CSE221

Lecture 16:
Multiway Search Trees

Acknowledgment: The content of this file is based on the slides of the textbook as well as the slides provided in former lectures at UNIST.

UNIST
ULSAN NATIONAL INSTITUTE OF SCIENCE AND TECHNOLOGY

Outline

- m-way search trees
- B-trees

ULSAN NATIONAL INSTITUTE OF
SCIENCE AND TECHNOLOGY

Memory Hierarchy

- Von Neumann model limitation
  - Memory is bottleneck
- Memory hierarchy

  CPU ↔ Register ← cache ← memory ← disk

  ← faster but smaller data fetch unit
- Performance implications
  - Performance is closely related to reducing the number of accesses to slow memory
  - Exercising data stored in faster memory is crucial

Improving Performance in Trees

- Number of memory accesses is closely tied to the <u>*height*</u> of the search tree
  - Balanced binary search tree has $\log_2 n$ height
  - Nodes accessed on the path from the root to a leaf are often located in slower memory

Can we break $\log_2 n$ barrier?

- Node data is contiguous and has spatial locality

Can we structure node

to use faster memory more efficiently?

UNIST
ULSAN NATIONAL INSTITUTE OF SCIENCE AND TECHNOLOGY

m-way Search Trees

- Root has **at least two** & **at most m** subtrees
- Node structure (n<m)

Contiguous space: spatial locality

n	A_0	E_1	A_1	E_2	…	E_n	A_n

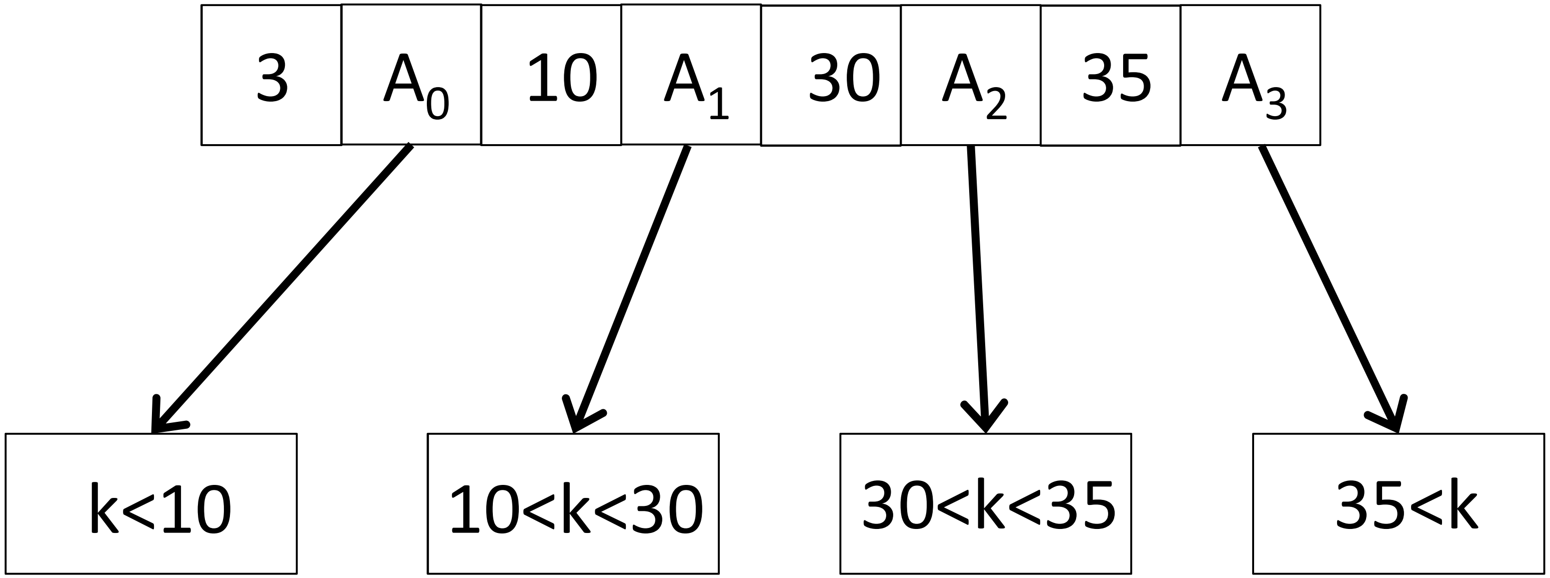
\# of elements　　pointer to subtree　　element

- E_i.key < E_{i+1}.key
- E_i.key < all keys in A_i < E_{i+1}.key

Tree is ordered!

