

Dados espaciais

Árvores R

Estrutura de Dados II

Jairo Francisco de Souza

(baseado no material do Prof Leonardo Guerreiro/UNIRIO)

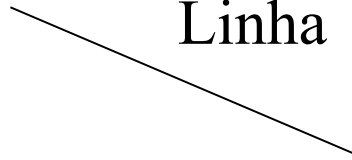
Dados espaciais

- Dados espaciais são dados que representam o espaço
- Exemplos:

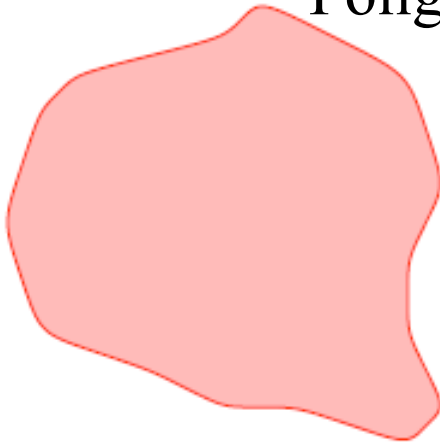
Ponto



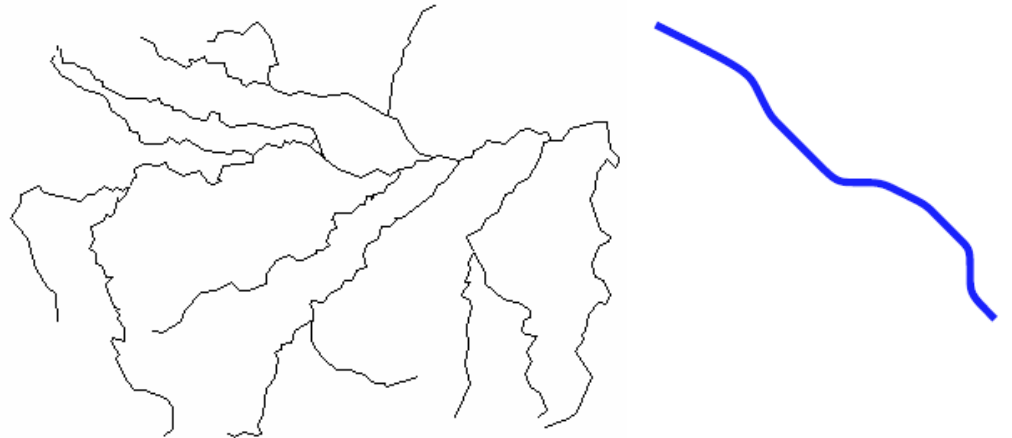
Linha



Polígono



Polilinha



Objetos espaciais

- Objetos espaciais são objetos que descrevem localizações ou formas geométricas
 - Exemplo: localização de um hidrante ou um poço, estradas, rios, redes de esgotos, florestas, parques, municípios, lagos.
- Os 3 tipos básicos de objetos espaciais são:
 - Ponto, Linha e Polígono
- Duas visões:
 - Objetos no espaço
 - Espaço em si

Objetos no espaço

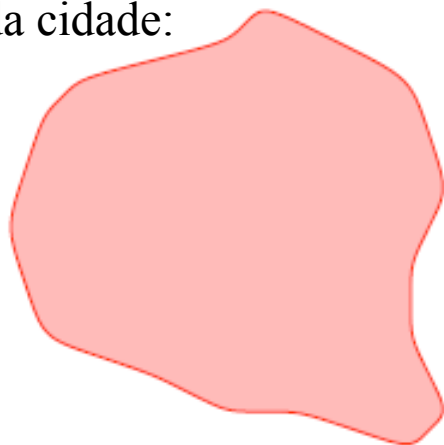
- Dados espaciais geralmente são acompanhados por dados não espaciais

Cidade: Bom
Jardim,

População: 30.000,

Localização:

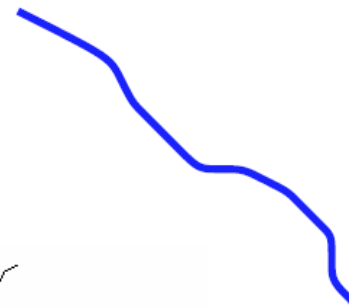
Cidade: Berlin,
População: 3.000.000
Área da cidade:



Rua: Av. Pasteur, número 380,
rota:

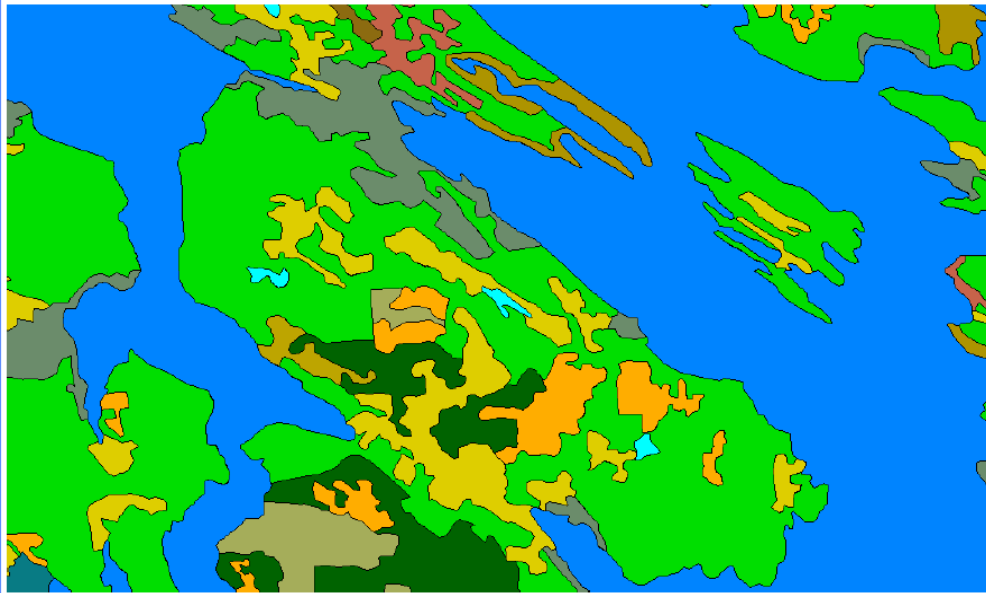
Rio:
Reno,
Rota:

Bacia:
Amazonas,
Leitos dos rios:



Espaço

- Alguma afirmação sobre cada ponto no espaço
 - Uso da terra (mapas temáticos)
 - Partições em cidades, estado e municípios,....




Classificações de Uso da Terra representadas como Polígonos

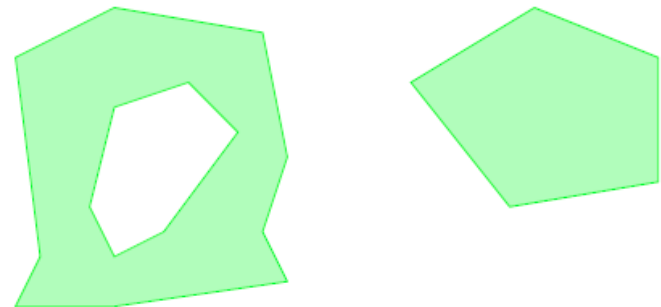


Modelagem

- Devemos considerar:
 1. Modelagem de um único objeto
 2. Modelagem de coleções de objetos relacionadas espacialmente

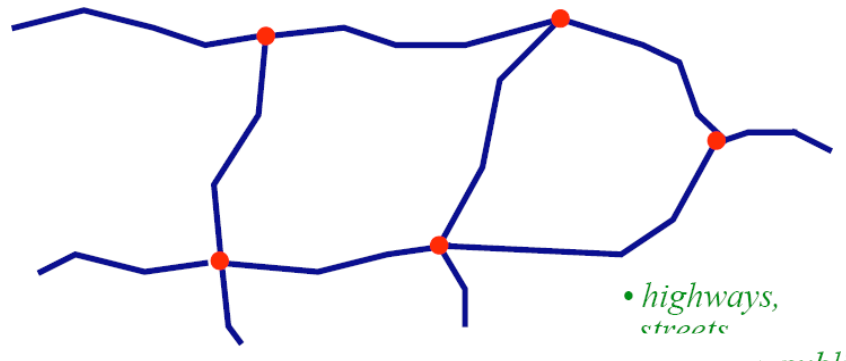
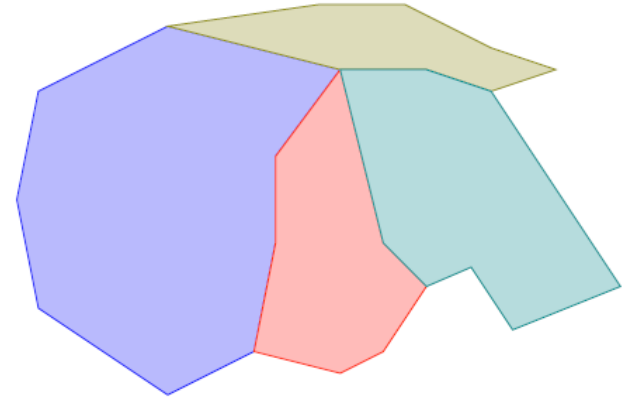
Modelagem

- Modelagem de um único objeto:
 - Ponto 
 - Ex: cidade
 - Aspecto geométrico de um objeto para o qual apenas a sua localização é relevante, e não sua extensão
 - Linha (polilinha)
 - Ex: rio, cabo, estrada
 - Movimento no espaço e conexões no espaço são relevantes
 - Polígono (região)
 - Ex: floresta, lago, cidade
 - Abstração de um objeto com extensão é relevante

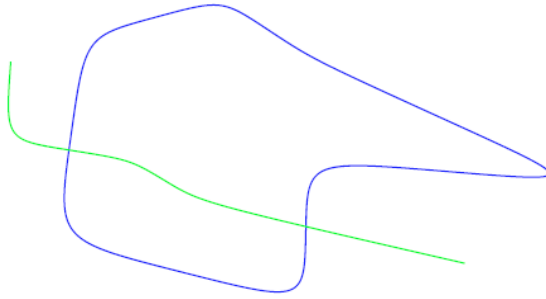


Modelagem

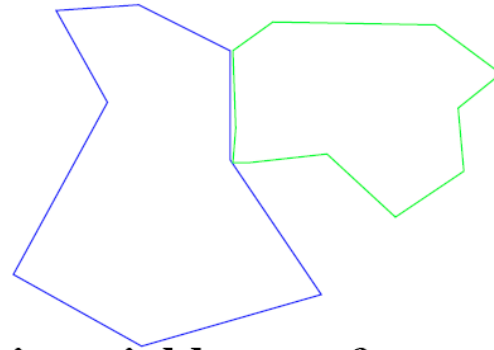
- Modelagem de coleções de objetos relacionadas espacialmente:
 - Partição:
 - Uso do solo, distritos, mapa de propriedade
 - Rede conectada espacialmente (grafo)
 - Autoestradas, estradas, ferrovias, transporte público, rios, mapa de redes elétricas, mapa de redes de telefone



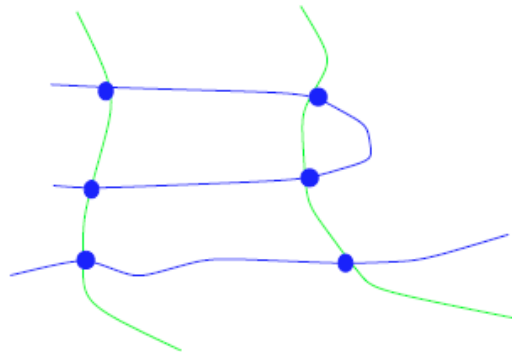
Operações



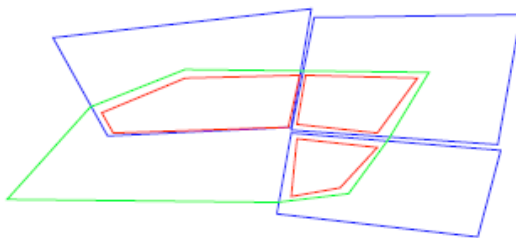
intersects



is_neighbour_of

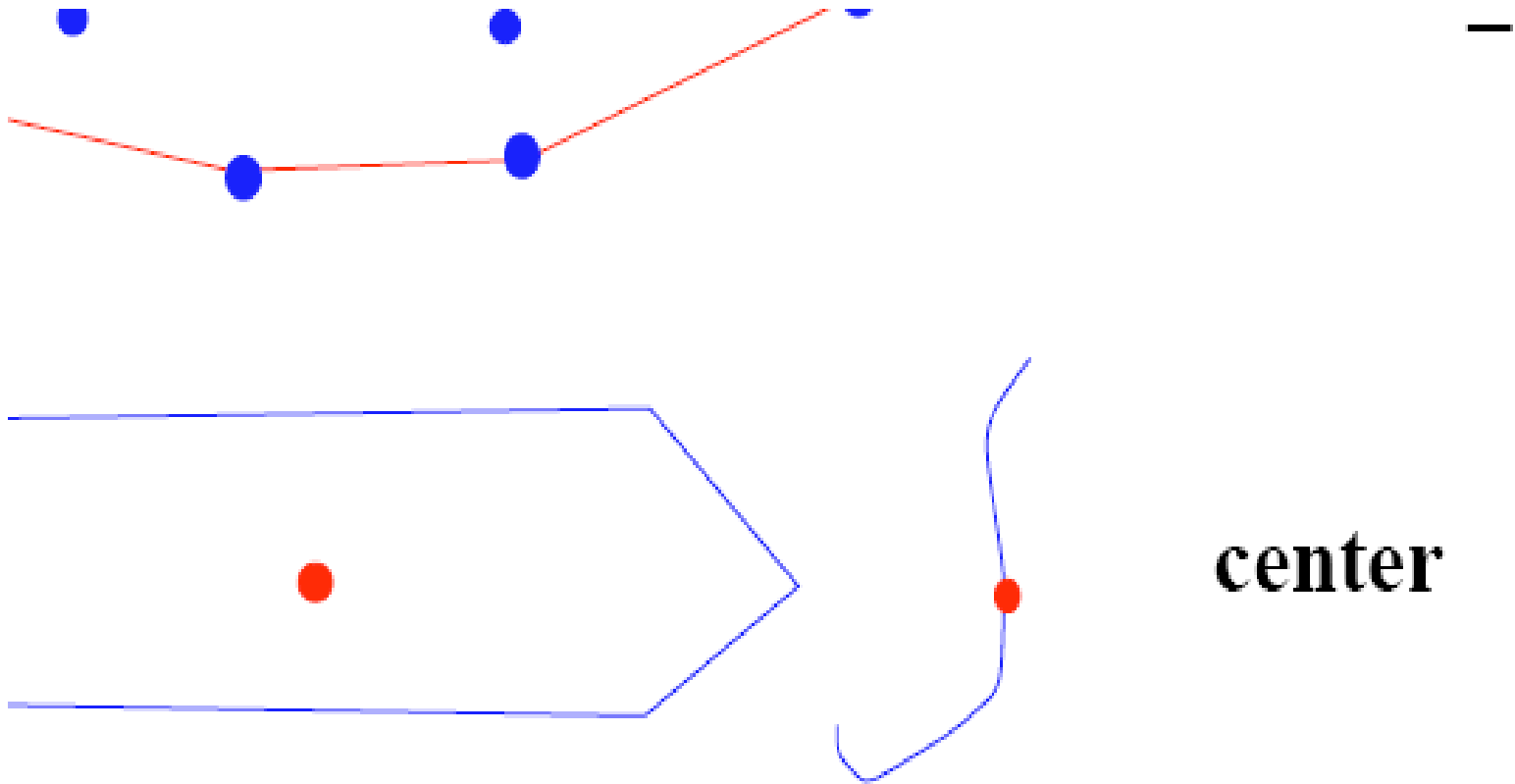


intersection

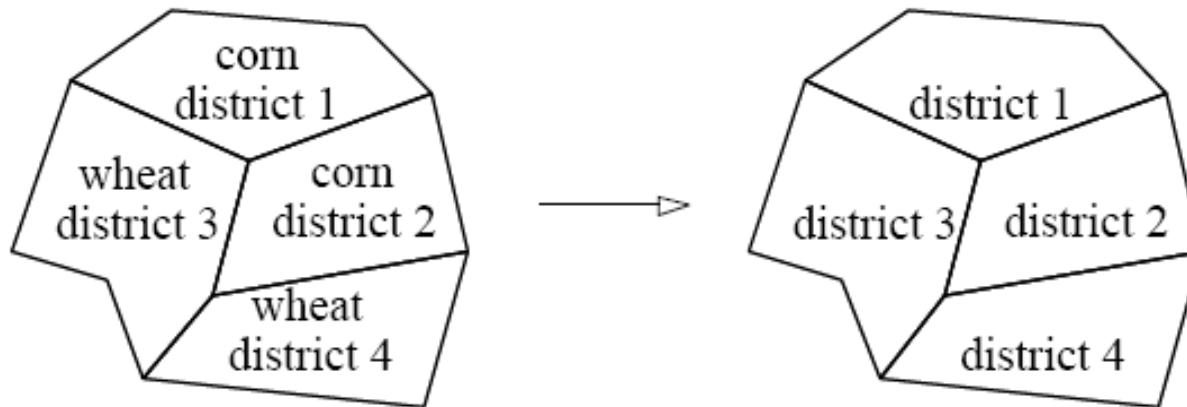


overlay

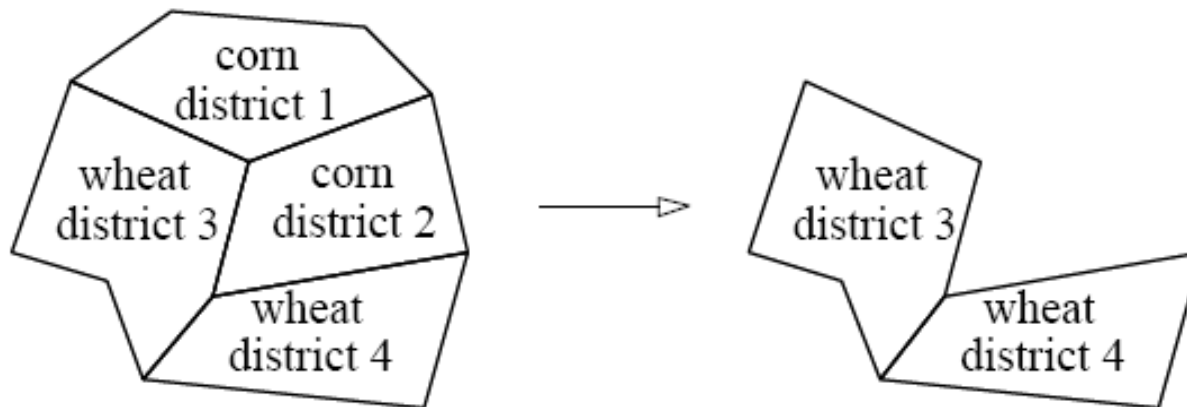
Operações



Operações

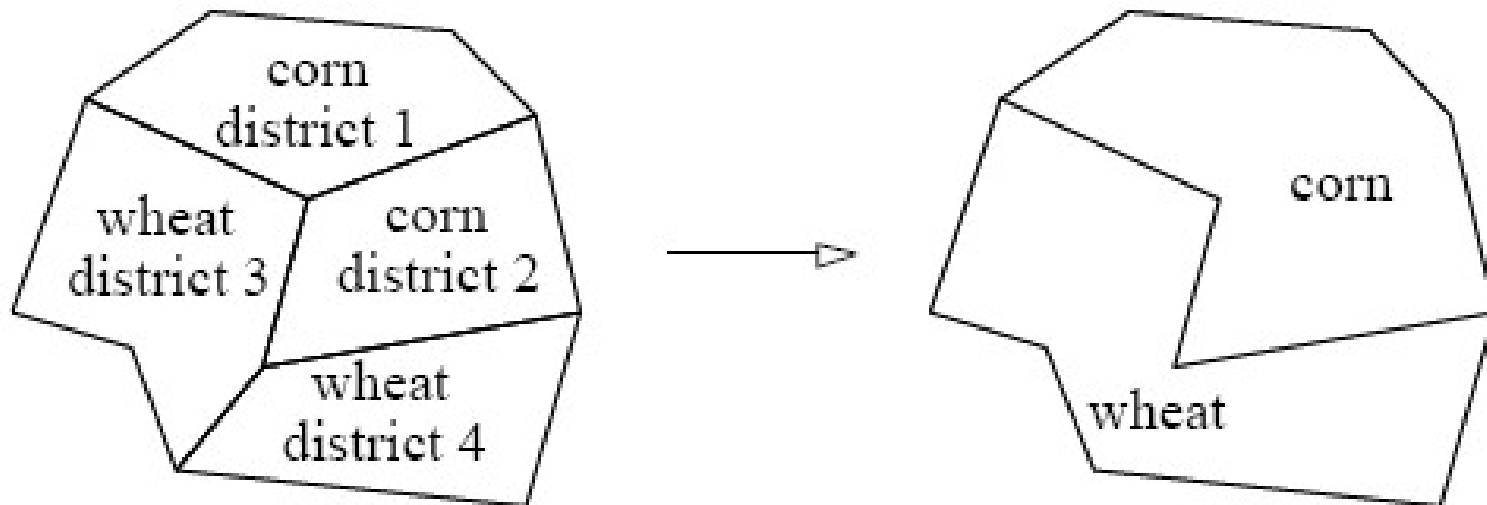


Projection



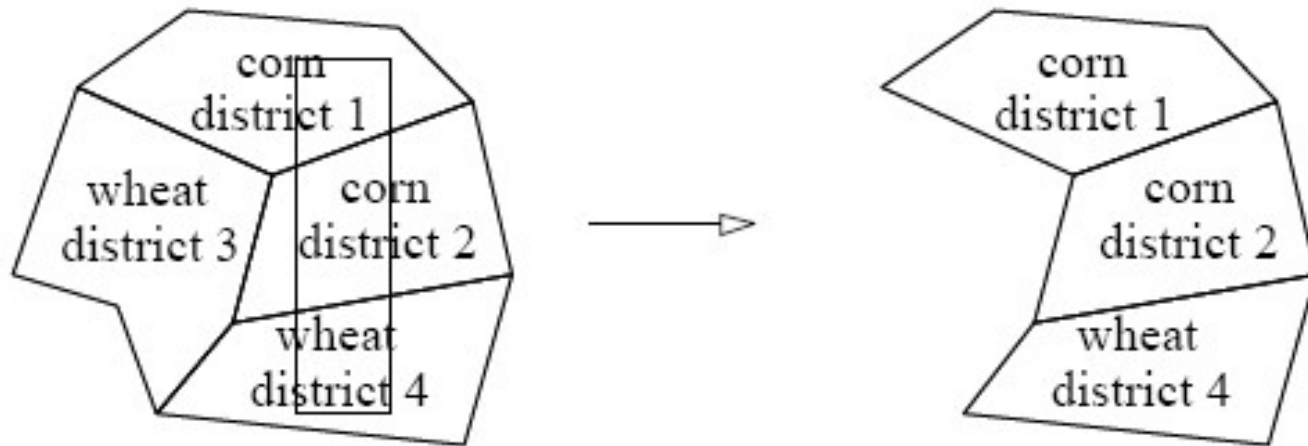
Selection

Operações

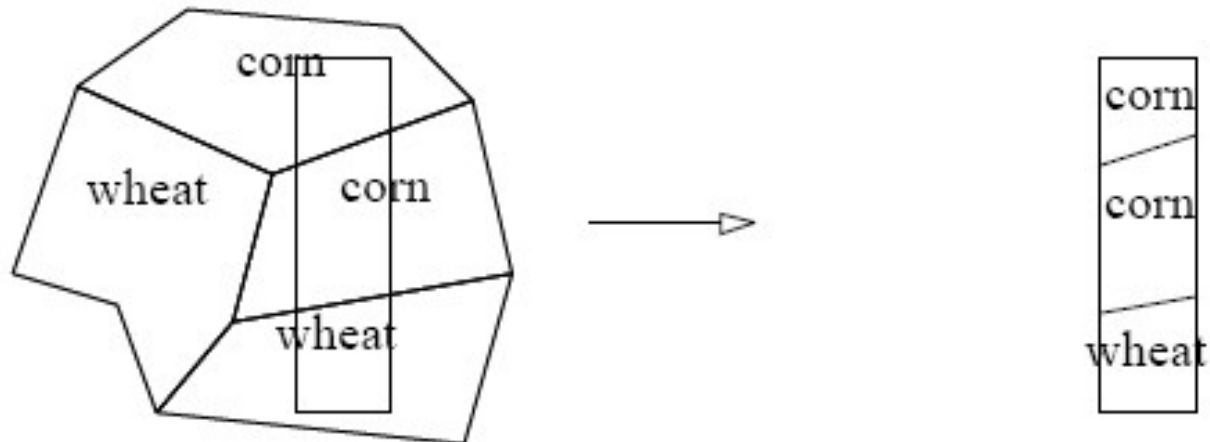


Projection with merging (“*fusion*”)

Operações



“Windowing”



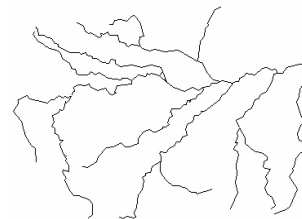
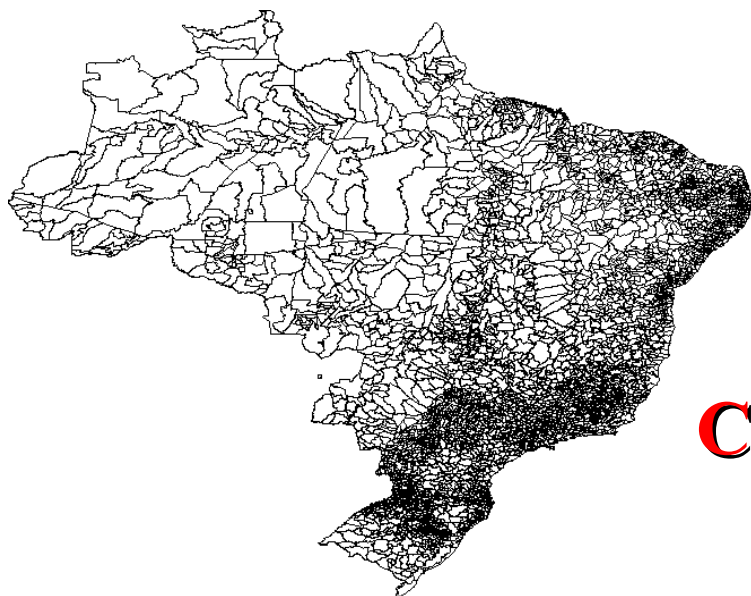
“Clipping”

Operações espaciais

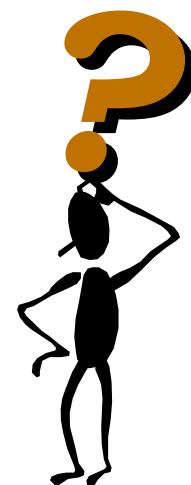


Operações espaciais

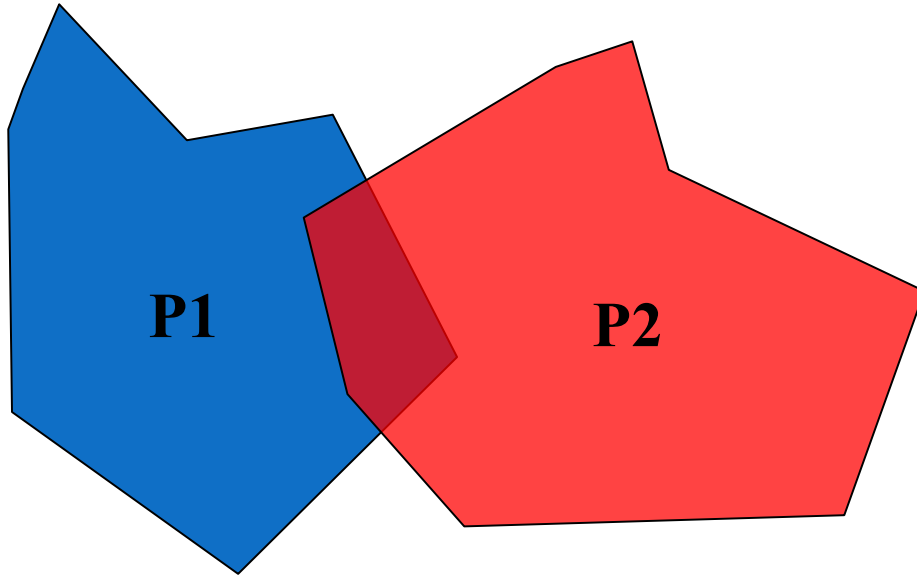
- É simples verificar a interseção entre bacia do amazonas e os municípios do Brasil?



