

Indexação para dados não convencionais



Prof. André Backes | @progdescomplicada

Dados Escalares

- Dizemos que um valor é escalar para denotar um número ou valor simples
- Mais fáceis de manipular
 - Qual a média das notas dos alunos que moram em Uberlândia?
 - Quantos alunos moram na mesma cidade?

Id	UF	Cidade	Rede	Redação
1	SP	Adamantina	Privada	624
2	SP	Adamantina	Pública	602
3	MG	Uberlândia	Privada	758
4	MG	Uberlândia	Pública	698
5	SP	São Carlos	Pública	702
6	MG	Uberaba	Privada	720

Dados Espaciais

- Possuem informações sobre elementos em um espaço
 - Este espaço pode ser tanto a representação de um local físico ou um espaço criado digitalmente
 - Podem representar locais de pontos no espaço, distâncias e formas geométricas
 - Permitem localização em um espaço
- Normalmente definidos por coordenadas (x,y,z ou latitude/longitude)
- Dados espaciais geralmente são acompanhados por dados não espaciais

Dados Espaciais

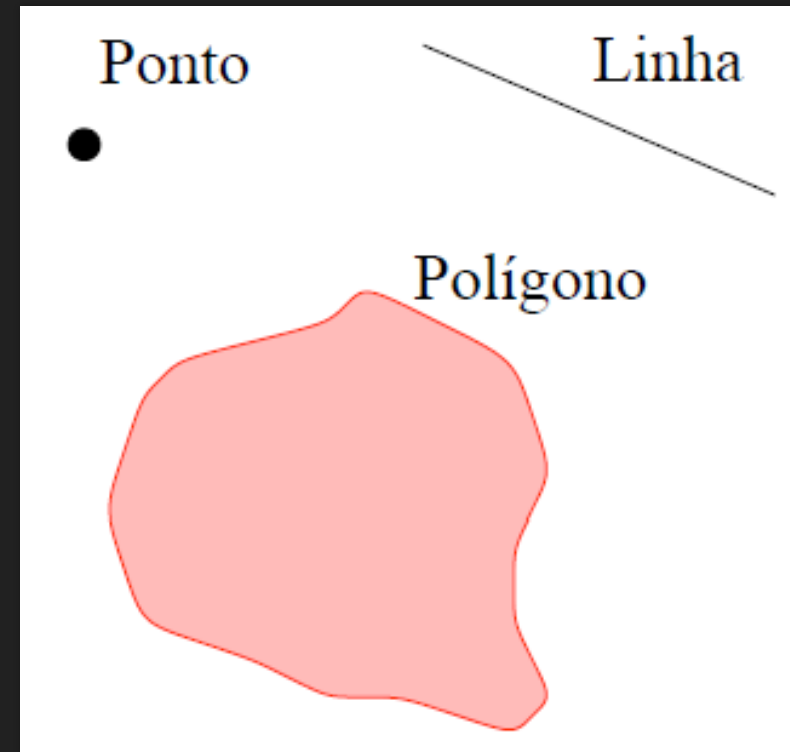
Dados Escalares					Dados Espaciais	
Id	UF	Cidade	Rede	Redação	Latitude	Longitude
1	SP	Adamantina	Privada	624	-21.6756	-51.063
2	SP	Adamantina	Pública	602	-21.6756	-51.063
3	MG	Uberlândia	Privada	758	-18.9113	-48.2622
4	MG	Uberlândia	Pública	698	-18.9113	-48.2622
5	SP	São Carlos	Pública	702	-22.0154	-47.8911
6	MG	Uberaba	Privada	720	-19.7502	-47.9325

Dados Espaciais

- São usados em muitas aplicações
 - Computação gráfica, visão computacional, realidade virtual
 - Sistemas de informações geográficas (google maps, google Earth)
 - Simulação / modelagem geométrica 2D ou 3D (AutoCAD, Google Sketchup)
 - Games

Objetos espaciais

- Objetos espaciais são objetos que descrevem localizações ou formas geométricas
 - 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
 - Polígono

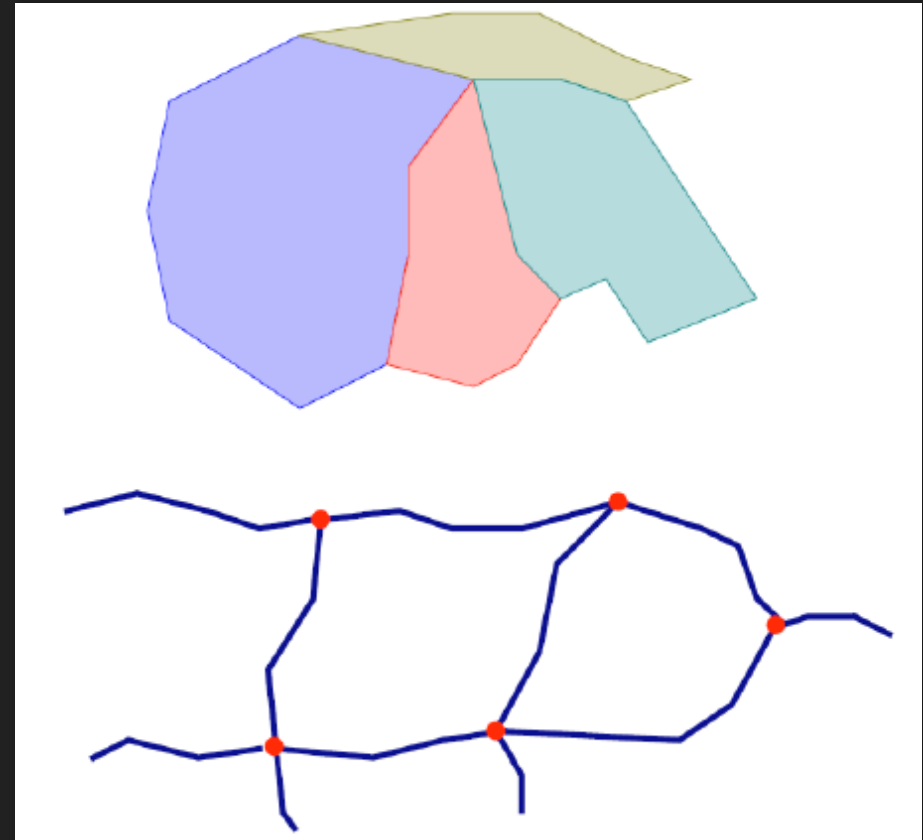


Objetos espaciais

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

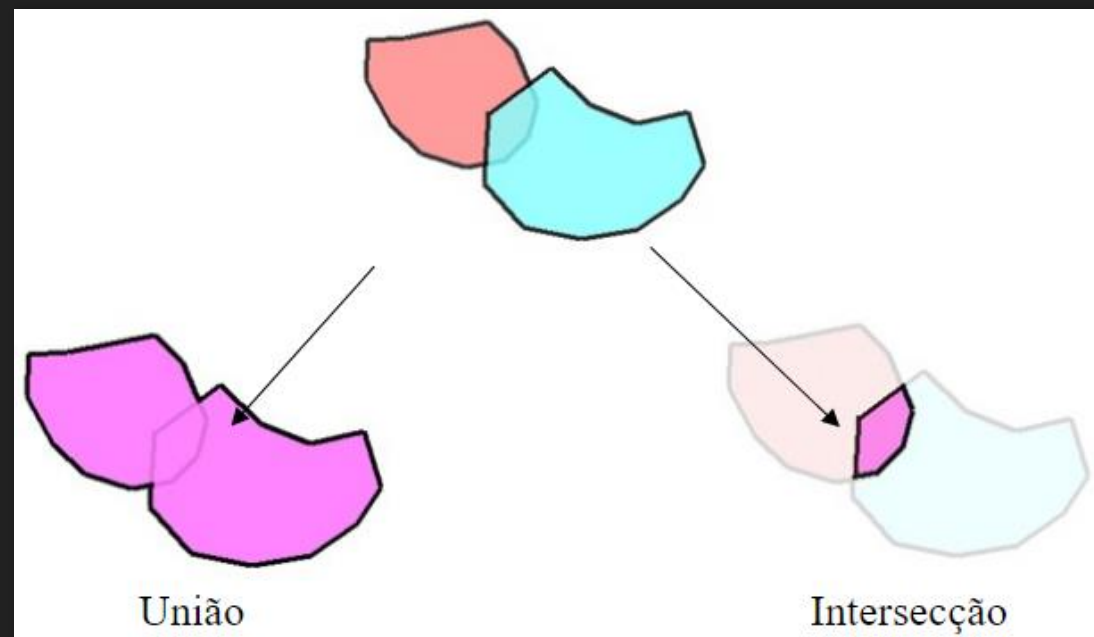
Objetos espaciais

- 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



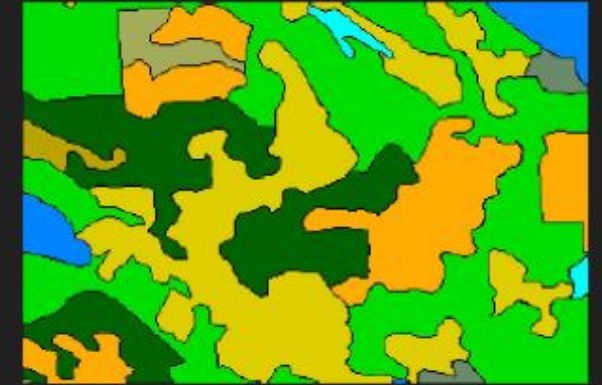
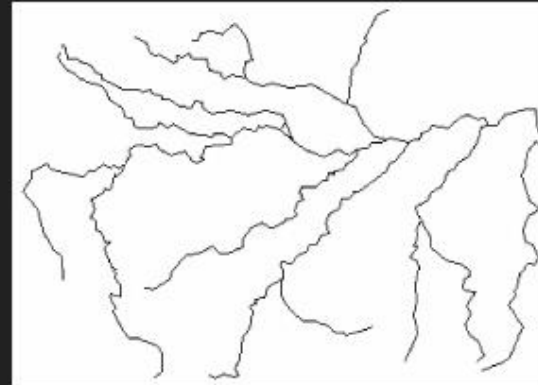
Operações com objetos espaciais

- Várias são as operações possíveis
 - União
 - Intersecção
 - Seleção
 - É vizinho de?
 - Menor retângulo envolvente



