



Dados Espaciais e Indexação

Cristina Dutra de Aguiar Ciferri

Arthur Emanuel de O . Carosia



Tipos de Dados Espaciais

Ponto: menor unidade possível para representar um objeto espacial.

Linha: seqüência de pontos conectados retilinearmente.

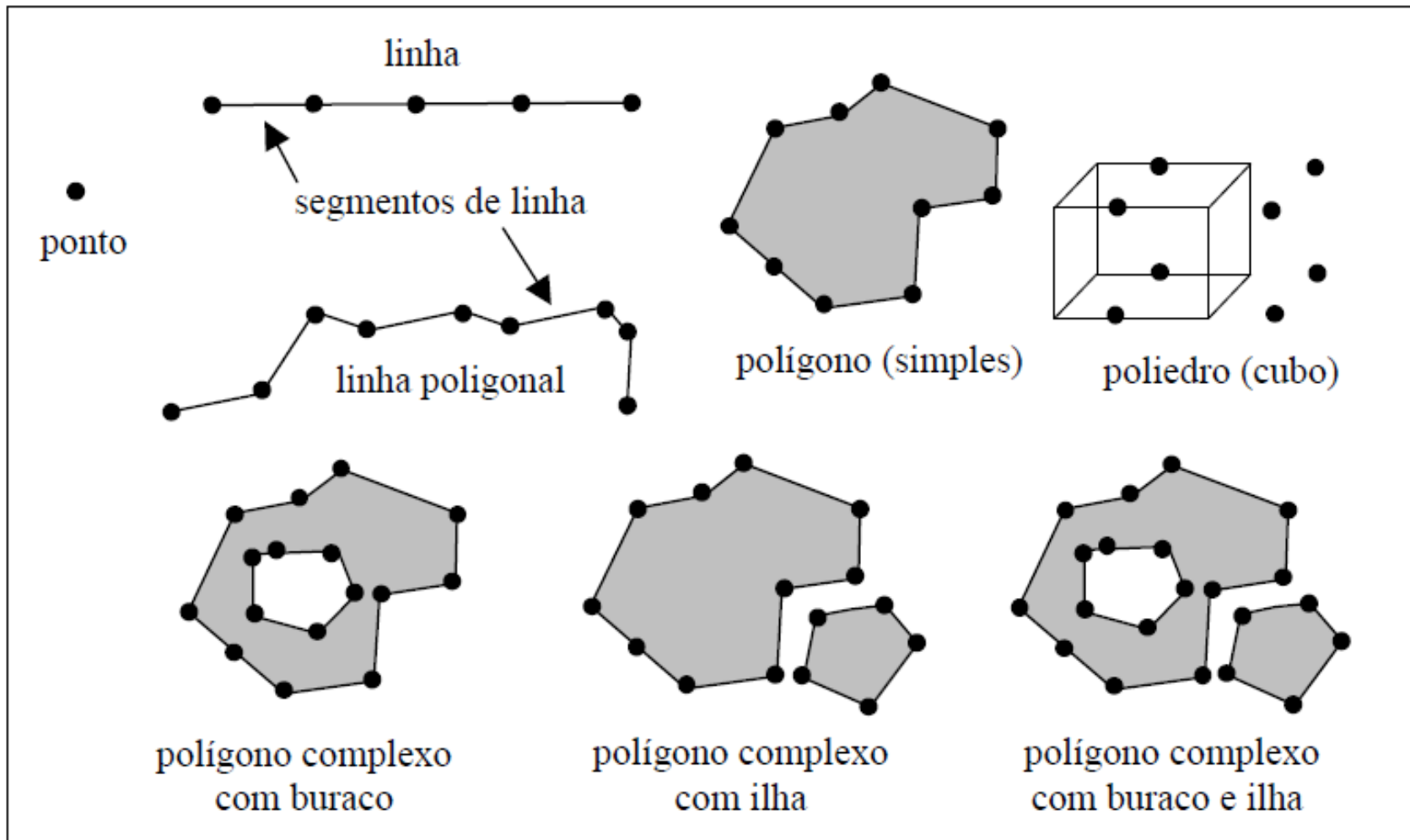
Linha Poligonal: seqüência de pontos que não estão dispostos de forma retilínea.

Polígono: seqüência de linhas ou de linhas poligonais, sendo que esta seqüência é fechada.

Polígonos Complexos: permitem buracos ou consistir de diversas partes disjuntas.

Poliedro: limitado por quatro ou mais polígonos, denominados faces, sendo que as interseções das faces formam as arestas e as interseções das arestas formam os vértices.

Tipos de Dados Espaciais





Estratégias de Representação dos Dados

Estratégias para armazenar objetos:

1. Representação dos objetos em memória secundária:
 - geometria exata do objeto.
2. Representação utilizada por MAM:
 - representação destes por formas geométricas mais simples (aproximadas);
 - diminuição dos requisitos de armazenamento e do custo para se determinar a satisfação de relacionamentos entre os objetos.



Estratégias de Representação dos Dados

Aproximação conservativa:

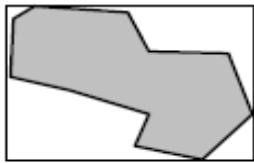
- cada ponto da geometria do objeto espacial deve estar contido na geometria da aproximação.

MBR (*Minimum Bounding Rectangle*):

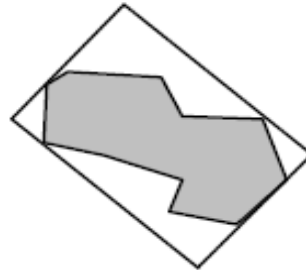
- “retângulo envolvente mínimo”;
- consiste do menor retângulo com lados paralelos aos eixos que contém completamente o objeto espacial;
- aproximação conservativa;
- requer somente alguns poucos *bytes* referentes a quatro coordenadas (i.e., intervalos $I_x = [x_L, x_U]$ e $I_y = [y_L, y_U]$).