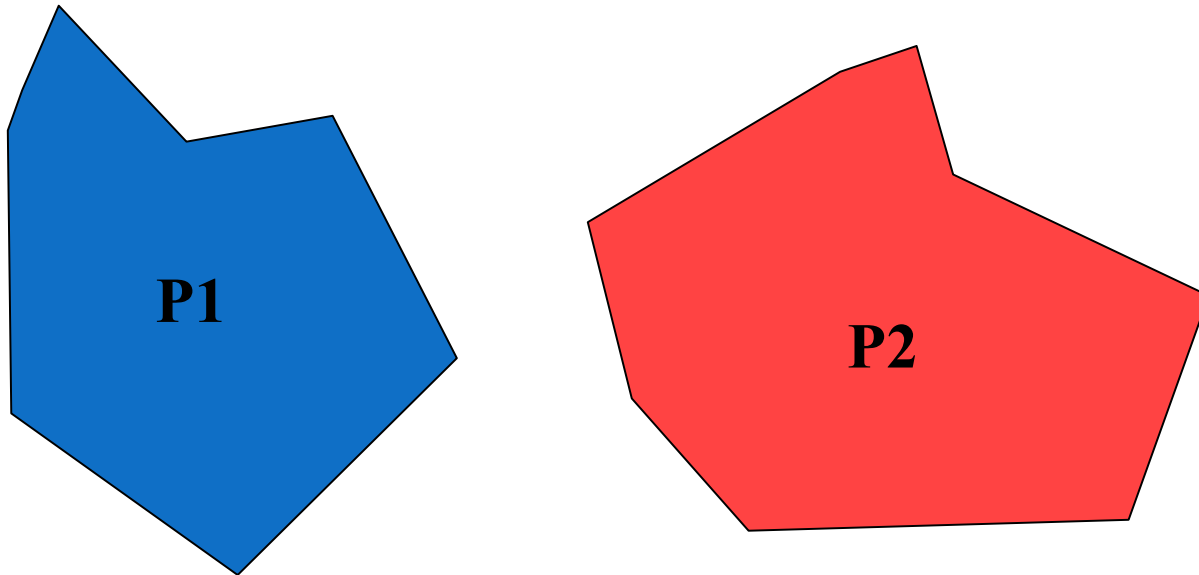
Como você faria o teste



Operações espaciais

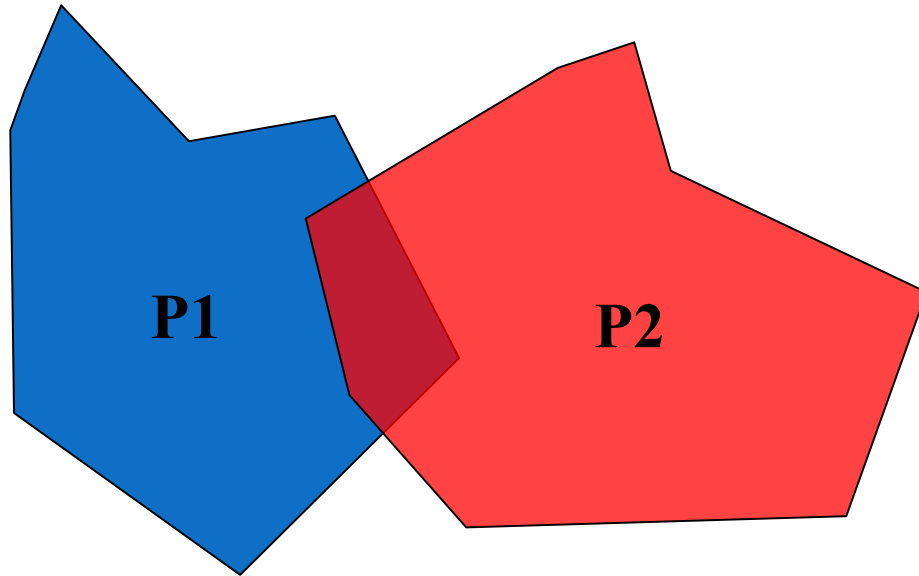


**Polígono 1 tem
interseção com
polígono 2!**

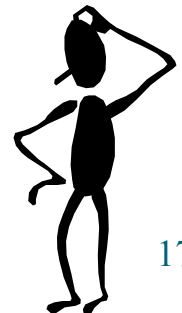
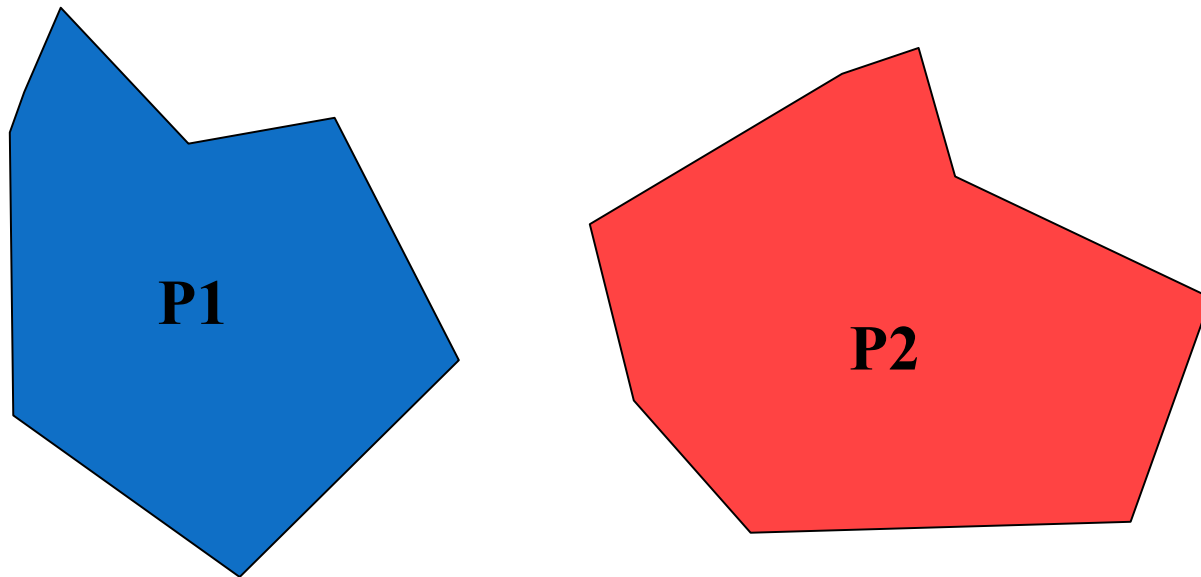


**Polígono 1 não
tem interseção
com polígono 2!**

Operações espaciais

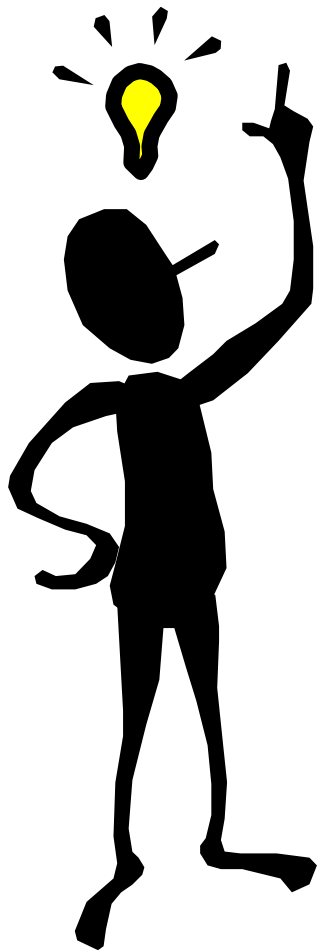


- Como fazer o teste de interseção?
- E se os polígonos tivessem milhões de pontos?
- Como fazer o teste de interseção mais rapidamente?

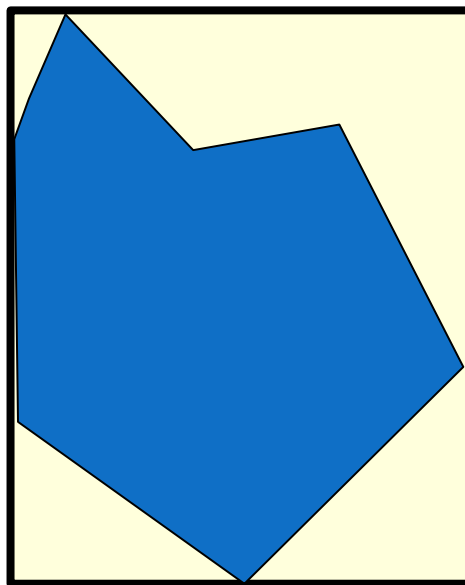


Operações espaciais

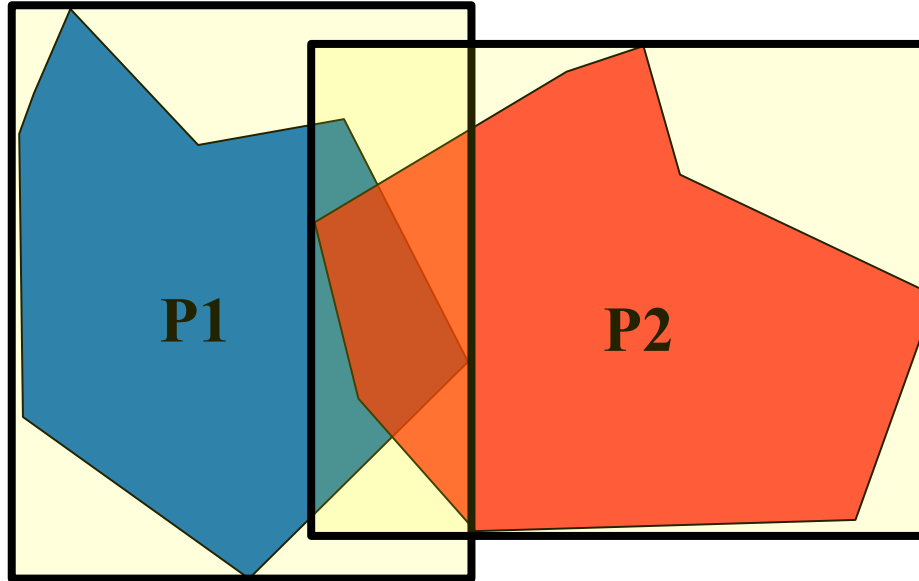
- Reduzindo a complexidade dos objetos
- Fazendo testes mais simples primeiro



Minimum bounding Rectangle (MBR)
ou
Menor Retângulo Envolvente (MBE)

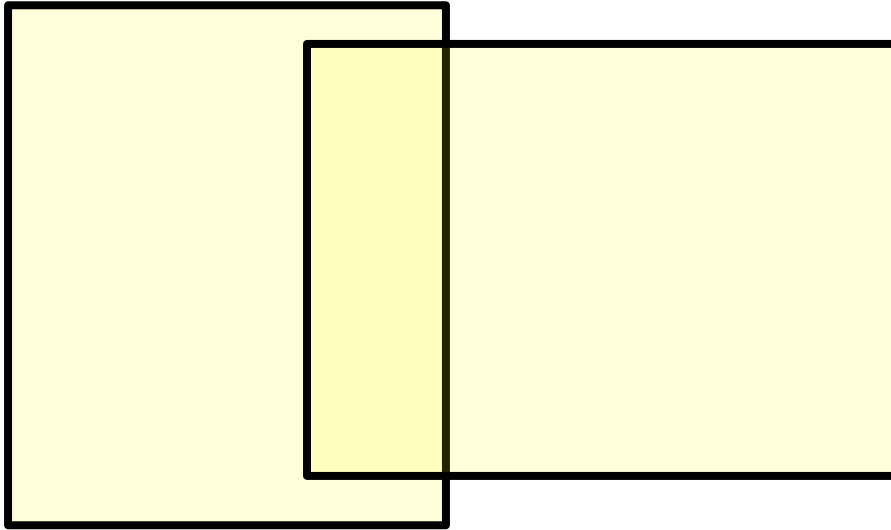


Operações espaciais

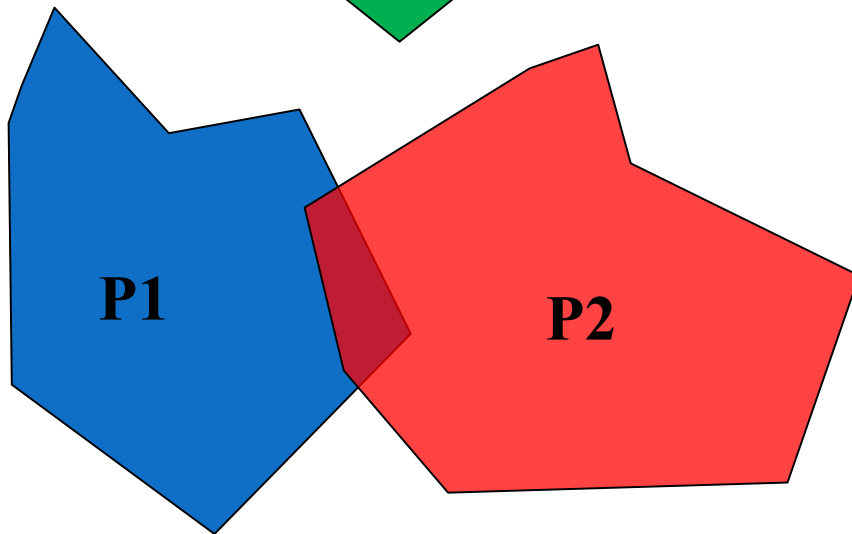


Comparando-se MBR podemos dizer que P1 e P2 têm interseção?

Operações espaciais



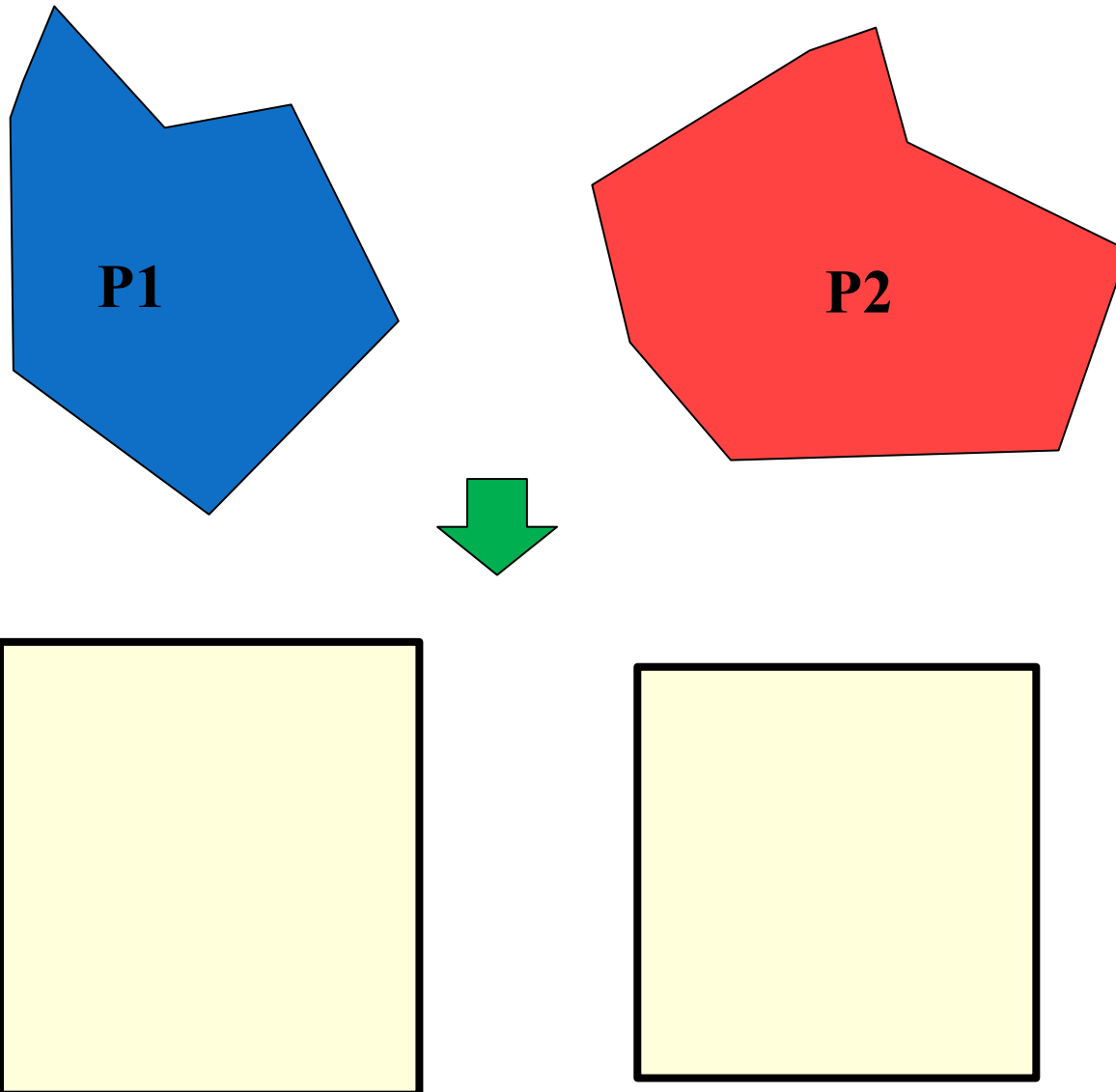
Se MBRs têm
interseção, em geral, não
se pode inferir que
objetos têm interseção.



Após o teste de MBR, é
necessário testar os
objetos reais.

Passo mais
custoso. \$\$\$\$

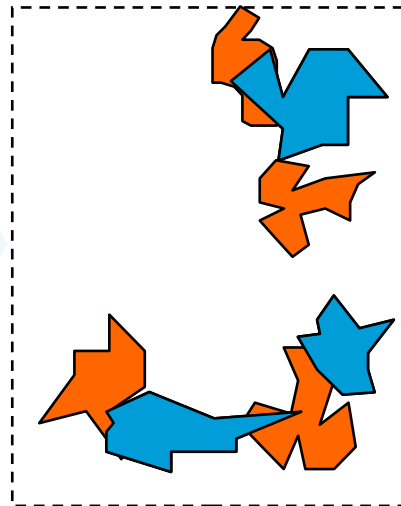
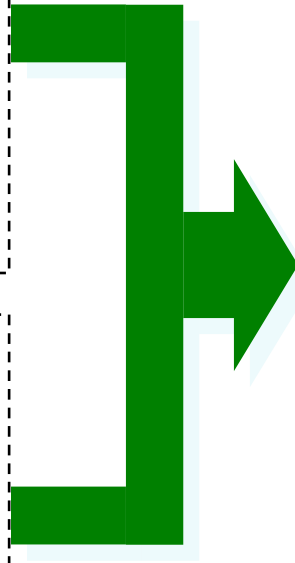
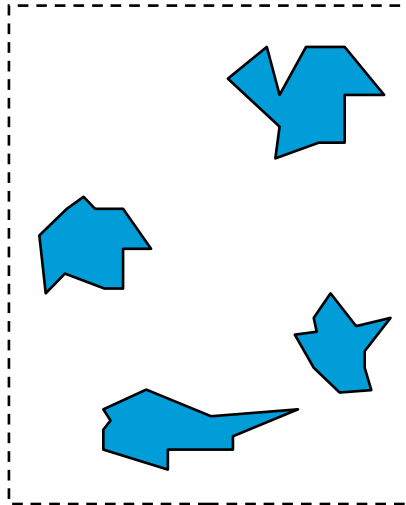
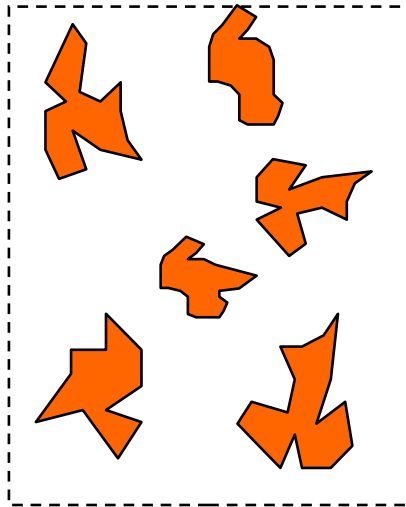
Operações espaciais



De outra forma, se objetos não têm interseção, então seus MBRs não têm intersecção.

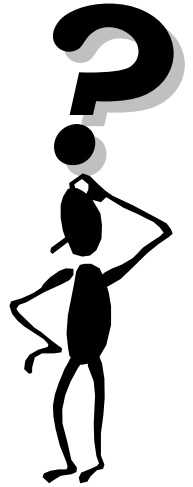
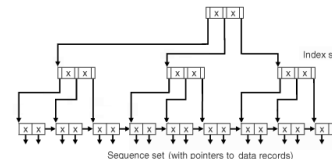
Logo, não é necessário testar os objetos reais.

