geográfico. A seguir, é descrito como que objetos primitivos do tipo ponto, linha, área ou superfície, são usados para representar as diferentes entidades da realidade.

2.3.1 Ponto

As entidades representadas por objetos do tipo ponto são aquelas que não possuem dimensões significativas, de acordo com a escala em uso. Entidades como postes elétricos, hidrantes, nascentes de rios, pontos de ônibus, normalmente são representadas pontualmente em mapas de escalas grandes (ex.: 1:5.000). Porém, em mapas de escalas um pouco menores (ex.: 1:20.000), os pontos são usados para representar a localização de escolas, hospitais, prédios públicos, entre outros. Já, em escalas bem pequenas, os pontos podem representar a localização de cidades no mapa. As coordenadas dos objetos tipo ponto podem ser armazenadas como dois atributos extras na tabela de atributos da entidade. Por exemplo, as coordenadas dos pontos que representam a localização de escolas municipais de um determinado bairro podem ser armazenadas na tabela de escolas, junto com os demais atributos descritivos(Figura 2.3.1).



Posição geográfica das escolas

ID	Coord.X	Coord.Y	Nome da Escola	Diretora	Fundação	N.alunos
1	4673000	252500	E.E.Sto Antonio	Maria José	01/05/35	240
2	4674000	254500	E.E. Prof. Rambo	Jose Silva	05/08/35	1100
3	4671000	253500	E.M. Imigrantes	Rita reis	07/06/57	740
4	4667000	253500	E.E.Gabriela Mistral	Rosa Maria	04/04/46	1250
5	4668000	254000	Instituto de Educação	Ana Maria	28/05/68	2600

Figura 2.1: Exemplo de Uso de Pontos

2.3.2 Linha

As entidades que são representadas por objetos do tipo linha são aquelas que possuem uma distribuição espacial linear (na escala em uso), como por exemplo: as ruas, rodovias, estradas de ferro, cabos telefônicos, rios, etc.

As linhas também são usadas, juntamente com os pontos, para representar estruturas de redes (Figura 2.3.2), que são usadas em aplicações do tipo redes de utilidades públicas (ex.: luz, telefone, gás e água), redes viárias (ex.: malha rodoviária, ferroviária, hidroviária, linhas aéreas) e redes naturais (ex.: hidrográfica).

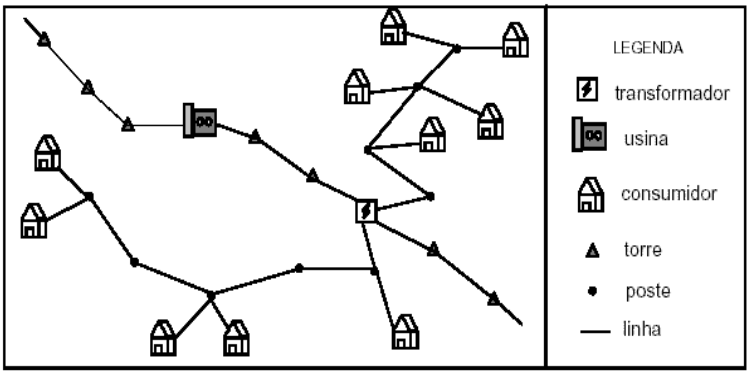


Figura 2.2: Exemplo de Uso de Linhas

Os atributos dos dados em uma rede podem estar relacionados aos nós ou às ligações. Como exemplo de atributos de ligações pode-se citar: direção do sentido do tráfego em uma rua,

distância entre duas cidades, diâmetro de uma tubulação, voltagem da rede elétrica, etc. Para atributos associados aos nós da rede pode-se citar: existência de semáforo em um cruzamento, tipo de válvula em um nó de rede de água, tipo do transformador de voltagem em uma rede elétrica, etc.

2.3.3 Polígono

Entidades com características bidimensionais são representadas no banco de dados por objetos do tipo polígono/área. Os limites das entidades podem ser definidos originalmente pelos próprios fenômenos (ex. limites de um lago, região costeira, etc), ou podem ter sido criados pelo homem (ex.: limites de um município, área de reserva florestal, etc).

Quanto à distribuição no espaço, as entidades podem ser representadas por polígonos isolados com possibilidade de sobreposição ou cada posição tem que pertencer a exatamente uma única entidade. A representação dos objetos do tipo polígono, segundo as diferentes visões quanto a distribuição no espaço, depende do modelo de dados suportado pelo sistema. Conforme a figura 2.3.3.

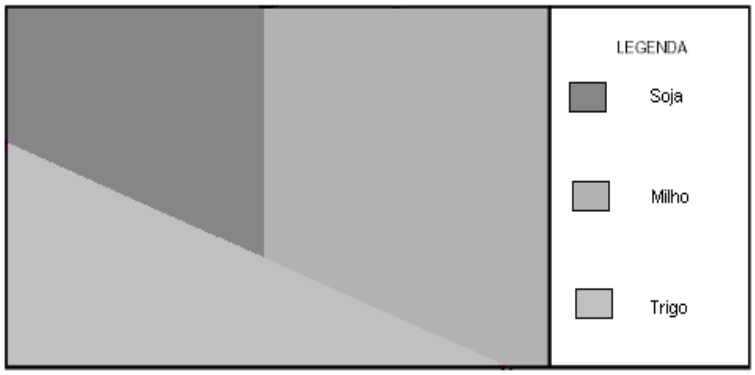


Figura 2.3: Exemplo de Uso de Poligono

2.3.4 Representação de Superfícies Contínuas

Alguns fenômenos da natureza, como por exemplo, elevação de terreno, pressão atmosférica, temperatura, densidade populacional, entre outras, são caracterizados por possuírem variação contínua no espaço.

Segundo Burrough [3], a variação da elevação sobre uma área pode ser modelada de diversas maneiras. Modelos de Elevação Digital, ou Modelos Digitais de Terreno podem ser represen-

tados tanto por superfícies definidas matematicamente (ex.: séries de Fourier) ou através de imagens de pontos/linhas.

As representações mais conhecidas, baseadas em imagens de pontos, são as matrizes de altitude, onde os dados são coletados em intervalos regulares de pontos (Figura 3.1d). Esta abordagem tem a desvantagem de introduzir redundância de dados, quando a área observada possui comportamento estável e pode perder informações, quando a área é muito acidentada.

Outra abordagem, também baseada em imagens de pontos, é o modelo de Grade Triangular ou TIN (Triangulated Irregular Network), onde os pontos são coletados mais densamente em áreas com maior variação acidental e mais esporadicamente nas outras áreas (Figura 3.1f). Os pontos são conectados formando faces triangulares, onde os valores coletados ficam associados aos vértices dos triângulos. Outro tipo de modelo de terreno, muito utilizado, é formado por um conjunto de linhas de contorno (linhas isométricas), que representam pontos de mesma elevação (Figura 3.1b).

Dentro de um SIG, os dados referentes à elevação podem ser convertidos de um modelo para outro, mas podem ocorrer perdas de informações, reduzindo os detalhes da superfície topográfica [1]. Projeções tridimensionais de superfícies contínuas podem ser usadas, por exemplo, para permitir uma melhor visualização do relevo da área observada.