Estratégias de Representação dos Dados

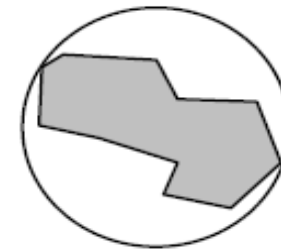
Exemplos de aproximações conservativas:



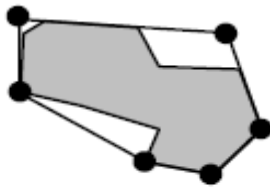
retângulo envolvente mínimo



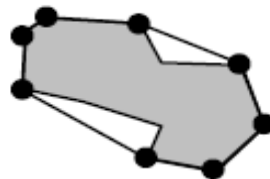
retângulo envolvente mínimo rotacionado



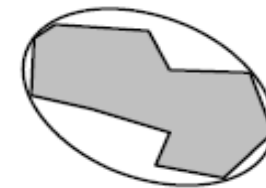
círculo envolvente mínimo



polígono envolvente mínimo com 6 vértices



casco convexo



elipse envolvente mínima

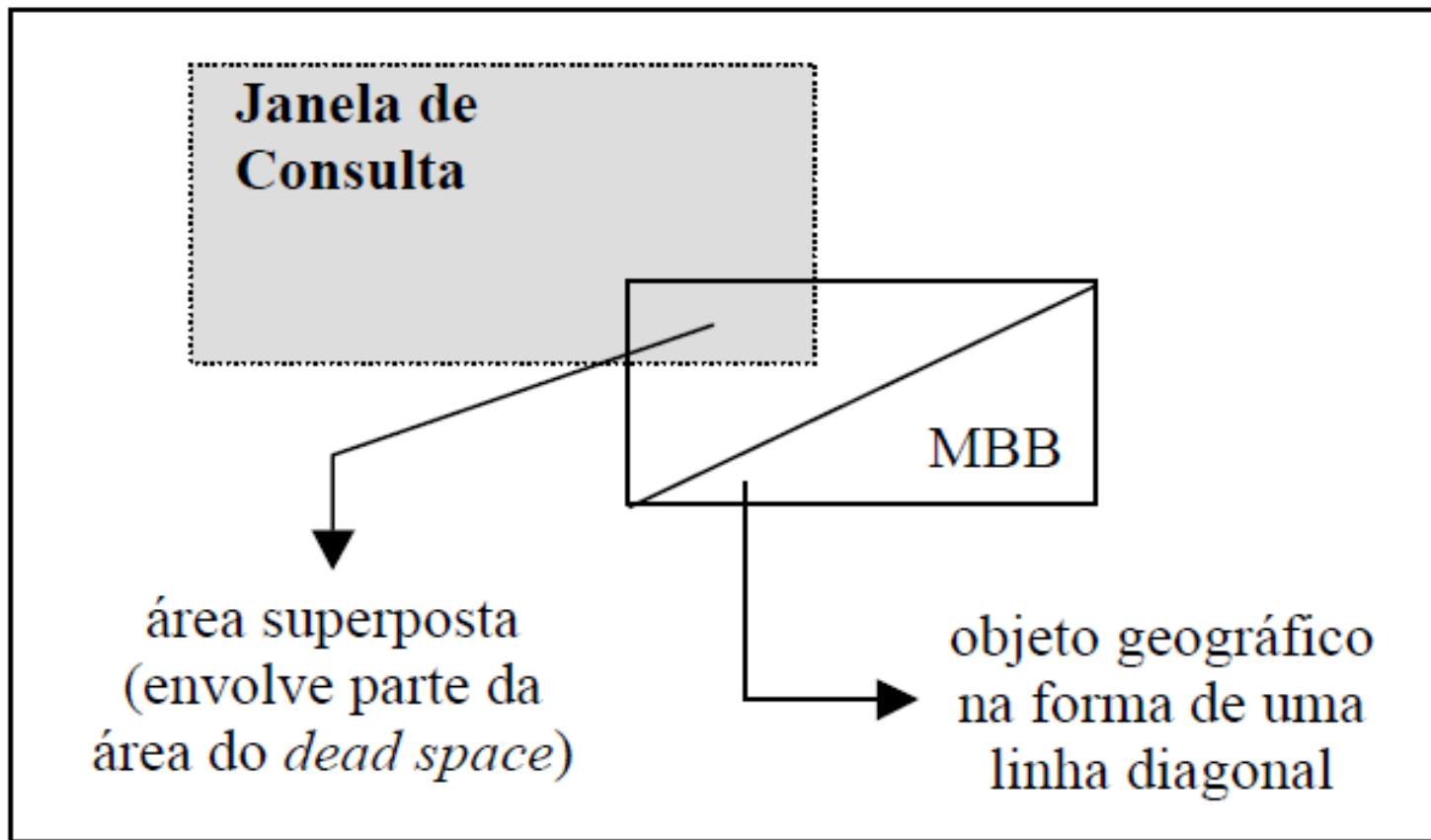


Fase de Filtragem e Refinamento

Uso de formas geométricas aproximadas:

- perda de precisão na representação da geometria;
- espaço contido dentro da aproximação que não faz parte da geometria do objeto espacial: *dead space*.
- A área de *dead space*, pode fazer com que o objeto espacial recuperado seja um candidato falso.
- Os MAMs retornam um superconjunto de candidatos. Para contornar este problema, existem as fases de filtragem e refinamento.
- O uso de aproximações garante que nenhum dos objetos espaciais que satisfaz um certo relacionamento espacial seja desconsiderado na resposta da consulta.