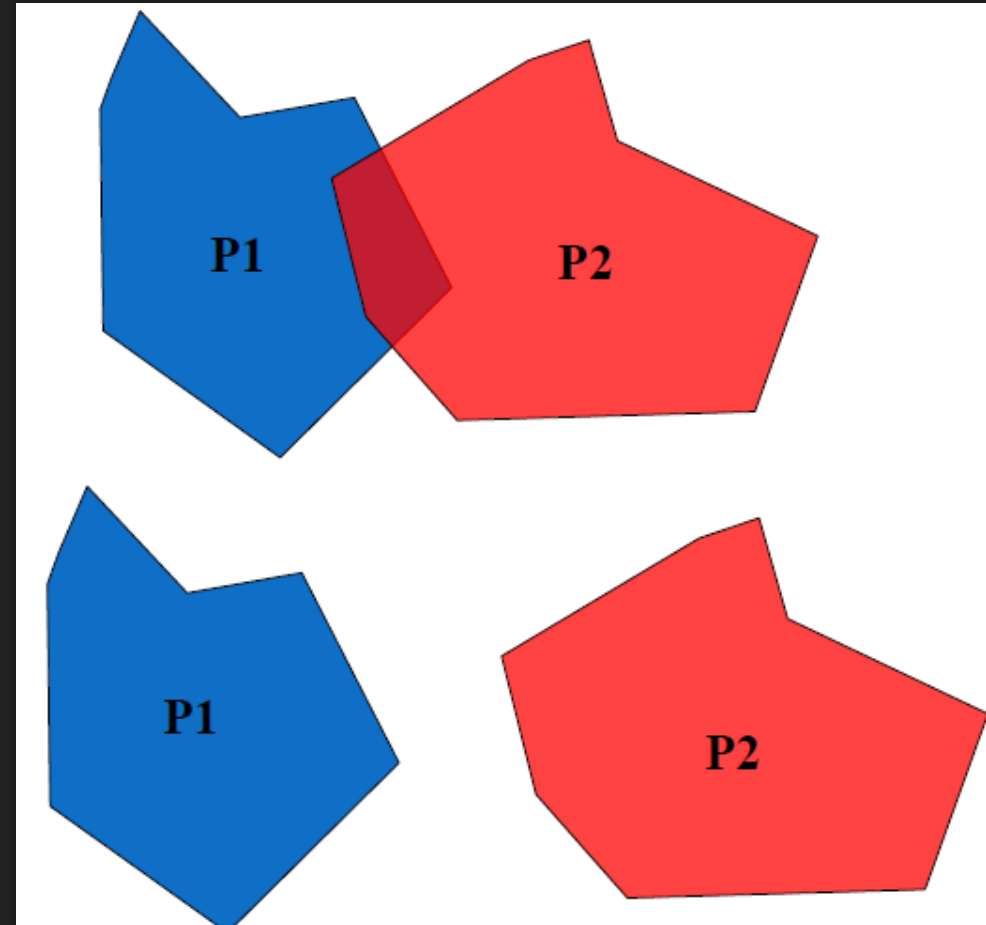
Operações com objetos espaciais

- Dado um Rio (ou bacia hidrográfica), um mapa dos municípios, e um mapa de relevo
 - Por quais municípios esse rio passa?
 - Quais áreas seriam inundadas se o nível do rio subir 1 metro acima do nível do solo?



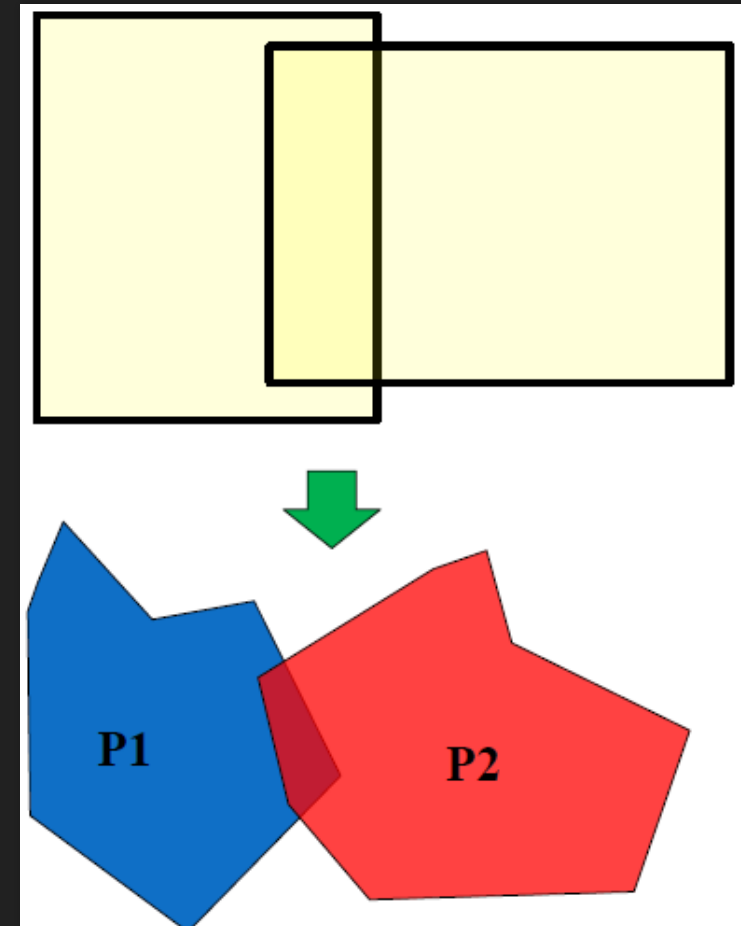
Operações com objetos espaciais

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



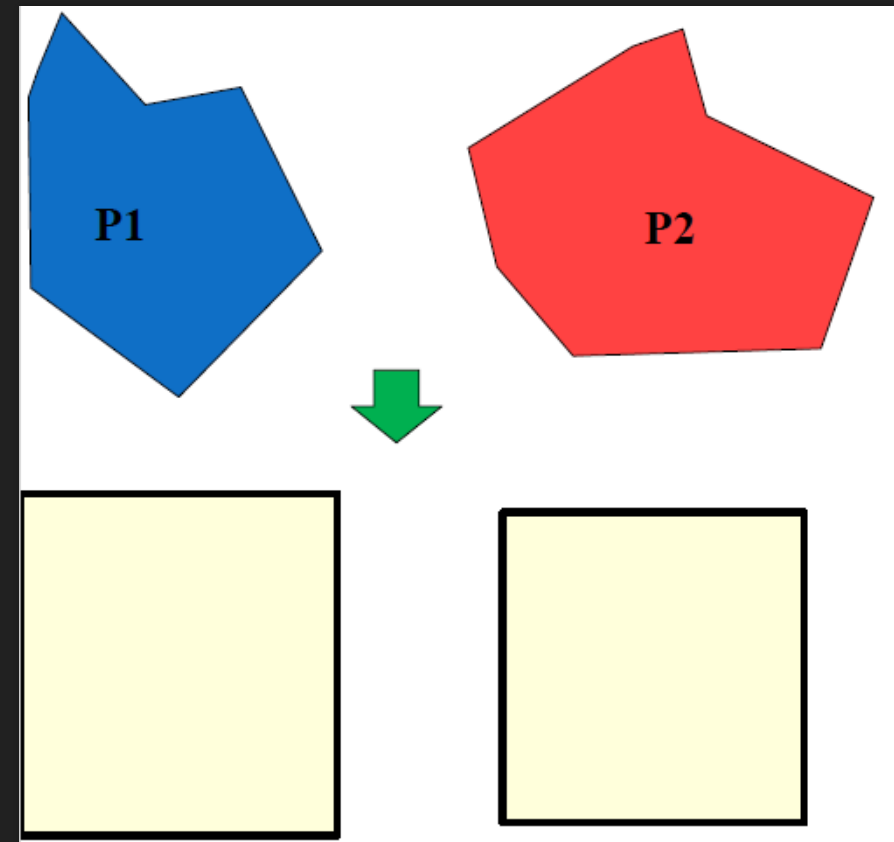
Operações com objetos espaciais

- Reduzindo a complexidade dos objetos
 - Fazendo testes mais simples primeiro
 - Menor Retângulo Envolvente
- Se os retângulos têm interseção, em geral, não se pode inferir que objetos têm interseção
 - Após esse teste é necessário testar os objetos reais
 - Passo mais custoso



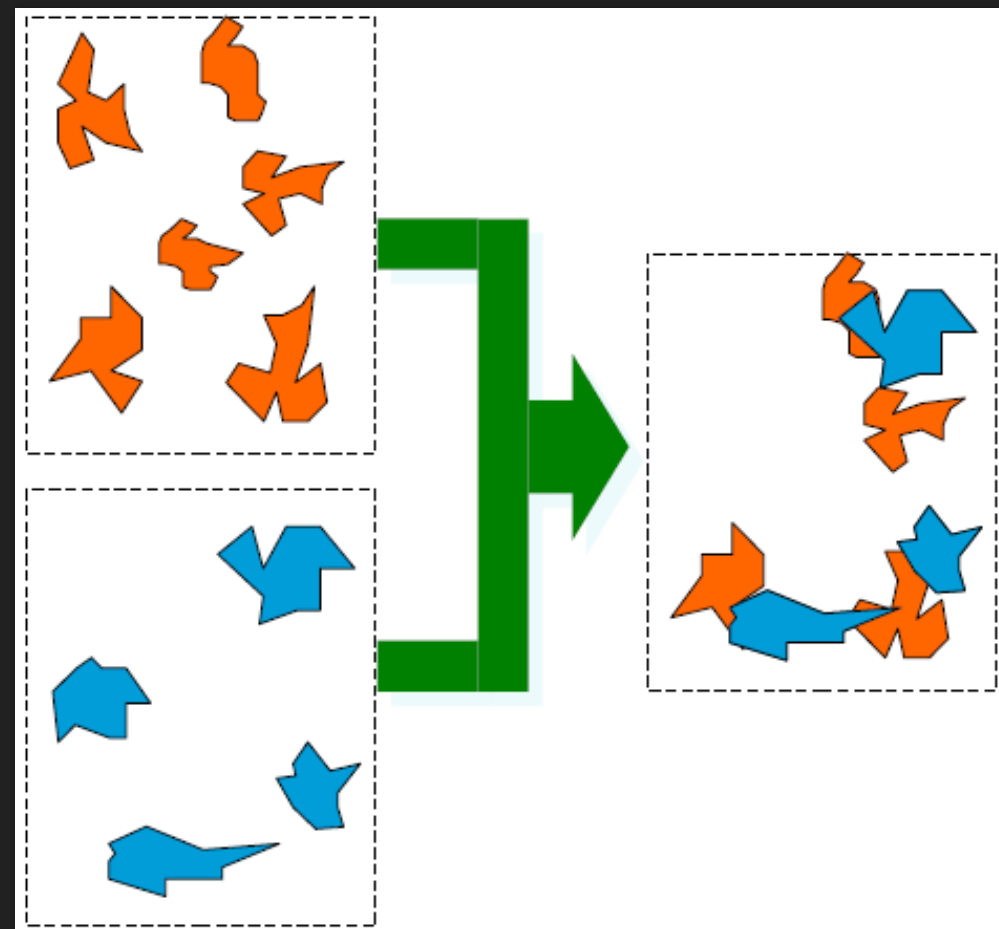
Operações com objetos espaciais

- De outra forma, se objetos não têm interseção, então seus retângulos não têm intersecção
- Não é necessário testar os objetos reais
 - Economia de tempo!