Capítulo 3

Modelagem e Armazenamento de Dados Geográficos

3.1 Modelagem de dados geográficos

Modelagem de dados geográficos é o processo de discretização que converte uma realidade geográfica complexa em um conjunto finito de registros ou objetos de um banco de dados [**?**].

Os modelos de dados existentes para SIG estão relacionados com as diferentes formas de percepção da realidade que podem ser empregadas. Estes modelos de dados podem ser divididos segundo duas visões: visão de campo e visão de objetos.

Na visão de campo, a realidade é modelada por variáveis que possuem uma distribuição contínua no espaço, como por exemplo, temperatura, tipo de solo ou relevo.

Todas as posições no espaço geográfico estão associadas a algum valor correspondente à variável representada. Os objetos definidos com uso do modelo são, na verdade, abstrações que representam fenômenos que acontecem na realidade (ex.: temperatura, pressão, umidade).

Por outro lado, na visão de objetos, entidades reais são observadas como estando distribuídas sobre um grande espaço vazio, onde nem todas as posições estão preenchidas e, além disso, mais de uma entidade pode estar situada sobre uma mesma posição geográfica.

Goodchild [7], identifica seis tipos diferentes de modelos de dados baseados na visão de campo, que são usados em SIG (Figura 3.1), são eles:

a) Amostragem Irregular de Pontos - o banco de dados contém um conjunto de tuplas <x,y,z> representando valores coletados em um conjunto finito de localizações irregularmente espaçadas. (ex.: estações de medição de temperatura)

b) Linhas de Contorno - o banco de dados contém um conjunto de linhas, cada uma com um valor z associado. (ex.: curvas de nível)

c) Polígonos - A área é particionada em um conjunto de regiões, onde a cada região está associado um valor que é único em todas as suas posições. (ex.: tipos de solos).

d) Amostragem Regular de Pontos - Como no item a, porém, com pontos distribuídos regularmente. (ex.: Modelos Numéricos de Terreno)

e) Grade Regular de Células - A área é dividida em uma grade regular de células, onde o valor da cada célula corresponde ao valor da variável para todas as posições dentro da célula. (ex.: imagens de satélites).

f) Grade Triangular - a área é particionada em triângulos irregulares. O valor da variável é definido em cada vértice do triângulo e varia linearmente sobre o triângulo. (ex.: TIN - Triangulated Irregular Network)

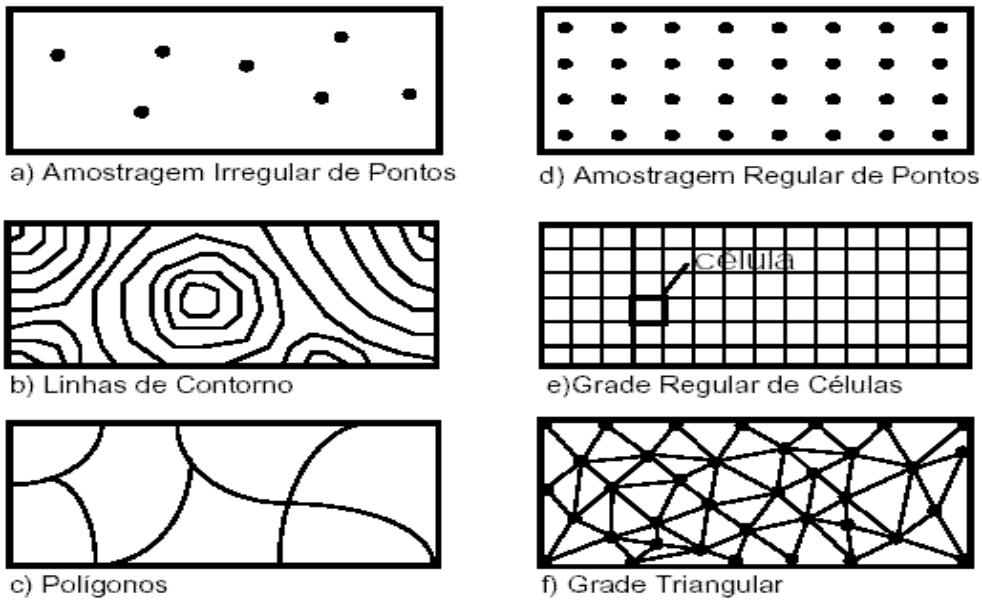


Figura 3.1: Modelos de dados baseado na visão de campos

Cada um desses modelos pode ser representado em um BD Geográfico como um conjunto de pontos, linhas, áreas ou células. Esses modelos, geralmente são confundidos, equivocadamente, com os modelos de representação de dados espaciais, matricial e vetorial. Cada um deles pode

ser mapeado em uma ou outra, daquelas representações, sendo que alguns modelos se adequam melhor a uma ou a outra. Por exemplo, os modelos Amostragem Regular de Pontos e Grade Regular de Células, são mapeados naturalmente para a representação matricial, enquanto que os demais modelos são mais bem representados numa estrutura vetorial [7].

No Modelo de Objetos, os objetos são representados como pontos, linhas ou áreas. Dois objetos podem estar localizados na mesma posição geográfica, ou seja, podem possuir coordenadas idênticas. Muitas implementações não fazem distinção no banco de dados, entre modelos de objetos e de campos. Por exemplo, um conjunto de linhas pode representar contornos (modelo de campos) ou estradas (modelo de objetos), embora as implicações das interseções sejam muito diferentes nos dois casos. O modelo de objetos é mais adequado para aplicações sócio-econômicas, que tratam com entidades criadas pelo homem (ex. rede de transporte, cadastro municipal, escolas, etc), enquanto que os modelos de campo são mais adequados para aplicações ambientais.

3.2 Armazenando topologia em banco de dados

Quando um mapa de uma região que está sobre a superfície curva da Terra, é projetado sobre uma superfície plana (ex.: folha de papel), algumas propriedades são alteradas (ex.: ângulos e distâncias), enquanto que outras permanecem (ex.: adjacências e pertinências). Estas propriedades que não se alteram quando o mapa sofre uma transformação são conhecidas como propriedades topológicas.

O termo topologia é atribuído às estruturas de relacionamentos espaciais que podem, ou não, ser mantidas no banco de dados. Um banco de dados espacial é dito topológico se ele armazena a topologia dos objetos. Por outro lado, um banco de dados é dito cartográfico se os objetos são vistos e manipulados somente de forma independente [6].

Banco de dados cartográficos são usados em muitos pacotes para confecção de mapas, onde as operações de análise são menos importantes do que as funções que auxiliam no posicionamento de rótulos, bibliotecas de símbolos cartográficos, etc.

Um banco de dados cartográfico pode ser convertido em um banco de dados topológico através do cálculo e identificação dos relacionamentos entre objetos. Este processo é conhecido como Construir Topologia (*Building Topology*) [8].

