

Indexação dos dados espaciais do Banco de Dados do Inventário de Minas Gerais

Thomaz Chaves de Andrade Oliveira¹

Samuel Rodrigues de Sales Campos²

Luca Araújo Egas Prieto³

Elias Bruno de Castro Lasmar⁴

¹ Universidade Federal de Lavras - UFLA/DCF
Caixa Postal 3037 - 37200-000 - Lavras - MG, Brasil
menininho@ufla.br

Abstract. This paper describes some of the Spatial Indexes present in the spatial indexing literature and their use in the “*Inventário Florestal de Minas Gerais*” project. The project is developing a GIS in which data gathered of many types that include field collected data and remote sensing data have been merged into a single database. One type of data in which is to be stored is spatial data. The spatial data includes city boundaries, forest fragments, parcels and other complex shapes related to the project that compose the GIS. The GIS requires spatial queries among these complex shapes. In order for these queries to return the results in the least possible time, efficient spatial indexing is required. In the present work we researched through the spatial indexing literature to choose the right spatial index and its tuning among all the best know spatial indexes that would best serve our current application development. A description of the choices that were made is present in this work.

Palavras-chave: Índice espacial , oracle, rtree, quadtree, Inventário Florestal de Minas Gerais, Banco de Dados.

1- Introdução

O inventário florestal de Minas Gerais consiste no mapeamento da flora nativa e dos reflorestamentos existentes no estado, etapa efetuada em 2004, e no monitoramento contínuo desta cobertura. A fase atual consiste na pesquisa quantitativa e qualitativa do estoque em volume e carbono, entre outros, além da lista de espécies da flora nativa e a caracterização dos reflorestamentos. Tais informações serão utilizadas como instrumento de política, planejamento e gestão florestal e ambiental pelo Instituto Estadual de Florestas, pela Secretaria de Estado do Meio Ambiente e Desenvolvimento Sustentável e por outras esferas do governo do estado de Minas Gerais. Na fase atual do projeto está sendo desenvolvido um SIG com os dados do inventário florestal de Minas Gerais. Esse SIG possui um banco de dados espaço-temporal que contém tanto dados georreferenciados como dados convencionais. No banco de dados do SIG, estão armazenados diversos dados espaciais, entre eles: fronteiras municipais do estado, contorno de bacias e sub-bacias hidrográficas, núcleos do IEF entre outros. O SIG proposto realiza buscas espaciais em tempo real e necessita um método de indexação espacial eficiente para que as buscas sejam eficientes e retornem os resultados em menor tempo hábil possível. Este presente trabalho apresenta alguns métodos de indexação espacial presentes na literatura e suas características. Com base nessas características e nas necessidades do projeto do SIG do Inventário de Minas Gerais foi feito um estudo das melhores opções de indexação espacial presentes na literatura para a futura indexação dos dados espaciais.

Referencial teórico

2.1 Bancos de dados espaciais:

O termo sistemas de informação geográfica (SIG) é aplicado a sistemas que realizam tratamento computacional de dados geográficos. A principal diferença entre um SIG e um sistema de informação convencional é a capacidade de armazenar tanto dados descritivos como as geometrias dos dados geográficos, Câmara et al (2006). Sistemas Gerenciadores de banco de dados convencionais, não trabalham com dados multidimensionais como: polígonos, caixas, linhas ou mesmo pontos de maneira eficiente Sellis (1987). Sistemas de Informações geográficas entre outros sistemas necessitam de um banco de dados que dê suporte a dados multidimensionais. O SIG devem ser capaz de manipular os dados espaciais e recuperar informações como proximidade, pertinência, adjacência, interseção etc. Lisboa (1996). Nos últimos anos, além dos bancos de dados comerciais, vem surgindo um grande número de tecnologias baseadas em Software livre para trabalhar com dados geográficos.[Queiroz 2006].