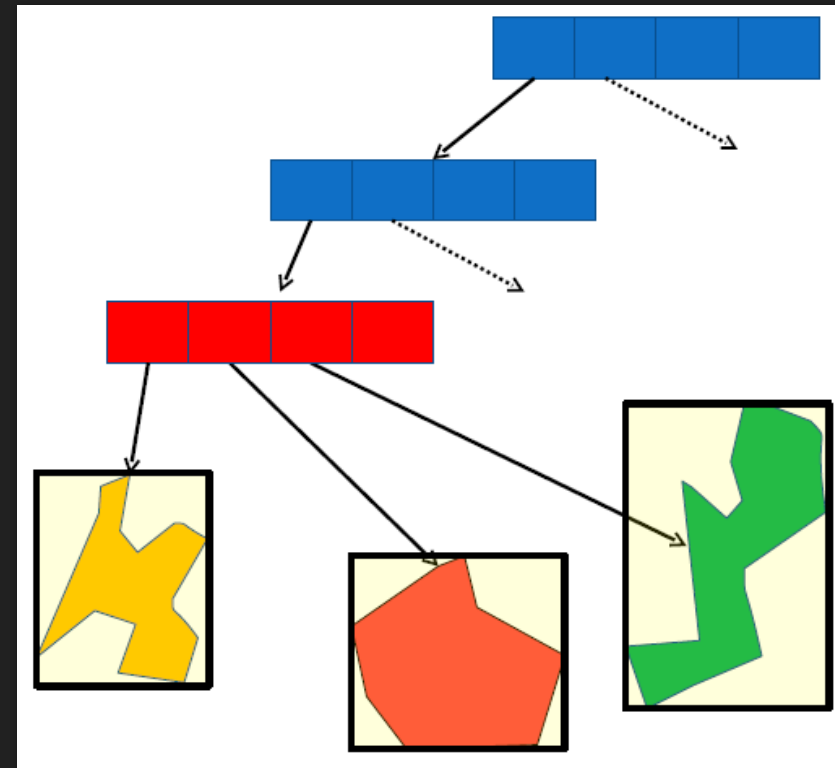
Operações com objetos espaciais

- Como comparar muitos objetos de forma eficiente?
- Solução: uso de índices



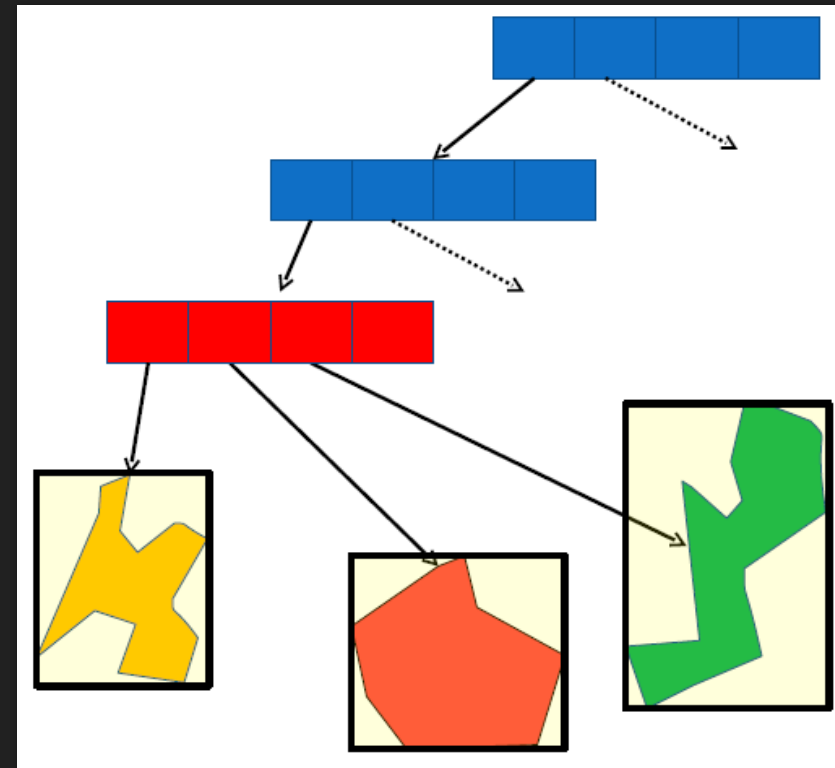
Árvores R

- Proposta por Guttman em 1984
- Indexação de dados em espaços métricos
- Entradas dos nós correspondem à retângulos em d -dimensões
- Cada entrada de nó folha contém
 - Ponteiro para objeto real
 - Menor Retângulo do objeto



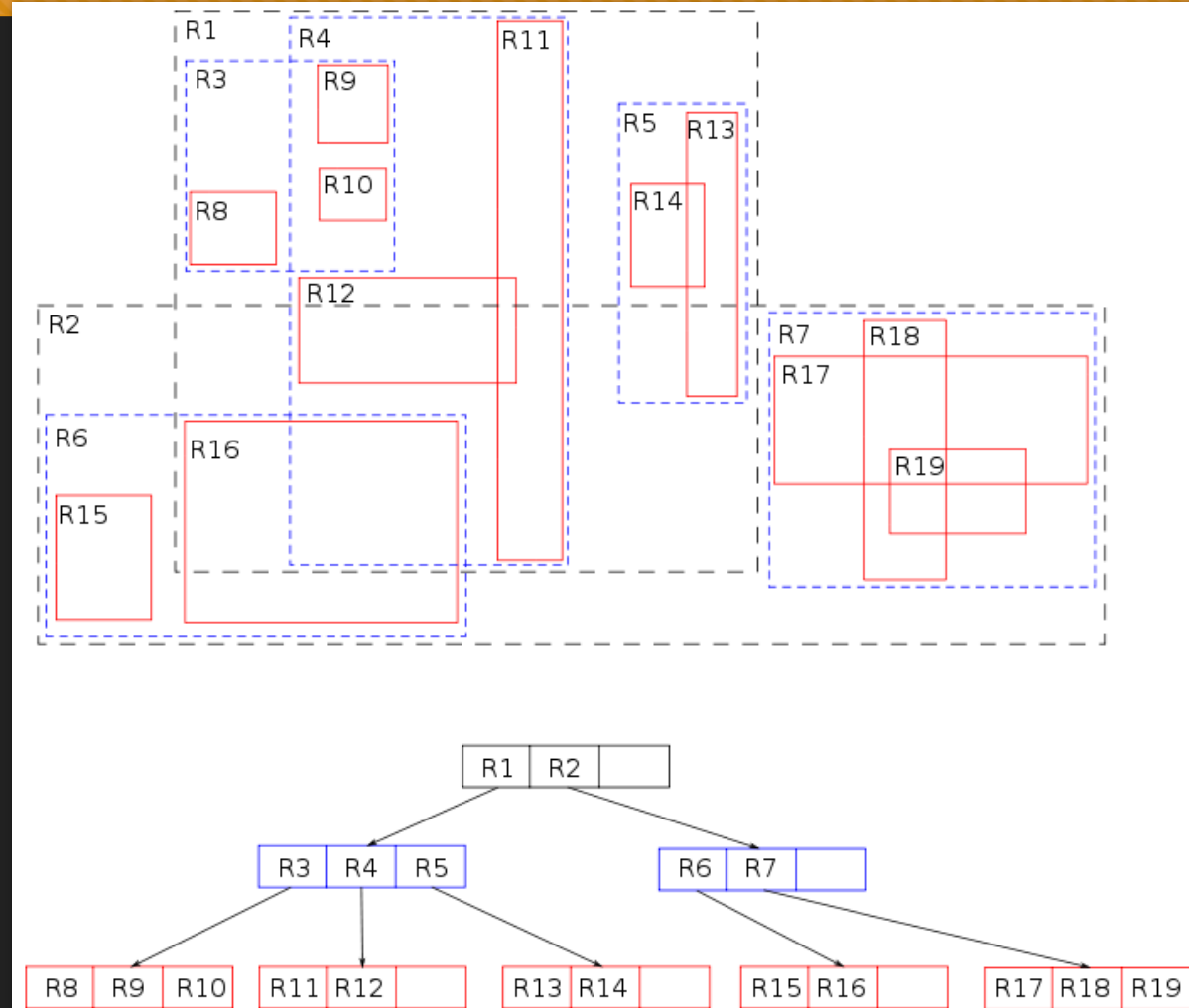
Árvores R

- Estrutura de dados hierárquica derivada da árvore B
 - Todos os nós folha aparecem no mesmo nível (árvore balanceada)
 - Organiza os dados em páginas
 - Projetado para armazenamento em disco (como usado em bancos de dados)
 - Cada página pode conter um número máximo de entradas, muitas vezes indicadas como M



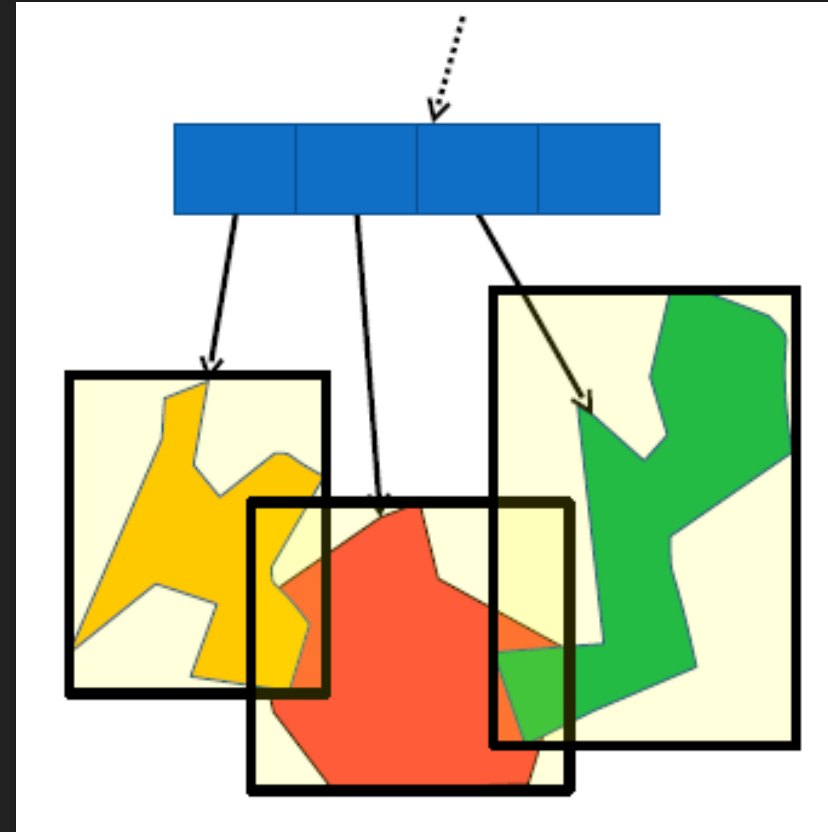
Árvores R

- Os maiores retângulos, ou seja, os nós mais próximos do nó raiz, contém todos os retângulos de seus nós filhos, e os nós-folha (os menores retângulos nesta hierarquia) são os objetos indexados
- O retângulo R1 engloba os retângulos R3, R4 e R5.
- Já estes, englobam os retângulos abaixo na hierarquia e assim sucessivamente



Árvores R

- 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, nós diferentes podem ter sobreposição
 - Ou seja, os retângulos mapeados podem ter sobreposição

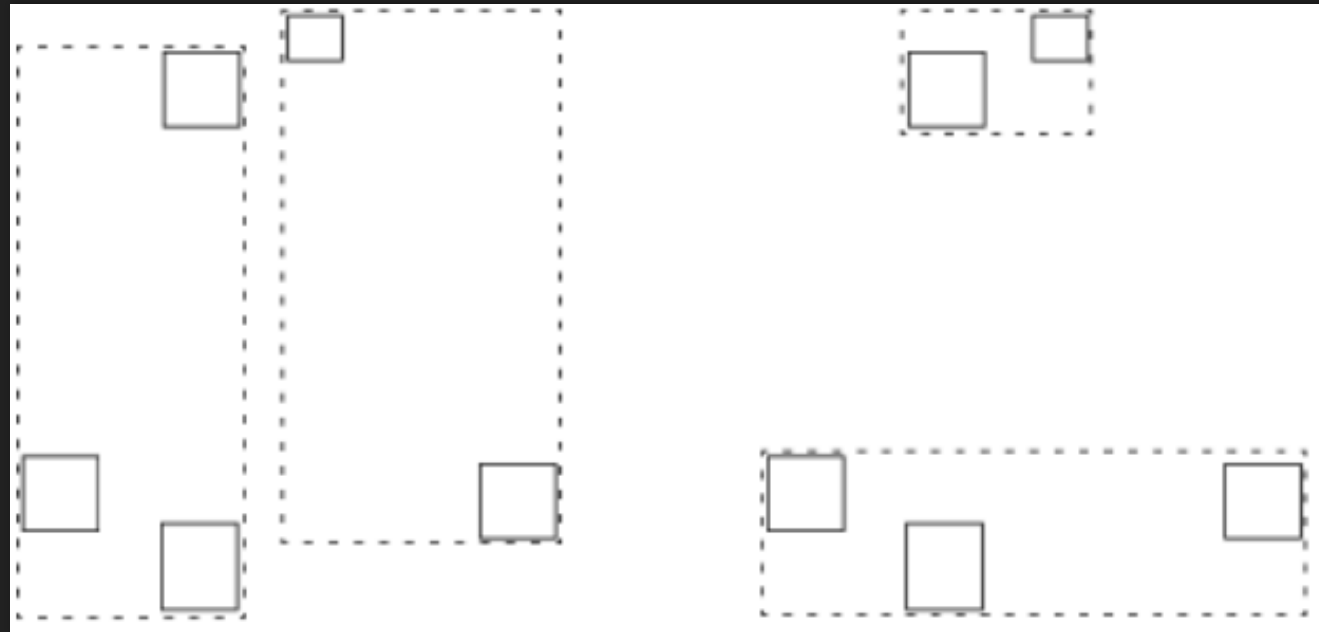


Árvores R - Inserção

- A estrutura da árvore-R depende da ordem em que os retângulos são inseridos e 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 sub-árvore 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 se propagar até a raiz