O processo de Construir Topologia é usado também, para identificar os objetos em um mapa, a partir das linhas digitalizadas. Este processo é feito empregando-se o conceito de Restrição Planar (*Planar Enforcement*), que consiste na aplicação de duas regras, sobre os objetos usados para descrever a variação espacial.

Basicamente, as regras de Restrições Planar são as seguintes:

- **Regra 1:** Dois objetos do tipo área não podem se sobrepor.
- **Regra 2:** Cada posição no mapa pertence a uma única área, ou a um limite entre áreas adjacentes.

O processo de Construir Topologia, exemplificado na figura 3.2, começa com um conjunto de segmentos de linha não relacionados (a). Cada interseção de linhas ou nodo terminal (nós) é identificado (b). Em seguida, cada segmento de linha existente entre dois nós consecutivos (arestas) é identificado. Finalmente, cada polígono resultante recebe um identificador, inclusive o polígono externo que pode receber um identificador diferenciado (c).

Os pontos, as linhas e os polígonos identificados durante o processo de construção da topologia, são armazenados em estruturas de dados, que variam de acordo com a implementação de cada sistema.

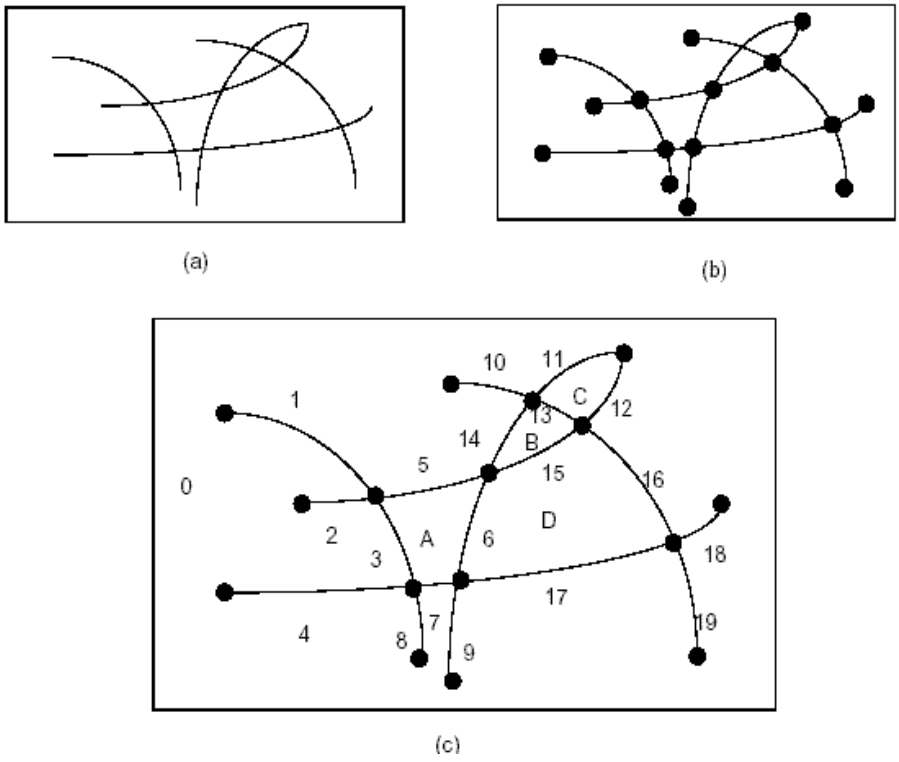


Figura 3.2: Processo de Construir Topologia

3.3 Armazenamento de Dados Geográfico

Uma estratégia comumente utilizada para descrever sistemas computacionais consiste em dividi-los em camadas funcionalmente cada vez mais complexas e, ortogonalmente, em sistemas especializados em determinada tarefa ou em determinado tipo de dado.

A primeira camada distingue as questões de visualização dos problemas de manipulação e corresponde ao nível conceitual do modelo. A segunda camada corresponde ao nível de representação do modelo. A terceira camada engloba os subsistemas que oferecem serviços de armazenamento e manipulação para atributos convencionais, representações matriciais e representações vetoriais.

A arquitetura baseada em camadas pode ser implementada de diversas formas, entre as mais simples estão a Estratégia Dual e a Estratégia Baseada em Campos Longos. "

- Estratégia Dual: Um SIG implementado segundo a estratégia dual utiliza um SGBD para armazenar em tabelas a componente convencional de todos os objetos, e arquivos normais para a componente espacial dos objetos. "

- Estratégia Baseada em Campos Longos: baseia-se no uso de SGBDs relacionais com suporte a campos longos, nos quais são armazenadas as componentes espaciais do objeto, em outras palavras, tanto os subsistemas para a manipulação e armazenamento dos atributos convencionais quanto subsistemas para armazenamento matricial e vetorial.

3.3.1 Armazenamento de Representações Matriciais

O método de armazenamento mais simples consiste em armazenar, nas páginas físicas em memória secundária, uma linha após a outra, ou uma coluna após a outra. Não é apropriado quando quando a operação exigir visitar elementos em uma ordem que não seja aquela usada no armazenamento, ou visitar um elemento e seus vizinhos em ambas as direções, ou mesmo recuperar os elementos da matriz coberta por uma janela definida por um intervalo de índices em cada dimensão.

No entanto, é possível minimizar os problemas decorrentes desse método através de um método de armazenamento que divide cada linha, ou coluna, em segmentos de comprimento fixo, submúltiplo do tamanho da linha, e que permita transferir tais segmentos da memória secundária para a memória principal e vice-versa.

Outro método de armazenamento consiste em dividir cada matriz em pequenos blocos, armazenados seqüencialmente em memória secundária segundo alguma ordem.

Em todos estes métodos, dado um elemento ou janela de elementos, é possível identificar os segmentos ou blocos envolvidos e capturar diretamente as páginas físicas em memória secundária que as contêm.

3.3.2 Armazenamento de Indexação de Representações Vetorias

Nesta subseção trataremos do armazenamento e indexação espacial de conjuntos de pontos, retângulos, linhas e regiões definidos em R2.

Pontos

Para o armazenamento e indexação de pontos são descritos a seguir dois métodos conhecidos como arquivos em grade e as árvores KD.

No método conhecido como arquivos em grade, uma grade particiona a região de interesse em células através de retas paralelas aos eixos espaçadas de forma irregularmente. Um arquivo

em grade consiste de duas escalas, uma para cada eixo, e um diretório. A escala associada ao eixo X é uma lista contendo as abscissas em que as retas paralelas ao eixo Y cruzam o eixo X (e similarmente para outra escala). O diretório é uma matriz bidimensional, com as mesmas dimensões da grade, tal que cada elemento do diretório corresponde a uma célula e aponta para a página física (ou *bucket*) que contém todas as coordenadas dos pontos localizados na célula. Vários elementos podem apontar para a mesma página física. As escalas e o próprio diretório se forem estruturas de dados pequenas, poderão residir em memória principal; caso contrário deverão por sua vez ser paginados.