Operações espaciais



**Como comparar
muitos objetos de
forma eficiente**

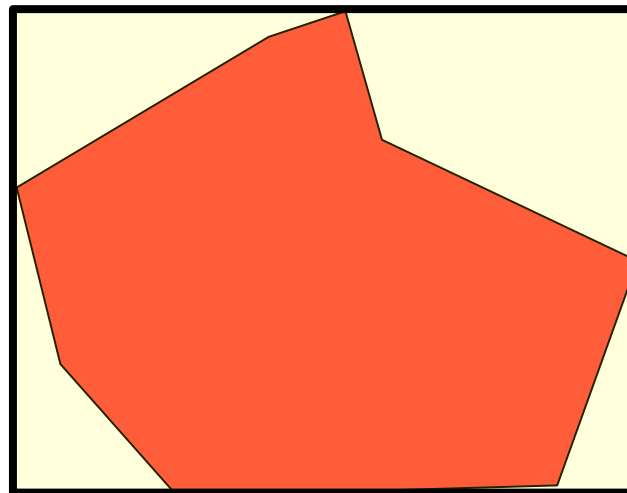
**Uso de
índices**



ÁRVORES R (R-Trees)

Árvore-R

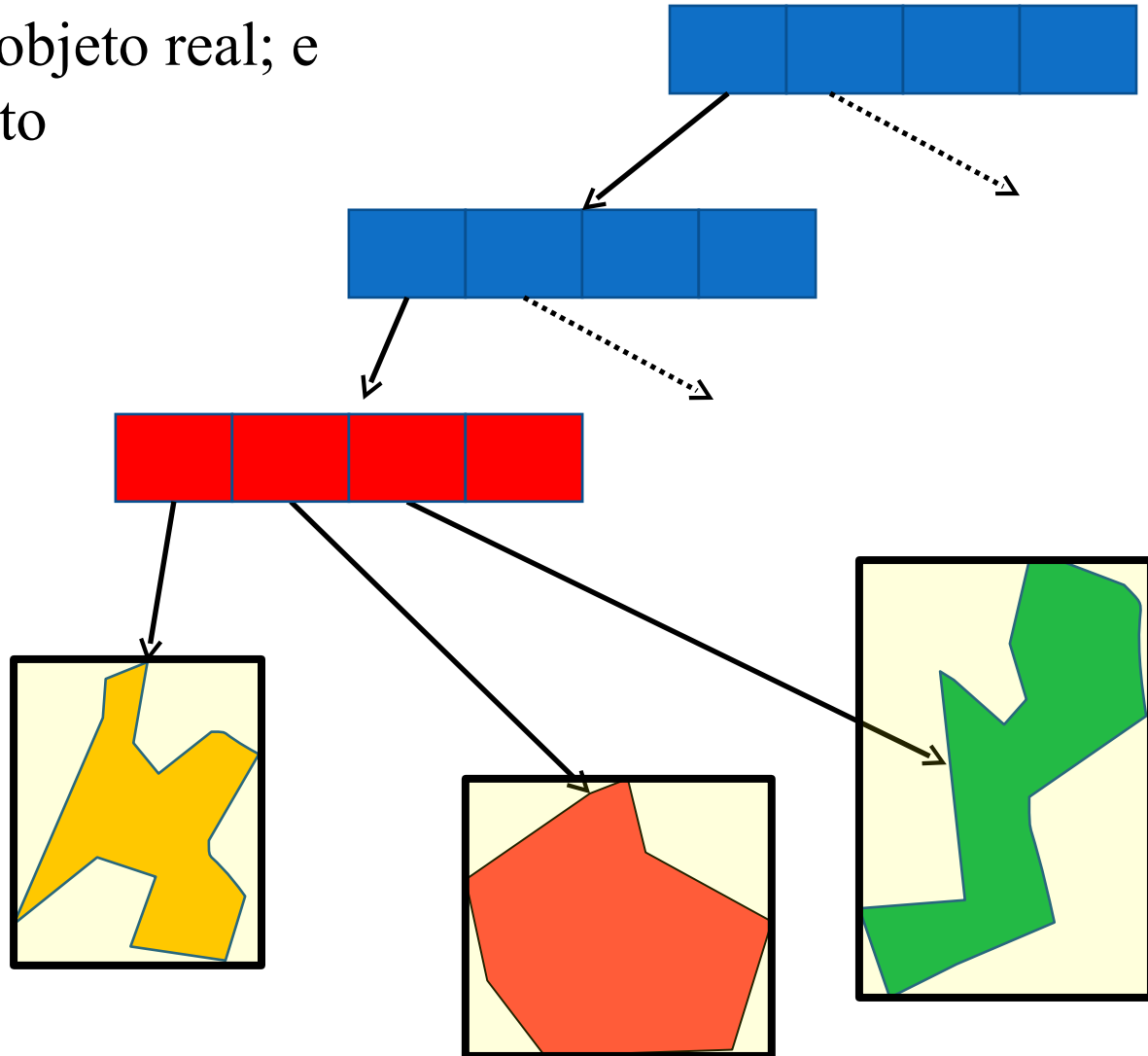
- Proposta por Guttman em 1984
 - A. Guttman, R-trees: a dynamic index structure for spatial searching, *Proceedings of the SIGMOD Conference*, Boston, June, 1984, pp. 47-57.
- Estrutura de dados hierárquica derivada da árvore B (Comer, 1979)
- Entradas dos nós correspondem à retângulos em d -dimensão



Árvore-R

Cada entrada de nó folha:

- Ponteiro para objeto real; e
- MBR do objeto

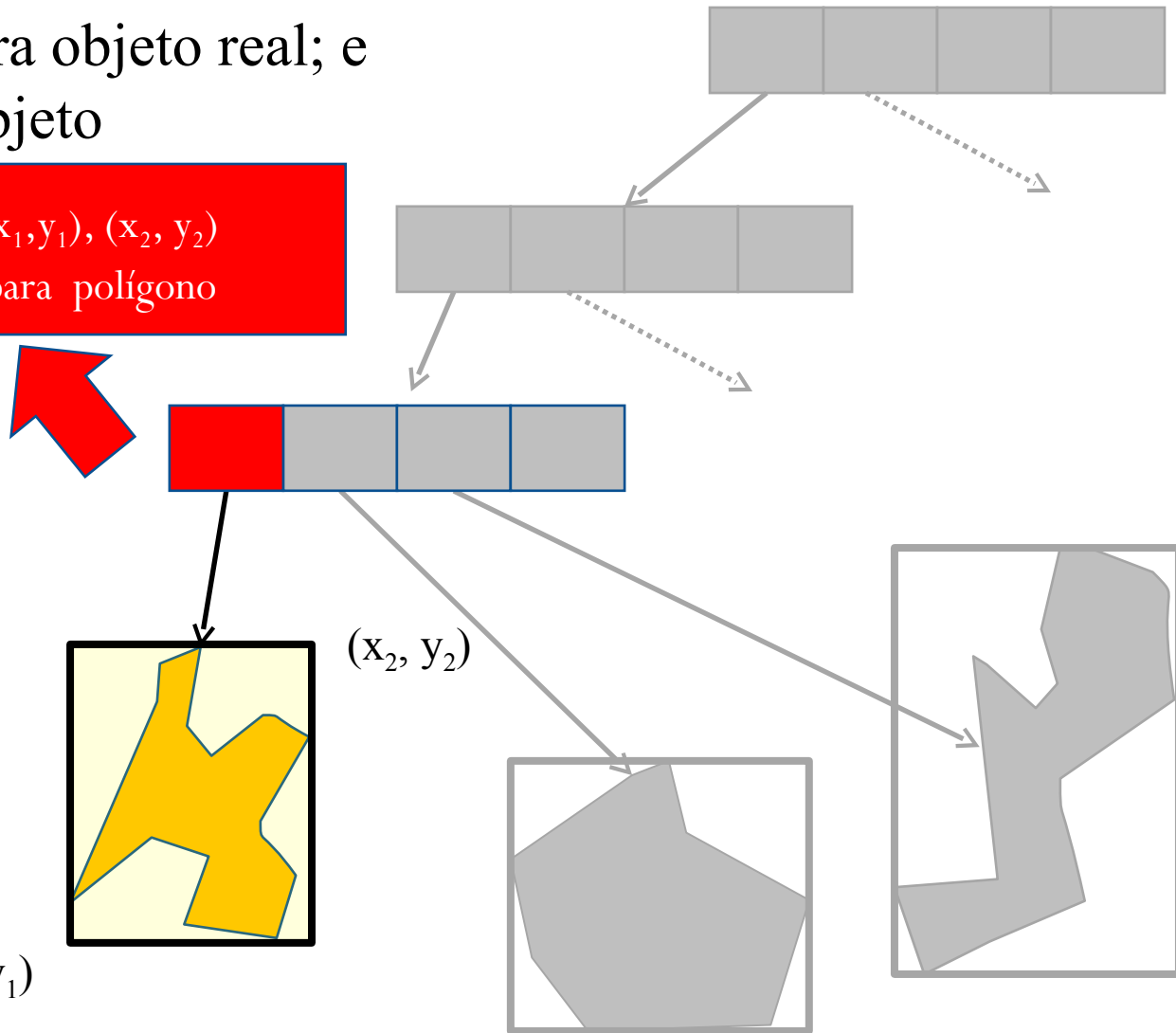


Árvore-R

Cada entrada de nó folha:

- Ponteiro para objeto real; e
- MBR do objeto

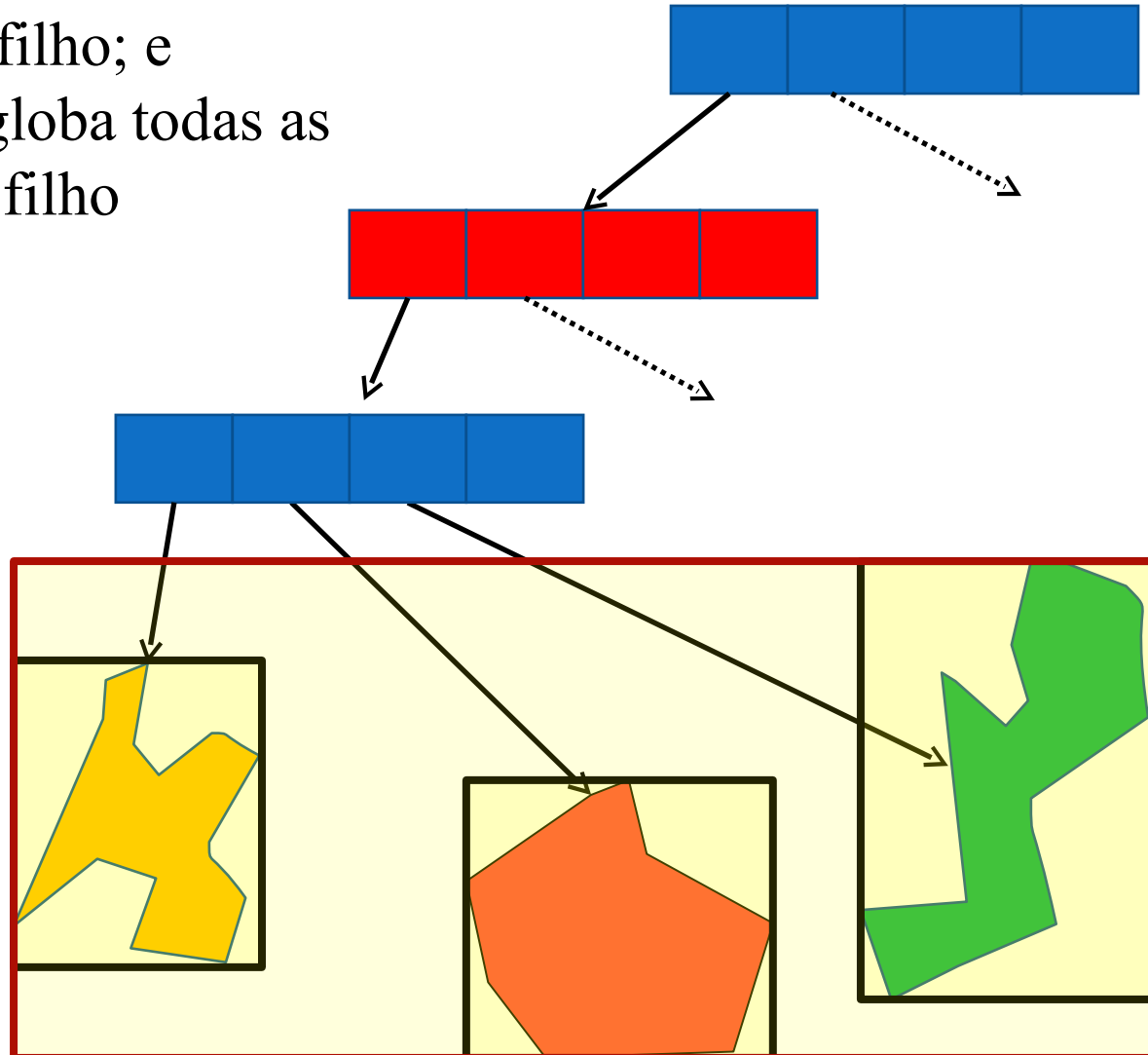
retângulo: $(x_1, y_1), (x_2, y_2)$
p*: aponta para polígono



Árvore-R

Cada entrada de nó interno:

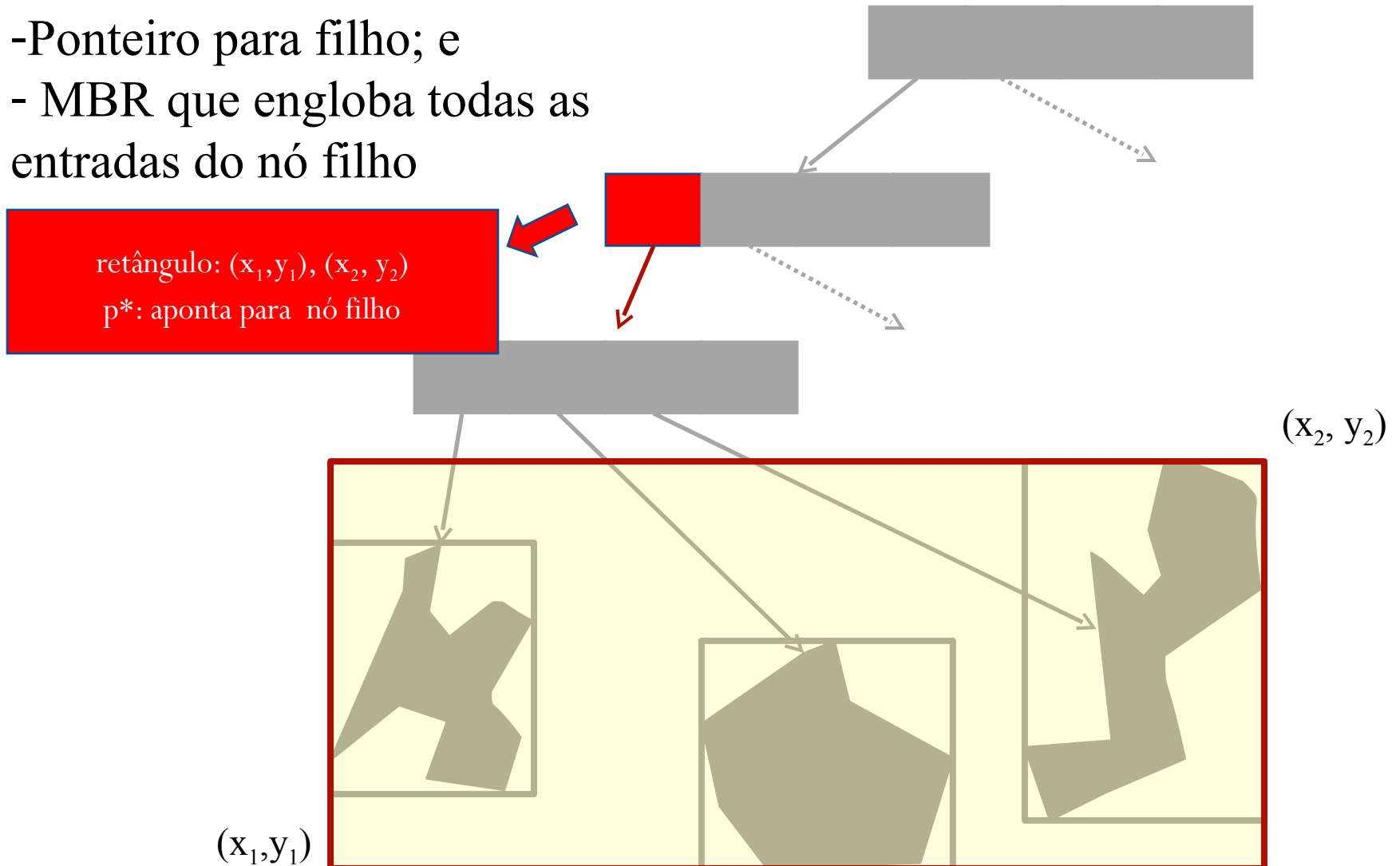
- Ponteiro para filho; e
- MBR que engloba todas as entradas do nó filho



Árvore-R

Cada entrada de nó interno:

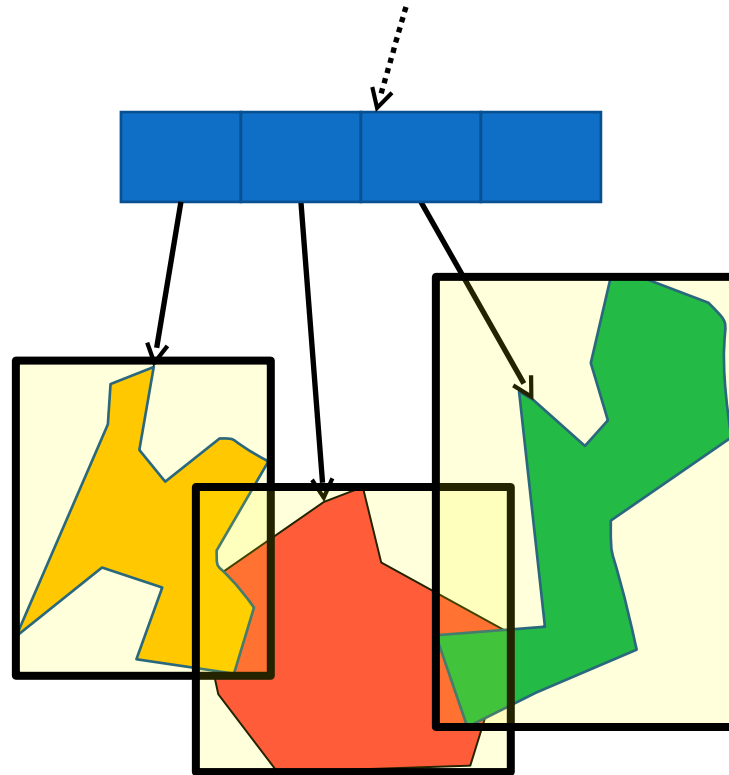
- Ponteiro para filho; e
- MBR que engloba todas as entradas do nó filho



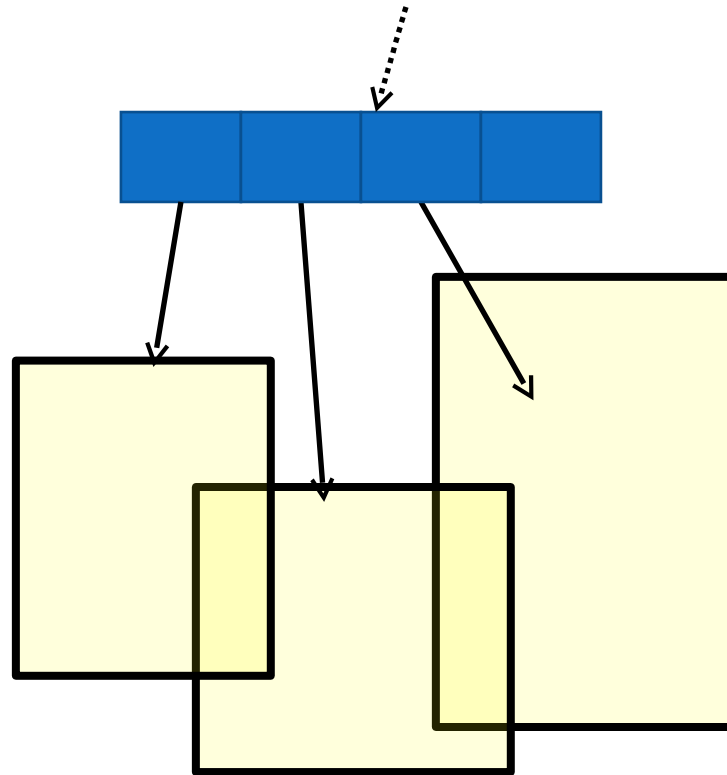
Árvore-R

- Frequentemente, nós correspondem à páginas do disco
- Parâmetros definindo o número mínimo e máximo de nós se aplicam
 - São escolhidos a fim do menor número de nós sejam visitados durante a busca.
- Importante:
 - Diferentemente da árvore B, entradas de nós diferentes podem ter sobreposição
 - Ou seja, retângulos das entradas podem ter sobreposição

Exemplo de sobreposição



Exemplo de sobreposição



A consulta espacial pode requerer que vários nós sejam visitados antes de garantir a presença ou ausência de um retângulo em particular.

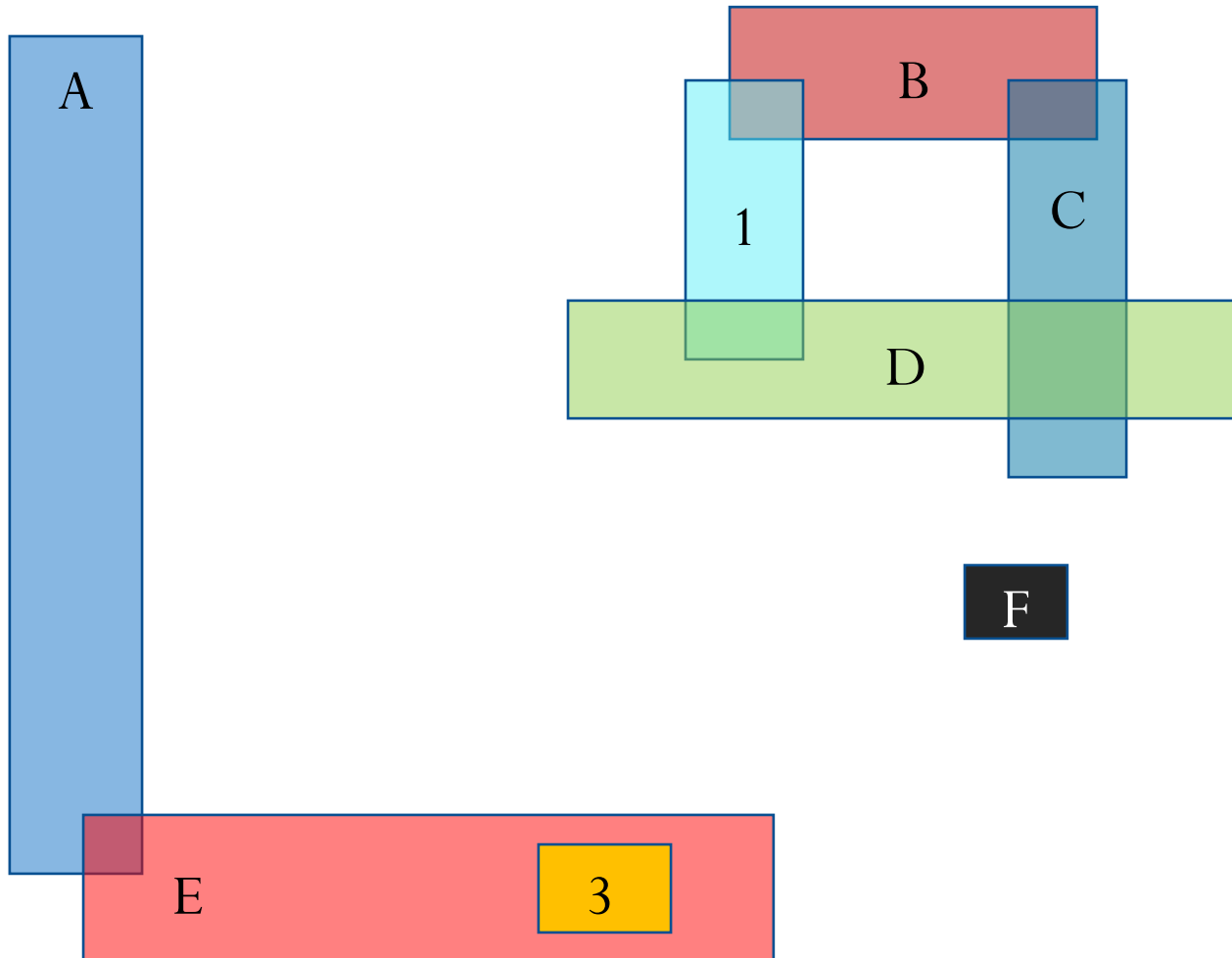
Árvore-R

- Regras semelhantes às da árvore-B
 - Todos os nós folha aparecem no mesmo nível
 - Cada entrada de um nó folha corresponde à tupla (R, O)
 - O é um ponteiro para o objeto real
 - R é o menor retângulo que espacialmente contém o objeto apontado por O
 - Cada entrada de um nó interno corresponde à tupla (R, P)
 - P é um ponteiro para um filho
 - R é o menor retângulo que espacialmente contém os retângulos no filho apontado por P .
- Uma árvore-R de ordem (m, M) é tal que
 - Cada nó contém entre $m \leq \text{Ceiling}[M/2]$ e M entradas, com exceção da raiz
 - A raiz tem pelo menos 2 entradas, ao menos que ela seja nó folha

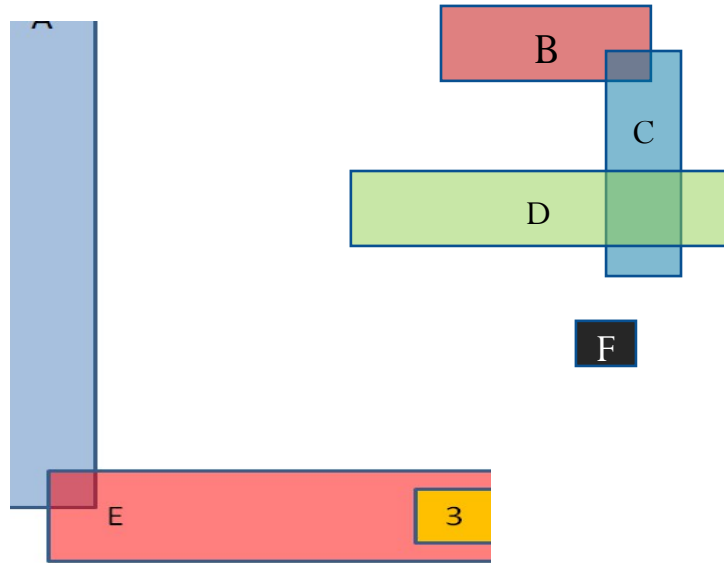
Árvore-R - Inserção

- Uma árvore-R não é única
 - Ela depende da ordem em que os retângulos são inseridos (e possivelmente removidos)
- Algoritmo para inserção de um nó é análogo ao algoritmo de inserção em árvore B
 - Novos retângulos são adicionados a nós folha
 - O nó folha apropriado é determinado pela
 - Navegação na árvore iniciando no nó raiz
 - A cada passo escolhe-se a subárvore cujo retângulo correspondente terá o **menor acréscimo de área possível**
 - Se ao inserir o nó ocorrer *overflow*, então executar *split*
 - $M+1$ registros devem ser distribuídos entre dois nós
 - *Split* pode propagar até a raiz