Árvores R - Inserção

- Problema
 - A operação de *split* pode ser feita de forma ineficiente
 - Comparação de divisões eficientes e ineficientes de cinco itens em dois grupos



Árvores R - Busca

- A busca na árvore R é semelhante à busca na árvore B
 - Nossa busca irá descer recursivamente do nó raiz até os nós folhas, adicionando todas as entradas de folha que correspondem a critérios no conjunto de resultados
- Problema
 - Uma grande quantidade de nós pode ter que ser examinada, pois um retângulo pode estar contido nos retângulos envoltentes de muitos nós
 - Mas o seu registro está contido em apenas um nó folha

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
 - A exclusão de uma entrada de um nó pode exigir a atualização dos retângulos delimitadores dos nós pai
 - No entanto, quando um nó está subcheio, ele não será balanceado com seus vizinhos.
 - Em vez disso, o nó será dissolvido e todos os seus filhos (que podem ser subárvores, não apenas objetos folha) serão reinseridos
 - A 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

Quadtree

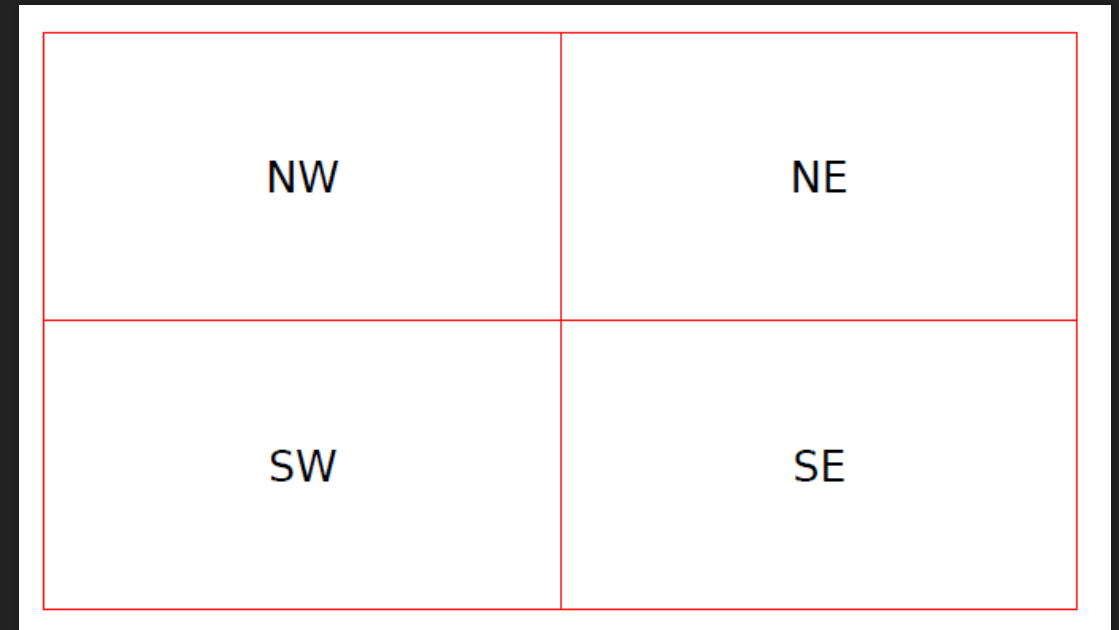
- A quadtree é uma estrutura de dados utilizada para representar dados espaciais
 - Técnica simples
 - Baseado no conceito de árvore binária
 - Generalização multidimensional (2D ou 3D)
 - Acelera o acesso a dados num espaço 2D ou 3D

Quadtree

- É uma árvore em que cada nó representa um quadrante da região espacial que o nó pai representa
 - Cada nó possui 4 (2D), 8 (3D) ou nenhum filho
 - Uma quadtree com 8 filhos é chamada octotree

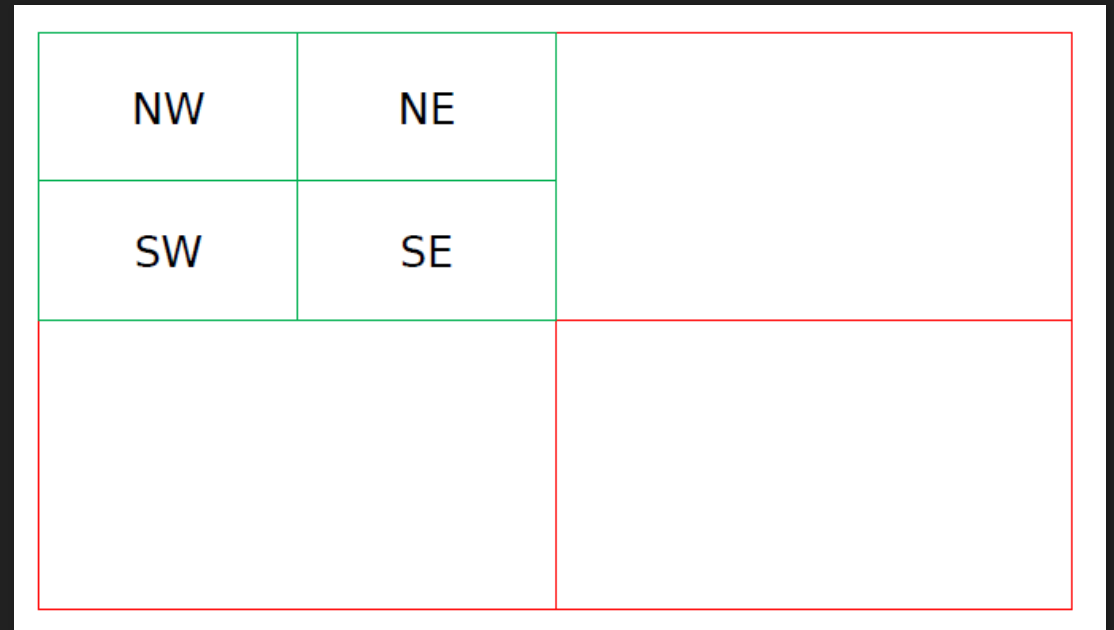
Quadtree

- O espaço é dividido em 4 quadrantes (direções)
- Os quadrantes são nomeados: Noroeste, Nordeste, Sudeste e Sudoeste



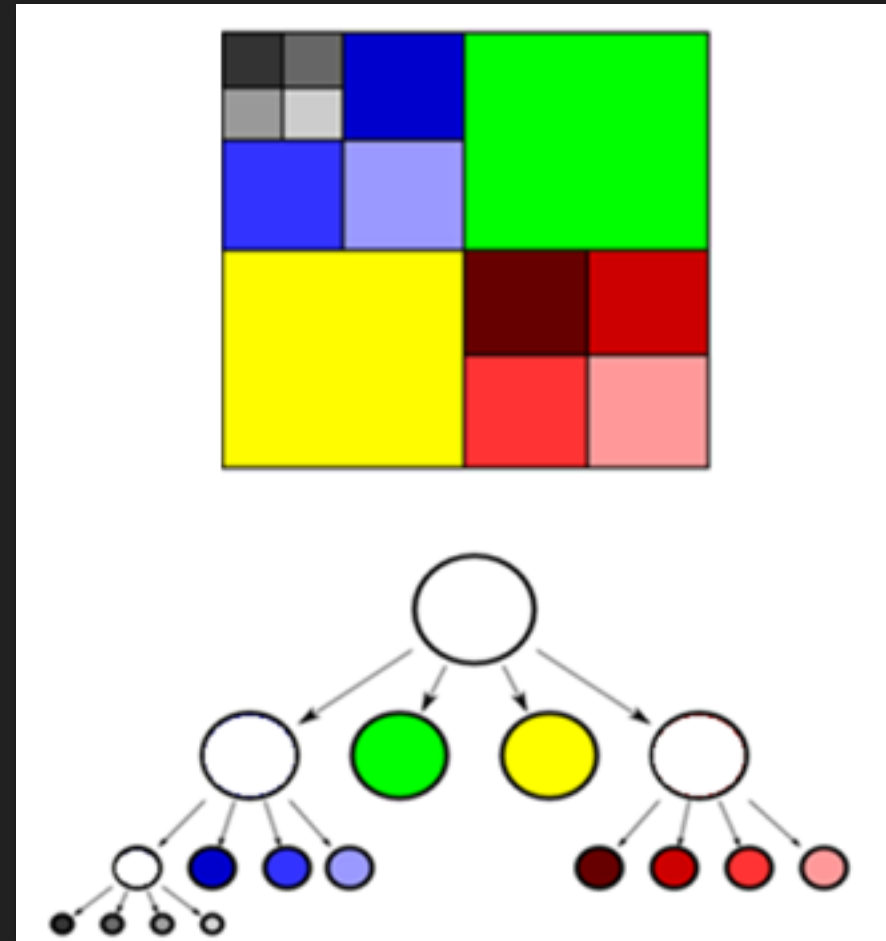
Quadtree

- Cada quadrante é dividido em quatro subquadrantes, e assim por diante, até que se chegue a uma profundidade pré-determinada ou até que cada quadrante contenha um número mínimo de pontos.



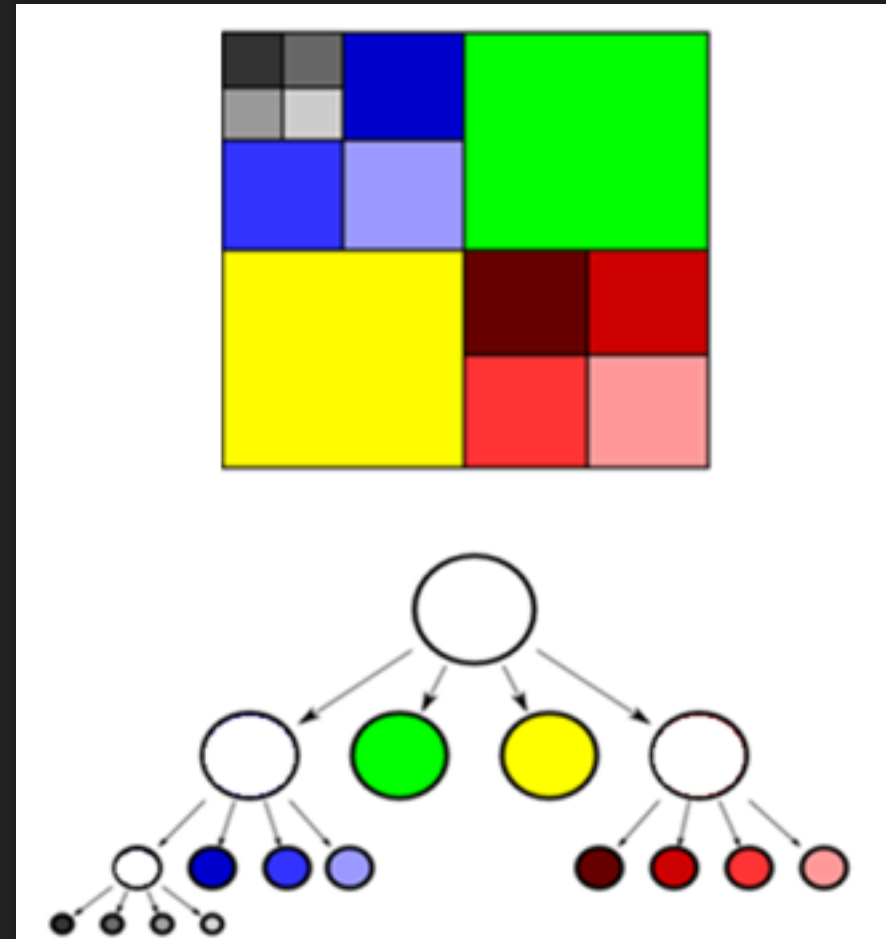
Quadtree

- A construção da quadtree começa com um nó raiz que representa toda a região espacial.
- Esse nó é dividido em quatro quadrantes, que se tornam os filhos do nó raiz.