Uma árvore kd particiona também o espaço de células, mas de forma diferente dos arquivos em grade. Uma árvore kd é uma árvore de busca binária, residindo em memória principal, tal que os nós interiores em cada nível contêm valores referentes a um único eixo, X ou Y, alternadamente, e as folhas apontam para páginas físicas. Varias folhas podem apontar para a mesma página física. Assim, o valor armazenado na raiz divide o espaço em dois subespaços através de uma reta perpendicular ao eixo dos X, digamos; o valor armazenado no filho à esquerda (ou direita) por sua vez divide o subespaço à esquerda (ou direita) em dois subespaços através de uma reta perpendicular ao eixo dos Y; e assim por diante, alternando as dimensões. O processo de subdivisão pára quando a célula contém um conjunto de pontos que possa ser armazenado em uma única página física.

Retângulos

Para o armazenamento de retângulos, o enfoque mais usual utiliza a noção de retângulo envolvente. O retângulo envolvente mínimo (r.e.m.) de um conjunto de objetos no R2 é o menor retângulo com lados paralelos aos eixos X e Y que contém todos os objetos no conjunto.

O melhor exemplo deste enfoque são as arvores R. Uma árvore-R é uma estrutura de dados hierárquica derivada da árvore-B. As diferenças estão centradas na natureza das chaves: valores numéricos ou alfanuméricos simples, no caso das árvores-B, e pontos extremos de retângulos, no caso das árvores-R. A árvore-R possui parâmetros para determinar a quantidade de chaves (retângulos) que poderão ocorrer em cada bloco de armazenamento, analogamente à árvore-B, em função do tamanho da página de armazenamento em disco. Todas as folhas aparecem sempre no mesmo nível da árvore. Nós internos contêm a delimitação de retângulos que englobam todos os retângulos dos nós nos níveis inferiores. Uma árvore-R de ordem (m, M) conterá entre m e M

entradas em cada nó da árvore (m < M / 2), exceto a raiz, que terá pelo menos dois nós (a menos que a árvore só tenha uma entrada). Pesquisas na árvore-R são relativamente simples de serem executadas. O único problema é que um grande número de nós pode ter que ser examinado, pois um retângulo pode estar contido em vários outros, mesmo que o objeto esteja contido em apenas um nó folha. Existem diversas variações das árvores-R, cada uma tentando aperfeiçoar um aspecto diferente. No entanto, muitas vezes estas variações introduzem desvantagens, ou uma maior complexidade de implementação, fazendo com que a árvore-R original acabe sendo a opção mais usual.

Linhas

O armazenamento e indexação espacial de conjuntos e linhas pode ser abordado adaptando-se os métodos utilizados para pontos e retângulos. Dado um conjunto de linhas, esta estratégia consiste em aproximar cada linha pelo seu retângulo envolvente mínimo e utilizar o método discutidos anteriormente para armazenar e acessar os retângulos assim gerados. Por exemplo, adotando-se árvores R, cada folha passa a conter como entradas um apontador para uma linha e seu retângulo envolvente mínimo.

Polígonos

O armazenamento e indexação de polígonos pode ser tratado de forma simplista, como um caso particular do tratamento de linhas. Baseia-se no uso de aproximações para a geometria dos polígonos e na decomposição dos polígonos em polígonos simples. Para que uma aproximação da geometria de um polígono seja útil ao processamento de consultas, ela deve ser conservativa, ou seja, deve conter completamente o polígono. Decompor um polígono significa dividí-lo em várias componentes simples - triângulos, trapézios, e outros - de tal forma que, durante o processamento de uma consulta espacial, apenas um pequeno número destas componentes precise ser examinado. Assim, cada polígono sofre inicialmente uma decomposição em elementos geométricos mais simples, que são então armazenados através de um método apropriado.

O método de decomposição por grades combina diretamente as idéias de aproximação e decomposição. Considere o espaço dividido hierarquicamente em células regulares. Cada polígono é aproximado pelo menor conjunto de células, de qualquer nível, que cobre completamente o polígono, até um certo nível de subdivisão definido pela precisão desejada. Os endereços das

células assim obtidas formam um conjunto de z-elementos associado ao polígono. Um conjunto de polígonos é então representado por um conjunto de conjuntos de z-elementos, armazenados em um índice convencional, como por exemplo, em uma árvore B.

Capítulo 4

Manipulação de dados geográficos

Neste capítulo serão apresentados como são realizadas as operações, a recuperação e a apresentação dos dados geográficos, elementos importantes na utilização de qualquer sistema de banco de dados.

4.1 Operações sobre os dados geográficos

O que se espera de um sistema de informação geográfico quanto a operações sobre os dados tende a variar conforme a função, área de aplicação e do tipo de usuário. Frente a isso, diversas são as alternativas para categorizar as operações de um SIG. As operações em banco de dados geográficos podem ser agrupadas segundo as classes dos objetos envolvidos, sejam eles geo-campos, geo-objetos ou ambos. Porém não serão apresentadas uma série ilimitada de operações, mas sim se fará uma análise a respeito das classes dos objetos envolvidos nas operações.

4.1.1 Operações sobre Geo-Objetos

Para operações em geo-objetos deve-se considerar como estes geo-objetos estão representados, pois um mesmo objeto pode ser representado por um ponto, ou por uma área, dependendo do nível de representação que se deseja, e conseqüentemente as operações a serem aplicadas sobre um ponto ou uma área são distintas. Desta forma, para se efetuar operações de construção e atualização sobre os geo-objetos, há a necessidade de operações sobre os relacionamentos que os representam.

Desta forma, existem diversos tipos de relacionamentos espaciais entre estes geo-objetos, que podem ser direcionados, topológicos ou métricos. Os direcionados são os relacionamentos

que indicam a direção de um objeto em relação ao outro, como por exemplo, um objeto esta acima, ou abaixo de outro. Ou seja, um relacionamento direcional sempre necessita de um ponto de referência. Por outro lado os relacionamentos topológicos, tambem necessitam de dois objetos e representa a relação que existe entre os mesmos, como por exemplo, contido ou adjacente. As relações topológicas são invariantes a transformações de escala, translação ou rotação, pois a referência é um outro objeto que se movimenta com o conjunto. E por fim tem-se as relações métricas que derivam das operações de distância e direção, também são operações que necessitam de uma referência.

Várias seriam os relacionamento e funções adimitidas no que se diz respeito a relacionamentos, porém alguns muito complexos e que só aumentariam a complexidade do sistema. Desta forma determina-se relacionamentos mais simples que em sua combinação podem formar relacionamentos mais complexos. A seguir serão relacionados alguns relacionamentos, tomando por exemplo w e x que denotam dois elementos topológicos simples, ou seja, são um ponto, ou uma reta, ou uma curva fechada ou aberta. Será utilizada a notação de $dim(x)$ para representar a quantidade de dimensões do objeto x e a notação $w\circ$ para indicar o interior delimitado pela linha, ou pela área w.