2.2 PostgreSQL:

O PostgreSQL, Stonebraker e Rowe (1986) é um sistema gerenciador de banco de dados objeto relacional gratuito e cujo o código fonte é aberto, desenvolvido a partir do projeto Postgree em 1986 em Berkeley, Universidade da Califórnia. A partir de 1995 o suporte a linguagem SQL foi incorporado e então um grupo de desenvolvedores vem mantendo e aperfeiçoando o código fonte sob o nome de PostgreSQL. Queiroz (2006). O postgresQL possui um grande suporte ao desenvolvimento de extensões. Essa característica possibilitou o desenvolvimento de uma extensão, o PostGIS, que foi desenvolvida pela empresa Canadense Refractions Research Inc. O código fonte é aberto (GNU GPL). Queiroz (2006)

2.3 Spring:

O Spring é um sistema de informações geográficas desenvolvido pelo DPI(divisão de processamento de Imagens) do INPE, é um software livre e é distribuído gratuitamente para os sistemas operacionais Linux e Windows através do endereço www.dpi.inpe.br. Queiroz (2006). O Spring inclui funções para o processamento de imagens, análise espacial, modelagem numérica de terreno, edição, importação, exportação e consultas a um banco de dados geográficos. O Spring utiliza a arquitetura dual, portanto em um bando de dados SIPRING, os dados geográficos são armazenados separadamente em arquivos formato ASCII-SPRING em um outro SGDB.

2.4 Oracle:

O Oracle Spatial é um conjunto integrado de funções e procedimentos que possibilita dados espaciais serem armazenados, acessados e analisados de forma rápida e eficiente num banco de dados ORACLE 9i. Murray (2001) O Oracle Spatial da suporte a dados de 2-4 dimensões usando um objeto do tipo SDO_GEOMETRY. Esse objeto modela todos os tipos de dados determinados pelo Open GIS Consortium (OGC) e da suporte a maioria das aplicações CAD/CAM e SIG. Kothuri (2002)

2.5 Escolha entre os SGDBs com extensões espaciais

Segundo Queiroz (2006) temos outras opções de SGDBs que possuem extensões espaciais, entre eles podemos citar: IBM DB2 Spatial Extender, o Informix Spatial e Geodetic Datablade e a extensão do MySQL. A Base de dados do inventário de Minas possui um

volume considerável de dados. Esses dados são de extrema importância e necessitam de um banco de dados que seja seguro, confiável e que armazene os dados de maneira consistente.

Recurso	Oracle Spatial	PostgreSQL com Tipos Geométricos	PostgreSQL com PostGIS	MySQL
Tipos espaciais	SFSSQL	Tipos simples	SFSSQL	SFSSQL
Indexação espacial	R-Tree e QuadTree	R-Tree nativa ou R-Tree sobre GiST	R-Tree sobre GiST	R-Tree
Operadores topológicos	Matriz 9-Interseções	Não	Matriz 9-Interseções DE	Em desenv.
Operadores de conjuntos	Sim	Não	Sim	Em desenv.
Operador de <i>buffer region</i>	Sim	Não	Sim	Em desenv.
Transformação entre sistemas de coordenadas	Sim	Não	Sim	Não
Tabelas de metadados das colunas geométricas	Sim (diferente do OGIS)	Não	Sim (conforme OGIS)	Não

Tabela 1 – Comparações entre SGDB, fonte Queiroz (2006)

Segundo Kothuri(2002) , o oracle possui para dados de poucas dimensões dois tipos de índices espaciais: R-tree e Quadtree. Esses índices utilizam das melhores propostas das pesquisas sobre índices espaciais. A tabela 1 faz um comparativo entre os SGDBs que possuem extensões espaciais. A equipe de tecnologia de informação do Inventário de Minas concluiu que o Oracle espacial entre os SGDB's, é que atende melhor entre os requisitos do Inventário de Minas Gerais .

3.1 Índices espaciais

Segundo Gutman (1984), dados espaciais que cobrem áreas em espaços multidimensionais não são bem representados por pontos no espaço. Dados de aplicações CAD, e aplicações SIG possuem essa característica. Uma operação comum desse tipo de aplicação é achar todos os objetos que estão dentro de uma determinada área, todos objetos que possuam algum ponto em comum ou pesquisas com algum predicado topológico. Para realizar esse tipo de operação é necessário um mecanismo de que acelere as operações espaciais. Um método de indexação baseado na localização do objeto é então necessário para que esse tipo de operação seja realizado de maneira eficiente. Estruturas clássicas de dados de apenas uma dimensão como tabelas hash ou árvores-B não funcionam para dados de mais de uma dimensão, pois essas estruturas trabalham com comparações de igualdade entre variáveis enquanto aplicações espaciais necessitam realizar comparações entre intervalo. Segundo Sellis (1987), Índices espaciais são estruturas de dados que permitem promover acesso a dados de mais de uma dimensão de maneira eficiente. Diversas estruturas de dados foram desenvolvidas com o propósito de indexar dados de mais de uma dimensão, entre elas estão: Quadtrees, R-trees, hB-trees, TV-trees, SS-trees. Kothuri (2002). Existem também índices espaciais híbridos QR-tree , Fu(2003).