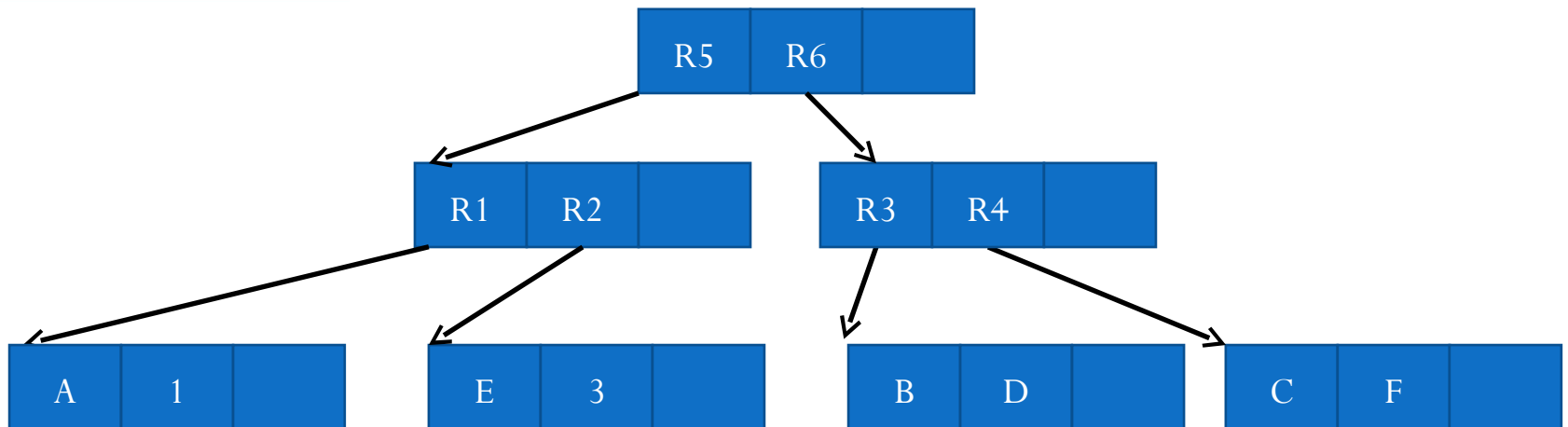
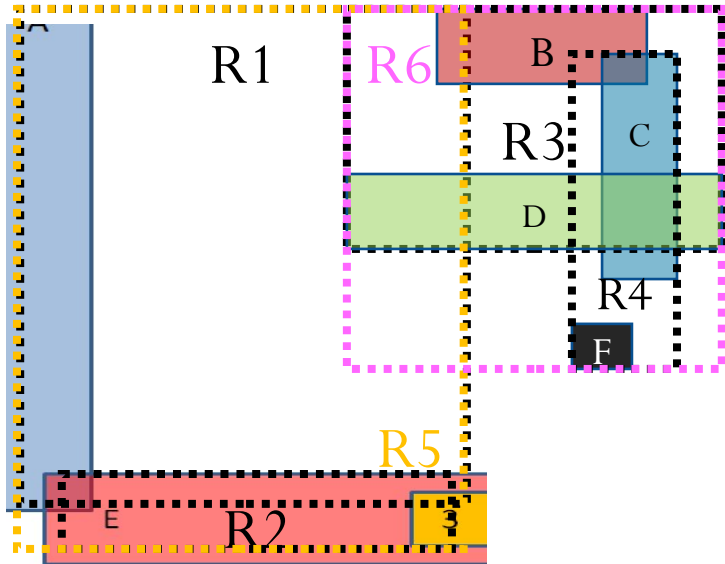
Exemplo ($M=3$ e $m=2$)



Exemplo ($M=3$ e $m=2$)



Exemplo (M=3 e m=2)



Insérer A

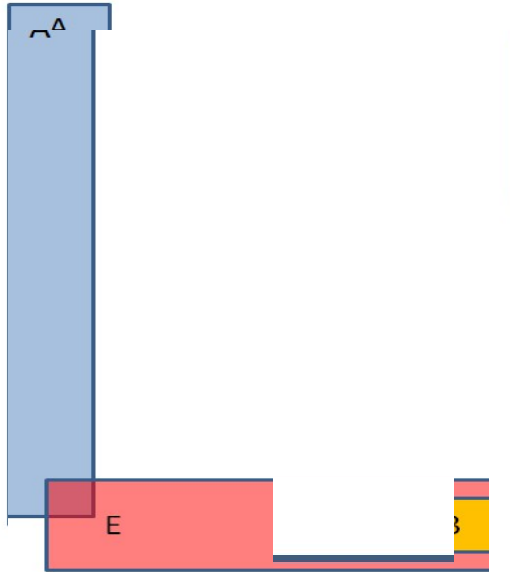


Insérer 1



| | | |
|---|---|--|
| A | 1 | |
|---|---|--|

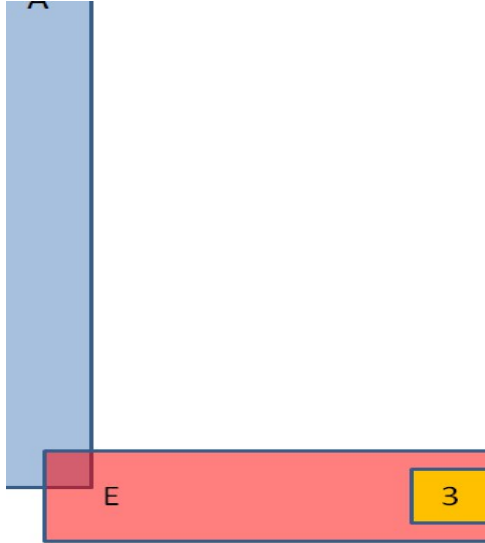
Insérer E



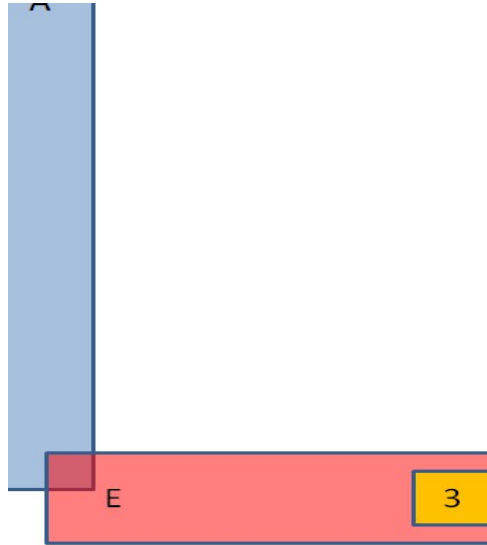
| | | |
|---|---|---|
| A | 1 | E |
|---|---|---|

Insérer 3

Overflow!



Inserir 3



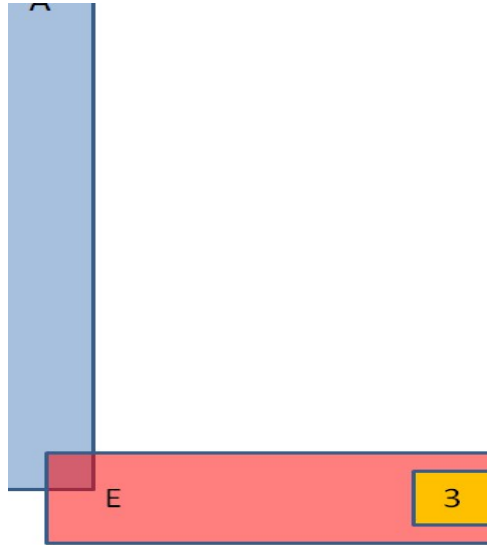
Overflow!

Dividir as entradas em dois nós.

Considerar na divisão as entradas cujo MBR envolvente das mesmas tenha a menor área.



Inserir 3



Overflow!

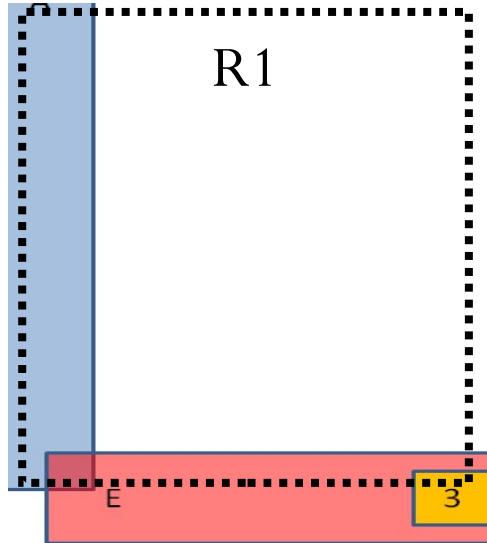
Dividir as entradas em dois nós.

Considerar na divisão as entradas cujo MBR envolvente das mesmas tenha a menor área.

Criar novo nó raiz com 2 entradas, contendo MBR para os nós folha.



Inserir 3

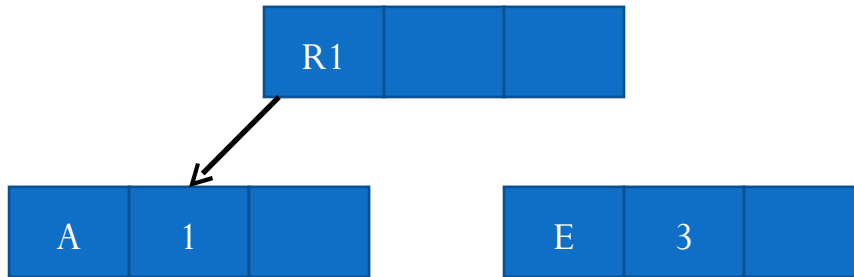


Overflow!

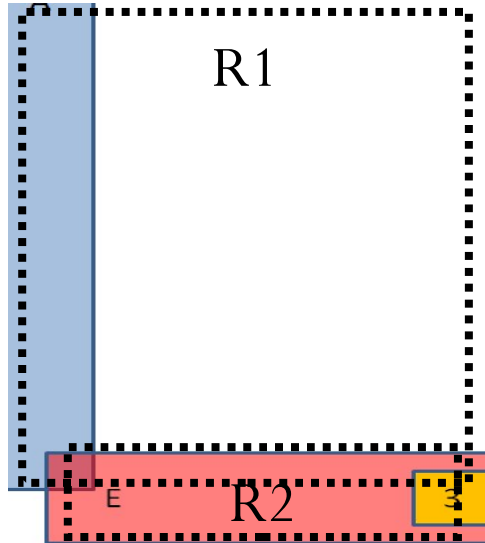
Dividir as entradas em dois nós.

Considerar na divisão as entradas cujo MBR envolvente das mesmas tenha a menor área.

Criar novo nó raiz com 2 entradas, contendo MBR para os nós folha.



Inserir 3

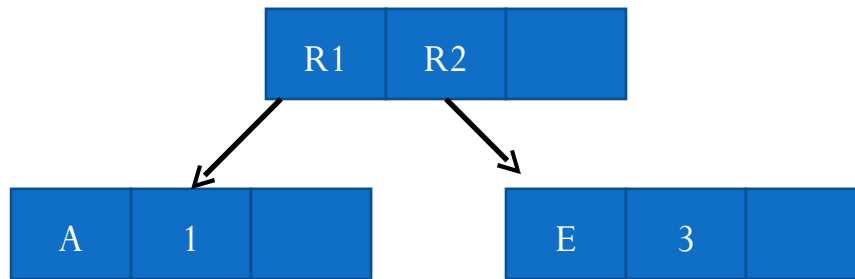


Overflow!

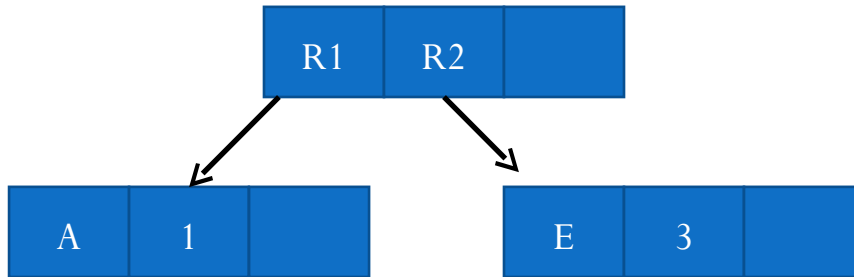
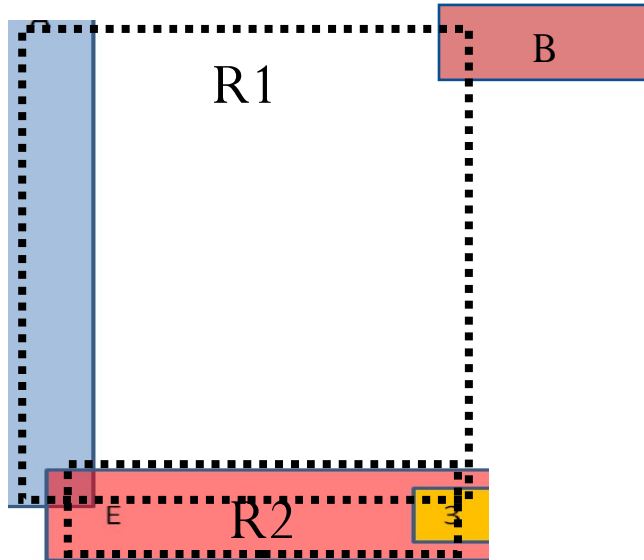
Dividir as entradas em dois nós.

Considerar na divisão as entradas cujo MBR envolvente das mesmas tenha a menor área.

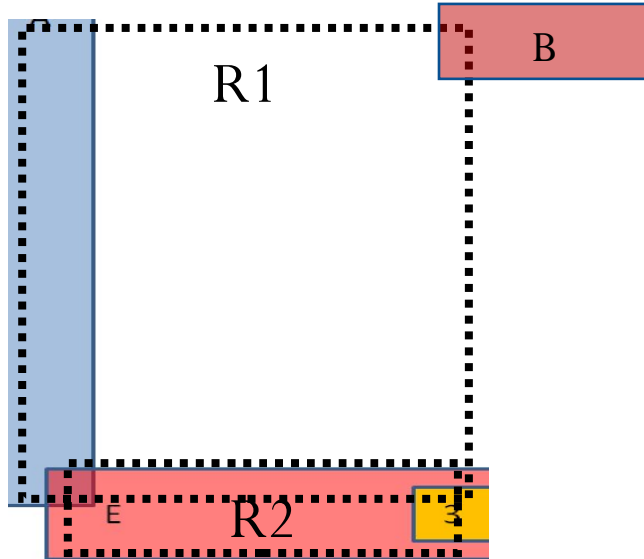
Criar novo nó raiz com 2 entradas, contendo MBR para os nós folha.



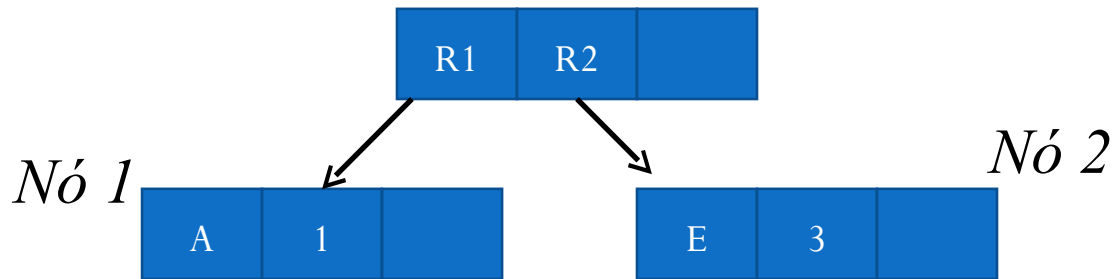
Insérer B



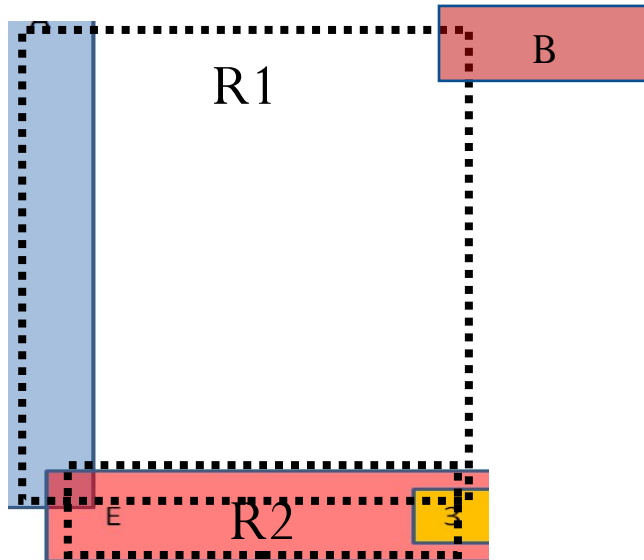
Inserir B



Em qual nó folha B deve ser inserido?
Nó 1 ou Nó 2?

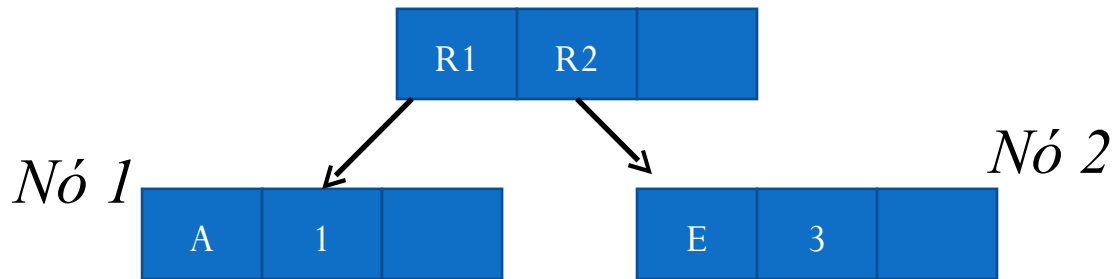


Inserir B

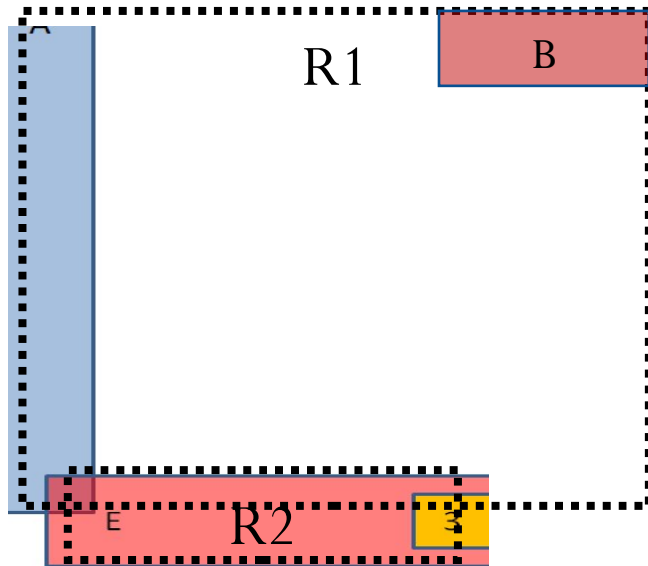


Em qual nó folha B deve ser inserido?
Nó 1 ou Nó 2?

Escolher o nó que cujo MBR das
entradas tem o menor acréscimo de
área → Nó 1

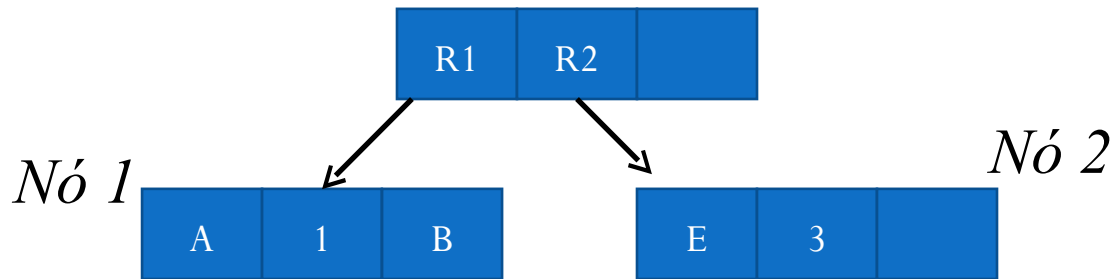


Inserir B

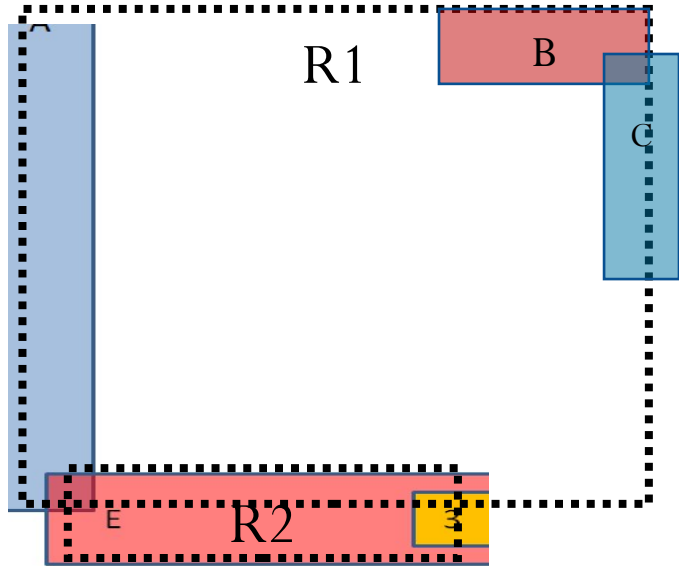


Em qual nó folha B deve ser inserido?
Nó 1 ou Nó 2?

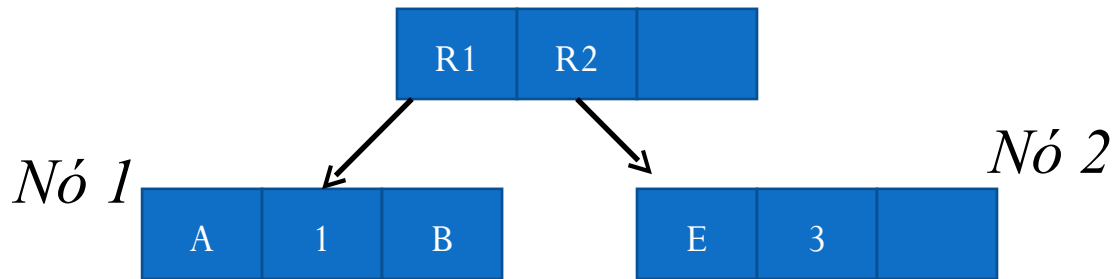
Escolher o nó que cujo MBR das
entradas tem o menor acréscimo de
área → Nó 1



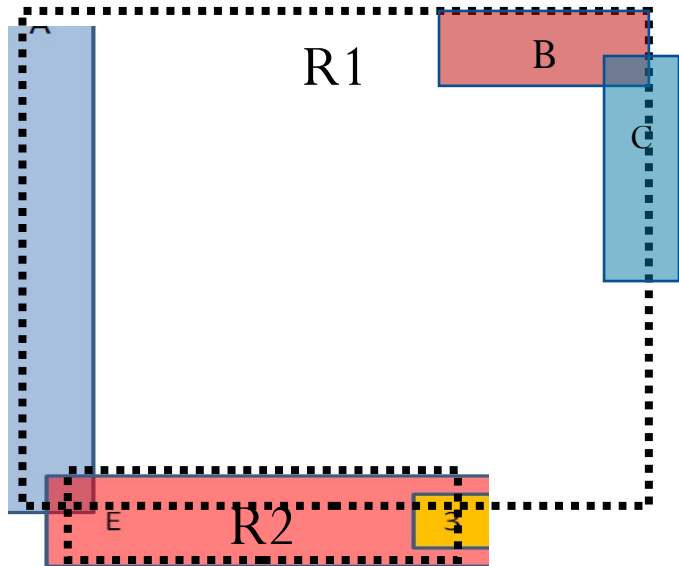
Inserir C



Em qual nó folha C deve ser inserido?
Nó 1 ou Nó 2?

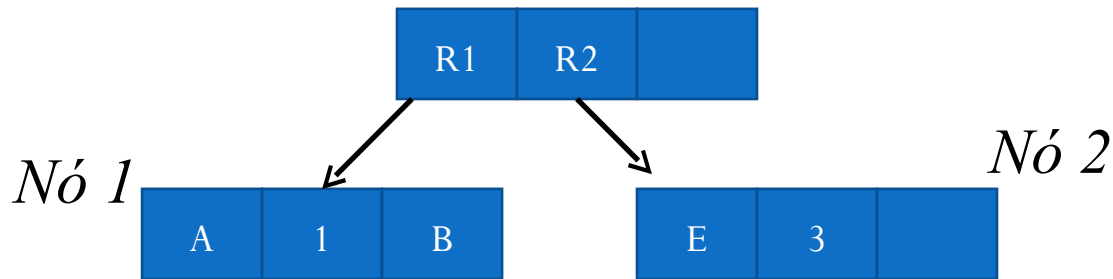


Inserir C

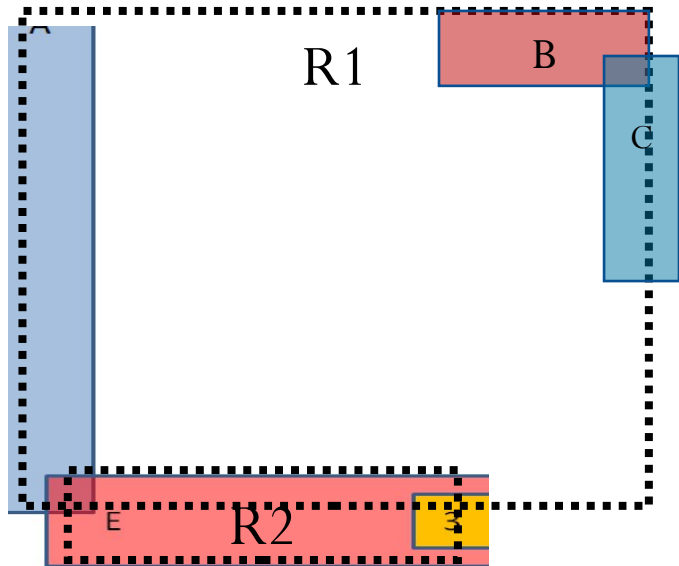


Em qual nó folha C deve ser inserido?
Nó 1 ou Nó 2?