O relacionamento *disjoint*, que verifica se um elemento esta dijunto do outro, é dado por:

$$w\ disjoint\ x \longleftrightarrow (w \cap x = \emptyset)$$

Já o relacionamento *touch*, que indica se uma região toca a outra, é dado por:

$$w\ touch\ x \longleftrightarrow (w \cap x \neq \emptyset) \wedge (w\circ \cap x\circ \neq \emptyset)$$

Tem-se o relacionamento *in* que indica se um elemento esta dentro do outro, dado por:

$$w\ in\ x \longleftrightarrow (w \cap x = w) \wedge (w\circ \cap x\circ \neq \emptyset)$$

Ainda temos o relacionamento de sobreposição, *overlap*, aplicável a dois elementos de mesmo tipo dado por:

$$w\ overlap\ x \longleftrightarrow (w \cap x \neq w) \wedge (w \cap x \neq x) \wedge dim(w\circ \cap x\circ) = dim(w\circ) = dim(x\circ))$$

O relacionamento *cross* indica se um elemento cruza o outro e é dado por:

$$w\ cross\ x \longleftrightarrow (w \cap x \neq w) \wedge (w \cap x \neq x) \wedge dim(w\circ \cap x\circ) = (max(dim(w\circ), dim(x\circ)) - 1$$

Também se tem a função *boundary* que retorna as fronteiras de uma região, as funções *from* e *to* que retornam o inicio e fim de uma linha não circular.

Com esses relacionamentos e com cálculos a partir desses relacionamentos, é permitido definir com precisão relacionamentos topológicos complexos, sem a necessidade de um conjunto muito extenso de conceitos. Porém algumas extensões podem ser necessárias para alguns elementos muito complexos, como regiões com buracos.

Desta forma temos que as operações sobre geo-objetos podem ser através de três aspectos como segue:

Restrições sobre atributos

Restrições sobre atributos, que são computados em cima dos atributos do objeto, como por exemplo, locais com um número x de habitantes. Uma operação realizada com base sobre os atributos é a seleção por atributos, que é semelhante a álgebra relacional. Exemplificando: "Recupere todas as cidades do Paraná com população acima de 100.000 habitantes."

Restrições espaciais

Quanto a restrições espaciais são as operações executas através das relações anteriormente citadas, referenciando-se a topologia, direção ou medidas. Aqui se englobam as operações de seleção espacial, por exemplo "recupere todas as cidades com distância menor que 150 km de Cascavel", e de junção espacial que é semelhante a junção em banco de dados relacional, por exemplo, "para cada rodovia do Paraná, localize os restaurantes e hotéis".

Propriedades dos Geo-Objetos

Toma-se por base as propriedade de geo-objeto, basicamente inclui-se os operadores matemáticos tradicionais sobre os atributos e também os operadores métricos. Exemplificando: "Para os lotes do jardim universitário, calcular a média de IPTU", ou, "liste quais as cidades de maior consumo de energia no Paraná".

4.1.2 Operações sobre Geo-Campos

Vericar-se-á nessa seção como definir operações sobre geo-campos, sabendo-se que geo-campos são funções como foi visto na descrição do modelo de dados. A atenção será dada

somente em duas categorias de operações, as pontuais e as de vizinhança.

Entende-se por operação pontual aquela que o resultado é um geo-campo cujo resultado de seus pontos depende somente do valor de cada ponto de entrada relacionado, ou seja, é um mapeamento entre os dados de entrada para os dados de saída. Podemos tomar, por exemplo, uma operação de ponderação entre tipos de solo e aptidão agrícola.

Por outro lado, a operação de vizinhança é quando o resultado depende de uma região de vizinhança ao ponto relacionado. Podemos tomar por exemplo o cálculo de máximos ou mínimos de uma região, como por exemplo o cálculo pluviométrico. Utilizado fortemente para a verificação de declividade, de diversidade e classificação de regiões.

4.1.3 Operações Mistas

Uma operação mista é aquela que envolve geo-campos e geo-objetos. Estas operações são comumente encontradas quando dado um conjunto de características, geo-campos, quer-se as regiões que estas podem ser encontradas, que são representadas por geo-objetos. Por exemplo, quer-se as regiões de mata atlântica do Brasil, isto é um dado de geo-campo, e o resultado é devolvido na forma de geo-objetos, mapas, que representam essas regiões.

4.2 Recuperação de dados geográficos

Nesta seção verificar-se-á as problemáticas para a recuperação de dados geográficos no que se diz respeito a definição de linguagens de consulta para o banco de dados geográficos.

As linguagens de consultas para banco de dados geográficos podem ser classificadas em textuais, visuais, e multimodais.

As linguagens textuais baseiam-se em SQL desta forma temos o problema de que o SQL foi concebido para banco de dados relacionais, que não são adequados aos problemas de SIG. Desta forma faz-se a inclusão de diretivas especiais para os banco de dados geográficos. Um exemplo que pode ser dado é a linguagem GEOQL que estende o SQL com nove operados espaciais (*intersects, adjacent, joins, ends-at, contains, situated-at, within, closest e furthest*), o que restringe o tipo de consulta que pode ser realizada[4].

Em oposição as linguagens textuais, as linguagens visuais permitem a construção de uma consulta através da combinação de símbolos, ícones e texto a fim de facilitar o uso. Esta abor-

dagem dificulta a formulação de uma semântica precisa para as consultas.

Já os ambientes multimodais combinam manipulação direta a consultas visuais e textuais, de forma a tirar proveito dos dois mundos, porém o que dificulta é conseguir coordenar conceitos distintos de interação e semântica.

4.2.1 Um exemplo de linguagem: LEGAL

A linguagem LEGAL - Linguagem Espacial para Geoprocessamento Algébrico - é uma linguagem de comandos interpretados para uso em análise geográfica que está em desenvolvimento no ambiente do sistema SPRING [2], proposta por Câmera Neto, com a finalidade de prover um ambiente geral para análise geográfica, operações de consulta espacial e operações de apresentação de resultados de consulta e manipulação. A linguagem LEGAL baseia-se no modelo definido pelo ODMG - *Object Data Management Group*.

Na linguagem LEGAL pode-se fazer a definição de esquemas conceituais através do comando CREATE, utilizando-se de especializações de classes já presente no banco de dados geográfico, como geo-objetos, temas, entre outros. A instanciação de objetos no LEGAL é feita através do comando INSERT, onde são criados os os geo-objetos e os geo-campos. Um exemplo de cada uso pode ser visto abaixo, onde se cria um estrutura de classe como especialização de *tematico* e é feita uma inserção de um objeto nessa classe.

```
CREATE Uso-do-Solo (nome CHAR(20), ano INTEGER) IS-A tematico (temas ={"Floresta",
"Cerrado", "Mangue"})
INSERT (nome = "acao antropia 1990", ano = 1990) IN Uso-do-Solo
```

As consultas em LEGAL podem ser formuladas envolvendo tanto geo-objetos como geo-campos. Um exemplo de consulta é o de selecionar os nomes de todos os municipios distantes a menos de 100 km de Cascavel:

```
SELECT m.nome FROM m Municipio, m Cidade WHERE c.nome = "Cascavel"and
distance(m,c) < 100
```

Neste exemplo as classes Municipio e Cidade são geo-objetos que representam os municipios e cidades. O que qualifica a consulta é a função espacial *distance*.

