

ginas folha em direção à página raiz de uma **árvore 2-3** (ou árvore B de ordem $m = 1$), obtido por Eisenbarth, Ziviani, Gonnet, Mehlhorn e Wood (1982, p. 159), esses autores propuseram a seguinte conjectura: a altura esperada de uma árvore 2-3 **randômica** (vide definição de árvore de pesquisa randômica na Seção 5.3) com N chaves é:

$$\bar{h}(N) \approx \log_{7/3}(N + 1).$$

Outras medidas de complexidade relevantes para árvores B randômicas são:

1. A utilização de memória é cerca de $\ln 2$ para o algoritmo original proposto por Bayer e McCreight (1972). Isso significa que as páginas têm uma ocupação de aproximadamente 69% da área reservada após N inserções randômicas em uma árvore B inicialmente vazia.
2. No momento da inserção, a operação mais cara é a partição da página quando ela passa a ter mais do que $2m$ chaves, desde que a operação envolva a criação de uma nova página, o rearranjo das chaves e a inserção da chave do meio na página pai localizada no nível acima. Uma medida de complexidade de interesse é $Pr\{j \text{ partições}\}$, que corresponde à probabilidade de que j partições ocorram durante a N -ésima inserção randômica. No caso da árvore 2-3:

$$Pr\{0 \text{ partições}\} = \frac{4}{7},$$

$$Pr\{1 \text{ ou mais partições}\} = \frac{3}{7}.$$

No caso da árvore B de ordem m :

$$Pr\{0 \text{ partições}\} = 1 - \frac{1}{(2 \ln 2)m} + O(m^{-2}),$$

$$Pr\{1 \text{ ou mais partições}\} = \frac{1}{(2 \ln 2)m} + O(m^{-2}).$$

No caso de uma árvore B de ordem $m = 70$, $Pr\{1 \text{ ou mais partições}\} \approx 0,01$. Em outras palavras, em 99% das vezes nada acontece em termos de partições durante uma inserção.

3. Considere o acesso concorrente em árvores B. Bayer e Schkolnick (1977) propuseram a técnica de aplicar um travamento na *página segura mais profunda* (Psm) no caminho de inserção. De acordo com o que foi mostrado na Seção 6.3.3, uma página é **segura** se ela contiver menos do que $2m$ chaves. Uma página segura é a mais profunda de um caminho de inserção se não existir outra página segura abaixo dela.

Já que o travamento da página impede o acesso de outros processos, é interessante saber qual é a probabilidade de que a página segura mais profunda esteja no primeiro nível. Essas medidas estão relacionadas às do item anterior. No caso da árvore 2-3:

$$Pr\{\text{Psm esteja no } 1^{\circ} \text{ nível}\} = \frac{4}{7},$$

$$Pr\{\text{Psm esteja acima do } 1^{\circ} \text{ nível}\} = \frac{3}{7}.$$

No caso da árvore B de ordem m :

$$Pr\{\text{Psm esteja no } 1^{\circ} \text{ nível}\} = 1 - \frac{1}{(2 \ln 2)m} + O(m^{-2}),$$

$$Pr\{\text{Psm esteja acima do } 1^{\circ} \text{ nível}\} = \frac{3}{7} = \frac{1}{(2 \ln 2)m} + O(m^{-2}).$$

Novamente, no caso de uma árvore B de ordem $m = 70$, em 99% das vezes a Psm está localizada em uma página folha, o que permite um alto grau de concorrência para processos modificadores. Esses resultados mostram que soluções muito complicadas para permitir o uso de concorrência de operações em árvores B não trazem grandes benefícios porque, na maioria das vezes, o travamento ocorrerá em páginas folha, o que permite alto grau de concorrência mesmo para os protocolos mais simples.

Mais detalhes sobre os resultados analíticos apresentados acima podem ser obtidos em Eisenbarth *et al.* (1982).

Observações Finais

Existem inúmeras variações sobre o algoritmo original da árvore B. Uma delas é a árvore B^* , tratada na Seção 6.3.2.

Outra importante modificação é a **técnica de transbordamento** (ou técnica de **overflow**) proposta por Bayer e McCreight (1972) e Knuth (1973). A idéia é a seguinte: assuma que um registro tenha de ser inserido em uma página cheia que contenha $2m$ registros. Em vez de particioná-la, olhamos primeiro para a página irmã à direita. Se a página irmã possui menos do que $2m$ registros, um simples rearranjo de chaves torna a partição desnecessária. Se a página à direita também estiver cheia ou não existir, olhamos para a página irmã à esquerda. Se ambas estiverem cheias, então a partição terá de ser realizada. O efeito dessa modificação é o de produzir uma árvore com melhor utilização de memória e uma altura esperada menor. Essa alteração produz uma utilização de memória de cerca de 83% para uma árvore B randômica.