Fase de Filtragem e Refinamento





Fase de Filtragem e Refinamento

Fase de filtragem

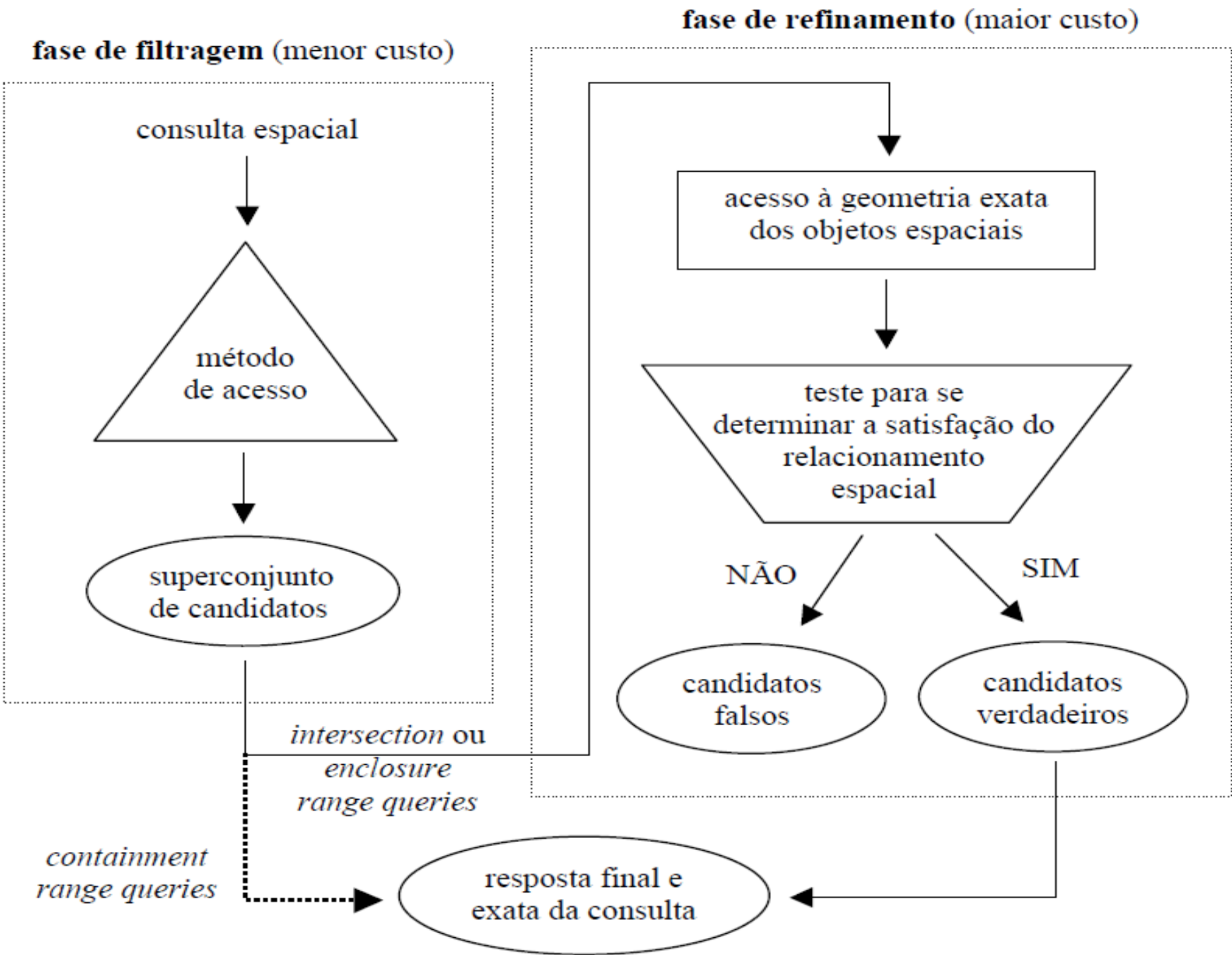
- Baseada em aproximações.
- Gera-se tanto candidatos verdadeiros quanto candidatos falsos.
- Etapa pouco custosa: manipula formas geométricas mais simples
- Baixo custo para se determinar a satisfação de um relacionamento espacial.
- Descarta os objetos espaciais que certamente não satisfazem um determinado relacionamento espacial.
- Usada para restringir a quantidade de objetos espaciais que serão analisados na fase de refinamento.
- Fase é menos custosa do que a fase de refinamento.



Fase de Filtragem e Refinamento

Fase de refinamento

- Presença de candidatos falsos no conjunto de objetos selecionados na fase de Filtragem
- Verificação, para os candidatos, se o relacionamento espacial é satisfeito com relação à geometria exata do objeto espacial
- Descartar candidatos falsos.
- Fase muito custosa:
 - acesso à geometria exata dos objetos espaciais;
 - cálculos geométricos complexos para se determinar a satisfação do relacionamento espacial.





R-Tree

- Mecanismo de indexação espacial estruturado hierarquicamente na forma de uma árvore balanceada (similar à B-tree).
- Usada comumente para a indexação de objetos espaciais armazenados em memória secundária
- É o mais importante dentre os métodos que aplicam o conceito de MBR.
- Provê suporte eficiente para *range queries*.

Exemplo:

Encontre todos os museus dentro de 2 quilômetros da minha localização atual.