Outra função da linguagem é a de permitir coleções de objetos, devido a ser baseada em ODMG, que possibilita a criação dessas coleções, desta forma operações de *next*, *first*, *last* entre outras podem ser usadas para a manipulação dos objetos presentes numa coleção.

4.3 Visualização de dados geográficos

A forma de visualização dos dados geográficos é muito importante em sistemas de informação geográficas. Minimamente um SIG deve oferecer visualização de campos e objetos geográficos em 2D e 3D, incluindo comandos para controlar padrões de preenchimento. A interface deve possibilitar a consulta a dados presente na tela, bem como prover formas de cruzamento de informações resultantes de várias consultas. Cuidados devem ser tomados na visualização de informações geográficas no que se diz respeito a exibição, pois geralmente a quantidade de dados é muito grande, desta forma há a necessidade de uma ampliação na projeção ou numa redução do conteúdo que deve ser gerenciada pelo visualizador. Aqui não se abordará maiores detalhes do sistema de visualização por estar não se tratar do escopo deste trabalho.

Capítulo 5

Processamento e Computação em Banco de Dados Geográficos

Neste capítulo serão abordados os aspectos do processamento de consulta e de gerenciamento de transações num banco de dados geográfico, o estudo aqui feito esta baseando no capítulo 7 de [5]. Discutir-se-á a seguir a respeito do processador de consultas e do gerenciador de transações que devem sofrer alterações quanto ao modelo tradicional de banco de dados devido as especificidades do banco de dados geográficos no que se diz respeito a como consultar e ao tempo de duração das transações.

5.1 Processamento de Consultas

O processamento de consultas é feito pelo processador de consulta, um dos principais componentes de qualquer SGBD. Ele deve traçar um "plano"para a execução da consulta, onde ele decompõe a consulta em pequenas operações de baixo nível. O principal componente do processador de consulta por sua vez é o otimizador, que busca fazer com que a ordem dessas pequenas operações sejam tais que minimizem o esforço computacional. O processo de consulta esta dividido em três fases conforme a figura 5.1.3 e que se segue:

5.1.1 Fase 1: Filtragem

A fase de filtragem baseia-se em localizar os dados que satisfaçam à consulta, este processamento é feito geralmente por índices espaciais ou estruturas de indexação espaciais. O desempenho depende diretamente de como é a distribuição espacial e como a estruturação do banco de dados decompõe o espaço. Devido a complexidade da informação espacial esses indí-

ces somente trazem uma aproximação, como o menor retângulo que possui o espaço geográfico, devido a isto é que se diz que é uma filtragem.

5.1.2 Fase 2: Refinamento

Após a filtragem, entra-se na fase de refinamento, onde os dados canditados são trazidos para a memória principal, e depois são aplicados os operadores espaciais específicos para a seleção do dado espacial requerido. O desempenho desse passo depende da quantidade de dados recuperados, de sua complexidade e dos operadores a aplicar, e que geralmente é elevado.

5.1.3 Fase 3: Pós-Processamento

Após a seleção dos dados que satisfazem a consulta, eles devem ser disponibilizados para que sejam passados as camadas superiores para o processamento ou a visualização. Desta forma deve existir um mecanismo eficiente para passar a geometria das componentes espaciais dos dados. Para isso devem haver métodos de armazenamento que governem a gerência de páginas de memória secundária de forma a armazenar as componentes exatadas separadamente dos índices, para a diminuição dos índices, mas se necessita de contiguidade, que é resolvida através do armazenamento das componentes espaciais vizinhas em páginas próximas.

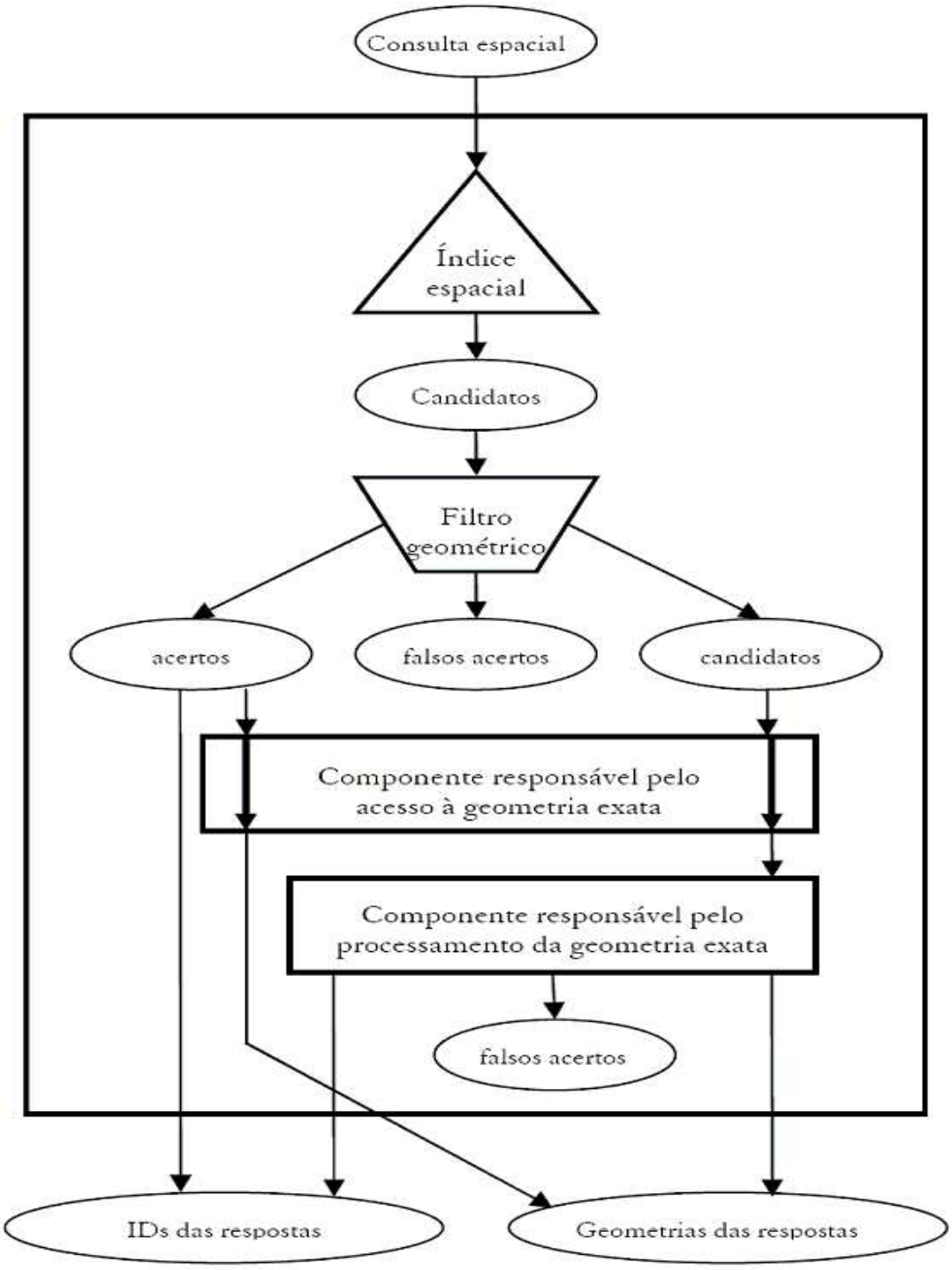


Figura 5.1: Processamento de Consultas Espaciais[5]

5.2 Computação de operações básicas

Nesta seção verificar-se-á as operações básicas de seleção como a sua computação é realizada, tendo por base o que foi discutido na seção que se falou a respeito de otimização de consultas. Em [5] há a discussão mais aprofundada, bem como a computação de junções espaciais e a computação de superposição de mapas temáticos.