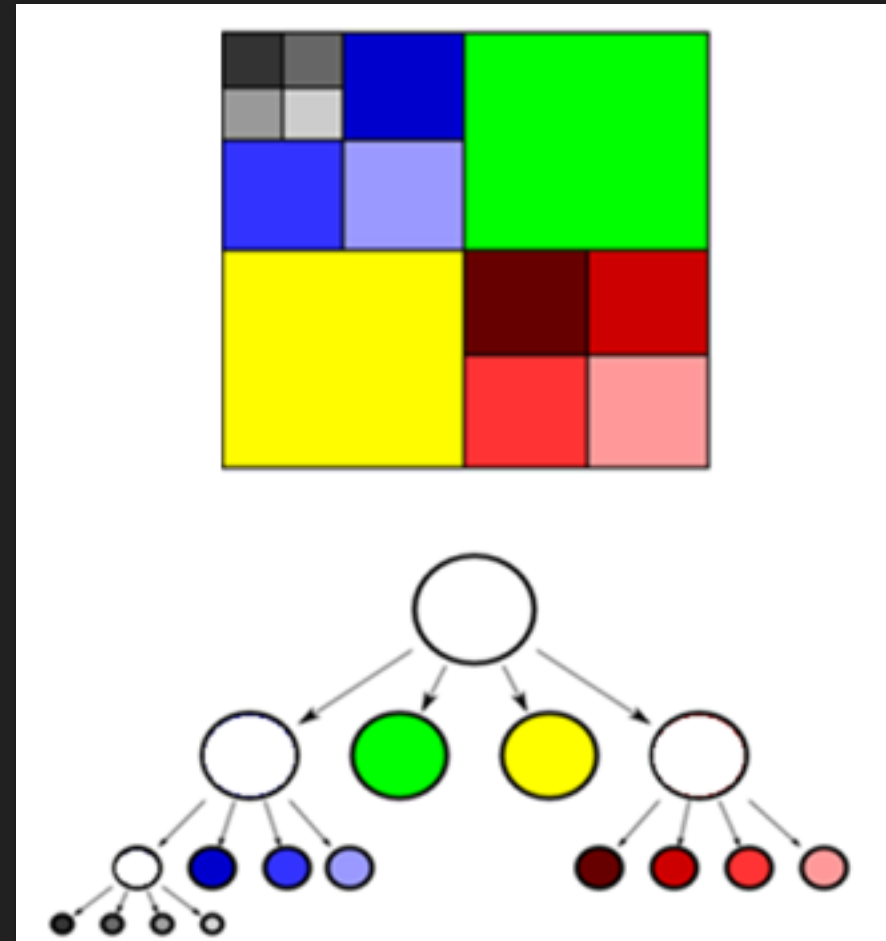
Quadtree

- Cada nó folha da quadtree representa uma região espacial que contém pontos
 - Os pontos podem ser armazenados diretamente no nó folha
 - Podemos usar uma estrutura de dados associada a ele, como uma lista encadeada, para armazenar os dados



Quadtree

- Dessa forma, a quadtree permite acesso rápido aos pontos que estão dentro de uma região espacial específica.
 - Consulta é simples em quadtree
 - Um único caminho é percorrido da raiz até a folha
 - Em cada nível, é escolhido um dos quadrantes que contém o ponto da consulta

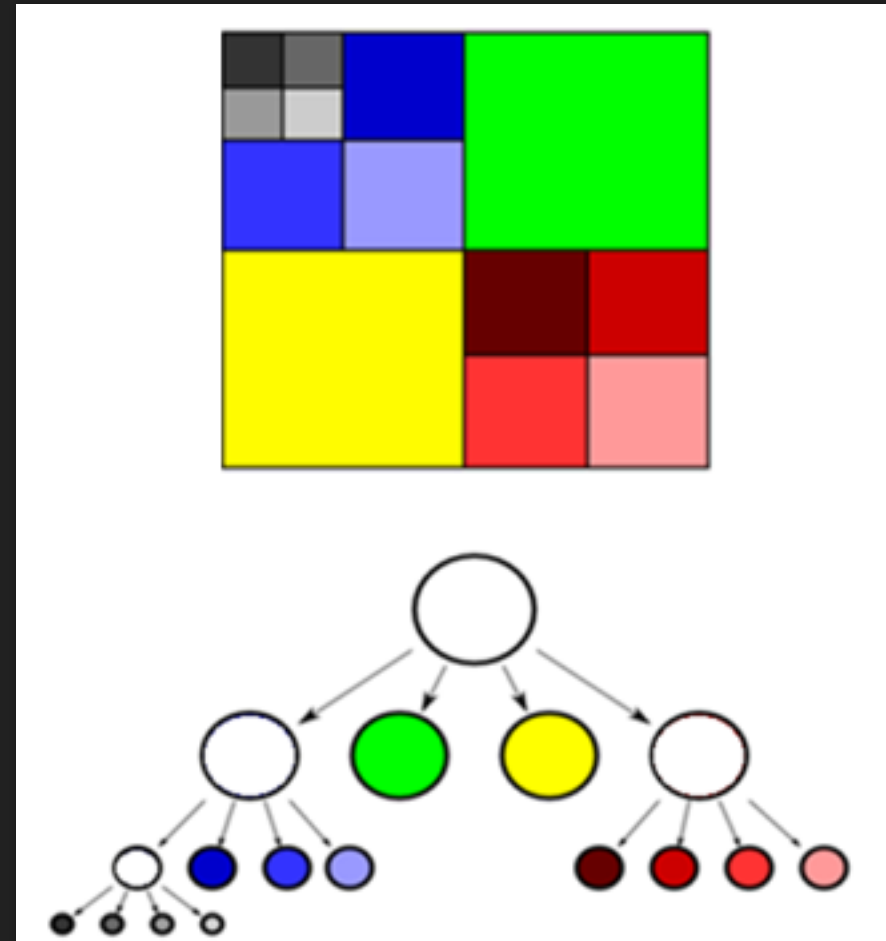


Quadtree

- Divisões do espaço devem ser realizadas segundo determinadas características ou regras
- Alguns tipos de quadtree
 - Region Quadtree
 - Point Quadtree
 - PointRegion Quadtree (PR Quadtree)
 - Quadtree para polígonos

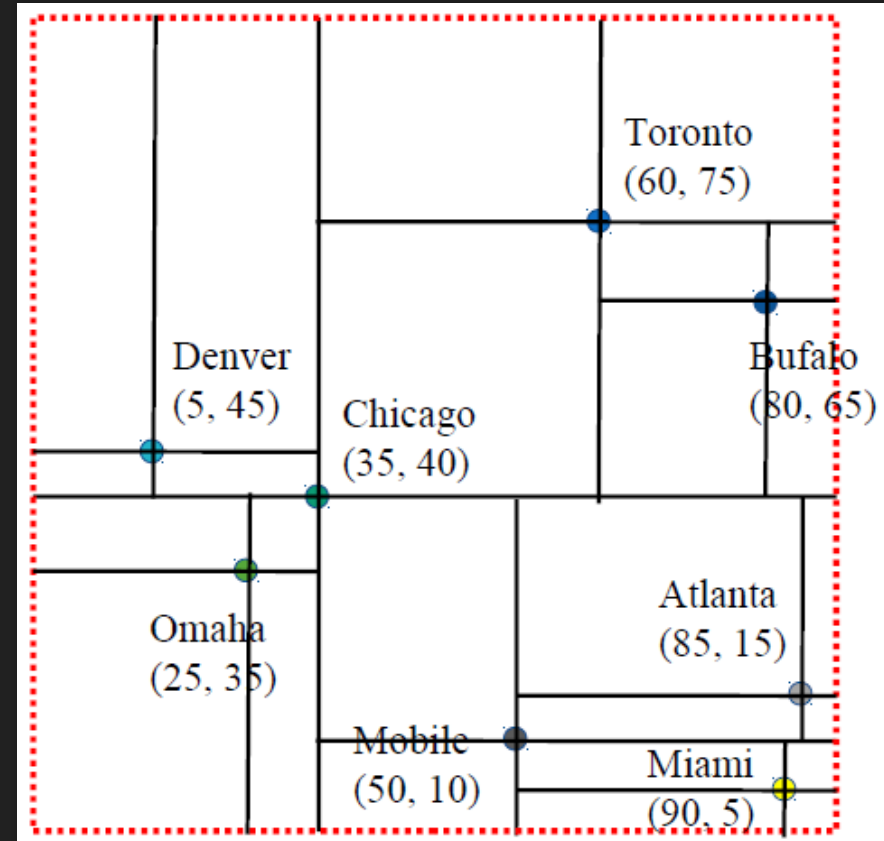
Region Quadtree

- Representa uma partição do espaço em duas dimensões, decompondo a região em quatro quadrantes iguais, subquadrantes e assim por diante.
- Cada nó folha contém dados correspondentes a uma sub-região específica.