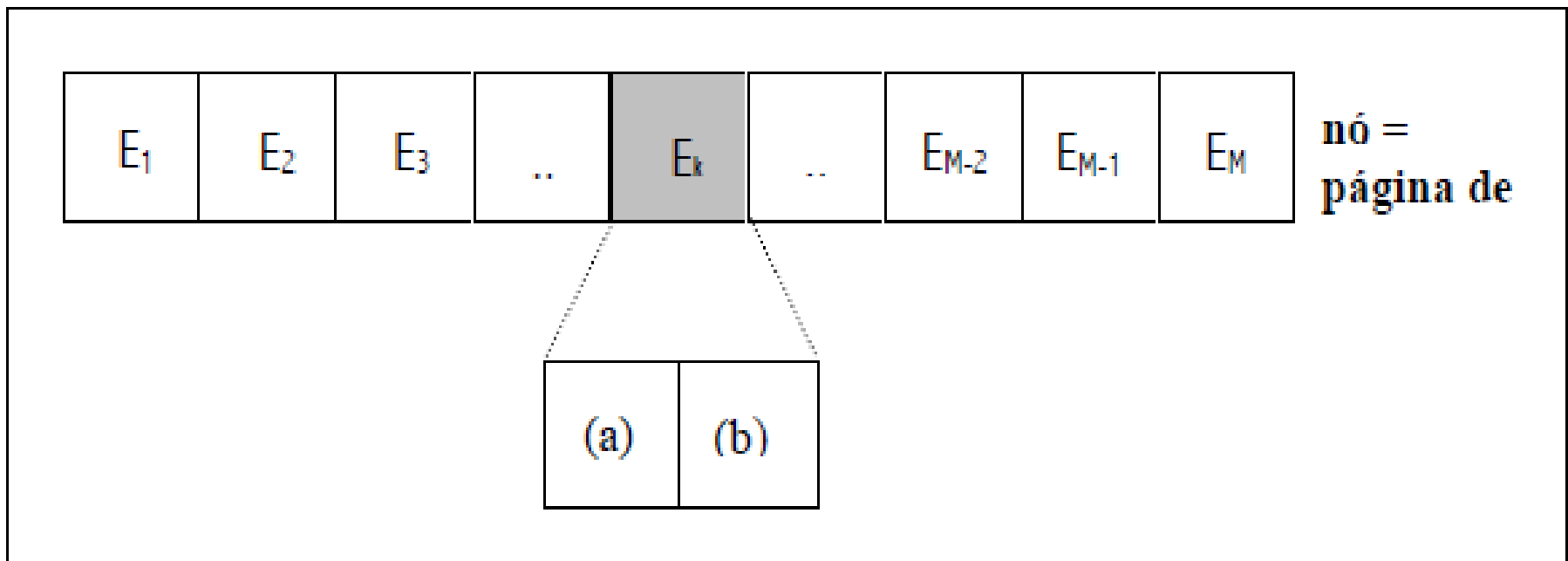
R-Tree

Estrutura de Dados

- A raiz tem no mínimo 2 nós;
- Composta por nós folhas e nós internos;
- Um nó corresponde a uma página de disco;
- Estrutura geral do nó:
 - Máximo de M entradas, mínimo de m entradas ($m \leq M/2$);
 - Contém informações sobre
 - (a) Localização espacial: MBR.
 - (b) Localização dos dados espaciais na memória: referência a um endereço de memória.

R-Tree

Estrutura de Dados



Estrutura dos nós da R-tree.

- (a) Localização espacial
- (b) Localização dos dados espaciais na memória



R-Tree

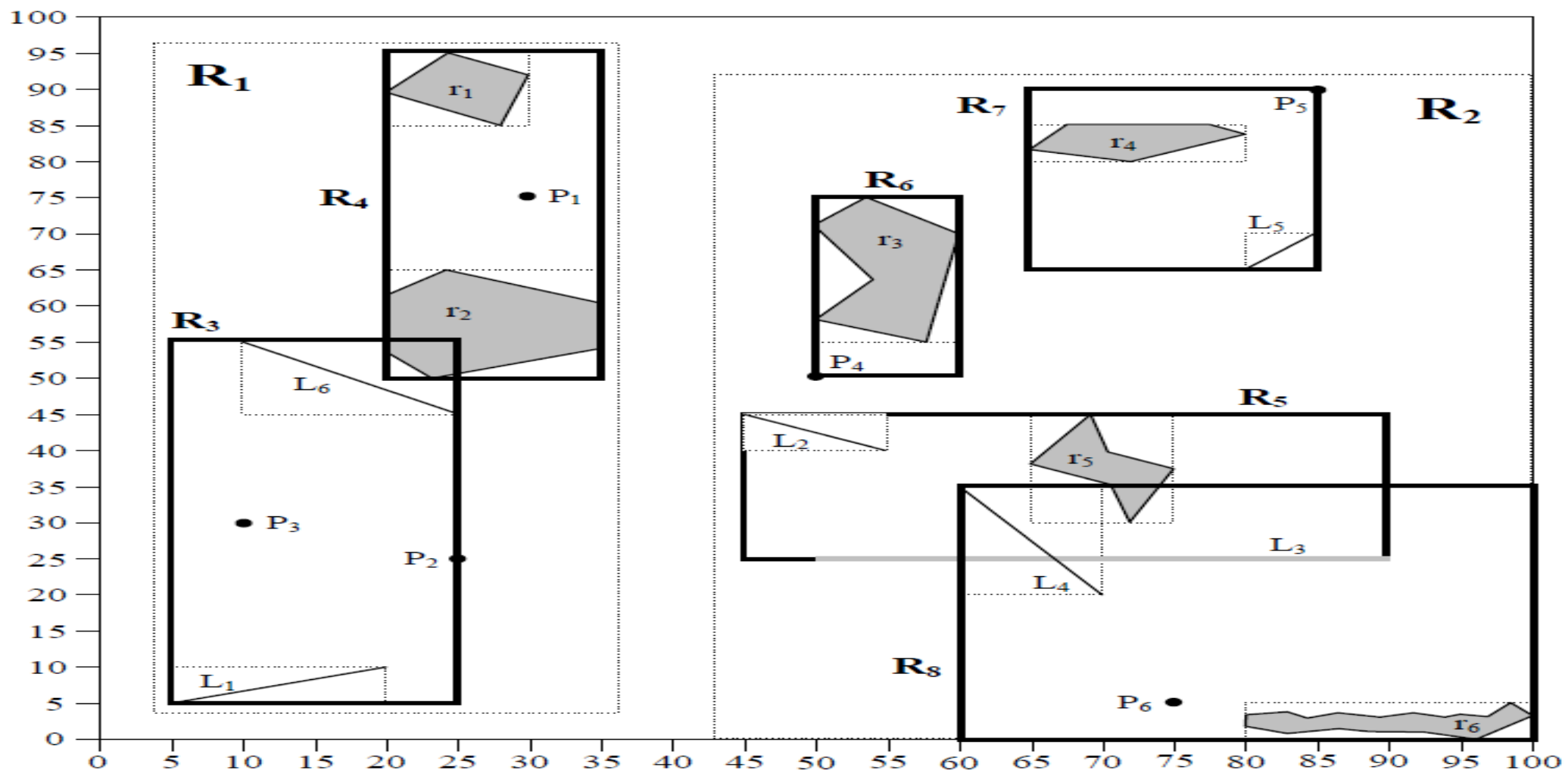
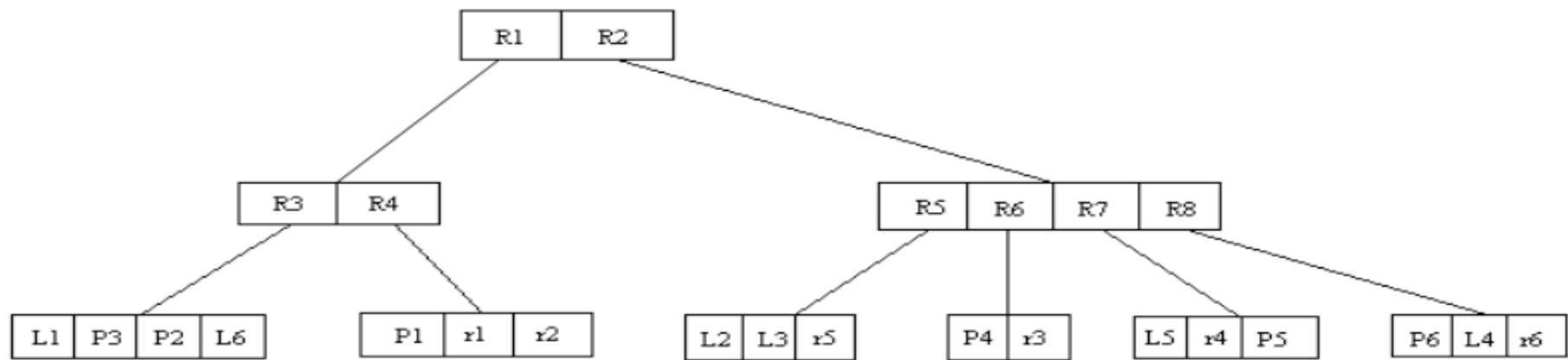
Estrutura de Dados