5.2.1 Computação de seleções espaciais

Uma seleção espacial é uma expressã da forma $\sigma[B(x)]$, onde B(x) é uma expressão boleana com uma variável livre, no caso x. Exemplificando: "Selecione todos os pontos x distantes do ponto p menos que 100 unidades", para este problema tem-se a expressão: $\sigma[distancia(x, p) < 100]$, onde p é um ponto fixo.

Duas estratégias podem ser aplicadas: a seleção por pesquisa exaustiva, que verifica-se em cada objeto a condição de B(x), possui desta forma um custo proporcional ao tamanho do conjunto de objetos a pesquisar; por outro lado usa-se a seleção por índice, representado pelo pseudo-código da figura 5.2.1, onde existe uma etapa de filtragem para selecionar parte dos objetos do conjunto S depois na etapa de refinamento é carregado para a memória as representações geométricas de cada objeto selecionado anteriormente, e então após a filtragem e o refinamento se aplica a seleção por pesquisa exaustiva.

```
   SELEÇÃO_POR_INDICE(S,B,I,C;R)
S  -  conjunto de geo-objetos
B  -  predicado de seleção
I  -  índice sobre S
C  -  expressão booleana envolvendo comparações
      espaciais com apenas uma variável livre
      que pode ser resolvida por I
R  - conjunto dos objetos de S que satisfazem B
Inicio
   R, P = ∅
   FILTRO(I,C;P)              % Etapa 1: Filtragem
   Para cada p em P faça      % Etapa 2: Refinamento
   Inicio
      Recupere o objeto s de S apontado por p
      Se s satisfizer B, acrescente s a R
   Fim
Fim
FILTRO(I,C;P)
Inicio
   P = ∅
   Acrescente a P todos os apontadores para
      objetos de S que satisfazem C, usando I
Fim
```

Figura 5.2: Seleção por índice[5]

O restante da seção refinará a função auxiliar FILTRO. A função auxiliar utiliza-se de uma outra expressão boleana C, para o filtro de forma que para todo geo-objeto s, se C(s) = falso entao B(s) = falso, desta forma pode-se concluir que haverão objetos s selecionados que não satisfazem B, porem todos os não selecionados não satisfazem B. Para a computação de $\sigma[B(x)](S)$ é guiada pelos retângulos armazenados na árvore-R, dependendo da expressão B. Tem-se que nas folhas da árvore-R são armazenados os retângulos que aproximam a geometria dos objetos, não a sua geometria real. Desta forma para que as árvores-R sejam utilizadas para a filtragem deve-se determinar uma expressão C' de forma que para todo o geo-objeto s, para todo retângulo r tal que r é igual ou contém o retângulo envolvente mínimo de s, se C'(r) = falso então C(s) = falso. Freqüentemente a expressão C' pode ser definida como a própria expressão C, como a expressão de *contem*. O inverso também pode ocorrer, pode ser que não haja como definir um C' para o uso de árvores-R como filtro, o que torna o filtro inútil, como por exemplo a disjunção, pois ter-se-ía que verificar os retângulos que são dijuntos ou que sobrepõe, o que aumenta o processamento, ao invés de simplificar. Finalmente o pseudo-código da figura 5.2.1 mostra a operação de filtro, e tem-se que o custo do algoritmo é proporcional ao número de nós

lidos somado ao número de nós apontados como resolvedores de C'.

```
FILTRO_ARVORE_R(A,C;P)

A  -  apontador não nulo para raiz de uma árvore-R
      sobre um  conjunto de geo-objetos S.
      Cada nó N é uma estrutura da forma:
      lt(N)   –  lista de entradas;
      rem(N)  –  r.e.m. das entradas em lt(N)
      Cada entrada E é uma estrutura da forma:
      Nó interior:
      pt(E)   –  apontador para a raíz de uma
                 sub-árvore de A
      rem(E)  –  retângulo envolvente mínimo da
                 sub-árvore apontada por E
      Folha:
      pt(E)   –  apontador para um objeto em S
      rem(E)  –  retângulo envolvente mínimo
                 do objeto apontado por E
C  -  expressão booleana envolvendo comparações
      espaciais com apenas uma variável livre
      que pode ser filtrada por A
P  -  conjunto de apontadores para objetos de S

Inicio
   Determine a expressão C' companheira de C
   P = ∅
   Q = {A}
   Enquanto Q ≠ ∅ faça:
   Inicio
      Selecione q em Q, removendo-o de Q
      Se q for uma folha,
         Para cada entrada E em ls(q),
            Se C'(rem(E)), acrescente pt(E) a P
      Senão
         Para cada entrada E em ls(q),
            Se C'(rem(E)), acrescente pt(E) a Q
   Fim
Fim
```

Figura 5.3: Filtro por árvores-R[5]

5.3 Exemplo de otimização de consulta

Agora será visto um exemplo de otimização de consulta que deve ser realizado para a melhor eficácia do sistema. Considere um banco de dados geográfico de cidades com um conjunto de objetos chamado CIDADES, onde para cada cidade existe um objeto no conjunto CIDADE, e que a localização da cidade é dado por um par de latitude e longitude e demais informações da cidade estão armazendas. Considere a consulta: *"Selecione o nome e a população de todas as cidades com menos de 20 mil habitantes próximas de Cascavel-Pr a 100 Km"*. Esta consulta pode ser feita através do comando:

```
SELECT d.nome, d.populacao
FROM d Cidade, c Cidade
WHERE c.nome = "Cascavel-Pr"
and distance(d,c) <= 100
adn d.populacao < 20.000
```

Existem três planos para a execução desta consulta, verificaremos aqui somente dois.

Considere o primeiro plano, onde as seguintes ações devem ser tomadas:

- 1. Determinar a posição b de Cascavel-Pr;
- 2. Determinar todos os conjuntos de C' de todas as cidades a 100Km de b;
- 3. Determinar o subconjunto C das cidades com menos de 20 mil habitantes do conjunto C'.

Baseado neste plano a execução dos itens 1 e 3 são determinações que não envolvem operações espaciais. Por outro lado, a operação 2 é subdividia em mais 3 operações, baseado na computação de seleção espacial visto na seção anterior, como segue:

- 2.1. Determinar as cidades que podem estar a menos de 100 km da posição, criando o conjunto I ;
- 2.2. Ler para a memória principal o conjunto I, criando o conjunto C";

- 2.3. Determinar os elementos de C"que estao a menos de 100km, formando o conjunto C.