O aumento de área de R1 ao inserir C
no Nó 1 é menor do que o aumento de
área de R2 ao inserir C no Nó 2.



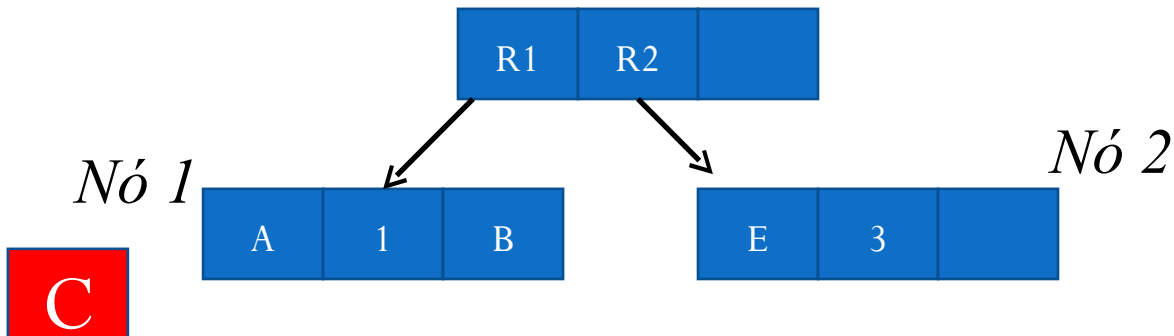
Inserir C



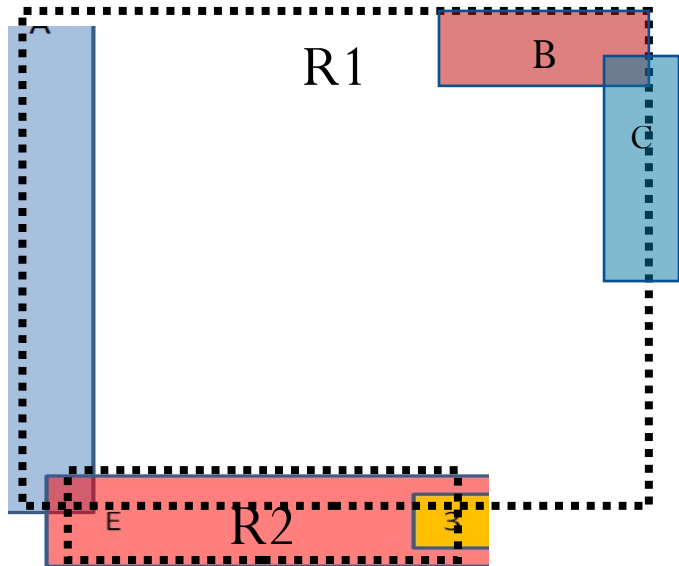
Em qual nó folha C deve ser inserido?
Nó 1 ou Nó 2?

O aumento de área de R1 ao inserir C
no Nó 1 é menor do que o aumento de
área de R2 ao inserir C no Nó 2.

Overflow!



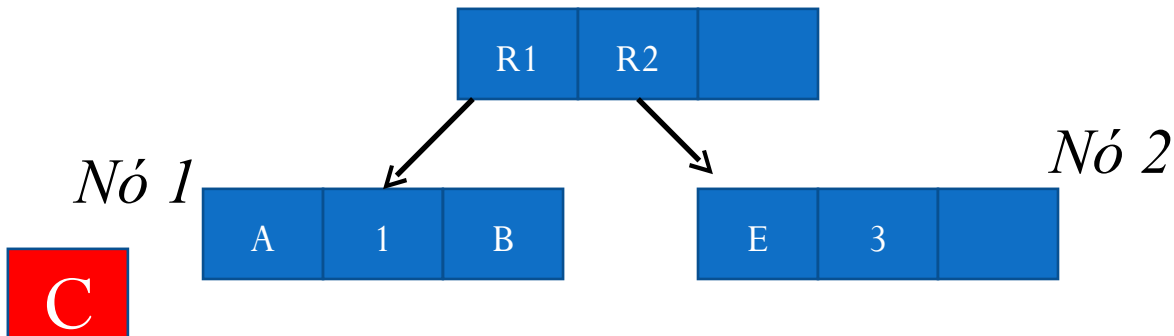
Inserir C



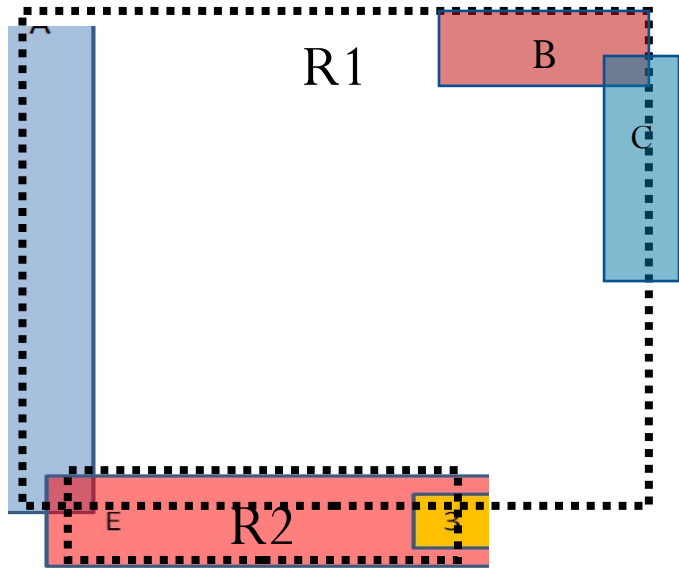
Em qual nó folha C deve ser inserido?
Nó 1 ou Nó 2?

O aumento de área de R1 ao inserir C
no Nó 1 é menor do que o aumento de
área de R2 ao inserir C no Nó 2.

**Overflow! → qual é a melhor forma
de distribuir entradas A1BC?**



Inserir C

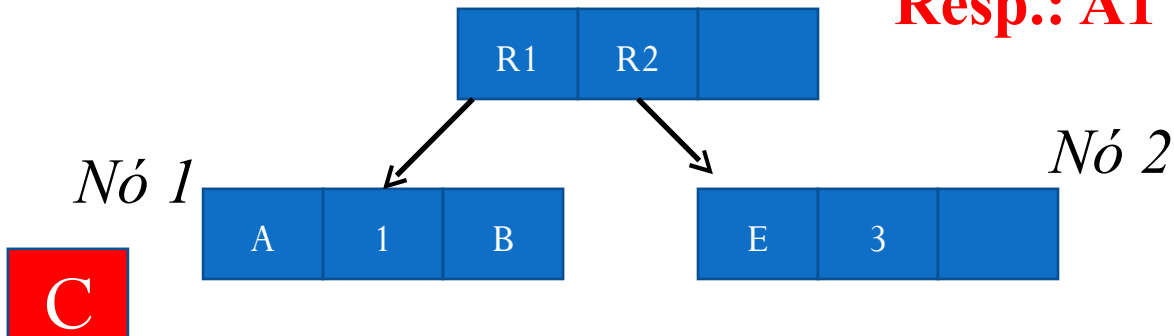


Em qual nó folha C deve ser inserido?
Nó 1 ou Nó 2?

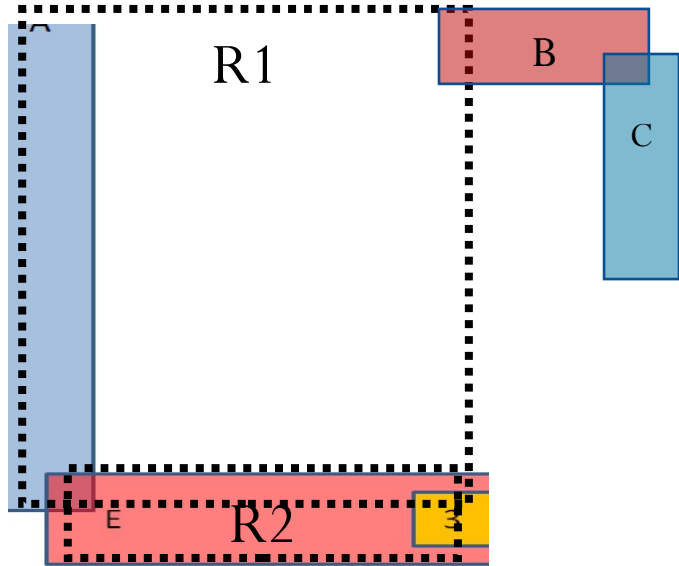
O aumento de área de R1 ao inserir C
no Nó 1 é menor do que o aumento de
área de R2 ao inserir C no Nó 2.

**Overflow! → qual é a melhor forma
de distribuir entradas A1BC?**

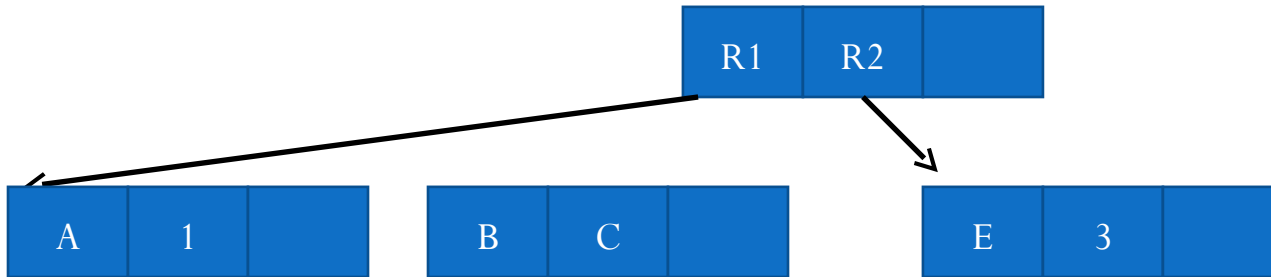
Resp.: A1 e BC



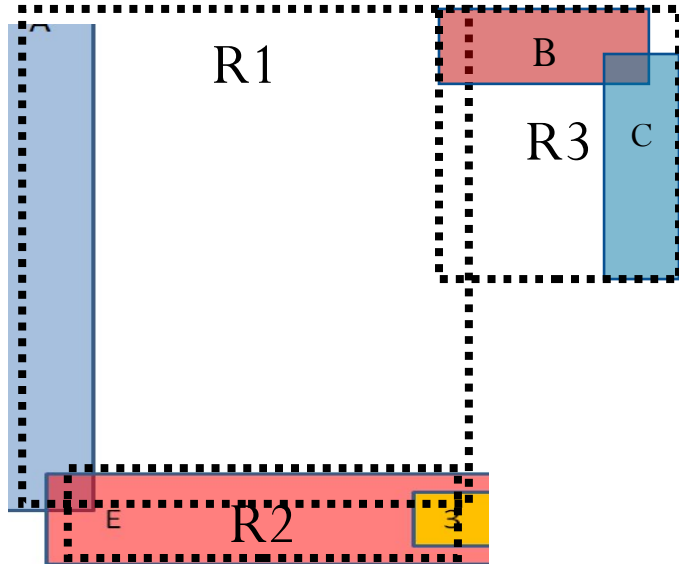
Inserir C



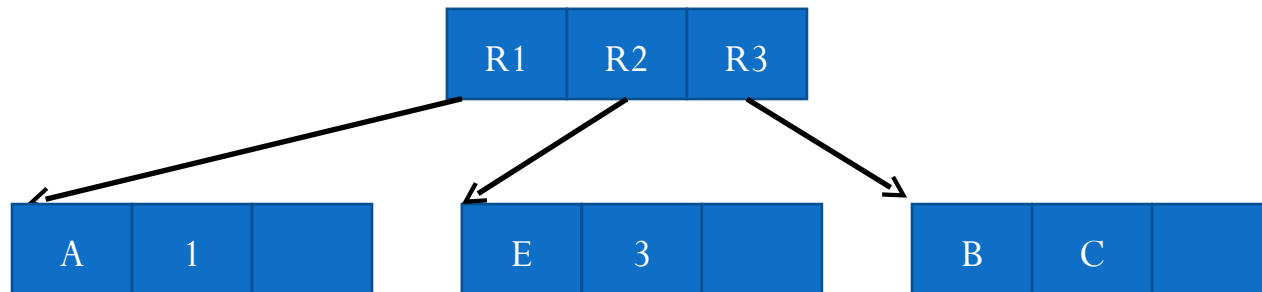
Nó com entradas A1 → R1 ser atualizado



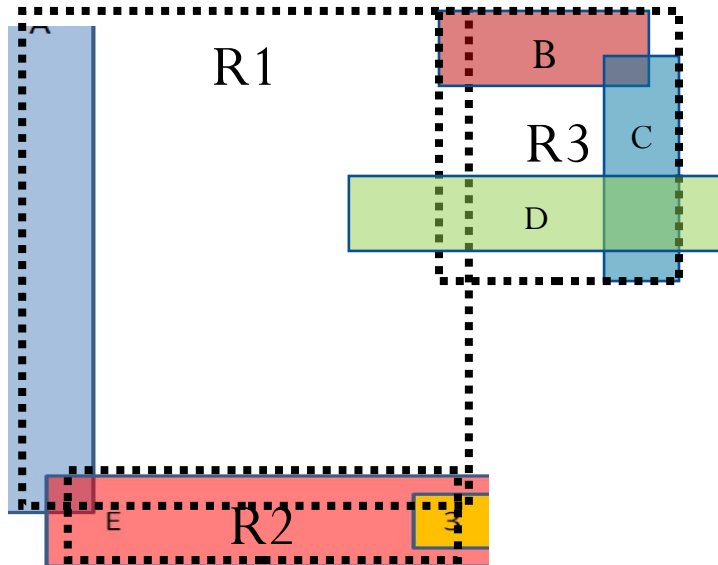
Inserir C



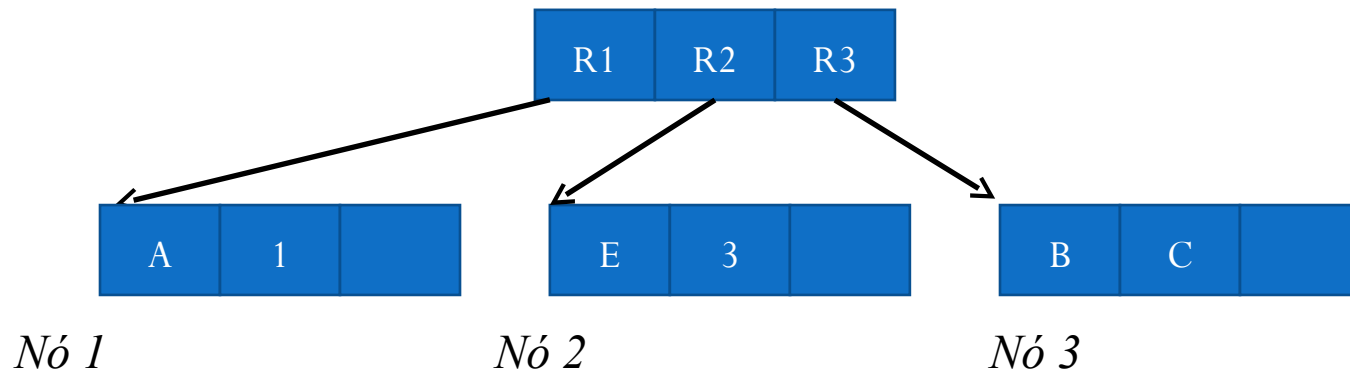
Nó com entradas A1 → R1 ser atualizado
Nó com entradas BC → incluir nova
entrada na folha (R3)



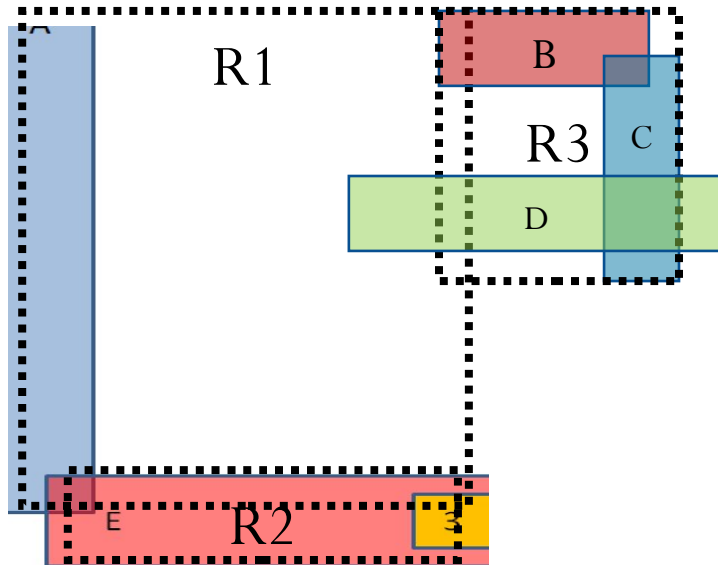
Inserir D



Em qual nó folha D deve ser inserido?
Nó 1 ou Nó 2 ou Nó 3?

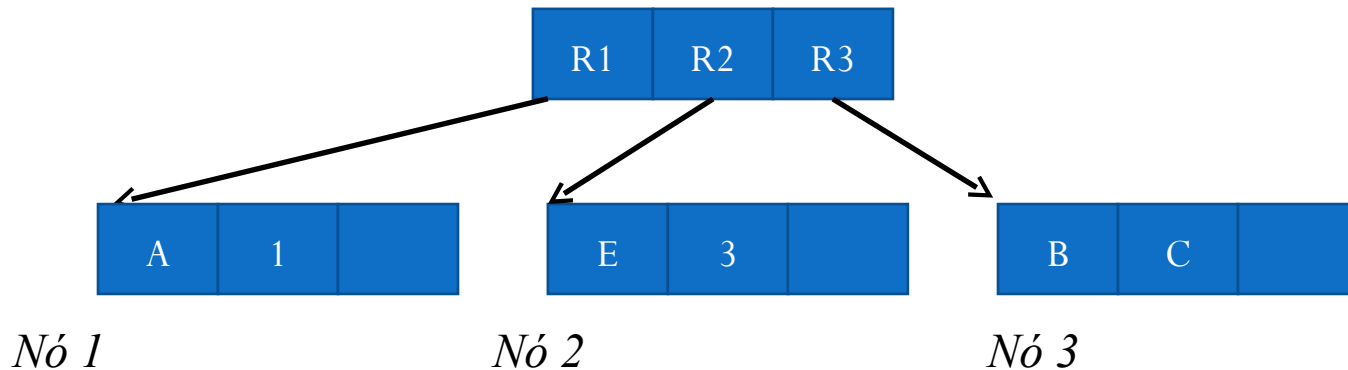


Inserir D

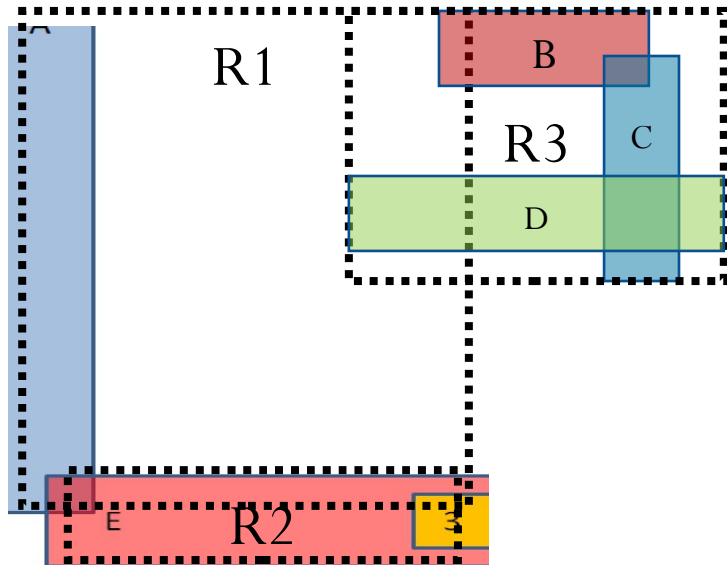


Em qual nó folha D deve ser inserido?
Nó 1 ou Nó 2 ou Nó 3?

R3 terá o menor acréscimo de área

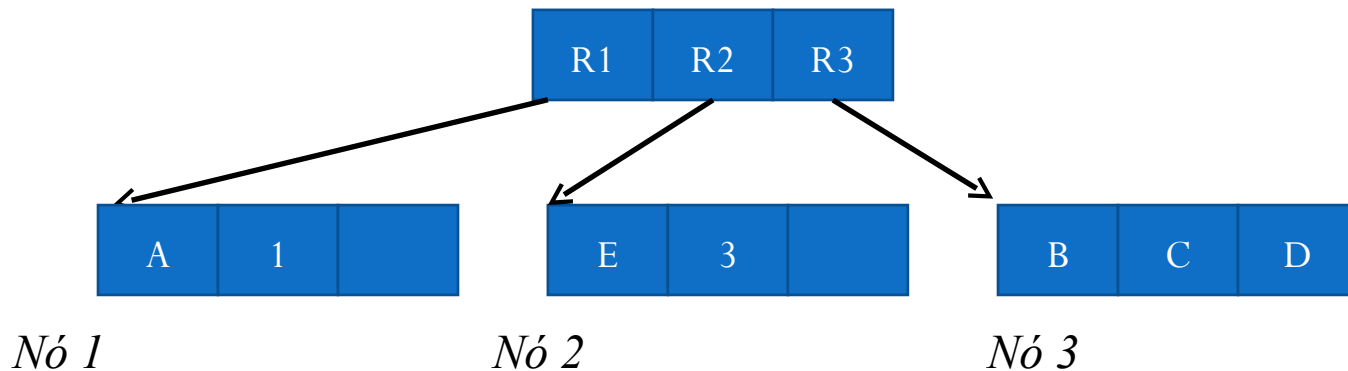


Inserir D

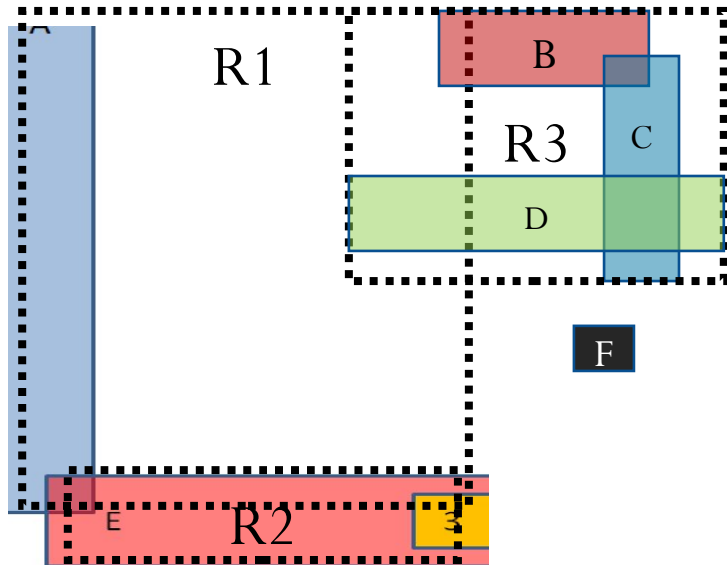


Em qual nó folha D deve ser inserido?
Nó 1 ou Nó 2 ou Nó 3?

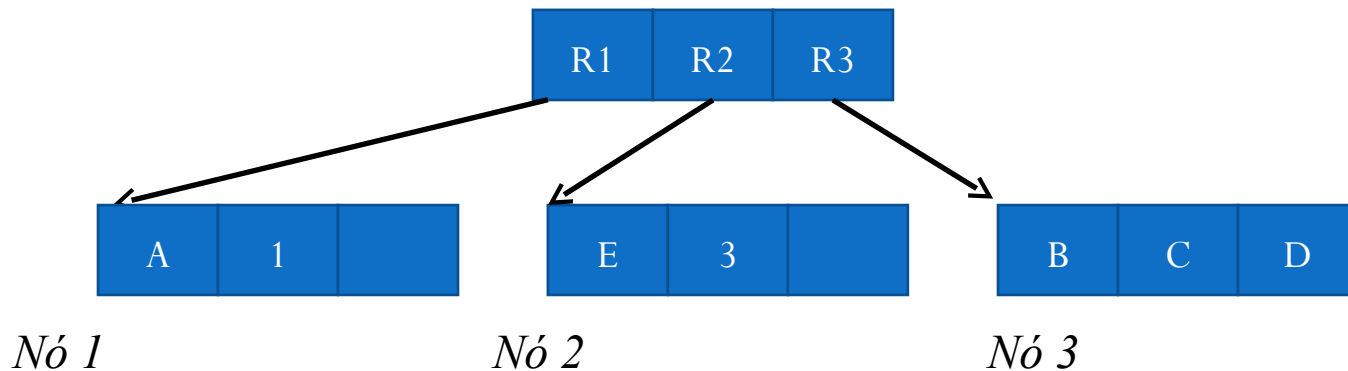
R3 terá o menor acréscimo de área



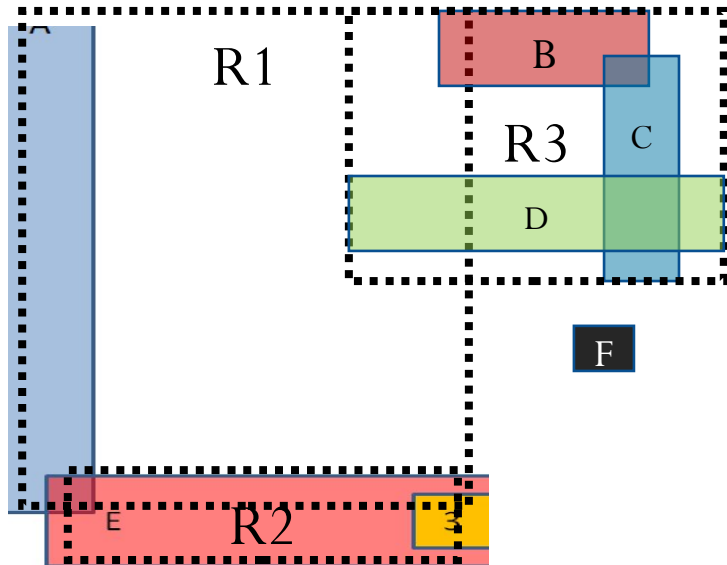
Inserir F



Em qual nó folha F deve ser inserido?
Nó 1 ou Nó 2 ou Nó 3?