- Subtrees A_i are also m-way search trees (recursive definition)

Example: 4-way Search Tree



UNIST
ULSAN NATIONAL INSTITUTE OF SCIENCE AND TECHNOLOGY

m-way Search Trees

- Maximum # of nodes: when all internal nodes are m-nodes (having m subtrees)
  - A tree of degree m and height h

$$1 + m + m^2 + m^3 + \ldots + m^h = (m^{h+1} - 1)/(m - 1)$$

- Each m-node has up to m – 1 elements
- Maximum # of elements = $m^{h+1} - 1$

Searching

```
// Search m-way search tree for an element with key x
E0.key=-MAXKEY;
for(p=root; p; p=Ai)
{
        Let p is a node (n, A0, E1, A1, .. , En, An)
        En+1.key = MAXKEY
        Determine i such that Ei.key <= x < Ei+1.key;
        if(x == Ei.key) return Ei; // x is found
}
// x is not found
return NULL;
```


UNIST
ULSAN NATIONAL INSTITUTE OF
SCIENCE AND TECHNOLOGY

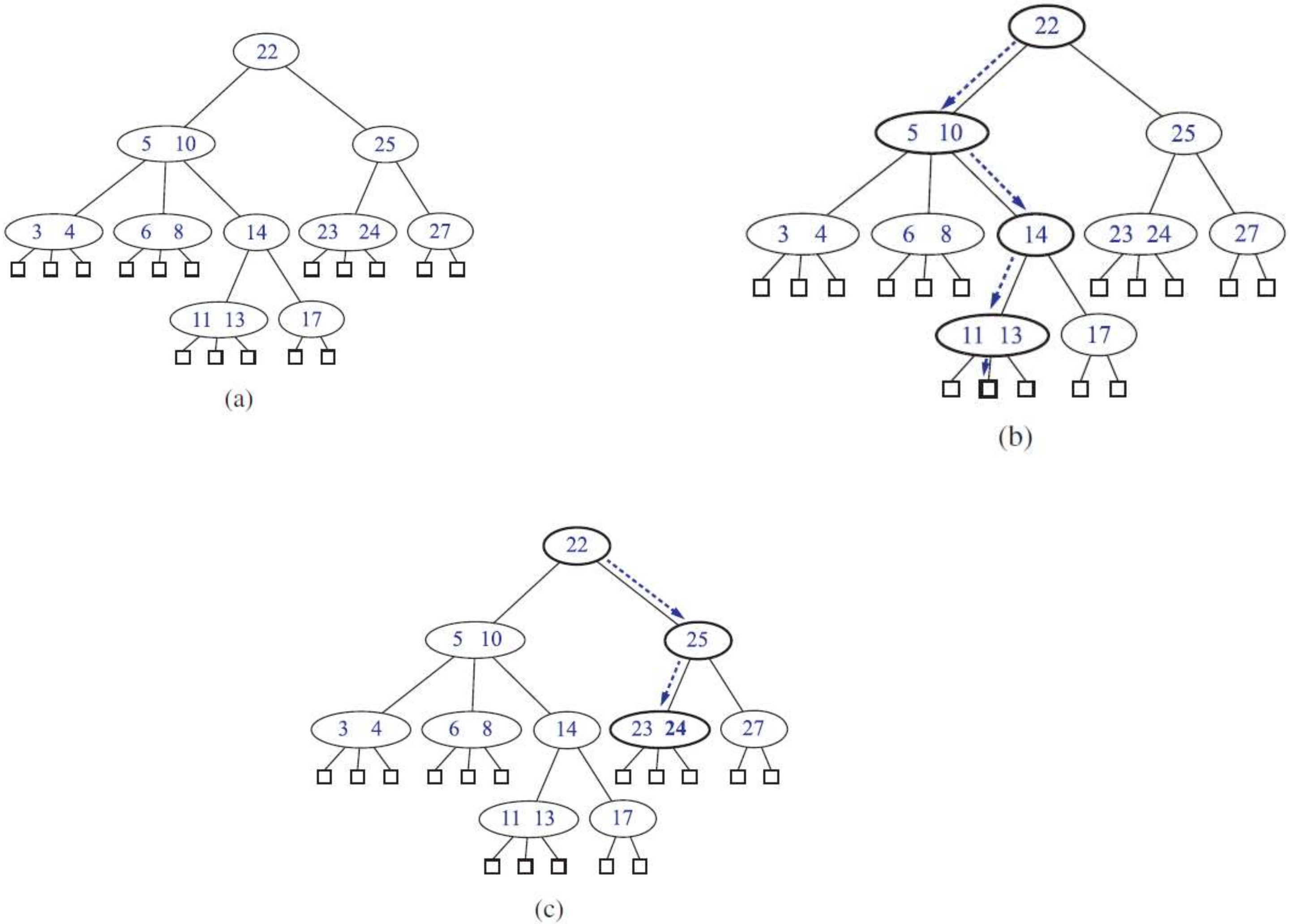


Figure 10.20: (a) A multi-way search tree T; (b) search path in T for key 12 (unsuccessful search); (c) search path in T for key 24 (successful search).

Outline

- m-way search trees
- B-trees

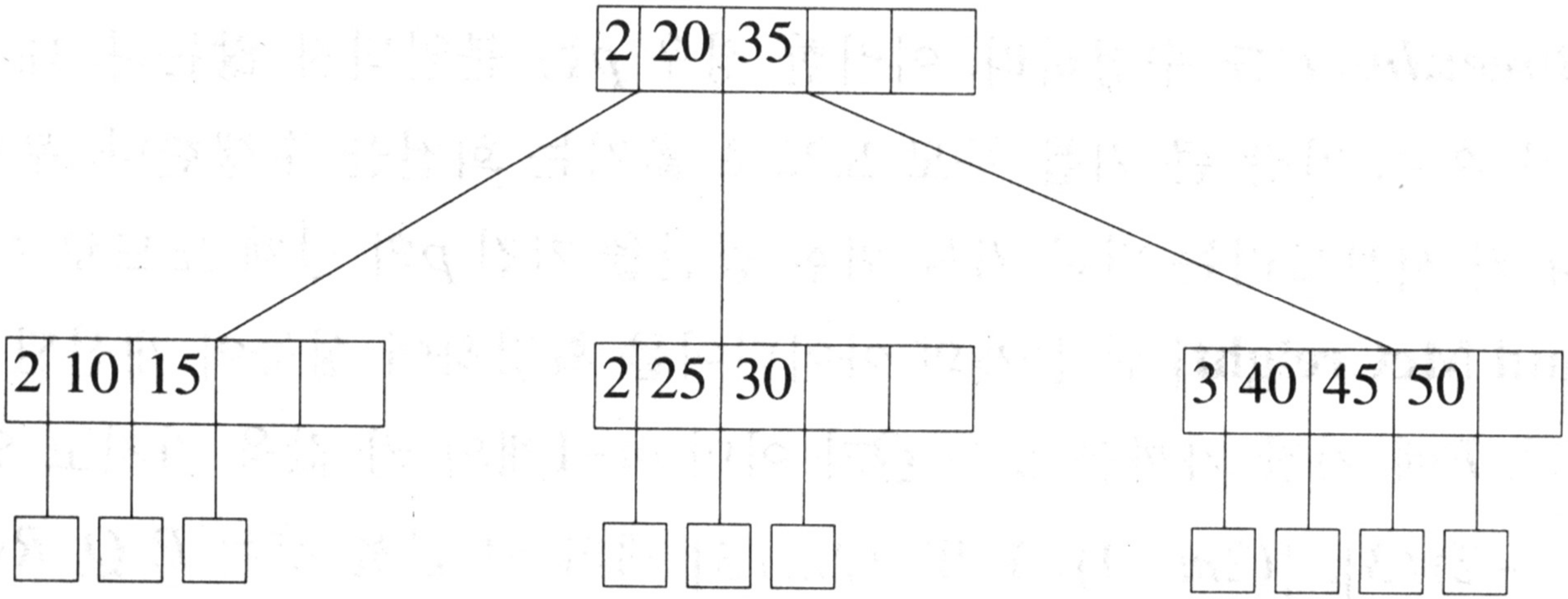
ULSAN NATIONAL INSTITUTE OF SCIENCE AND TECHNOLOGY