Nós internos armazenam informações sobre nós de nível imediatamente inferior.

- Um nó interno possui entradas da forma (I, p) , onde:
 - I corresponde ao MBR que engloba o MBB de todas as entradas do nós inferiores;
 - p é o endereço de um nó filho.

Nós folhas armazenam exclusivamente informações sobre os objetos espaciais.

- Um nó folha contém, entradas da forma (I, id) , onde:
 - I corresponde ao MBR do objeto espacial identificado por id ;
 - id é o identificador de um objeto espacial, sendo uma referência a um endereço de memória que possui os dados do objeto.





R-Tree

Algoritmo de Pesquisa

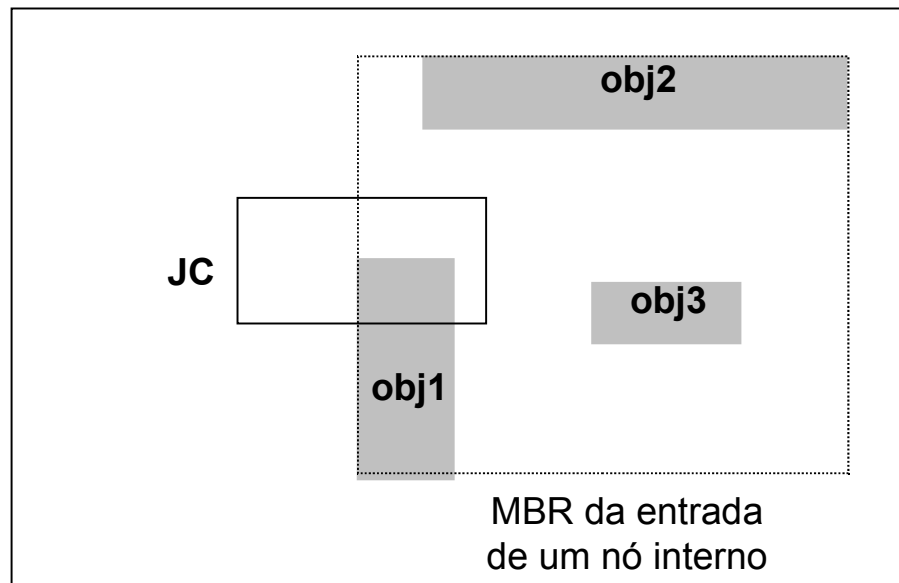
Dado uma R-tree cujo nó raiz é T , encontrar todos os registros cujos MBR sobreponham a janela de busca S .

1. [Procurar nas sub-árvores] Se T não é folha, chegar cada entrada E para determinar se E sobreponha S . Para todas as entradas que sobreponham S , chamar o Algoritmo de Pesquisa na árvore cuja raiz é apontada por E .
2. [Procurar nos nós folhas] Se T é folha, checar todas as entradas E para determinar se E sobreponha S . Se sobrepor, E é um registro a ser retornado.

R-Tree

Algoritmos de Pesquisa - Intersection range query

- Permite a pesquisa de intersecções entre os objetos espaciais e a janela de consulta.



Relacionamento de intersecção entre a janela de consulta e um objeto (obj1)