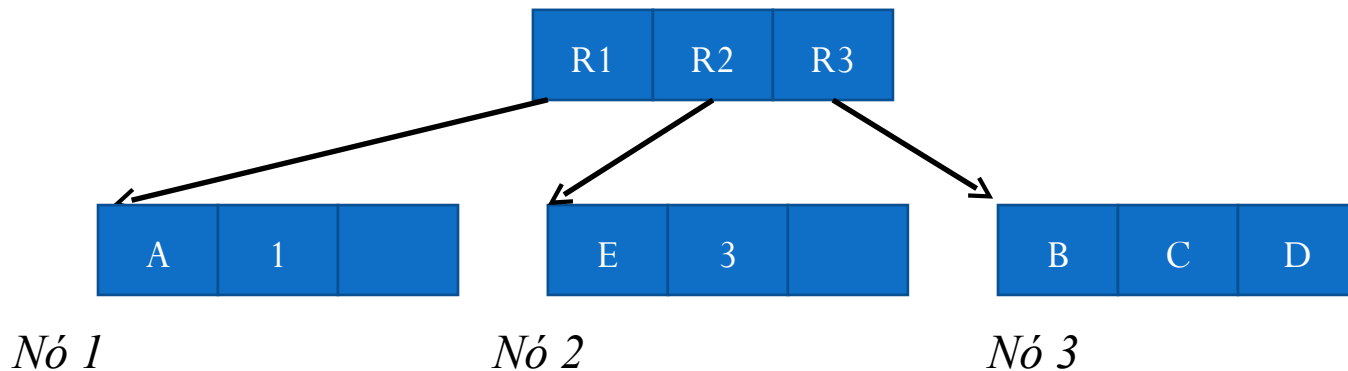


Inserir F

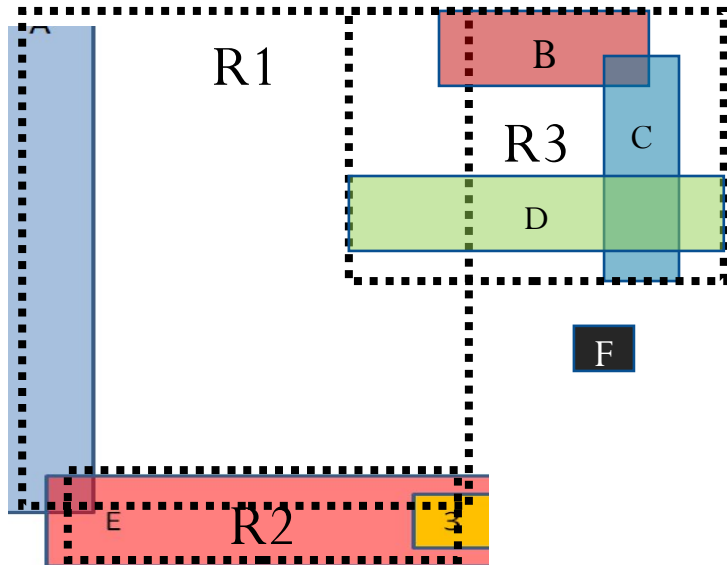


Em qual nó folha F deve ser inserido?
Nó 1 ou Nó 2 ou Nó 3?

R3 terá o menor acréscimo de área.



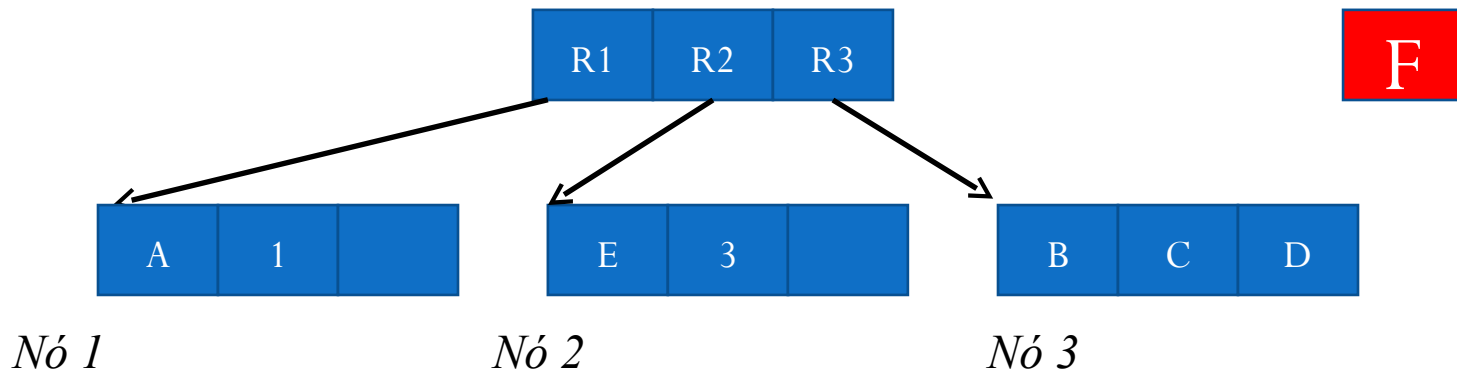
Inserir F



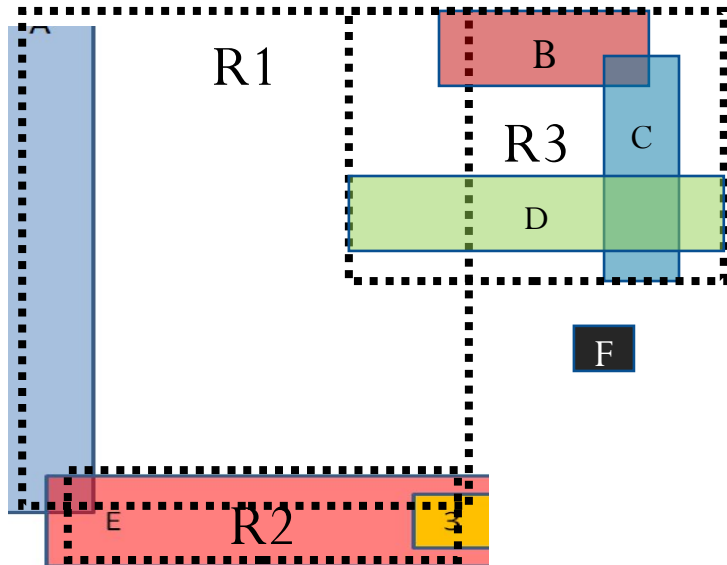
Em qual nó folha F deve ser inserido?
Nó 1 ou Nó 2 ou Nó 3?

R3 terá o menor acréscimo de área.

Overflow!



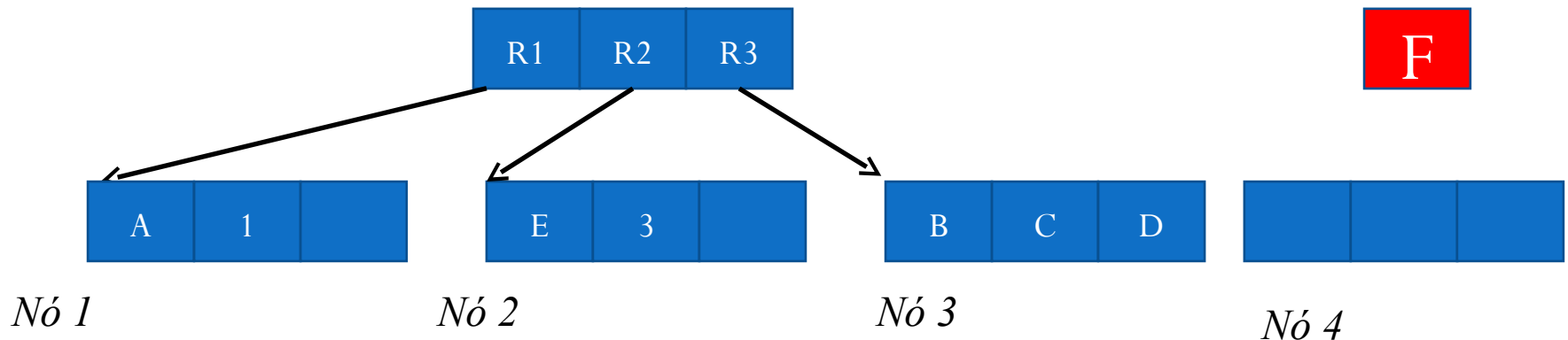
Inserir F



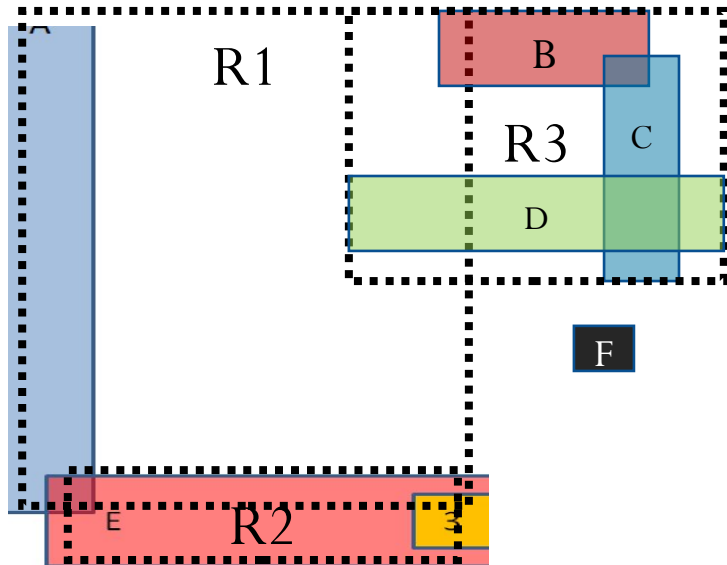
Em qual nó folha F deve ser inserido?
Nó 1 ou Nó 2 ou Nó 3?

R3 terá o menor acréscimo de área.

Overflow! → Qual é a melhor distribuição para BCDF?



Inserir F

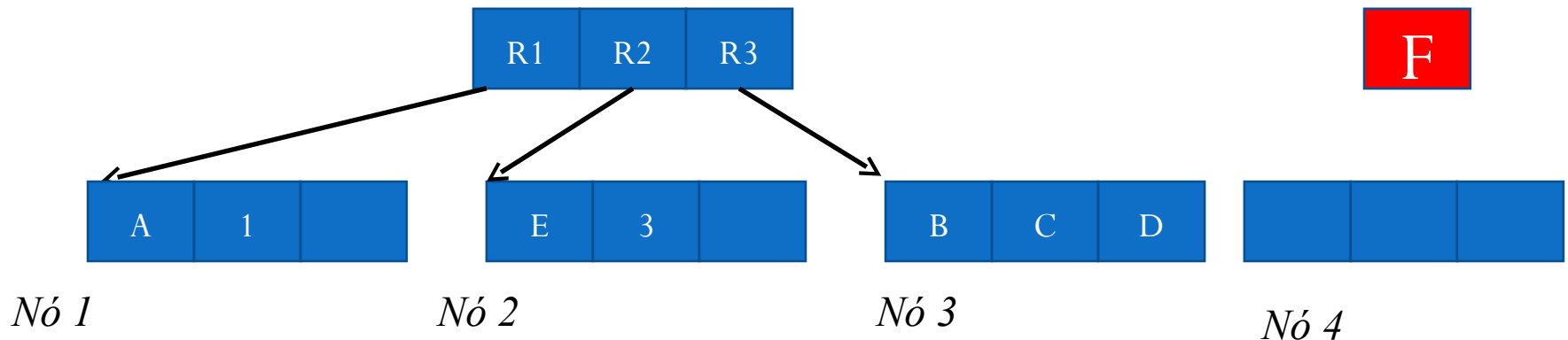


Em qual nó folha F deve ser inserido?
Nó 1 ou Nó 2 ou Nó 3?

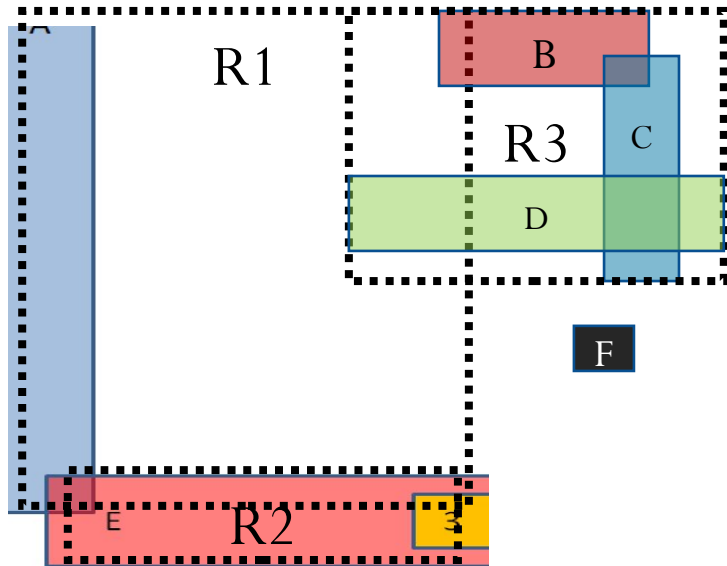
R3 terá o menor acréscimo de área.

Overflow! → Qual é a melhor distribuição para BCDF?

Resp.: BD e CF



Inserir F

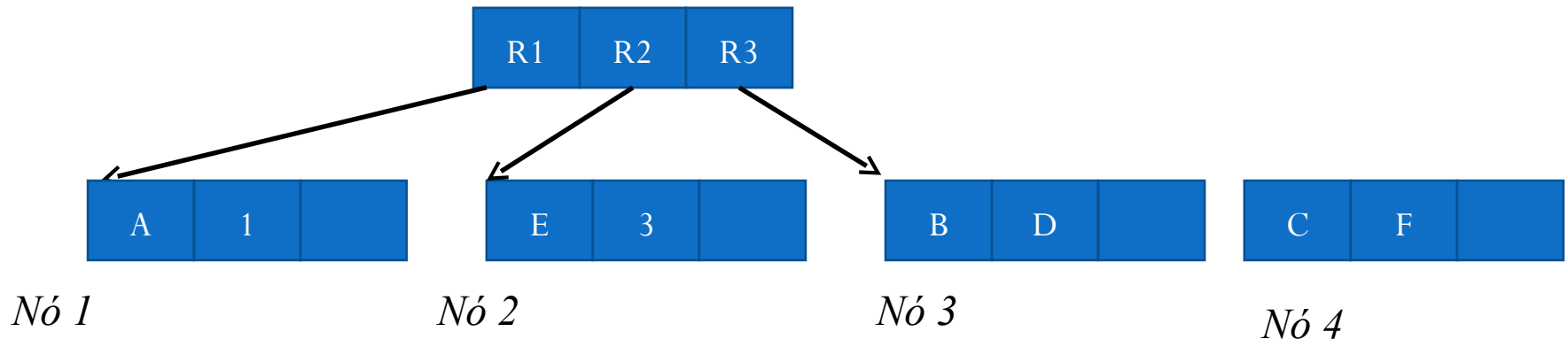


Em qual nó folha F deve ser inserido?
Nó 1 ou Nó 2 ou Nó 3?

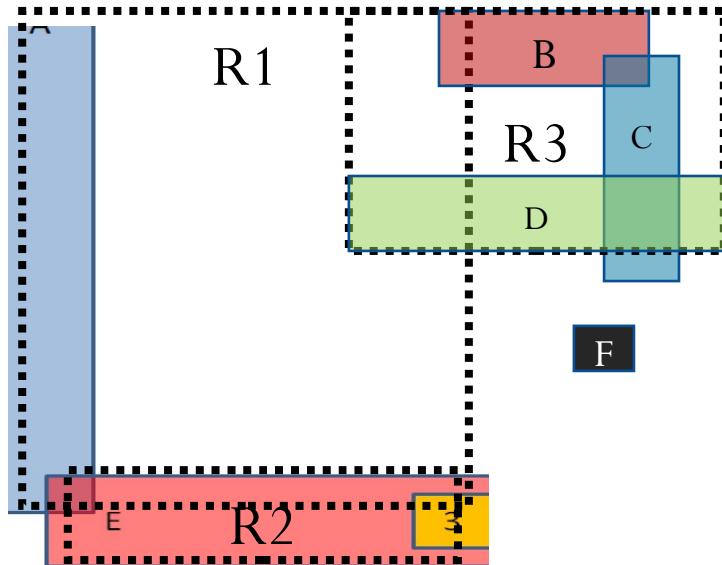
R3 terá o menor acréscimo de área.

Overflow! → Qual é a melhor distribuição para BCDF?

Resp.: BD e CF



Inserir F



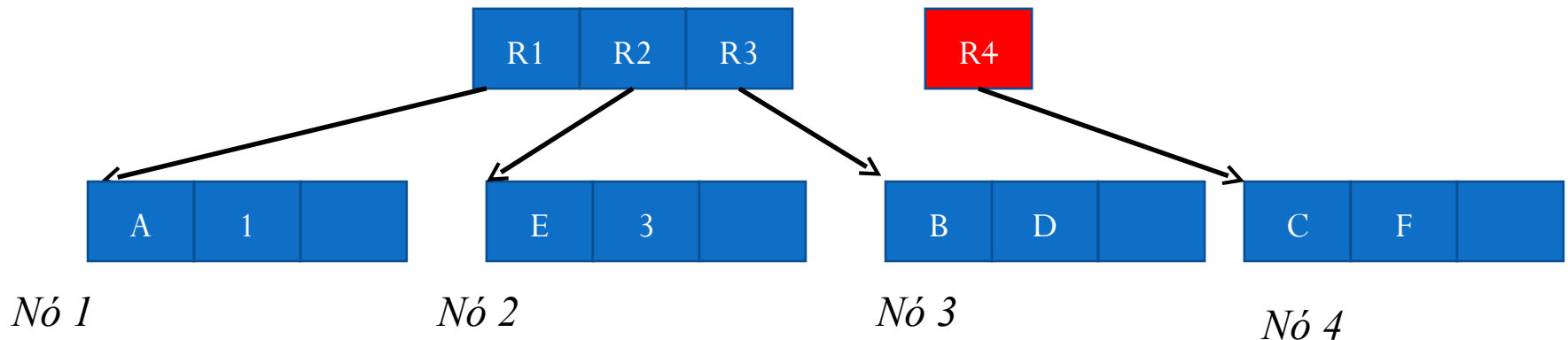
Em qual nó folha F deve ser inserido?
Nó 1 ou Nó 2 ou Nó 3?

R3 terá o menor acréscimo de área.

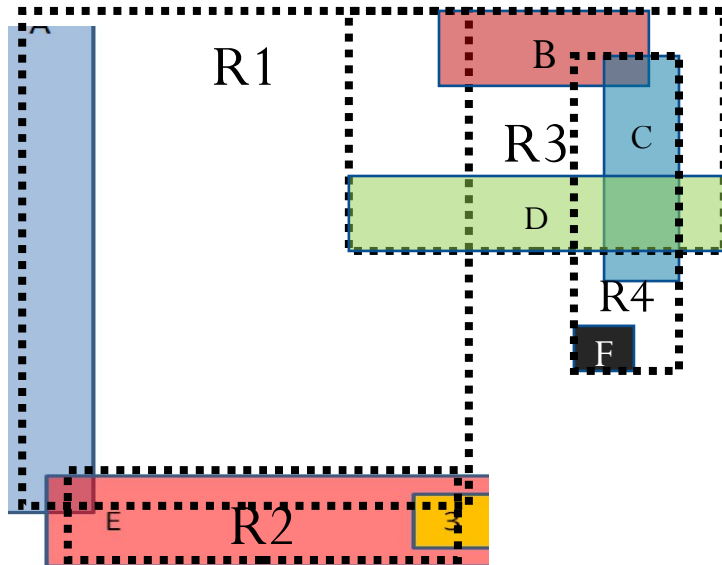
Overflow! → Qual é a melhor distribuição para BCDF?

Resp.: BD e CF

Atualizar R3 e incluir nova entrada na raiz.



Inserir F



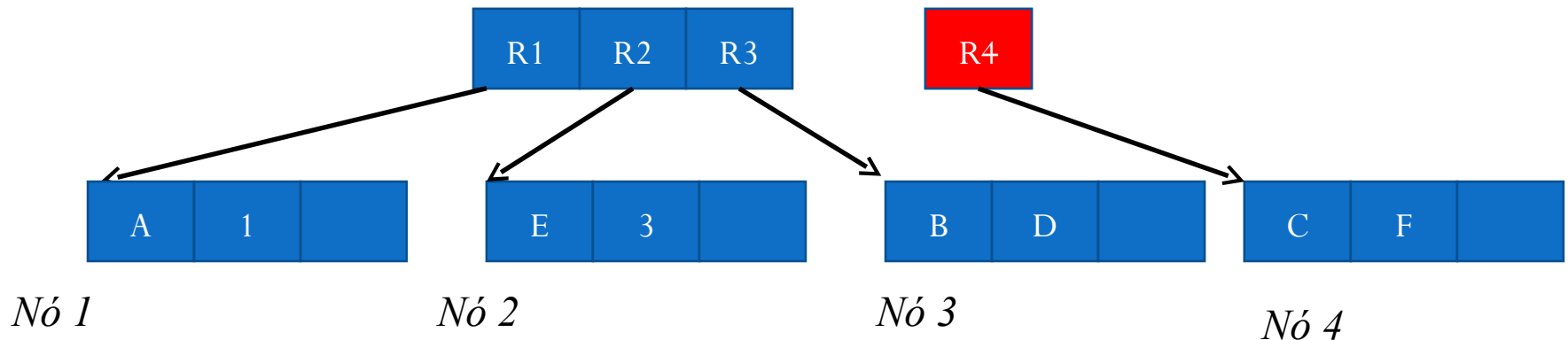
Em qual nó folha F deve ser inserido?
Nó 1 ou Nó 2 ou Nó 3?

R3 terá o menor acréscimo de área.

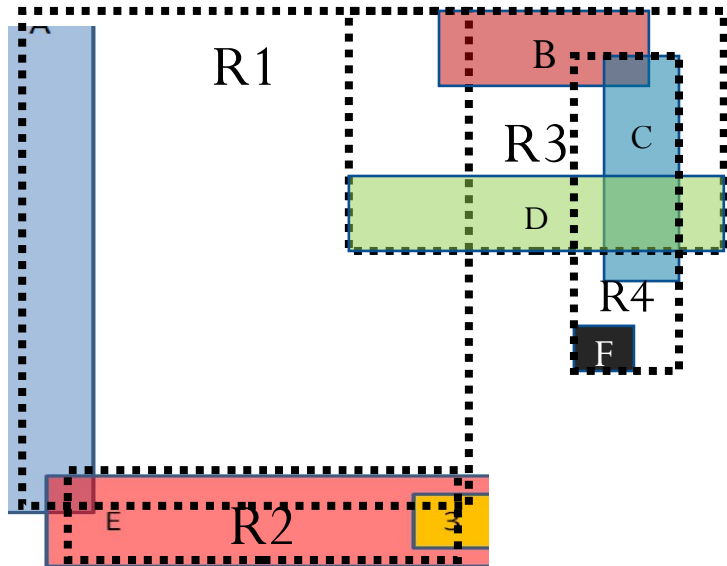
Overflow! → Qual é a melhor distribuição para BCDF?

Resp.: BD e CF

Atualizar R3 e incluir nova entrada na raiz.

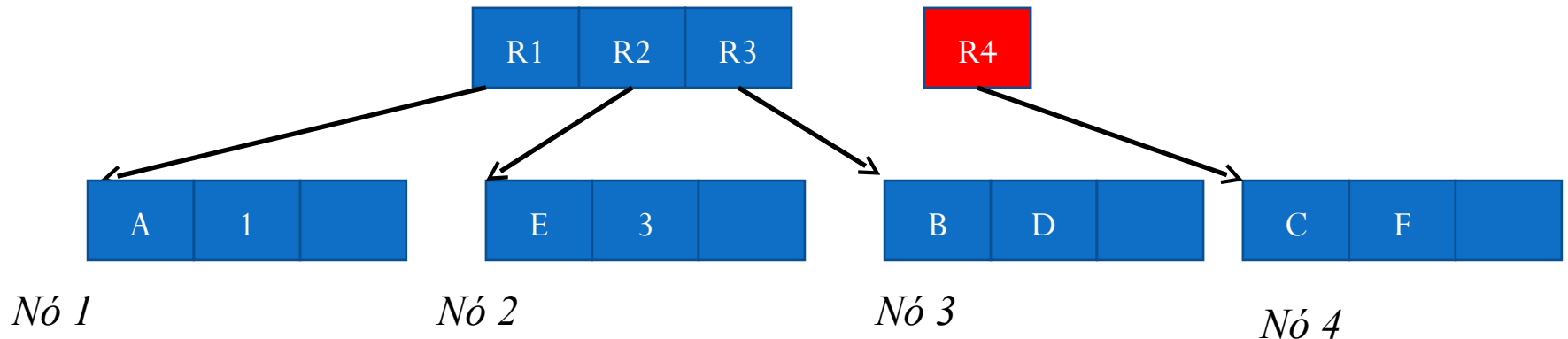


Inserir F

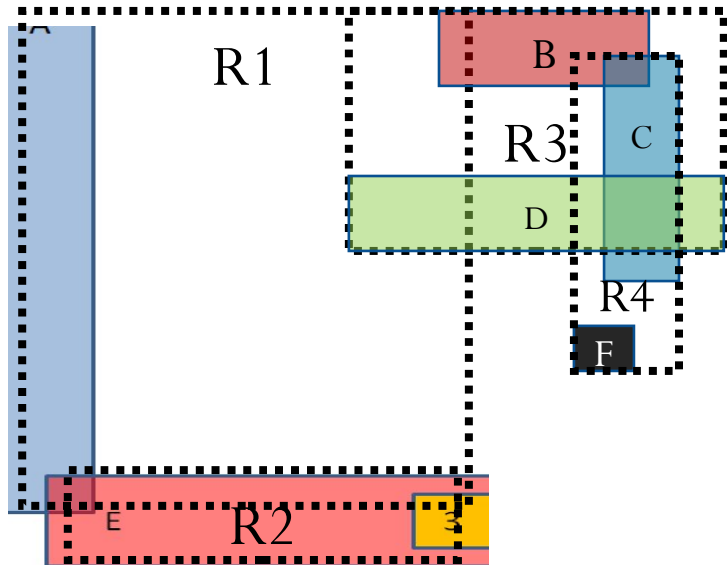


Overflow na raiz!