Nesse artigo, entraremos em detalhes apenas nas Quadtrees e R-trees que são os índices espaciais inclusos no framework Oracle Spatial. Queiroz (2006) sugere que uma idéia fundamental para entendermos a indexação espacial, que é o uso de aproximações das geometrias, isto é, as estruturas do índice trabalham com representações mais simples dos objetos, como por exemplo, o menor retângulo envolvente. Essas aproximações nos

permitem a realizar operações espaciais de maneira mais eficiente. A maneira mais comum de se trabalhar com as consultas espaciais é realizar a busca em duas etapas: uma etapa de filtragem e uma etapa de refinamento. O principal objetivo deste método é o de reduzir e selecionar rapidamente os candidatos que satisfaçam um consulta ou operação espacial. A redução do espaço de busca é muito importante, pois a etapa seguinte, a de refinamento utiliza algoritmos geométricos computacionalmente caros que são aplicados a geometria exata dos objetos espaciais.

Quadtree

Segundo FU (2003), as quadtrees foram uma das primeiras estruturas de dados para indexar dados de mais de uma dimensão. Finkel e Benley desenvolveram a mesma em 1974 e a partir de então surgiram centenas de artigos e livros que tratam as suas aplicações e variações.

De acordo com Lisboa(1996), uma quadtree é uma família de estrutura de dados hierárquicas, todas baseadas no princípio de divisão recursiva do espaço.conforme a figura 2.

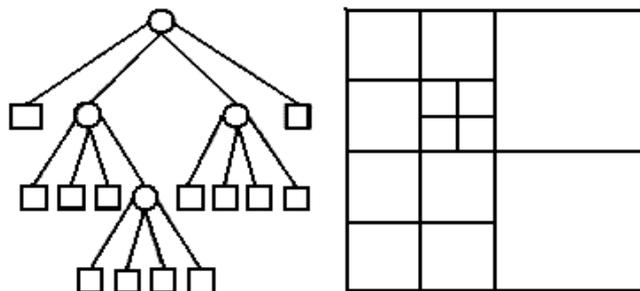
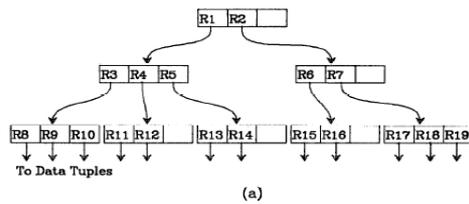


Figura 2. Quadtree e suas decomposições - Fonte YU-CHEN FU (2003)

Rtree

A Rtree foi criada por Guttman (1984), que a definiu como sendo uma árvore balanceada similar a uma árvore-B com os índices nos nós folhas apontando para os objetos espaciais. Os objetos espaciais são aproximados pelo menor retângulo envolvente. Os nós correspondem a páginas em disco se a estrutura de dados esta armazenada em disco. A estrutura é projetada de forma a visitar um número pequeno de nós para se chegar a um nó folha. Algumas propostas com melhorias sobre a Rtree original foram propostas ao longo dos anos, como por exemplo a R+-tree Sellis (1987) .



(a)

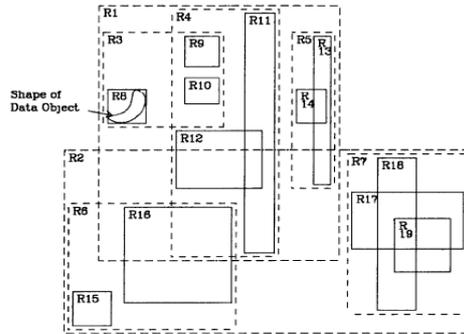


Figura 3 Rtree- Fonte Gutmann (1984)

Índices espaciais no Oracle Spatial:

O Oracle Spatial da suporte a criação de índices para dados espaciais. Esses índices espaciais podem ser criados com extensões de sintaxe SQL. As Rtrees no Oracle spatial logicamente estão implementadas como árvores e internamente estão implementadas em tabelas do Banco de dados. As buscas envolvem SQL recursivos para se chegar da raiz aos nós da árvore. Kothuri (2002) Essa abordagem resulta em buscas mais eficientes devido a uma melhor preservação de aproximações espaciais, mas pode tornar-se lenta para atualizações e para criar-se os índices.