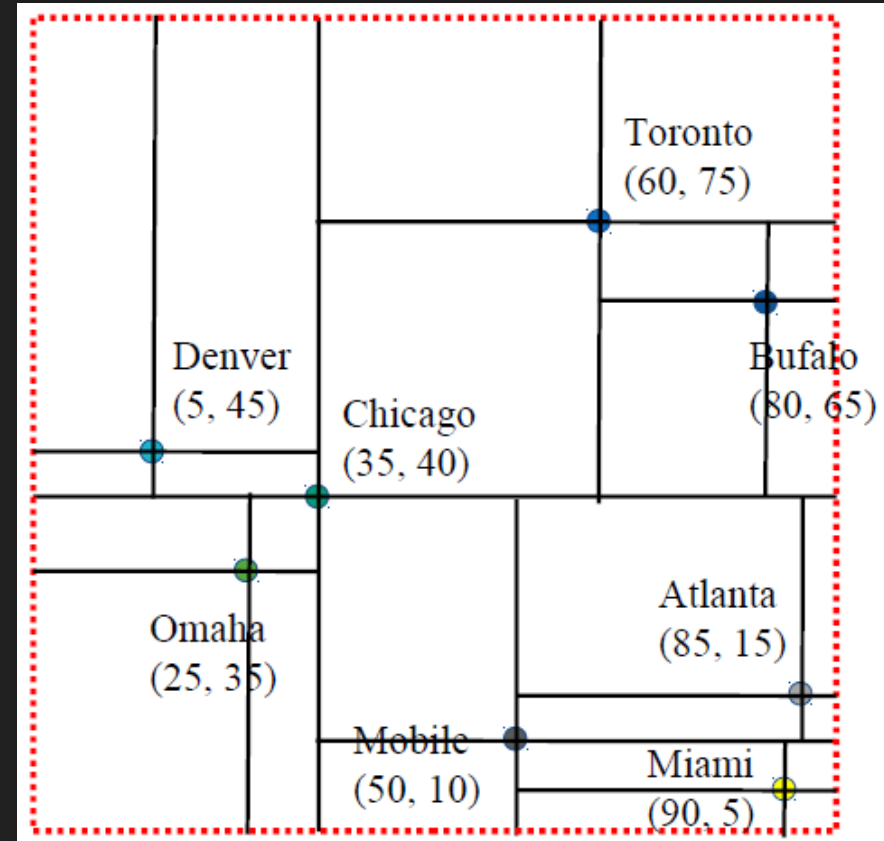
Point Quadtree

- É uma adaptação de uma árvore binária
- Usada para representar dados de pontos bidimensionais
- Compartilha todas as características da quadtree, mas é uma verdadeira árvore, pois o centro de uma subdivisão está sempre em um ponto



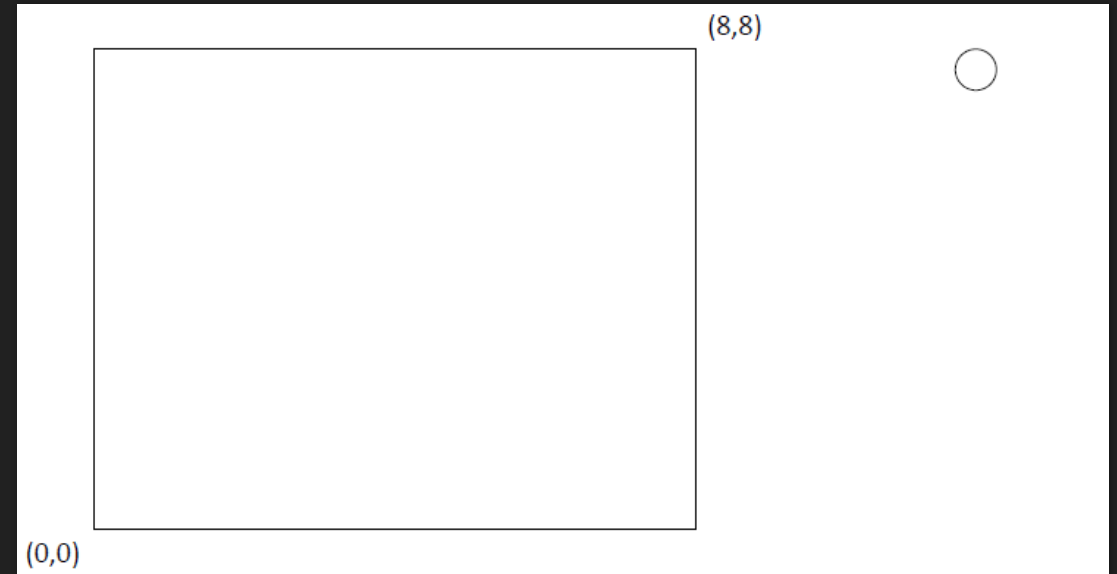
Point Quadtree

- É muito eficiente na comparação de pontos de dados bidimensionais e ordenados, geralmente operando em tempo $O(\log n)$



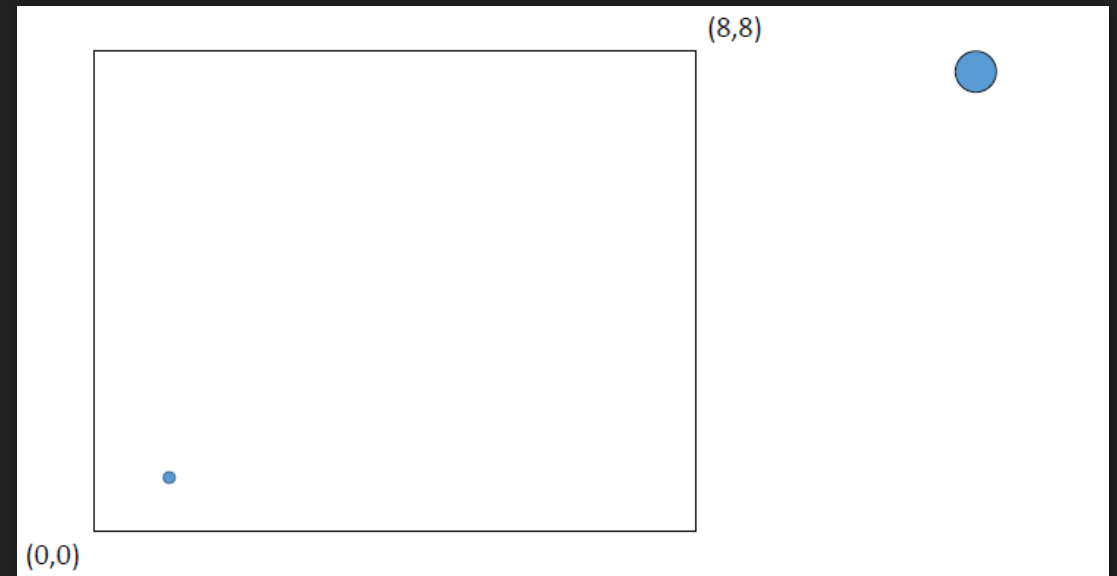
Point Quadtree

- Ponto sobre a linha de um quadrante
- Regra adotada
 - Limites inferior e esquerdo de cada bloco são fechados: NÃO pertence se estiver sobre a mesma linha
 - Limites superior e direito são abertos: pertence se estiver sobre a mesma linha



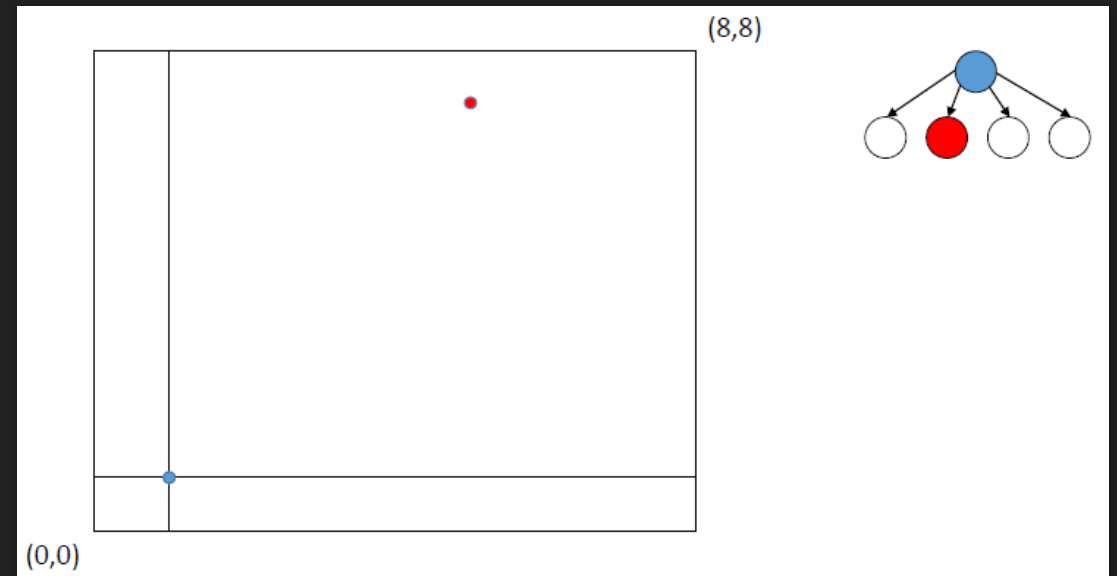
Point Quadtree

- Inserir coordenada (1,1)



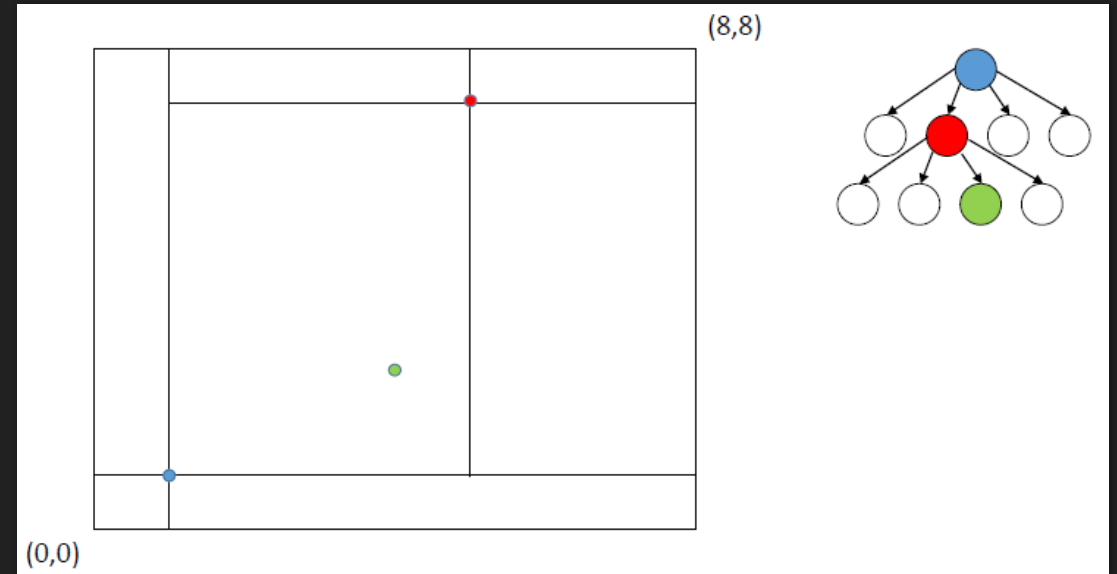
Point Quadtree

- Insere coordenada (1,1)
- Insere coordenada (5,7)



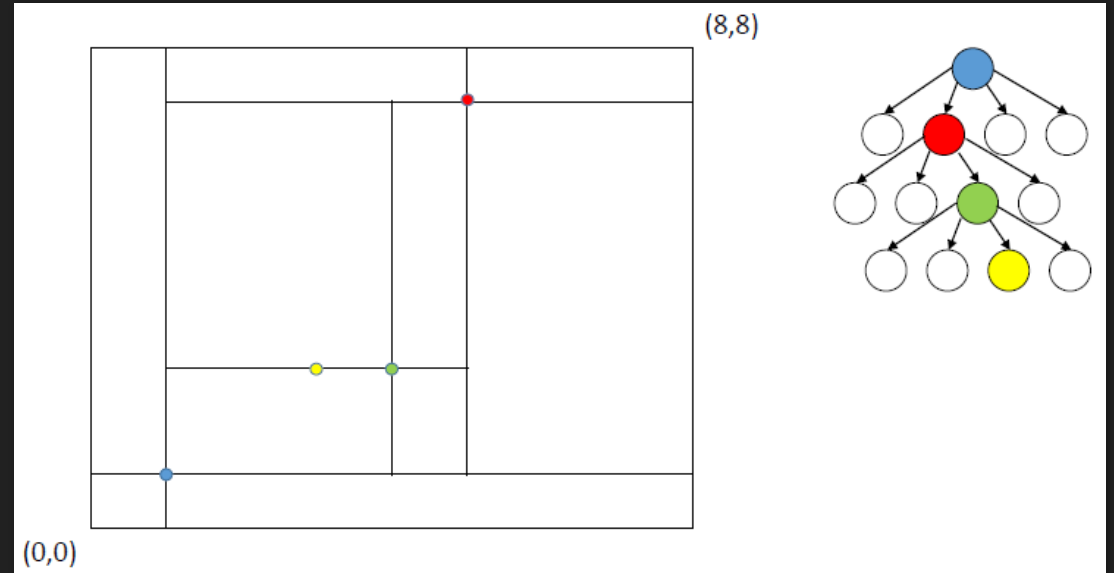
Point Quadtree

- Inseire coordenada (1,1)
- Inseire coordenada (5,7)
- Inseire coordenada (4,3)