Criar novo nó e redistribuir entradas de acordo com a menor área dos MBR resultantes.



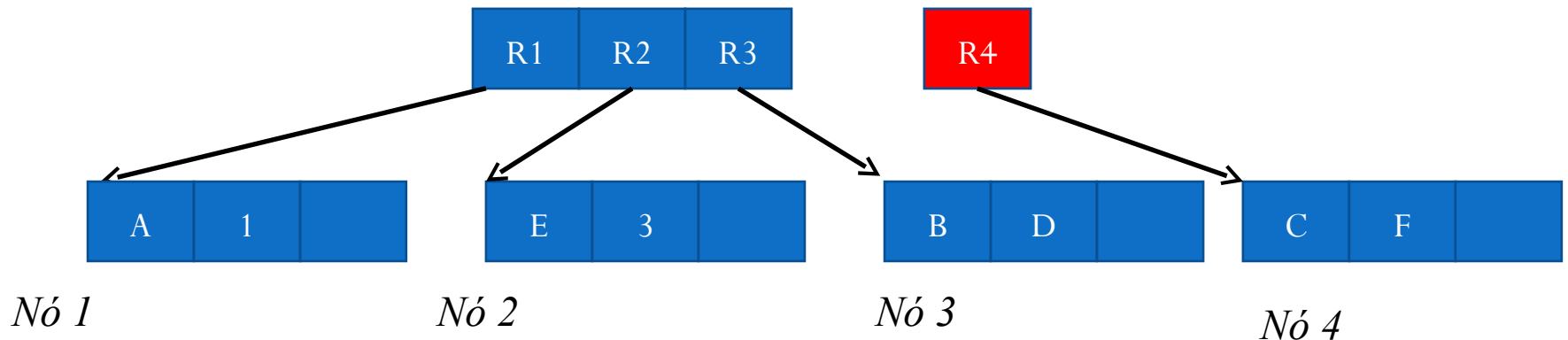
Inserir F



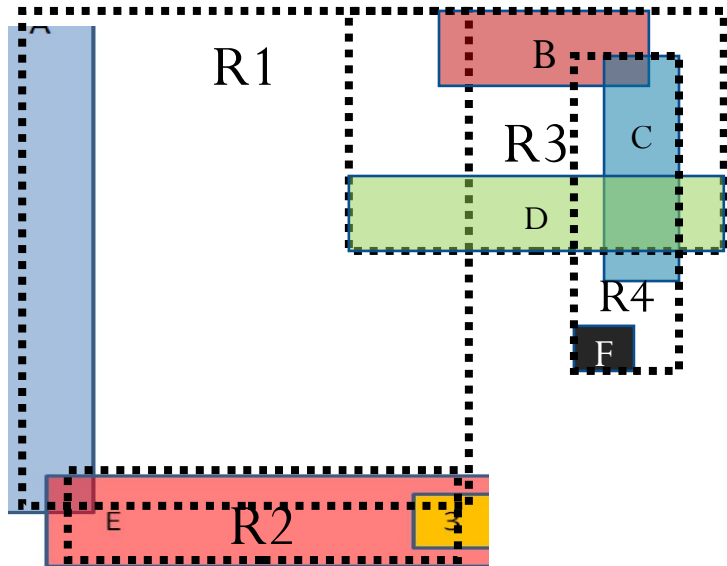
Overflow na raiz!

Criar novo nó e redistribuir entradas de acordo com a menor área dos MBR resultantes.

Nó com R1 R2 e nó com R3 e R4



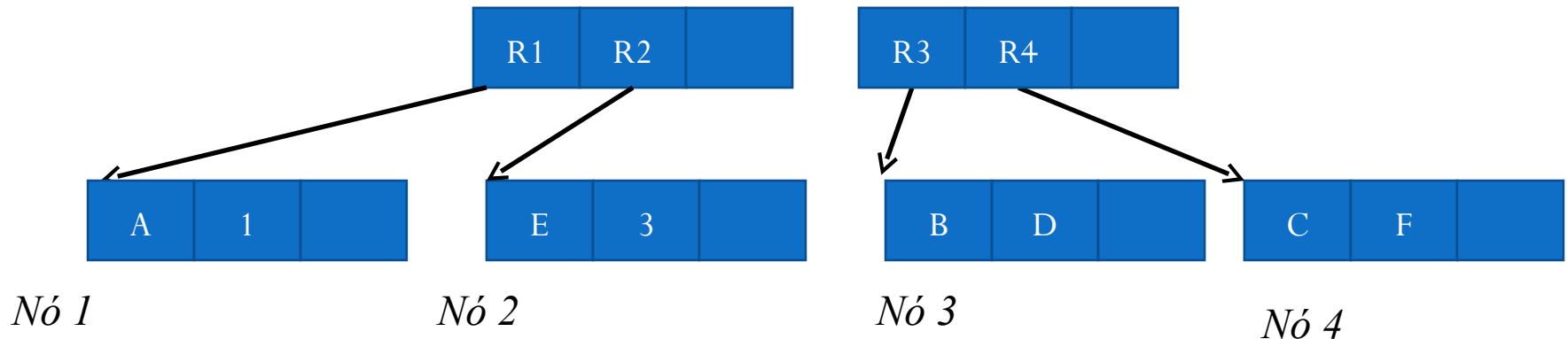
Inserir F



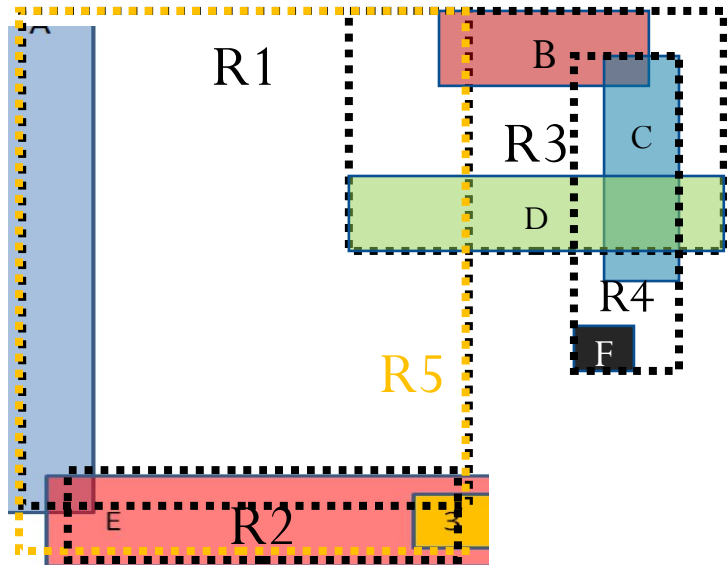
Overflow na raiz!

Criar novo nó e redistribuir entradas de acordo com a menor área dos MBR resultantes.

Nó com R1 R2 e nó com R3 e R4



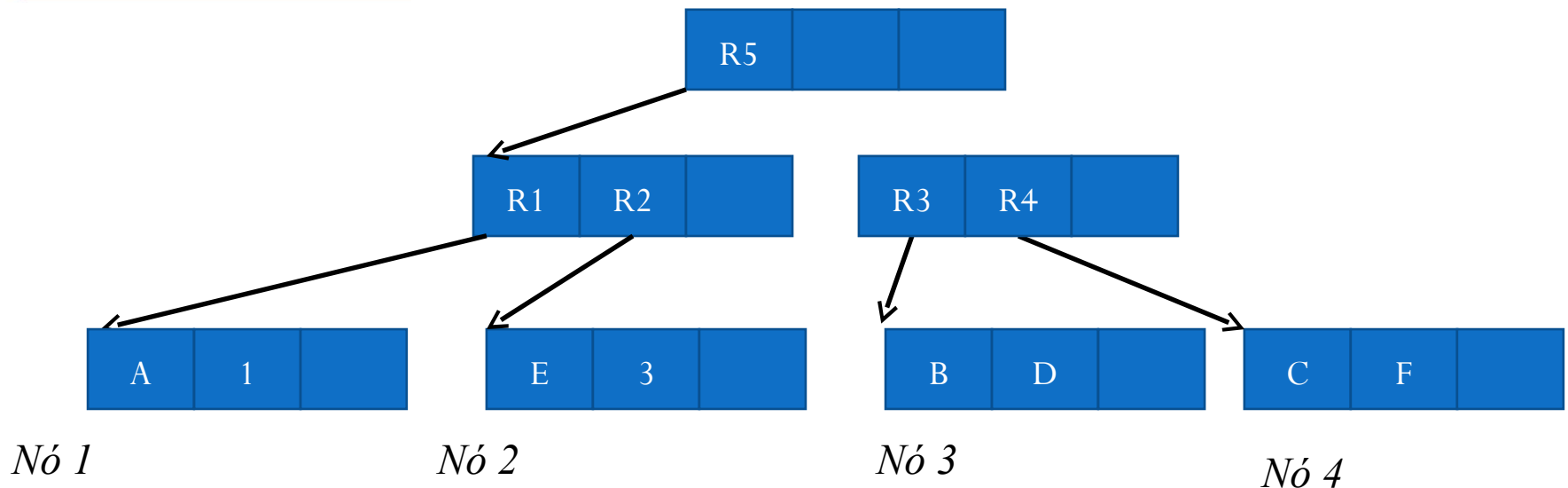
Inserir F



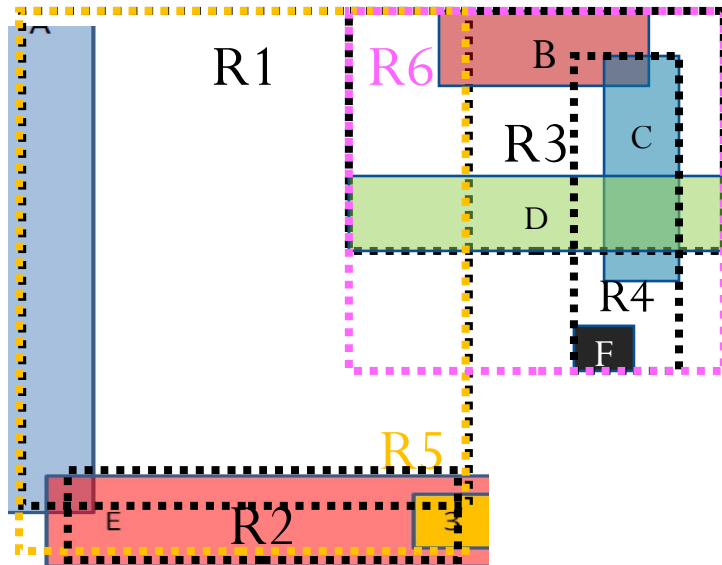
Overflow na raiz!

Criar novo nó e redistribuir entradas de acordo com a menor área dos MBR resultantes.

Nó com R1 R2 e nó com R3 e R4



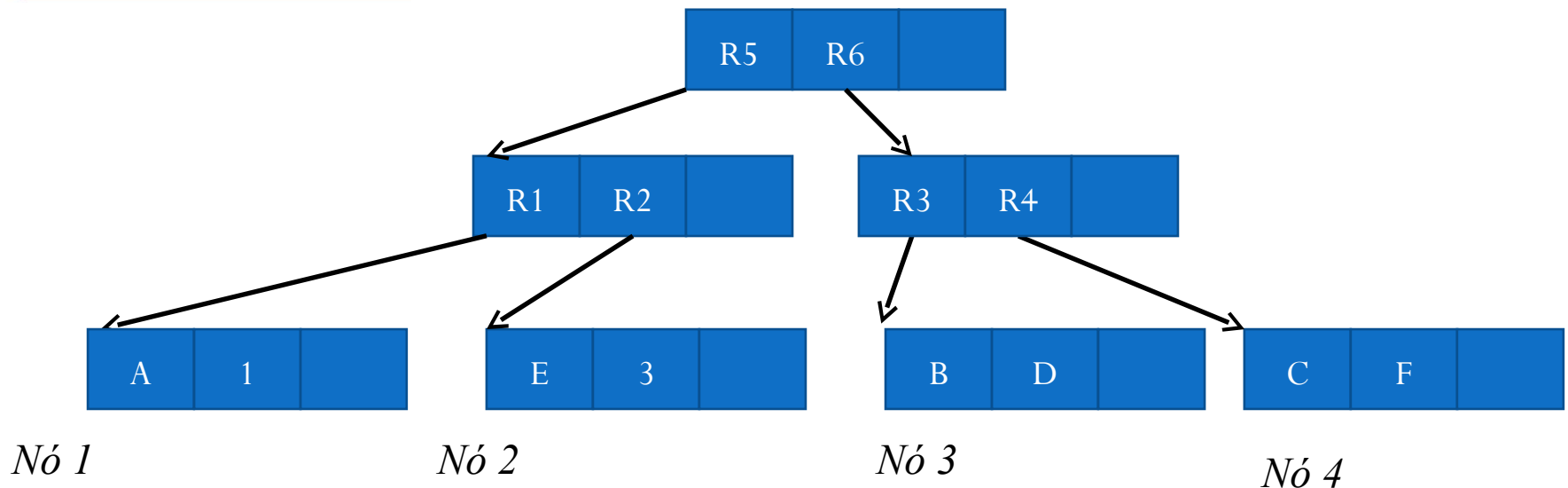
Inserir F



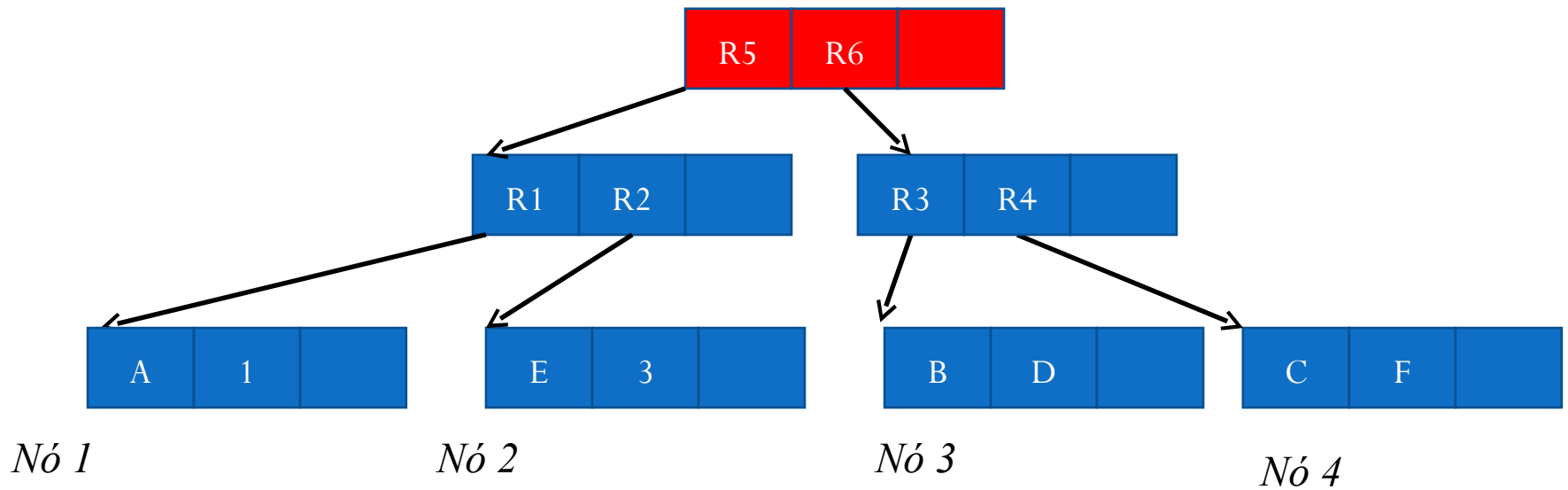
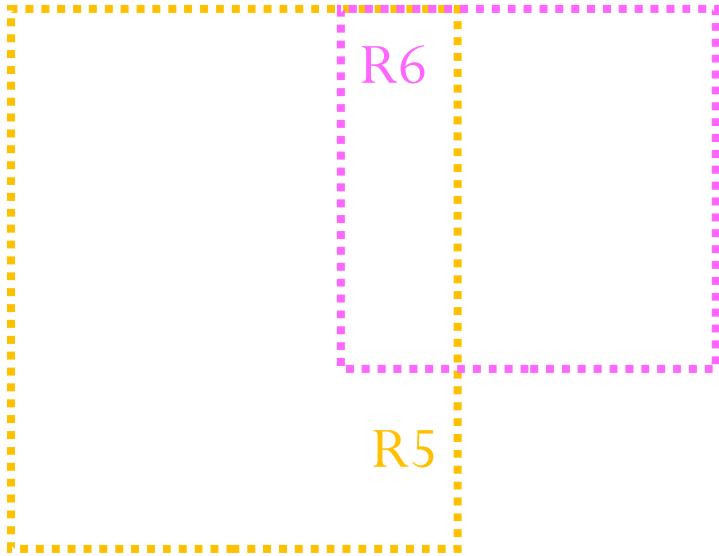
Overflow na raiz!

Criar novo nó e redistribuir entradas de acordo com a menor área dos MBR resultantes.

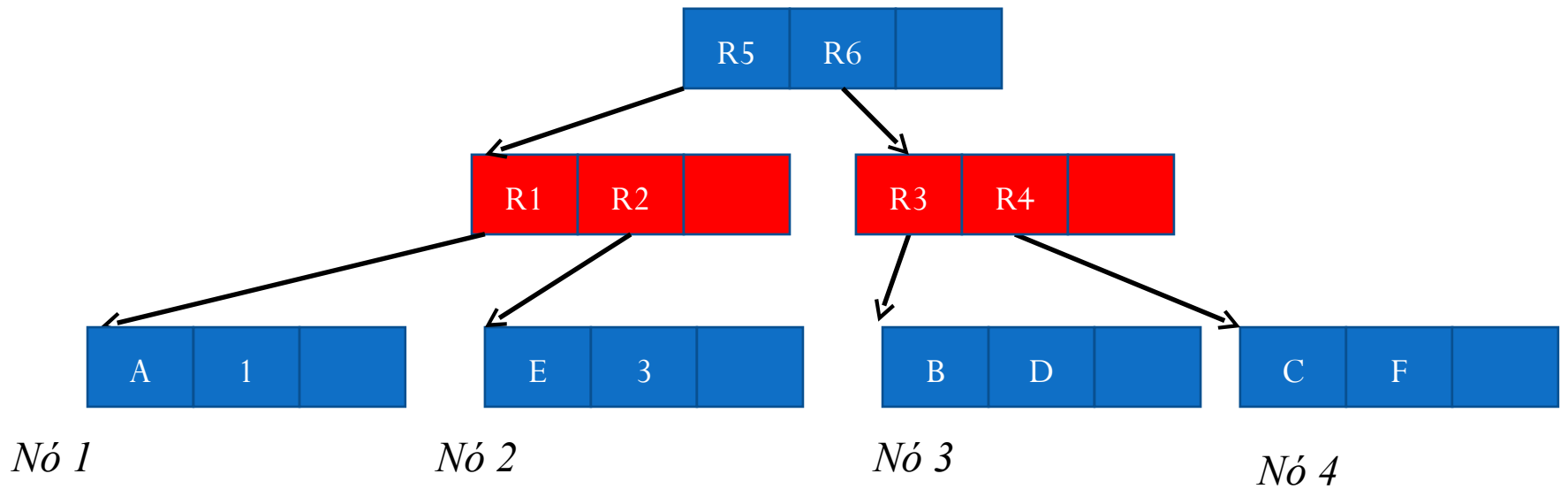
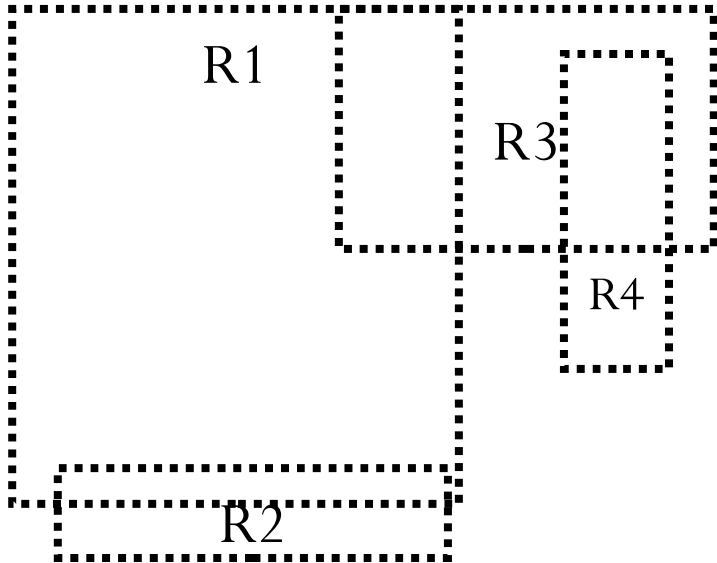
Nó com R1 R2 e nó com R3 e R4



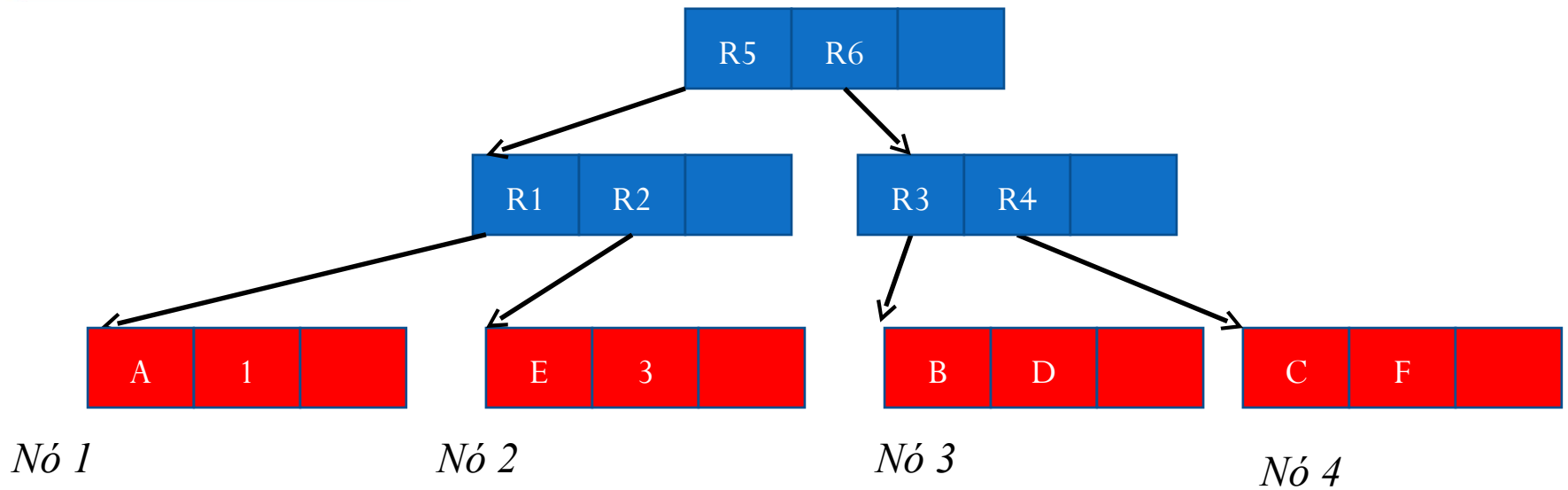
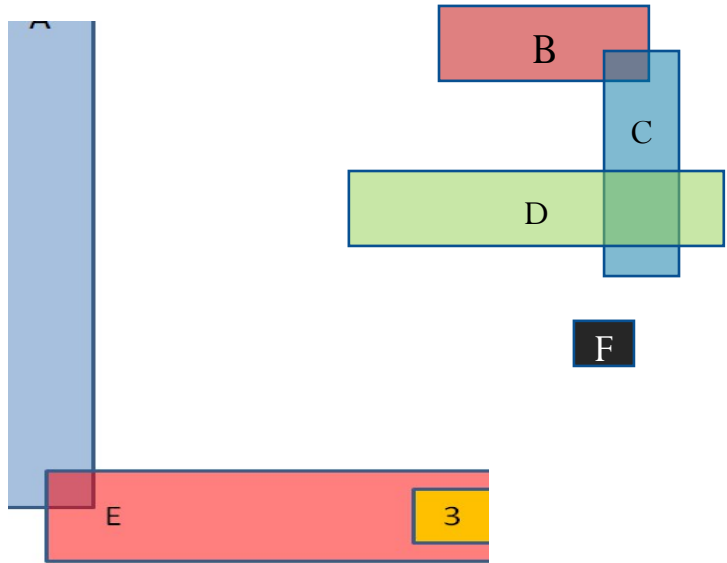
Nível 1