A Quadtree realiza aproximação das geometrias através de aproximações com quadrantes, que são resultantes da subdivisão recursiva do espaço (tesselamento). A quadtree utiliza índices e árvores-B para realizar buscas espaciais e outras operações DML(*Data Manipulating Language*). Essa abordagem acarreta em: Uma criação simplificada para os índices, atualização rápida dos índices e uma herança de um controle de concorrência de árvore B. Kothuri (2002)

Processamento de buscas espaciais no Oracle:

Para realizar as buscas espaciais de forma eficiente, o Oracle Spatial utiliza o modelo de busca multi-estágios. No primeiro filtro, denominado filtro primário os índices são utilizados como filtro de busca Kothuri(2002). Para realizar uma busca espacial, no qual se busca encontrar determinadas geometrias que satisfaçam um determinado predicado geométrico-espacial, as geometrias candidatas a satisfazerem o predicado são selecionadas primeiramente através dos índices espaciais.

No filtro intermediário, as geometrias candidatas são comparadas utilizando aproximações interiores das geometrias de busca e as candidatas. Algumas são selecionadas e outras não dependendo do critério de busca.

O ultimo estágio de busca, denominado filtro secundário é onde são realizados cálculos computacionais com as geometrias exatas. Esses cálculos são computacionalmente dispendiosos e determinam o resultado exato da busca que eventualmente são entregues ao usuário.

Funcionamento da Quadtree como filtro primário:

O Quadtree necessita que o usuário defina um *tiling level*, (número de subdivisões recursivas do espaço). Cada geometria é então aproximada por um conjunto mínimo de quadrantes. Kothuri (2002). Durante o de subdivisão do espaço, também conhecido como *tesselation*, os quadrantes são divididos entre quadrante interiores e de fronteiras, dependendo se todos estão no interior da geometria ou não. Todos os quadrantes que envolvem a geometria são então introduzidos uma tabela de índices espaciais. Como resultado, uma determinada geometria poderá ter múltiplas linhas na tabela de índices espaciais, onde cada linha guarda diferentes códigos de quadrantes interior|exterior e o rowID da geometria (linha da geometria na tabela). Uma árvore-B é então montada contendo o código do quadrante e o rowID da geometria.

Quando se realiza uma busca que envolve uma geometria de busca como parâmetro, a geometria de busca ocorre uma *tesselation* para a mesma. Usando cada quadrante da geometria de busca e o RowID chegamos a todos os geometrias que possuam quadrantes com a mesma posição dos quadrantes da geometria de busca.

Quanto maior o *Tiling level*, mais precisas são as aproximações das geometrias, conseqüentemente mais refinadas são as buscas, porém existe compromisso armazenamento em disco. Encontrar o *tiling level* correto é crucial pois os quadrantes podem ser pequenos demais ou grande demais. Kothuri (2002) Quanto maior o *tiling level* mais seletivo é o filtro primário, conseqüentemente menos geometrias são submetidas ao filtro secundário. Existe uma função do Oracle que estima o *tiling level* correto para as geometrias do Banco de dados. Essa função é a `SDO_ESTIMATE_TILING_LEVEL()`;

Funcionamento da Rtree como filtro primário:

No Oracle Spatial, as Rtree mantém a sua estrutura lógica de árvore e são implementadas como uma tabela onde cada nó da Rtree corresponde a uma linha em uma tabela e um ponteiro filho na Rtree corresponde a um RowID na mesma tabela. A Raiz da árvore esta presente nos metadados o que permite uma navegação da raiz a folha da mesma árvore. Kothuri (2002)As Rtree no Oracle Spatial, mantém a sua estrutura lógica de árvore, porém estão implementadas como uma tabela onde cada nó da árvore corresponde a uma linha da tabela. Cada nó filho corresponde a um rowID na mesma tabela. A raiz da tabela está na tabela dos metadados o que permite uma navegação da raiz aos nós folhas da árvore. Cada folha possui contém o rowID de cada geometria e também possui o seu mínimo retângulo envolvente. Kothuri(2002).

Indexação espacial para os dados do Inventário de Minas.

Os dados que são espaciais no Banco de dados do inventário de minas foram criados através do software da ESRI ArcGIS 9.1. A parte espacial dos dados, também conhecida como '*shapes*' são importados para o banco de dados oracle através da ferramenta ESRI Arc Catalog via ArcSDE. Os shapes que foram importados são shapes precisos e devido a essa

precisão demandam espaço em disco considerável para o seu armazenamento.