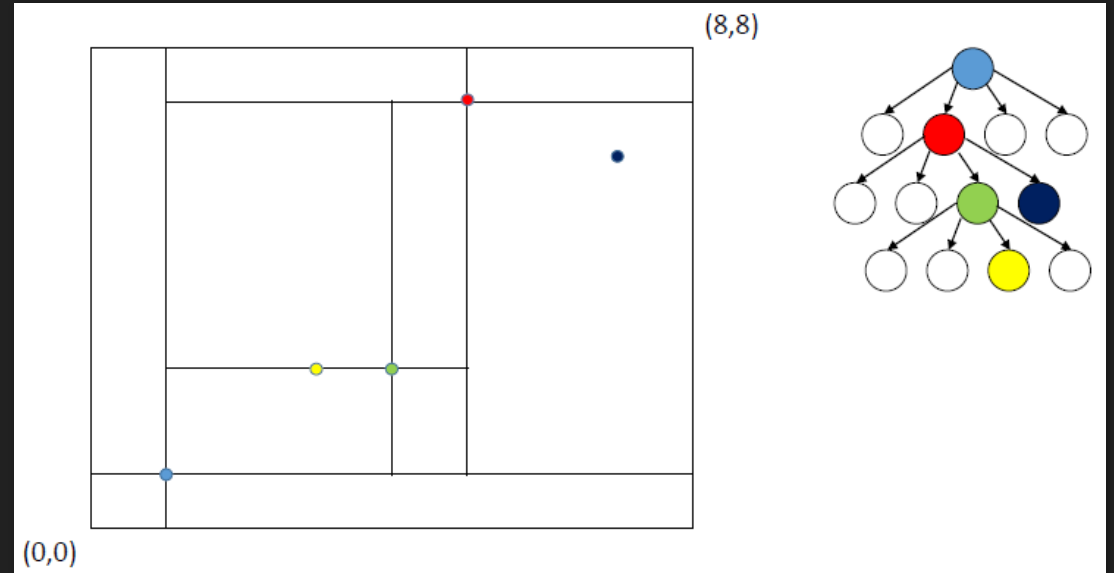
Point Quadtree

- Inserir coordenada (1,1)
- Inserir coordenada (5,7)
- Inserir coordenada (4,3)
- Inserir coordenada (3,3)



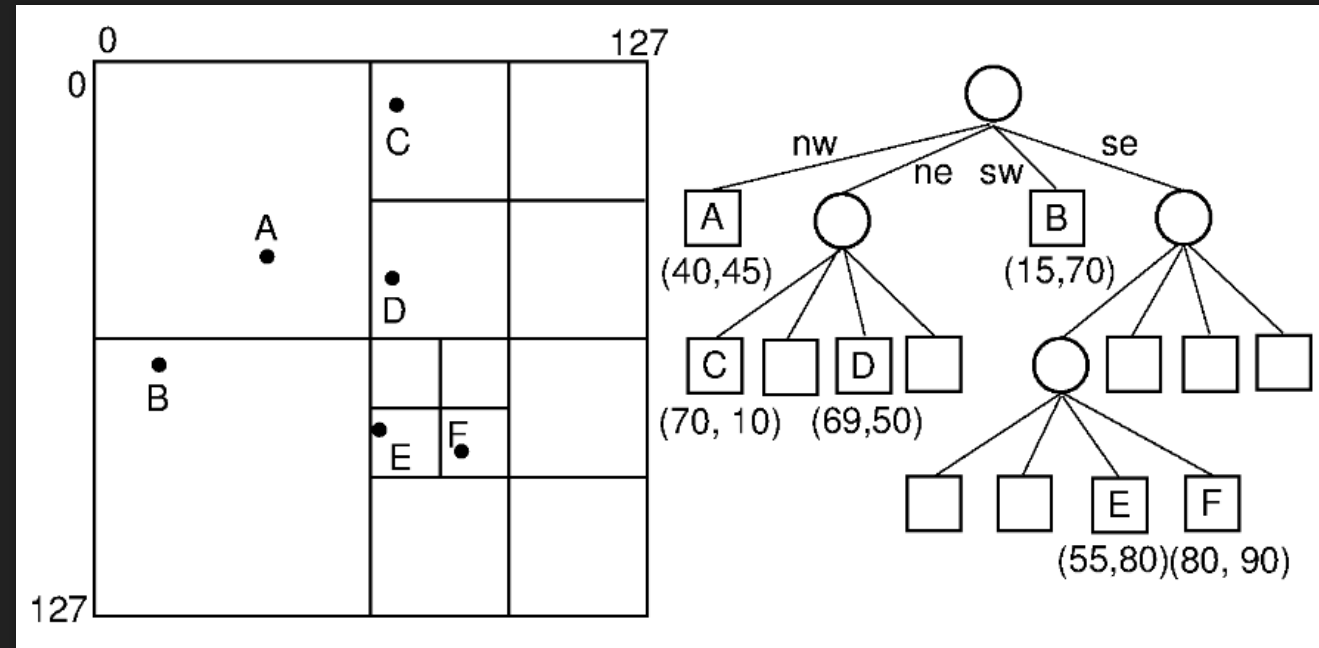
Point Quadtree

- Inere coordenada (1,1)
- Inere coordenada (5,7)
- Inere coordenada (4,3)
- Inere coordenada (3,3)
- Inere coordenada (7,6)



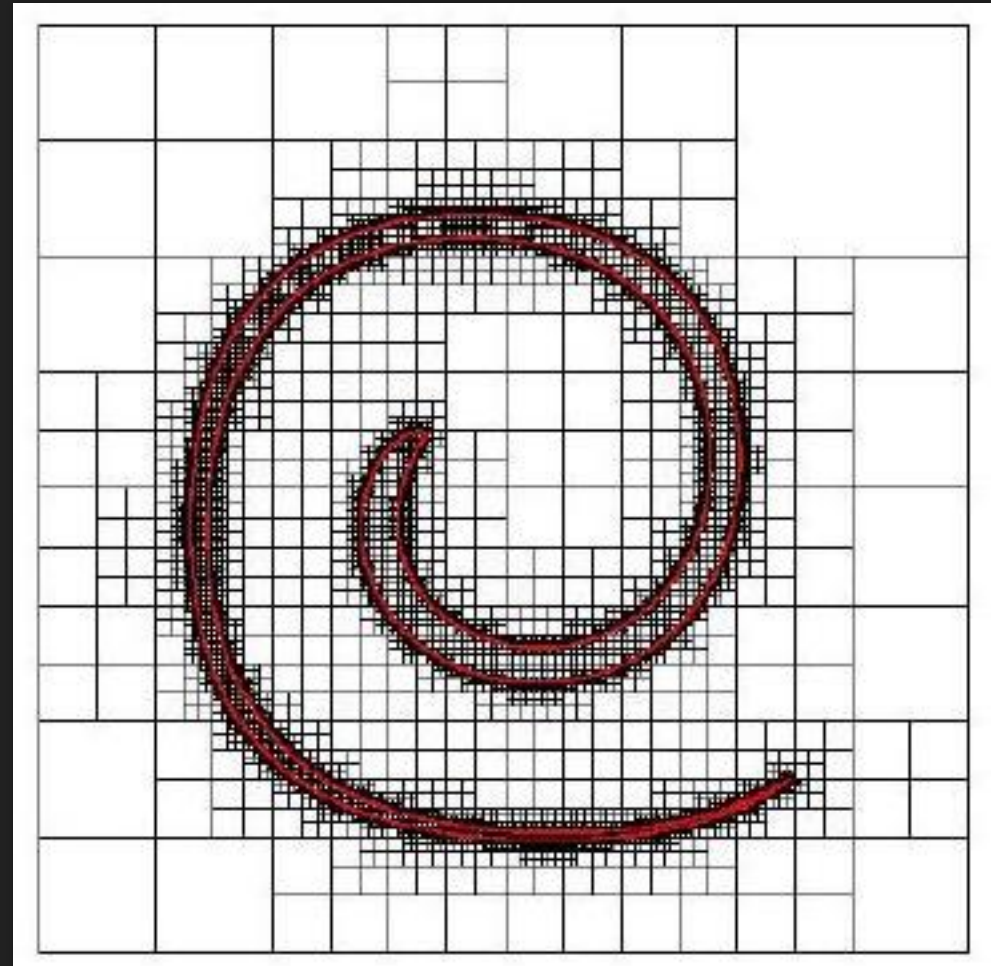
PointRegion Quadtree (PR Quadtree)

- Semelhante a Region Quadtree
- Representa uma coleção de pontos em duas dimensões, decompondo a região em quadrantes iguais, até que nenhum nó folha contenha mais de um único ponto

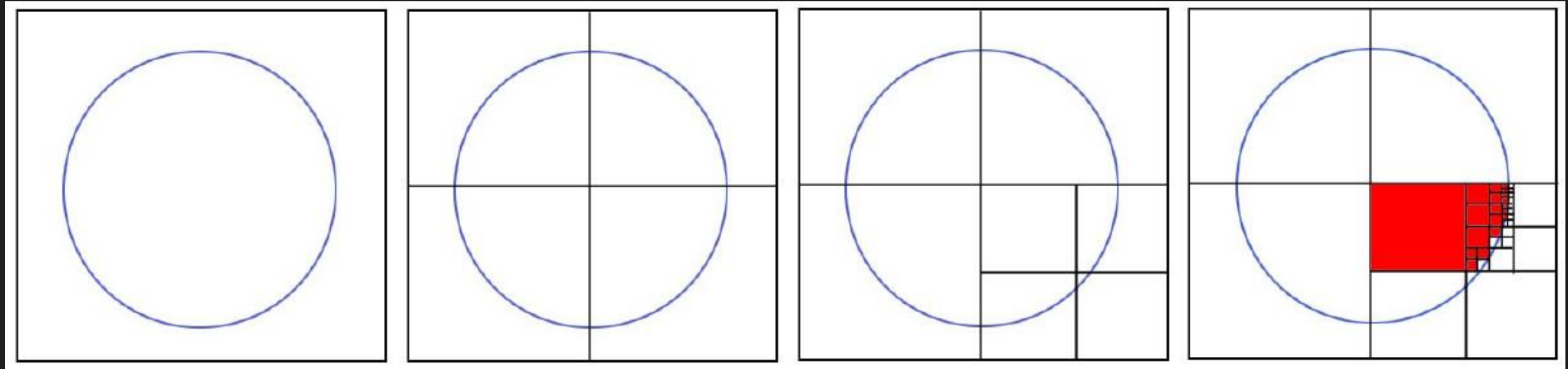


Quadtree para polígonos

- Uma QuadTree pode também ser utilizada para preencher polígonos
 - Subdividir o espaço
 - Se o quadrante for homogêneo ou se o nível de detalhe foi alcançado, para a divisão
 - Caso contrário, continue dividindo