R-Tree

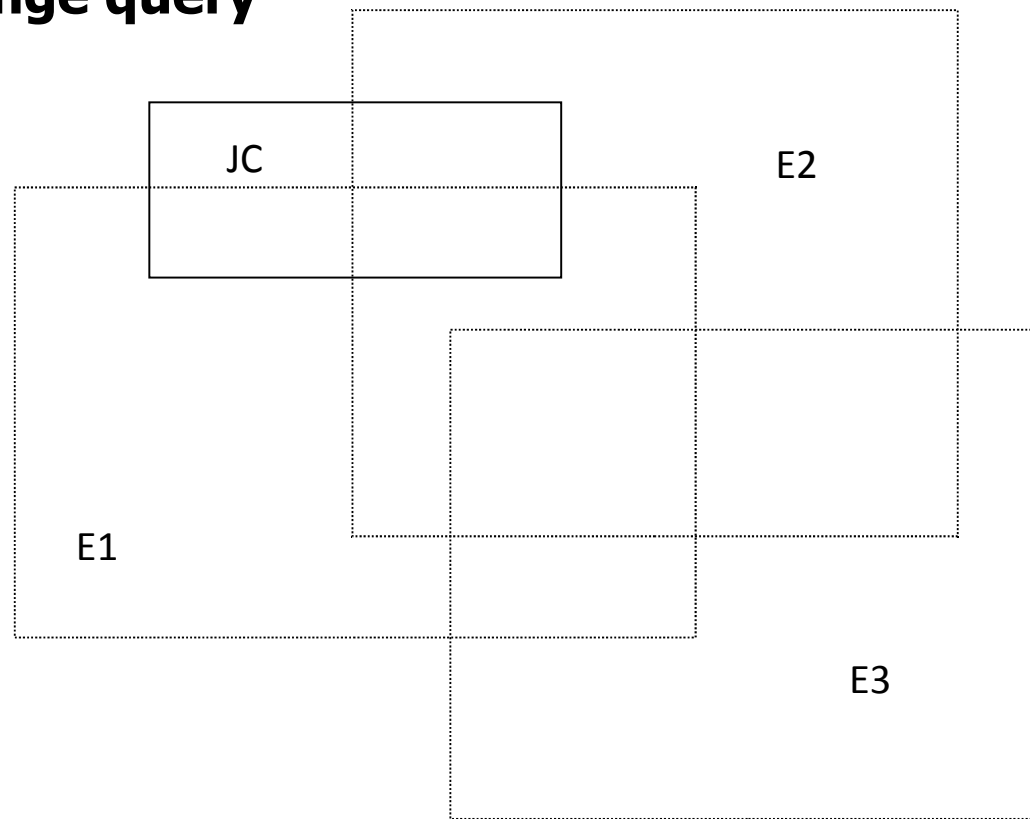
Intersection range query

- A ramificação do percurso inicial reduz o desempenho da pesquisa, pois significa que um número maior de nós terá de ser visitado.
- Essa ramificação é acentuada se a JC intersectar muitas entradas de algum nó interno. O que pode ocorrer devido a:
 - a) Sobreposições entre os MBR's das entradas de um nó interno.
 - b) Armazenamento de MBR's grandes.
 - c) Escolha de uma JC muito abrangente.



R-Tree

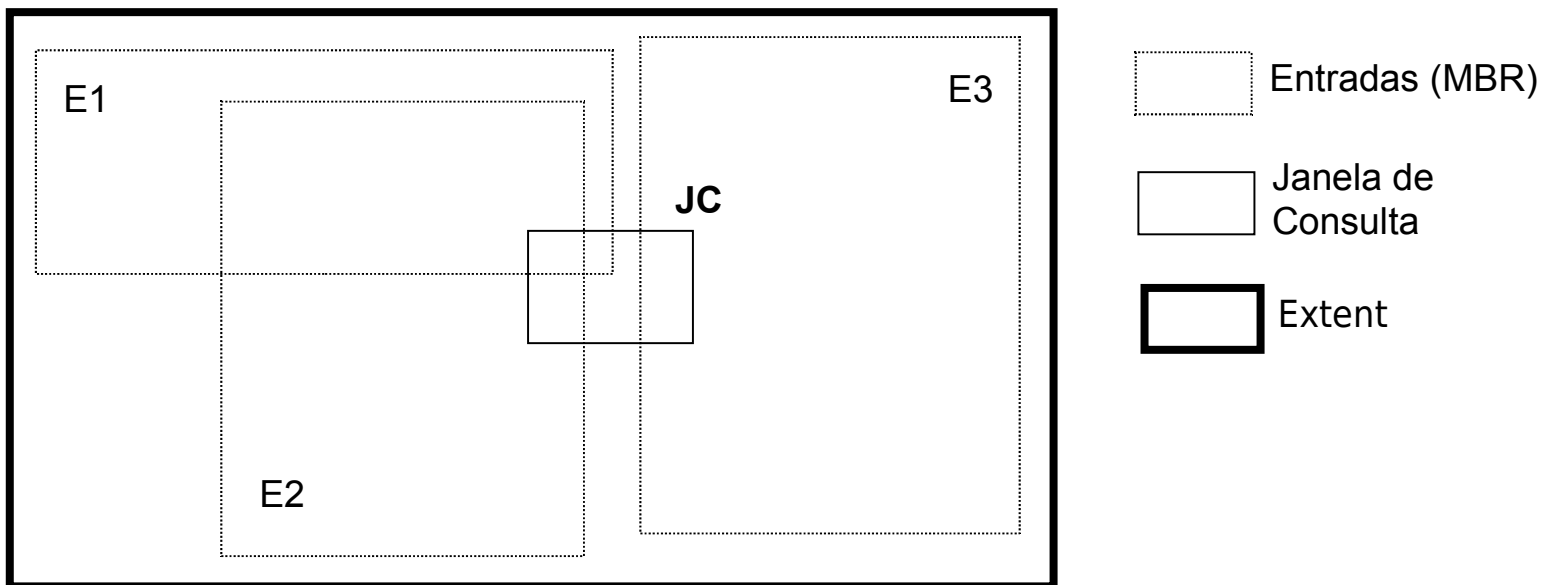
Intersection range query



a) Ramificação devido à sobreposição entre os MBR's das entradas de um nó interno