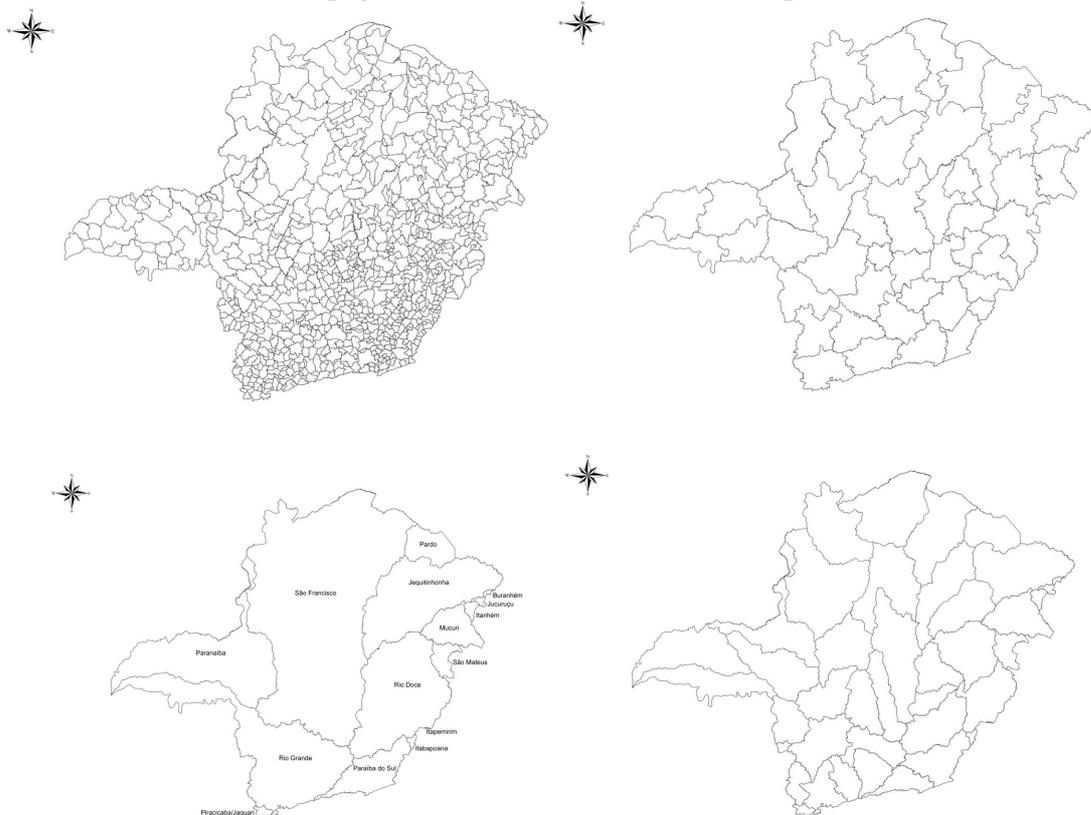


Figura 4 - Municípios, Regionais IEF, Bacias e Sub-Bacias de Minas Gerais

A figura 3 ilustram os municípios do estado de minas gerais, Regionais do IEF, Bacias e Sub-Bacias hidrográficas, sendo que todos esses shapes estão armazenados no Banco de dados em tabelas diferentes. O Oracle utiliza o objeto SDO_Geometry para modelar dados espacial de 2-4 dimensões. Kothuri (2002) Todas as tabelas que armazenam dados georreferenciados possuem um objeto do tipo SDO_GEOMETRY. Cada parcela é um retângulo e necessita de um polígono simples para o seu armazenamento. Os *shapes* dos Municípios, Bacias, Sub-Bacias, Núcleos-IEF, e Regionais-IEF representam subdivisões do espaço bem precisas e complexas. Esses objetos não possuem como características mobilidade e transformações constantes ao longo do tempo.



Figura 5 – Fragmento Fazenda Galheiros, visualização no Google Earth

Comparação entre Rtree e Quadtree

Existem vantagens e desvantagens em cada um dos métodos de indexação espacial do Oracle. A tabela 2 compara os dois tipos de indexação.