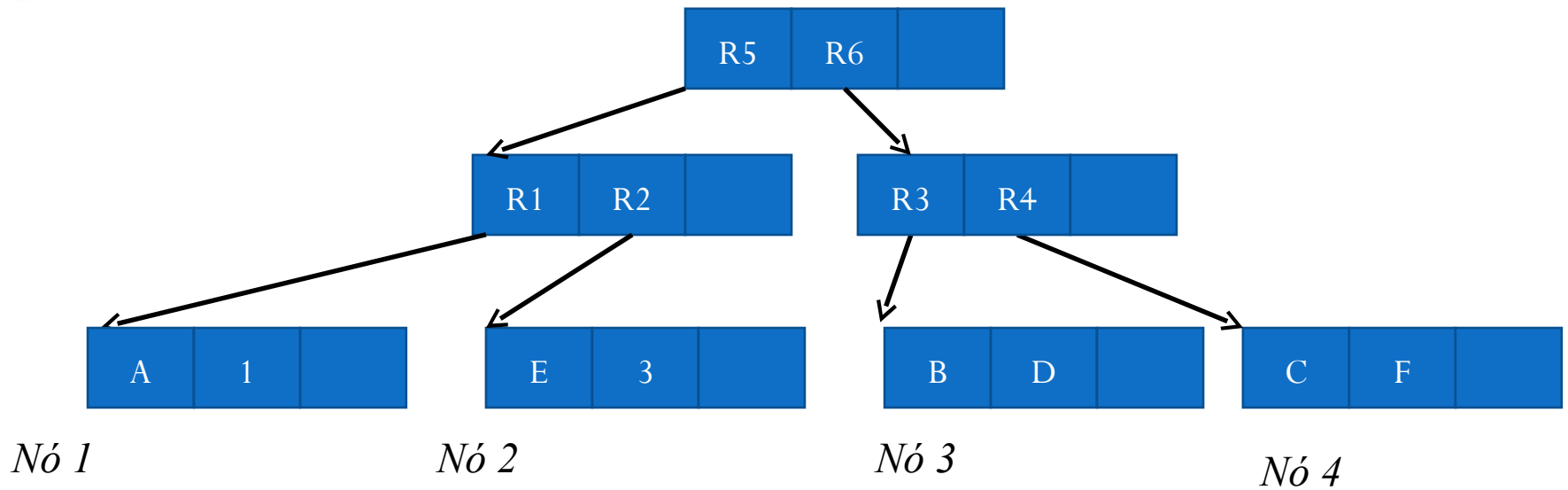
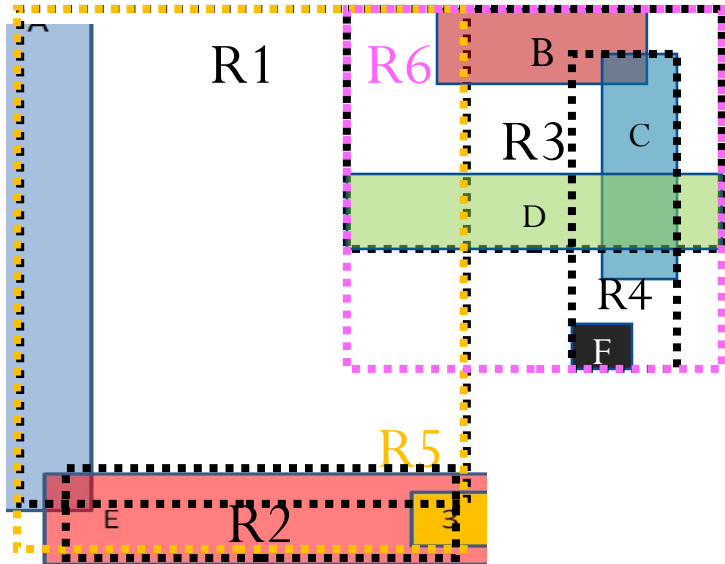
Nível 2



Nível 3

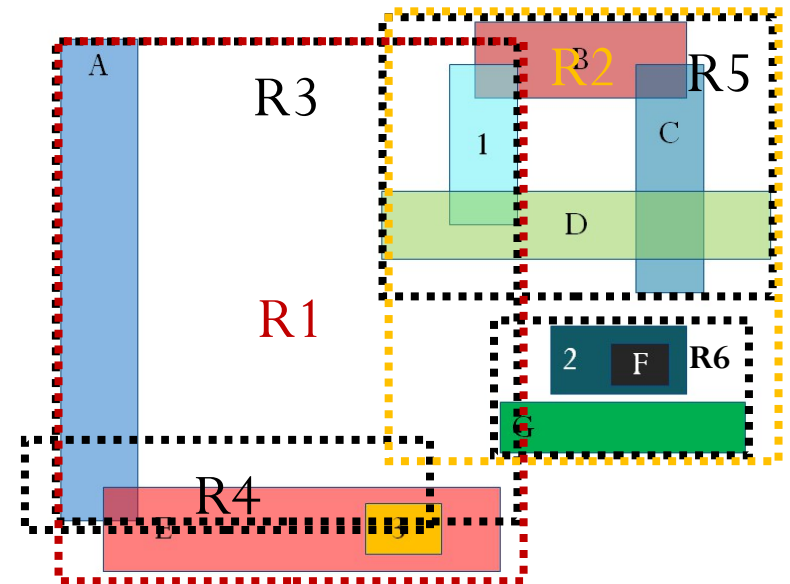


Níveis 1, 2 e 3.

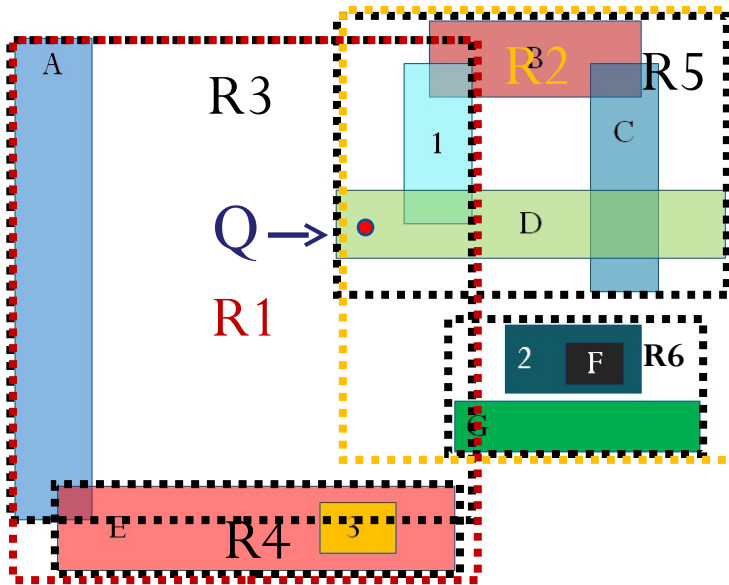


Árvore R - Busca

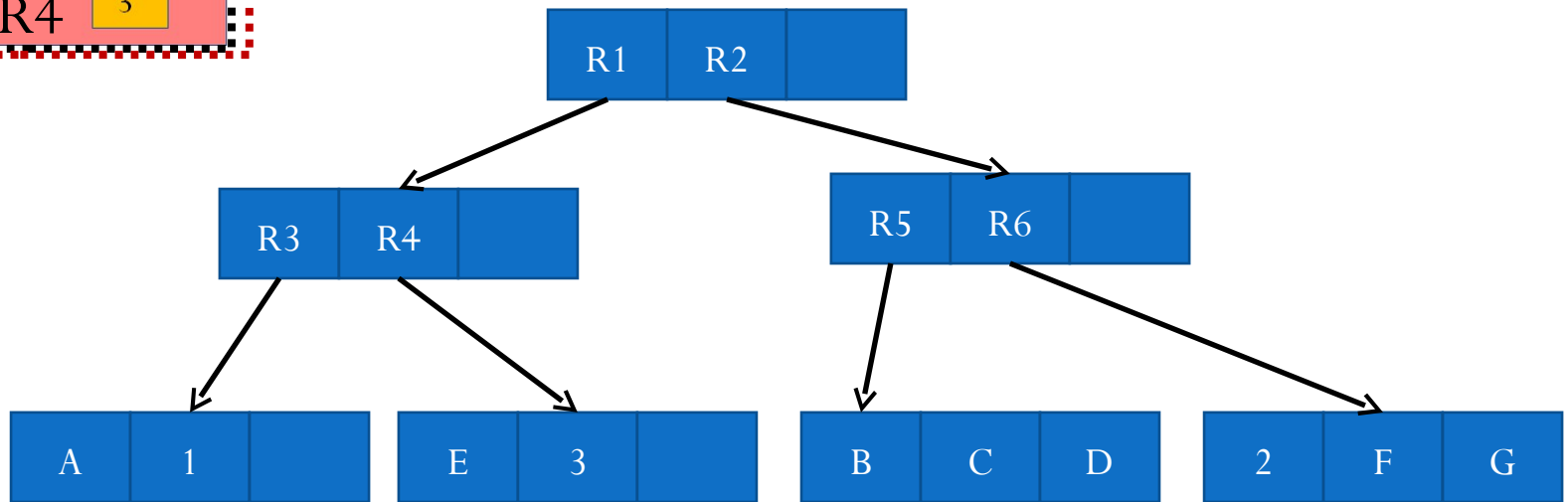
- A busca em uma árvore-R é semelhante à busca em árvore-B
- Problema:
 - Uma grande quantidade de nós pode ter que ser examinada, pois um retângulo pode estar contido nos retângulos envolventes de muitos nós
 - Mas o seu registro está contido em apenas um nó folha.
 - Exemplo: Retângulo 1 está contido nos retângulos:
 - R1
 - R2
 - R3
 - R5



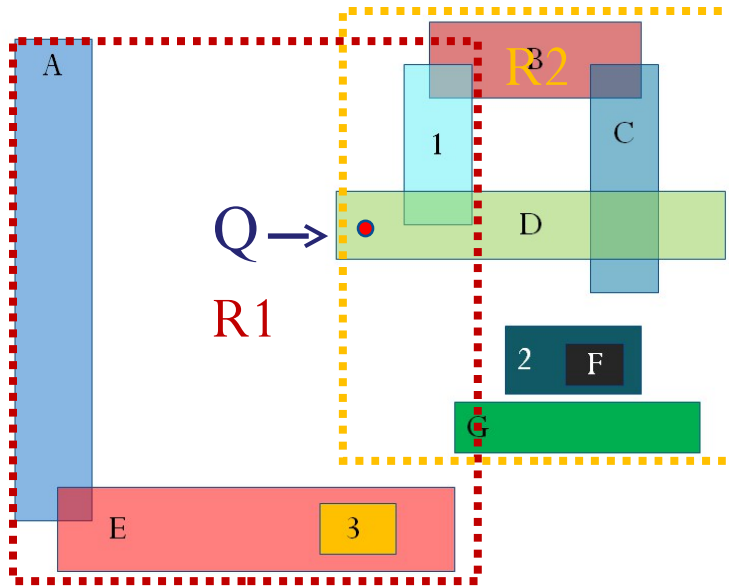
Árvore R - Busca



Qual seria o caminho percorrido na árvore R a seguir para encontrar o ponto Q?

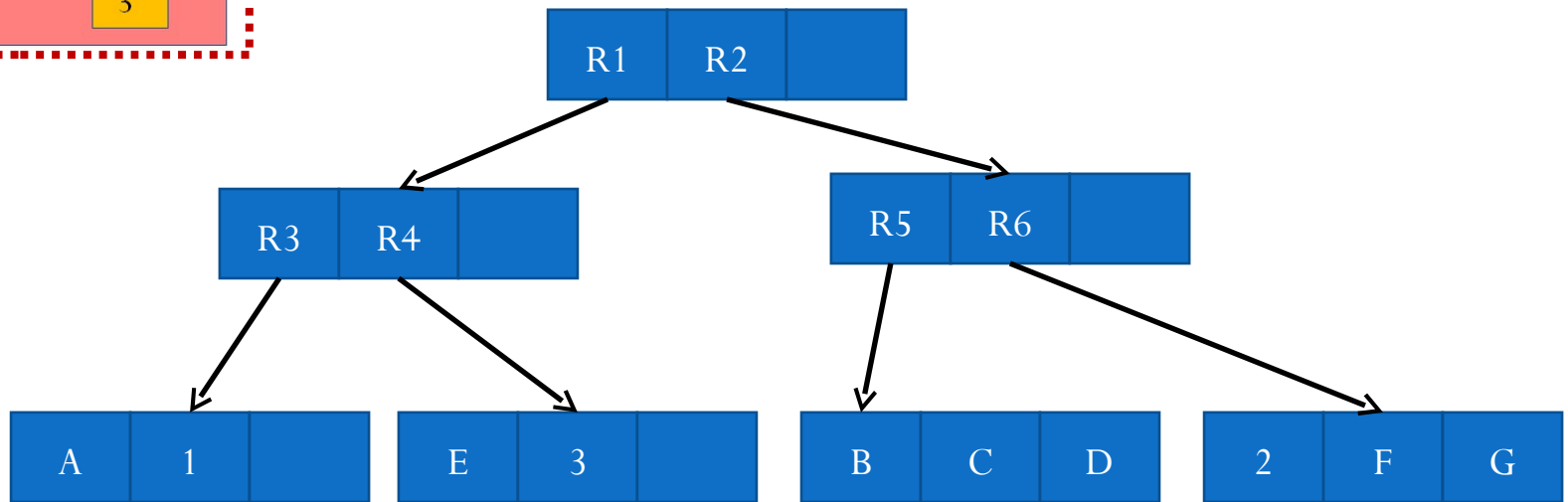


Árvore R - Busca

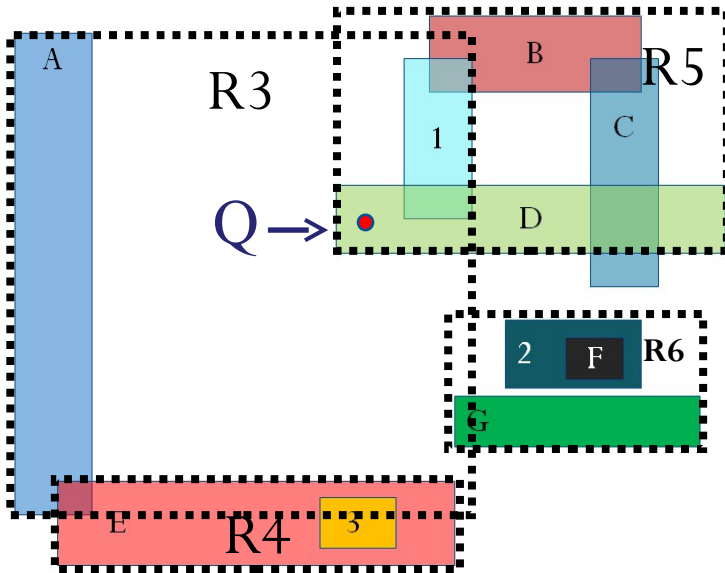


Qual seria o caminho percorrido na árvore R a seguir para encontrar o ponto Q?

1) Raiz ($\sqrt{R1}$ e $\sqrt{R2}$)



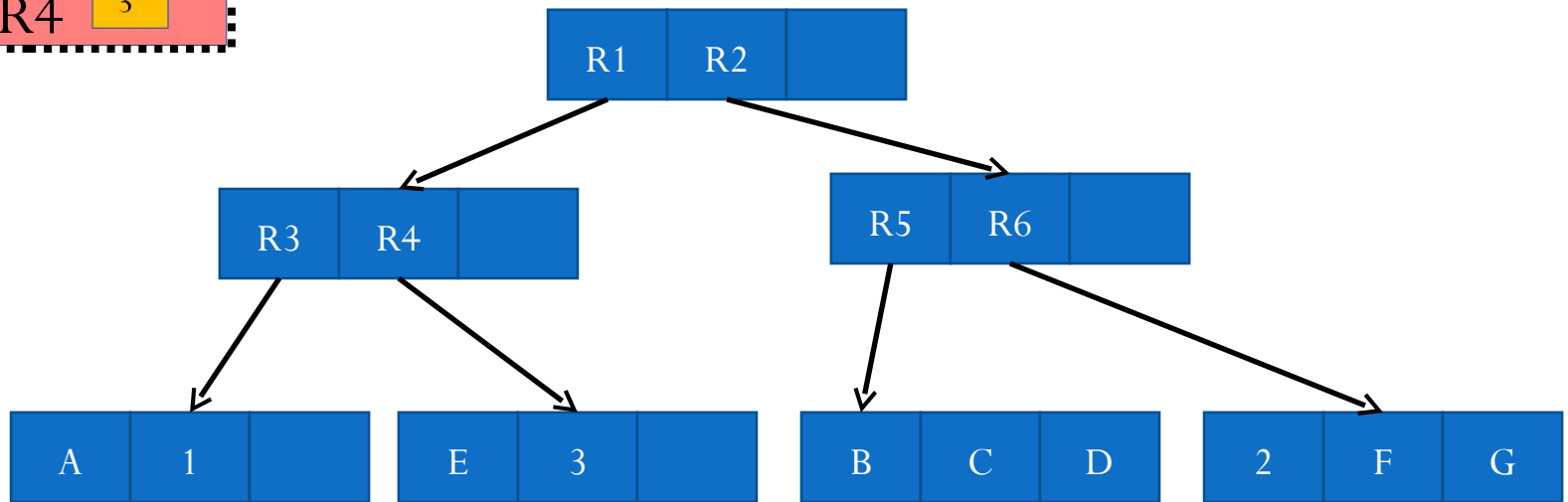
Árvore R - Busca



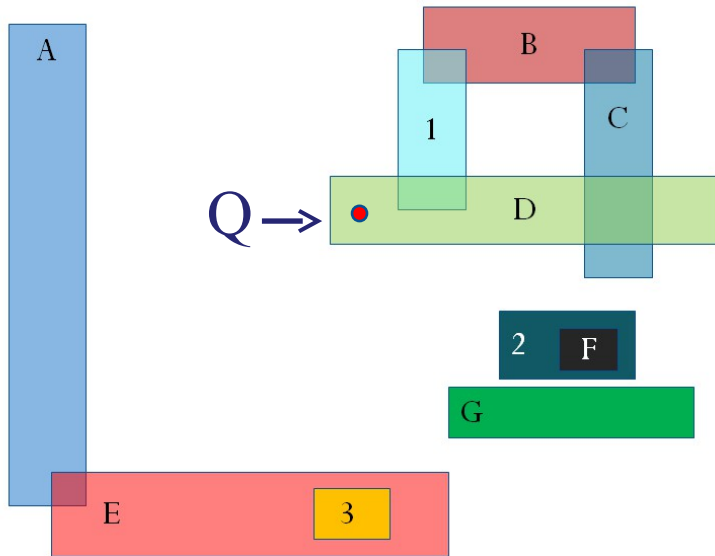
Qual seria o caminho percorrido na árvore R a seguir para encontrar o ponto Q?

1) Raiz (\checkmark R1 e \checkmark R2)

2) Segundo nível (\checkmark R3, \emptyset R4, \checkmark R5, \emptyset R6)



Árvore R - Busca

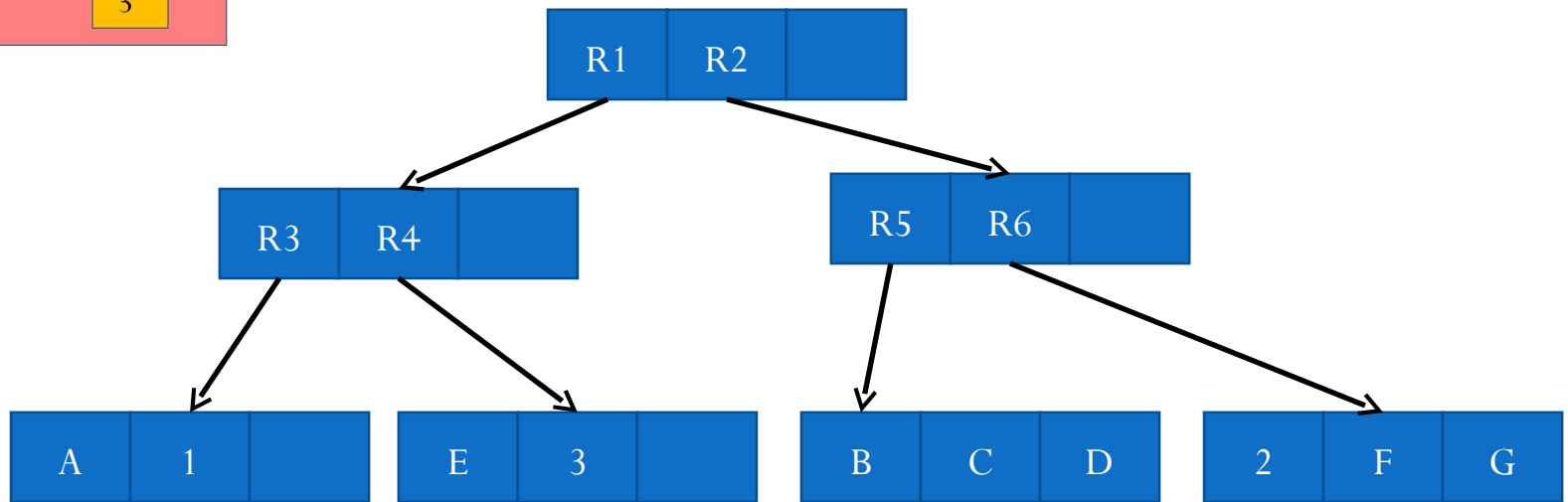


Qual seria o caminho percorrido na árvore R a seguir para encontrar o ponto Q?

1) Raiz (\checkmark R1 e \checkmark R2)

2) Segundo nível (\checkmark R3, \emptyset R4, \checkmark R5, \emptyset R6)

3) Terceiro nível (\emptyset A, \emptyset 1, \emptyset B, \emptyset C, \checkmark D)



Árvore-R - Remoção

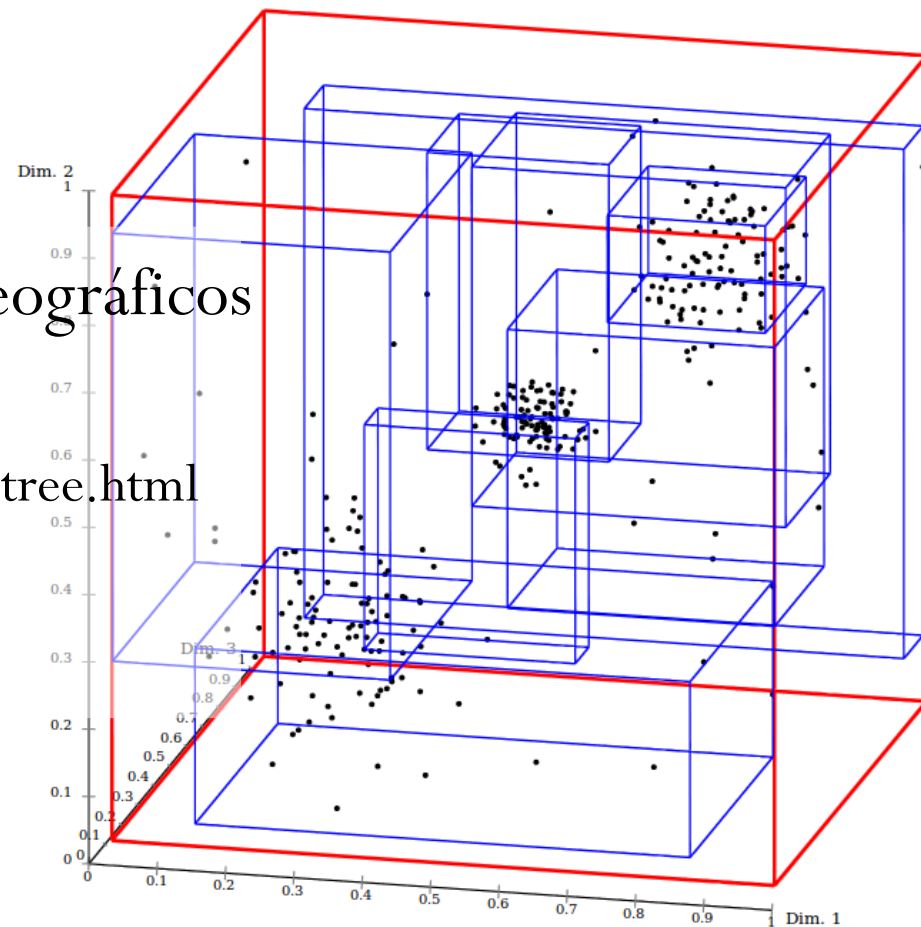
- Remover retângulo R de uma Árvore-R
 - Localizar nó folha L que contém R e remover R de L .
 - Ajustar os retângulos envolventes no caminho de L até a raiz
 - Todos os nós onde ocorrer *underflow* são armazenados em um conjunto U .
 - Quando a raiz é alcançada
 - Se ela tem apenas um único filho, então o filho se torna a nova raiz
- Os nós onde ocorreu *Underflow* são reinseridos na árvore
 - Entradas de U que correspondem a nós folha são incluídas em nós folha
 - Outros nós, têm suas entradas posicionadas no nível que faça com que seus nós folha continuem no mesmo nível dos demais.

Remoção: Árvore-R x Árvore-B

- Na árvore B nós sofrem merge com nós adjacentes ou é feita redistribuição
- Na árvore-R nós são reinseridos
- Árvore-R
 - Não existe conceito de adjacência
 - Merge de nós e redistribuição poderiam ser feitas. Entretanto,
 - Reinserção permite que a árvore dinamicamente reflita a mudança de estrutura espacial ao invés de gradualmente sofrer degradações o que poderia ocorrer mantendo parentesco durante o ciclo de vida da árvore.

Aplicações

- Mais direta: banco de dados geográficos
- Consultas por faixa de valores
 - SQLite: <http://www.sqlite.org/rtree.html>



```
CREATE VIRTUAL TABLE demo_index USING rtree(  
    id,                -- Integer primary key  
    minX, maxX,        -- Minimum and maximum X coordinate  
    minY, maxY         -- Minimum and maximum Y coordinate  
);  
  
SELECT id FROM demo_index  
WHERE minX>=-81.08 AND maxX<=-80.58  
    AND minY>=35.00  AND maxY<=35.44;
```