

Árvore-B*

Anderson Canale Garcia

Material gentilmente cedido por:
Cristina D. Aguiar

Árvore-B*

- Proposta por Knuth em 1973
 - variação de árvore-B
 - Característica
 - cada nó contém, no mínimo, $2/3$ do número máximo de chaves
 - Posterga o *split*
 - estende a noção de redistribuição durante a inserção para incluir novas regras para o particionamento de nós
-

Redistribuição durante Inserção

- Funcionalidade
 - permite melhorar a taxa de utilização do espaço alocado para a árvore
- *split*
 - divide **uma** página com *overflow* em **duas** páginas semivazias
- redistribuição
 - chave que causou *overflow* (e outras chaves) é colocada em outra página
 - **não** requer a criação de novas páginas

X

Redistribuição durante Inserção

- Opção interessante
 - a rotina de redistribuição já está codificada para prover suporte à remoção
 - a redistribuição evita, ou pelo menos adia, a criação de novas páginas
 - tende a tornar a árvore-B mais eficiente em termos de utilização do espaço em disco
 - garante um melhor desempenho na busca, desde que um número menor de nós pode reduzir a altura da árvore, por exemplo
-

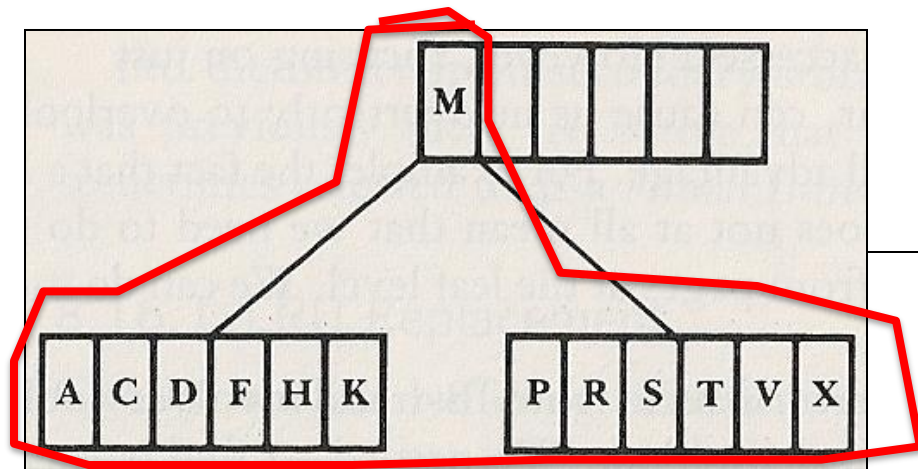
Split x Redistribuição

- Somente *split* na inserção
 - no pior caso, a utilização do espaço é de cerca de 50%
 - em média, para árvores grandes, o índice de ocupação é de ~69%
 - Com redistribuição na inserção
 - em média, para árvores grandes, o índice de ocupação é de ~86%
-

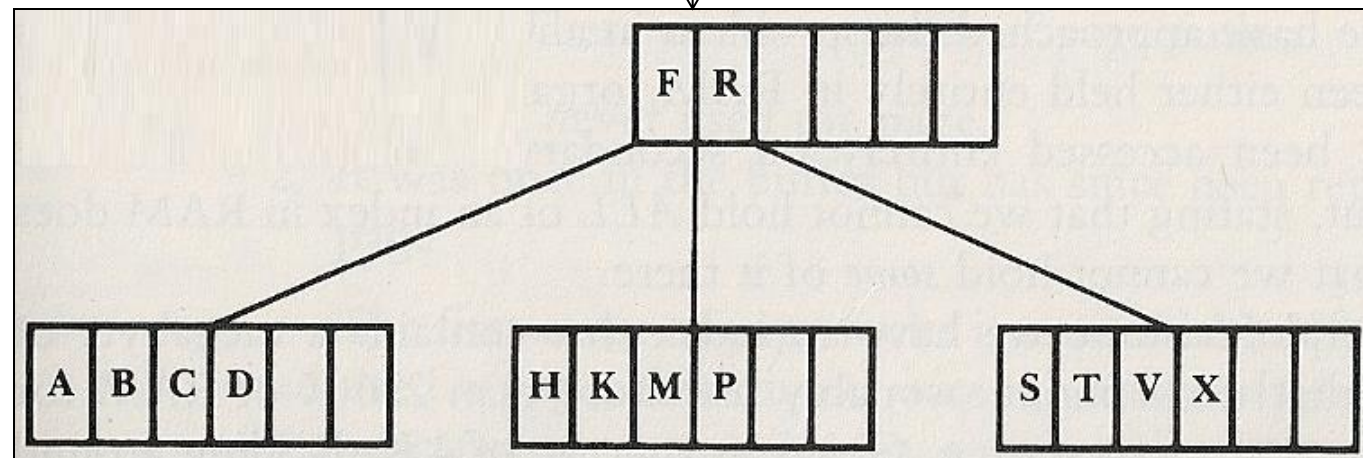
Árvore-B*

- Geração
 - utiliza uma variação do processo de subdivisão
 - Características
 - a subdivisão é adiada até que duas páginas irmãs estejam cheias
 - na sequência, a divisão do conteúdo das duas páginas em três páginas (*two-to-three split*) é realizada
-

Split 2-to-3: Exemplo



split 2-3 após
inserção de B



Definição Formal

- Propriedades de uma Árvore-B*
 - cada página possui um máximo de m descendentes
 - cada página, exceto a raiz e as folhas, possui no mínimo $(2m-1)/3$ descendentes → taxa de ocupação
 - a raiz possui pelo menos 2 descendentes, a menos que seja um nó folha
 - todas as folhas aparecem no mesmo nível
 - uma página interna com k descendentes contém $k-1$ chaves
 - uma folha possui no mínimo $\lfloor (2m-1)/3 \rfloor$ chaves e no máximo $m - 1$ chaves → taxa de ocupação
-

Observações

- Mudança na taxa de ocupação
 - afeta as rotinas de remoção e redistribuição
 - Particionamento da raiz
 - problema
 - raiz não possui nó irmão
 - soluções
 - dividir a raiz usando a divisão convencional (*1-to-2 split*); ou
 - permitir que a raiz seja maior
-

Exercício

- Considere as seguintes chaves

15	13	2	8	7	18	23	25	38	43	57	9	12	17	0	5	6	4	3
----	----	---	---	---	----	----	----	----	----	----	---	----	----	---	---	---	---	---

- Criar uma árvore-B*
 - Ordem da árvore: 5
 - Obedecer ordem de entrada dos dados
 - Considere implementadas as rotinas:
 - *Split*, concatenação e redistribuição
 - Ilustre graficamente o índice gerado
 - Indique o RRN de cada página e o endereço do nó raiz no registro de cabeçalho
-