Rtree	Quadtree
Aproximação das geometrias não pode ser otimizada	Aproximação das geometrias pode ser otimizada
A criação dos índices é mais fácil, não necessita de 'fine tune' os índices.	Acerto e otimização dos índices é mais complicada, necessita de 'fine tune' os índices.
Menos espaço em disco é utilizado	Mais espaço em disco é utilizado
Se a aplicação envolve o operador SDO_NN Rtree é mais rápida.	Se aplicação necessita do operador SDO_NN a Quadtree é mais lenta
Se a aplicação necessitar de muitas atualizações a Rtree pode não ser uma boa escolha	Atualizações constantes não afetam a performance da quadtree
Pode indexar até 4 dimensões	Somente duas dimensões podem ser indexadas

Tabela 2 – comparação entre Quadtree e Rtree – Fonte adaptada de Murray (2002)

Quando muitas inserções e remoções ocorrem na Rtree, a qualidade da mesma pode degradar, o que pode influenciar a sua performance Murray (2002). Devido ao conteúdo da Base de dados do Inventário Florestal de Minas Gerais ser relativamente estático isso não se torna um problema.

Segundo Kothuri (2002), para aplicações que são atualizadas constantemente, com polígonos simples, com alto acesso de concorrência ou quando as aplicações envolvem algumas buscas espaciais específicas como, o uso da quadtree pode ser recomendado. Porém o 'tiling level' adequado a aplicação deve ser encontrado. As Rtrees não necessitam de acerto de parâmetros e supera as Quadtree em quase todos os outros casos, podendo ter uma performance equivalente ou superior na maioria das ocasiões. Devido dos métodos de indexação espacial e dos dados do Inventário, o índice Rtree é a melhor opção, devido a sua performance em velocidade, facilidade de criação e pela não necessidade do 'fine tune' dos índices. Todos os dados do inventário possuem indexação Rtree e foram criados automaticamente pelo ferramenta ArcCatalog via ArcSDE.

Conclusão:

O Inventário Florestal de Minas Gerais é um projeto que em seu escopo um banco de dados geográfico. Esse banco de dados necessita de SGDB que atenda os requisitos do projeto entre os quais podemos destacar robustez e segurança e devido escolhemos o Oracle Spatial que por trabalhar com a arquitetura integrada, possui funções internas para se trabalhar com os dados espaciais. Para que essas funções espaciais trabalhem de maneira eficiente existe a necessidade da criação de um índice espacial que funciona como um filtro primário em um processo de seleção de dois estágios. Os índices do Oracle são de dois tipos Quadtree e Rtree. Pelo fato de os objetos espaciais do banco de dados serem objetos com formatos complexos, pela facilidade de criação do índice, performance em velocidade e o fato dos dados espaciais não serem atualizados constantemente, o método de indexação do banco de dados é o Rtree para todas as tabelas do banco de dados do Inventário Florestal de Minas Gerais.

Referências:

Guttman Anthonin. R-trees: A dynamic index structure for spatial searching. *Proc. A CM SIGMOD Int. Conf. on Management of Data*, pages 47-57, 1984.

Sellis, Timos; Roussopoulos, Faloutsos Christos: The R+-Tree: A Dynamic Index for Multi-Dimensional Objects 13th VLDB Conference, Pages 507-517 Brighton 1987.

Kothuri, Ravi Kanth V.; Ravada Siva ; Abugov Daniel: Quadtree and R-tree Indexes in Oracle Spatial: A Comparison using GIS Data, 546-557 ACM SIGMOD '2002

Lisboa, Jugurta ; Iochpe Cirano Introdução a Sistemas de Informações Geográficas com Ênfase em Banco de Dados 10^a Escuela de Ciência Informáticas, Departamento de Computación, Universidad de Buenos Aires, Argentina, 22 a 27 Julho 1996 XV JAI – **Jornada de Atualização em Informática, XVI Congresso da SBC**, Recife-PE, 4 a 9 de agosto de 1996

Queiroz, Gilberto Ribeiro ; Ribeiro, Karine Reis: Tutorial sobre Banco de Dados Geográficos. GeoBrasil 2006 **Instituto Nacional de Pesquisas Espaciais**

Hellerstein M. Joseph ; Naughton Jeffery F. ; Pfeffer Avi Generalized Search Trees for Database Systems 21th VLDB Conference Zurich, Switzerland, 1995.

Murray Chuch Oracle Spatial User's Guide and Reference, Release 9.01

StoneBraker, Michael ; Rowe Lawrence A; The design of Postgres