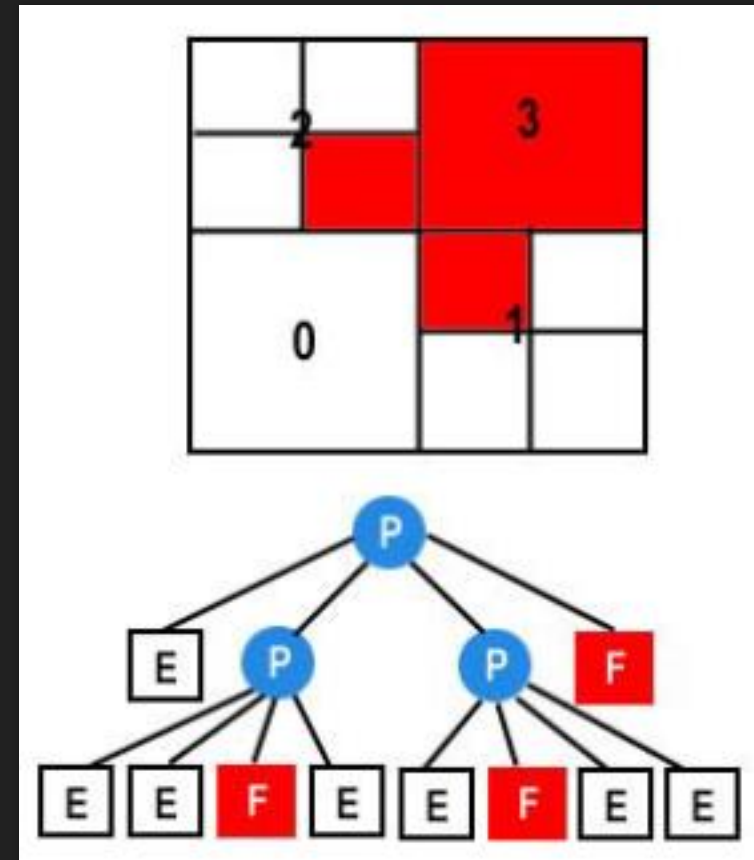


Quadtree para polígonos



Quadtree para polígonos

- As divisões são representadas na árvore como nós caso sejam quadrantes heterogêneos e como folhas caso sejam quadrantes homogêneos



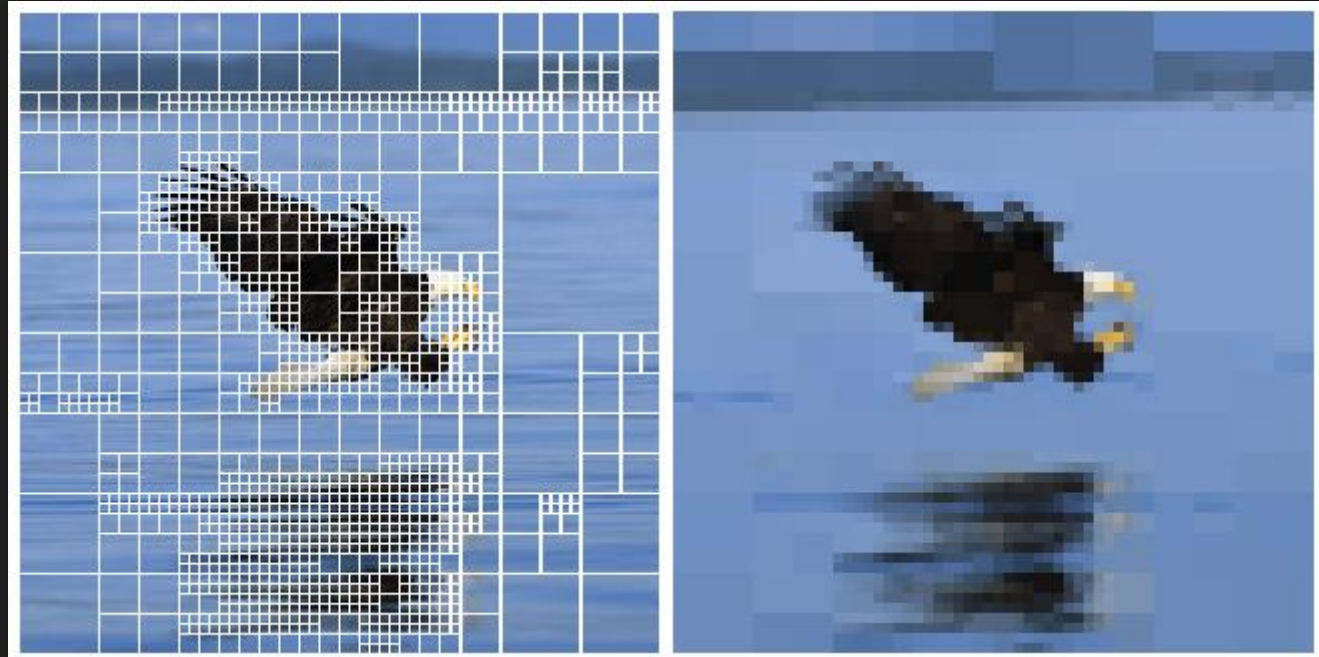
Quadtree

- Aplicações

- Análise de imagens satélite e o processamento de dados de sensores em sistemas de monitoramento
- Processamento de imagens, algoritmos de compressão de imagens
- Medicina ecografias, ajudando a identificar tumores pela cor na imagem
- Videochamadas ajuda na melhora da velocidade da transmissão do vídeo ao transmitir só o que foi alterado na imagem
- Jogos
 - Detectar se um tiro de um personagem atingiu o adversário
 - Detecção de colisões.

Quadtree

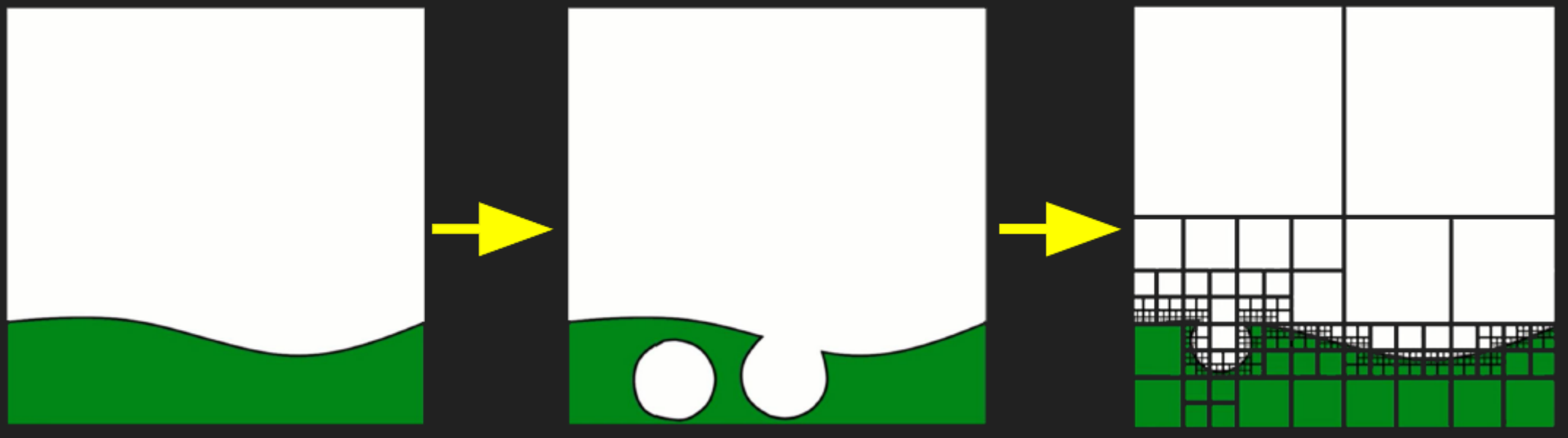
- Aplicações
 - Decomposição de uma imagem em regiões de cores homogêneas
 - Útil em aplicações como compactação ou segmentação de imagens



Quadtree

- Aplicações
 - Se eu der um tiro em determinada direção, quais objetos do ambiente vou acertar?
 - Qual objeto vou acertar primeiro?
 - Qual dano vou causar no cenário?

Quadtree



Quadtree

- Vantagens
 - Muito útil em compactação de imagens
 - Pode ser utilizada no processo de rotacionar imagens
 - Tem estrutura mais enxuta e robusta que árvores binárias (tamanho menor, menos nodos a se percorrer para chegar ao que se procura)
 - Inserções constantes não afetam a performance da quadtree (se não precisar de rebalanceamento)

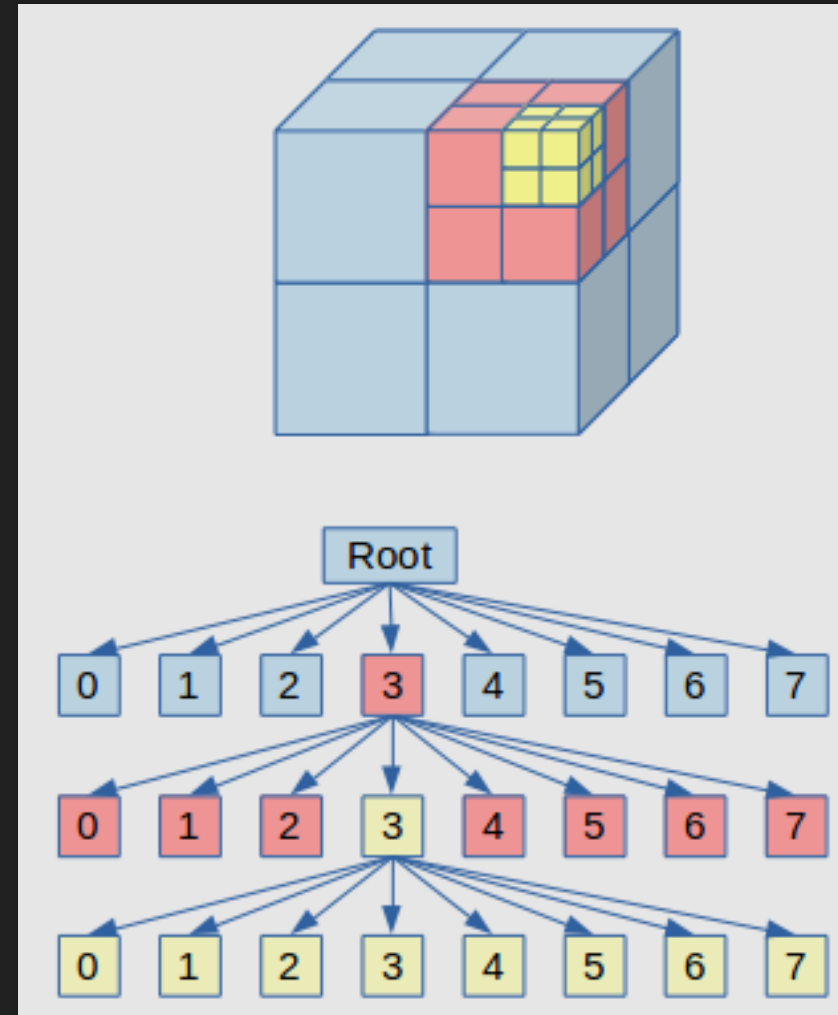
Quadtree

- Desvantagens

- Se a imagem tiver muitas cores diferentes, a árvore pode ficar tão complexa que a imagem compactada ficará maior que a original
- Imagens complexas geram uso alto de CPU para geração de uma quadtree
- Somente imagens em duas dimensões (2D) podem ser indexadas com quadtrees (a árvore R, por exemplo, trabalha com 4 dimensões)
- Desperdício de memória, vários ponteiros podem não ser usados

Octree

- A mesma ideia da QuadTree para polígonos, mas no 3D
 - Cada região é dividida em 8 novas regiões



Octree

- Aplicações
 - Nível de detalhes de renderização
 - Indexação espacial
 - Detecção de colisão eficiente em três dimensões
 - Seleção de *frustum*
 - Estimativa de conjuntos