O que se pode otimizar é criar uma espécie de *pipeline* das operações 2.1, 2.2, 2.3 e 3 de forma a cada elemento que passar pela etapa ser computado diretamente para o conjunto C, evitando o processo de armazenamento nos resultados intermediários I, C' e C".

Esta consulta será vantajosa se o número de cidades selecionadas no item 2 for razoavelmente pequeno. Outra concepção que se pode fazer é inverter as ações 2 e 3, podendo também serializar algumas operações para evitar os resultados intermediários. Esta solução será melhor se a quantidade de cidades com menos de 20 mil habitantes for suficiente menor.

5.4 Gerência da transações

Transações em banco de dados são operações que devem ser atômicas e levam o banco de dados de um estado consistente para um novo estado também consistente. Porém transações em banco de dados convencional geralmente são curtas e utilizam-se de poucos dados, o que não ocorre em banco de dados geográfico.

Geralmente em banco de dados geográficos as transações são longas e aninhadas, ou seja, duram muito tempo e vários usuários participam das alterações com a finalidade de formar um todo. Normalmente alterações podem durar semanas e devido a isso a transação geralmente é chamada de sessão de trabalho. Por fim essas transações podem estar organizadas em subtransações aninhadas. Ou seja a subtransação pode ser liberada para que outros usuários vejam suas alterações, ou pode ser feito e desfeito. Devido a isso serão verificados os mecanismos para a implementação de transações. Porém, para atualizações convencionais podem ser utilizados métodos diretos de controle de transações.

Capítulo 6

Aplicações e Sistemas Gerenciadores de Banco de Dados Geográficos

Várias são as aplicações que se utilizam de banco de dados geográficos. Neste capítulo apresentar-se-á o sistema Spring, o projeto Sagre e exemplos de banco de dados geográficos livres.

6.1 O Sistema Spring

O SPRING é um SIG (Sistema de Informações Geográficas) com funções de processamento de imagens, análise espacial, modelagem numérica de terreno e consulta a bancos de dados espaciais. O SPRING é um projeto do INPE / DPI (Divisão de Processamento de Imagens) com a participação da EMBRAPA/CNPTIA, IBM Brasil, TECGRAF - PUC Rio e PETROBRÁS/CENPES[10].

O SPRING tem por objetivo a construção de um SIG para aplicações na agricultura, controle de florestas, gestão ambiental, na geografia e geologia, e no planejamento urbano e regional. Busca ser um sitema de rápido aprendizado num ambiente unificado de geoprocessamento e sensoriamento remoto. Além disso o SPRING é um sistema multiplataforma.

6.2 O Projeto Sagre

O SAGRE - Sistema Automatizado de Gerência de Rede Externa é um sistema de operações desenvolvido pelo CPqD para gerenciar a expansão, modernização e operação da rede externa, atuando como elemento de integração e padronização para os diversos departamentos

das empresas operadoras de telecomunicação. O SAGRE é capaz de controlar e operar automaticamente mapas urbanos, fazer estudos de mercado e demanda, registro de redes de canalização e cabos, bem como o planejamento empresarial e manutenção de rede.

6.3 Exemplo de Banco de Dados Geográficos Livres

Acompanhando as novas necessidades dos usuários de Geotecnologias, os Sistemas Gerenciadores de Banco de Dados (SGBD) criaram módulos, ou extensões, específicos para o armazenamento e análise de dados geográficos.

O PostgreSQL foi o primeiro SGBD de código aberto a trabalhar com um módulo específico para o tratamento dos dados geográficos vetoriais. Este módulo denominado de PostGIS[9] foi desenvolvido por uma empresa canadense chamada *Refractions Research* e segue a especificação SFS *(Simple Features Specification)* do OGC *(Open Geospatial Consortium)*. Para que o PostGIS contemple toda a SFS, é necessário que ele seja compilado juntamente com a biblioteca GEOS *(Geometry Engine - Open Source)*. Com isso, o PostGIS passa a possuir mais de 130 funções e operadores para o tratamento de dados geográficos vetoriais, podendo atender todas as demandas presentes numa instituição onde antigamente somente era possível com o *Oracle Sptial* ou *Microsoft SQL Server com ArcSDE*.

O PostgreSQL suporta três tipos de indexação nativos: *BTree, RTree e GiST (Generalized Search Trees)*. O Btree é usado para ordenação de dados em um eixo somente, logo ele não tem muita utilidade para o tratamento de dados geográficos. Já o Rtree divide os dados em retângulos que, por sua vez, podem ser novamente divididos em novos retângulso, e assim sucessivamente. Apesar do Rtree ser utilizado em alguns bancos de dados espaciais para a indexação de dados em SIG, a implementação do Rtree não é tão robusta quanto a implementação do GiST. Esta por sua vez, pode ser entendida como uma divisão dos dados em: objetos ao lado de, objetos que sobrepõe a, objetos que estão dentro de, entre outros. Assim como as outras indexações, ela é utilizada para acelarar pesquisas, porém ela pode tratar uma variedade de estruturas de dados irregulares, o que não é possível com o BTree. Devido as restrições do Rtree do PostgreSQL, o PostGIS emprega a Rtree construída sobre o GiST.

O SGBD MySQL 4.1 também contém uma extensão para dados geográficos, também dentro dos padrões da OGC. Com todas essas opções fica claro que não existe justificativa para a

compra de um SGBD proprietário, como Oracle ou Microsoft SQL Server, que também contam com extensões para SIG.

Referências Bibliográficas

[1] ARONOF, S. **Geographic Information Systems: a management perspective**. Canada: WDL Publications, 1989.

[2] BARBOSA, C. C. F. **Álgebra de mapas e suas aplicações emsensoriamento remoto e geoprocessamento**. 1997.

[3] BURROUGH, P. **Principles of Geographical Information Systems for Land Resources Assessment.** Oxford: Clarendon Press, 1986.

[4] CAMARA, G. E. A. **Anatomia de Sistemas de Informação Geográfica**. UNICAMP, 1996.

[5] CAMARA, G. E. A. **Bancos de Dados Geográficos**. http://www.dpi.inpe.br/livros/bdados/capitulos.html, Acessado em Agosto de 2005.

[6] GOODCHILD, M. F. **Geographical data modeling.** In: FRANK, A.U.; GOODCHILD, M.F. Two Perspectives on Geographical Data.

[7] GOODCHILD, M. F. **Integrating GIS and environmental modeling at global scales.** In: GIS/LIS '91. Proceedings... Atlanta, 1991. V.1.

[8] LAURINI, R.; THOMPSON, D. **Fundamentals of Spatial Information Systems.** San Diego: Academic Press, 1992.

[9] **POSTGIS**. http://postgis.refraction.net, Acessado em Agosto de 2005.

[10] **SPRING - Sistema de Processamento de Informações Georeferenciadas**. http://www.dpi.inpe.br/spring, Acessado em Agosto de 2005.