R-Tree

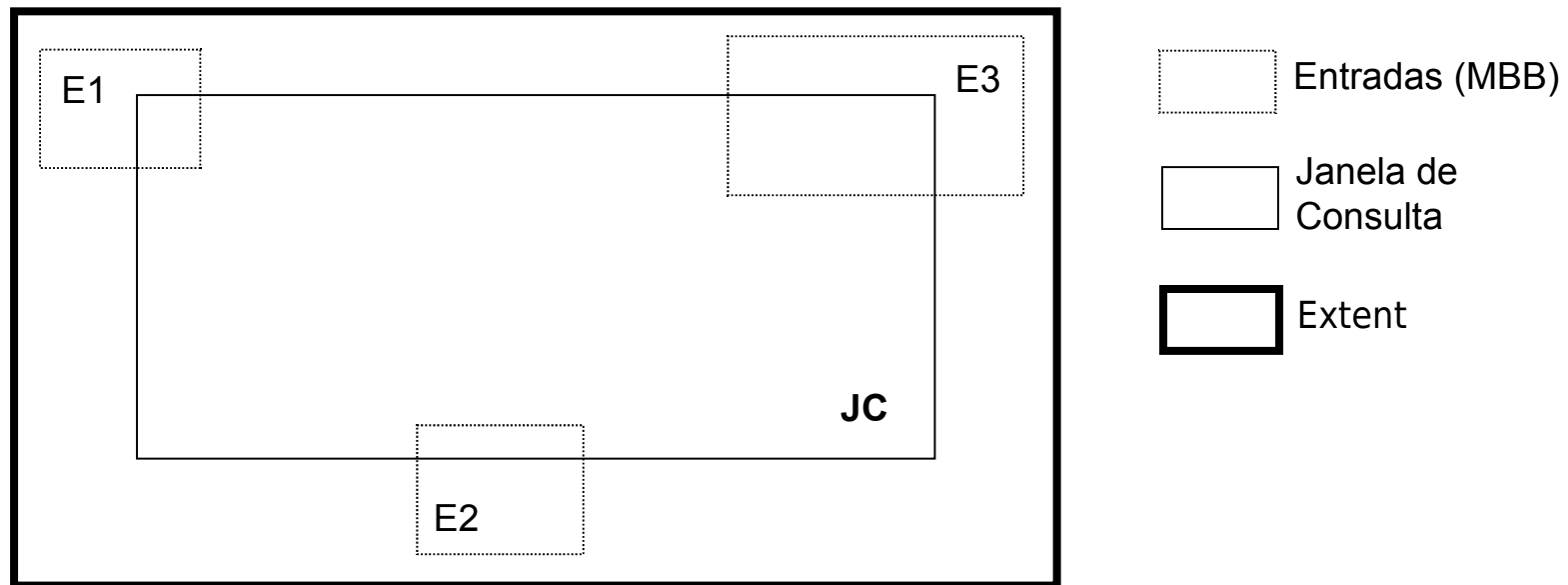
Intersection range query



b) Ramificação devido ao armazenamento de MBR's grandes

R-Tree

Intersection range query

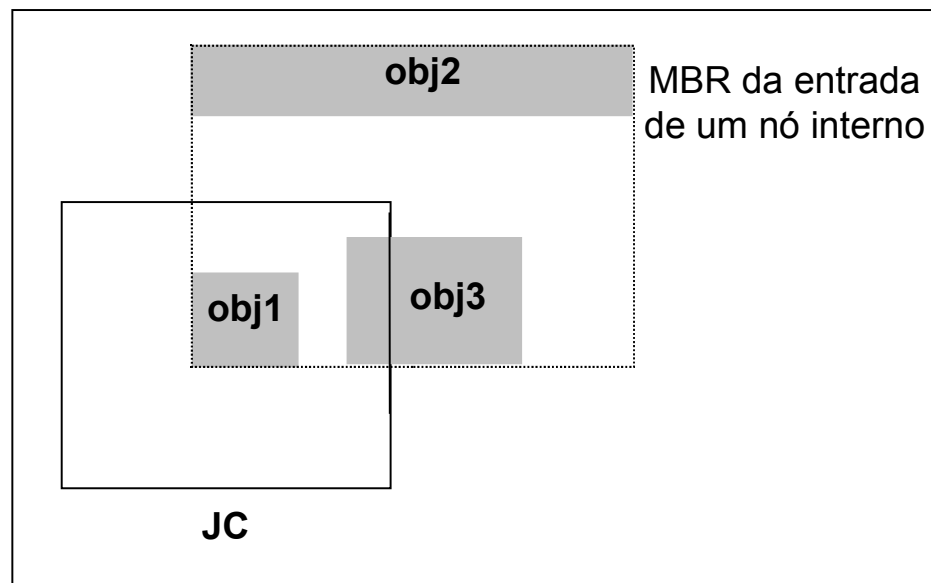


c) Ramificação devido à escolha de uma JC muito abrangente

R-Tree

Containment range query

- São pesquisas por relações de inclusão de objetos espaciais na janela de consulta.

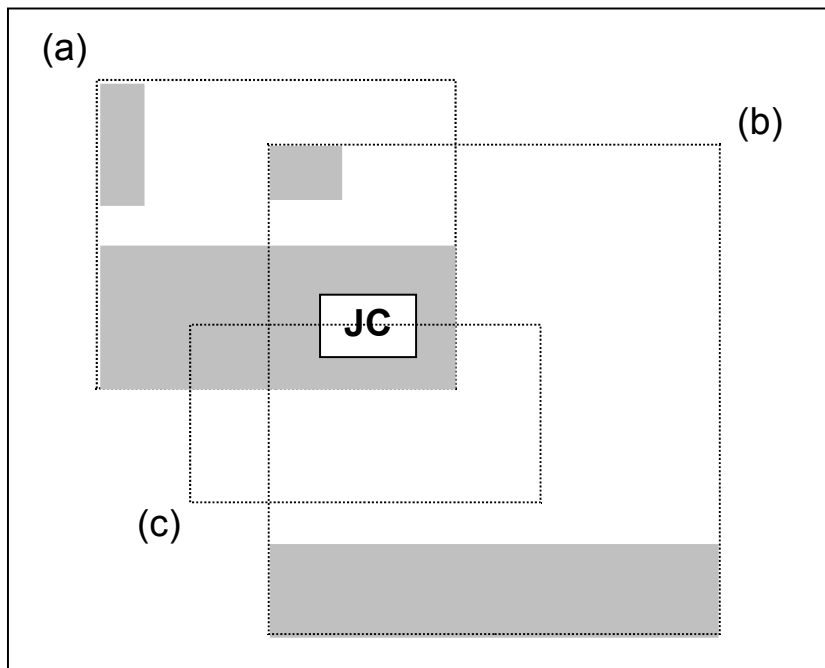


Relacionamento de inclusão entre a janela de consulta e um objeto (obj1)

R-Tree

Enclosure range query

- São pesquisas por relações de inclusão da janela de consulta em objetos espaciais.



O MBR de entrada do nó interno:

- a) Engloba a JC e possui um MBR de objeto espacial que contém a JC.
- b) Não possui nenhum MBR de objeto espacial que contém a JC, mas a entrada engloba esta última.
- c) Não engloba a JC e portanto não possui nenhum MBR de objeto espacial que possa contê-la.



R-Tree

Algoritmos de Pesquisa

- **Intersection e enclosure range queries:**

Necessitam da fase de refinamento para eliminar falsos candidatos selecionados.

- **Containment range queries:**

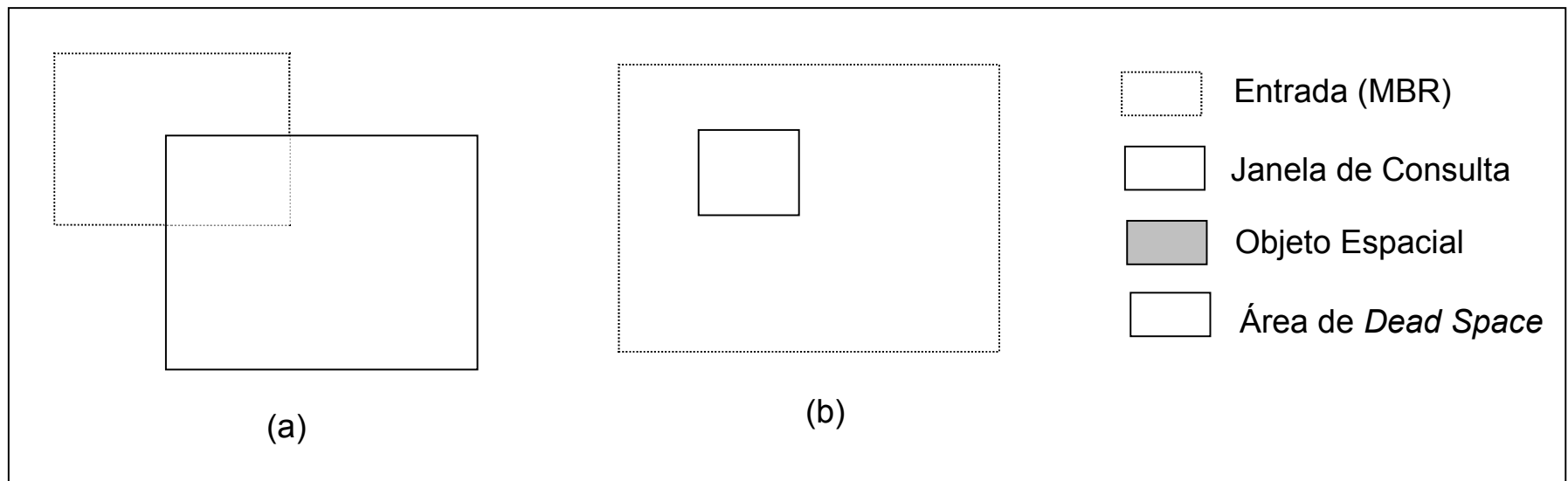
Não precisam de refinamento, pois se a JC contém o MBR de uma entrada e o MBR desta entrada contém um objeto espacial, então a JC contém esse objeto.

- **Enclosure range queries:**

Não costumam requerir uma descida até os nós folhas, pois o fato do MBR da entrada de um nó não englobar a JC já elimina a possibilidade de um MBB de objeto espacial desta entrada contê-la.

R-Tree

Algoritmos de Pesquisa



Candidatos falsos e influência da área de dead space na determinação dos relacionamentos topológicos de (a) Interseção e (b) Inclusão (contém).



R-Tree

Algoritmo de Inserção

1. Escolha do nó folha onde o objeto será inserido.
 - Escolhe-se o caminho que fará a área total do nó folha sofrer o menor aumento (assegura-se que elementos próximos ficarão no mesmo nó sempre que isso for possível).
2. Inserção propriamente dita.
 - Se não houver espaço para novas entradas no nó, o mesmo é particionado, e as $M+1$ entradas são distribuídas entre as partes.
3. Propagação dos efeitos da inserção e do particionamento aos níveis superiores.
 - Os MBR's das entradas dos nós superiores vão sendo atualizados a partir do nó folha no qual ocorreu a inserção.
 - Como isso envolve a criação de novas entradas nos nós superiores, podem ocorrer novos particionamentos.
4. Particionamento do nó raiz.
 - Cria-se um novo nó raiz que terá como filhos os dois nós resultantes do particionamento da antiga raiz.



R-Tree

Algoritmo de Remoção

1. Busca do nó folha que contém a entrada a ser removida.
2. Remoção propriamente dita.
 - Se um nó interno ficar com um número de entradas menor que o mínimo permitido (m), esse nó é eliminado e suas entradas são realocadas na árvore (redistribuição, agrupamento), respeitando-se o nível.
3. Propagação das mudanças em níveis superiores, do nó folha em direção ao nó raiz.
 - Novos agrupamentos e redistribuições podem ocorrer.
4. Ajuste do nó raiz (caso fique com apenas uma entrada, o nó filho passa a ser a nova raiz).



R-Tree

Algoritmo de Modificação de Dados

- Modificações na geometria do objeto espacial indexado pela R-Tree implicam na necessidade de se alterar o MBR da entrada do nó folha utilizado para representar o objeto.
- Um simples ajuste no MBR da entrada pode fazer com que a área total do nó folha aumente excessivamente, caso não haja um controle.
- A modificação é então feita através dos algoritmos de remoção e inserção de entradas.



R-Tree

Demonstração

<http://gis.umb.no/gis/applets/rtree2/jdk1.1/>



Referências

- Antonin Guttmann: R-Trees: A Dynamic Index Structure for Spatial Searching, Proc. 1984 ACM SIGMOD International Conference on Management of Data, pp. 47–57. ISBN 0-89791-128-8
- Apresentação R Tree e R* Tree. UFSCAR. Carlos Henrique Villa Pinto.
- V. Gaede and O. Günther, Multidimensional Access Methods, ACM Computing Surveys, 30 (1998), pp. 170-231.
- CIFERRI, R. R. Análise de Influência do Fator Distribuição Espacial dos Dados no Desempenho de Métodos de Acesso Multidimensionais no Suporte às Consultas Espaciais de Seleção. 2002. Tese (Doutorado em Ciência da Computação). Centro de Informática, Universidade Federal de Pernambuco, Recife, PE, Brasil, 2002.