B-trees

- m-way search trees with special properties:
  - Replace a NULL pointer to an external node
- Definition
  - If not empty, root node has **at least two** children
  - All internal nodes (except root) have **at least ceil(m/2) children**
  - All external nodes are at the same level
- Balanced m-way search trees

UNIST
ULSAN NATIONAL INSTITUTE OF SCIENCE AND TECHNOLOGY

Example

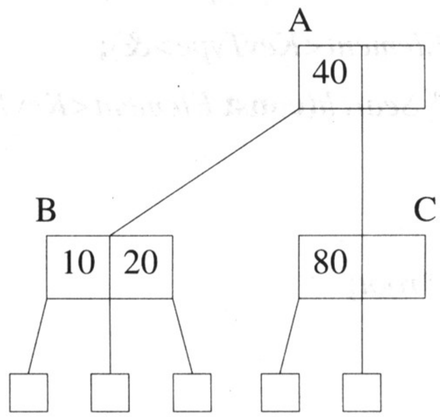


5-way B-tree example, ceil(5/2) = 3

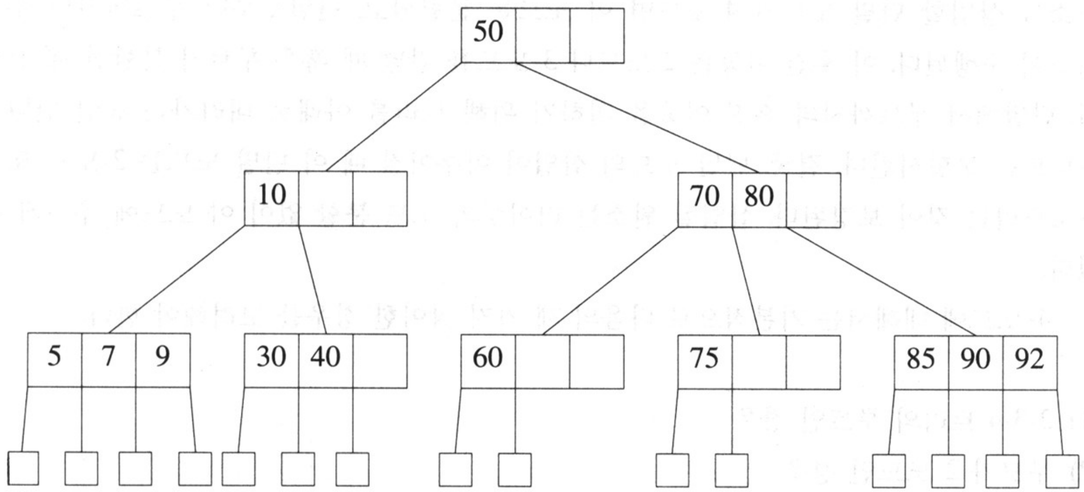
ULSAN NATIONAL INSTITUTE OF SCIENCE AND TECHNOLOGY

2-3 and 2-3-4 Trees

- 2-3 tree is B-tree of order 3
- 2-3-4 tree is B-tree of order 4
  - Also called (2,4) tree or 2-4 tree



2-3 tree



2-3-4 tree

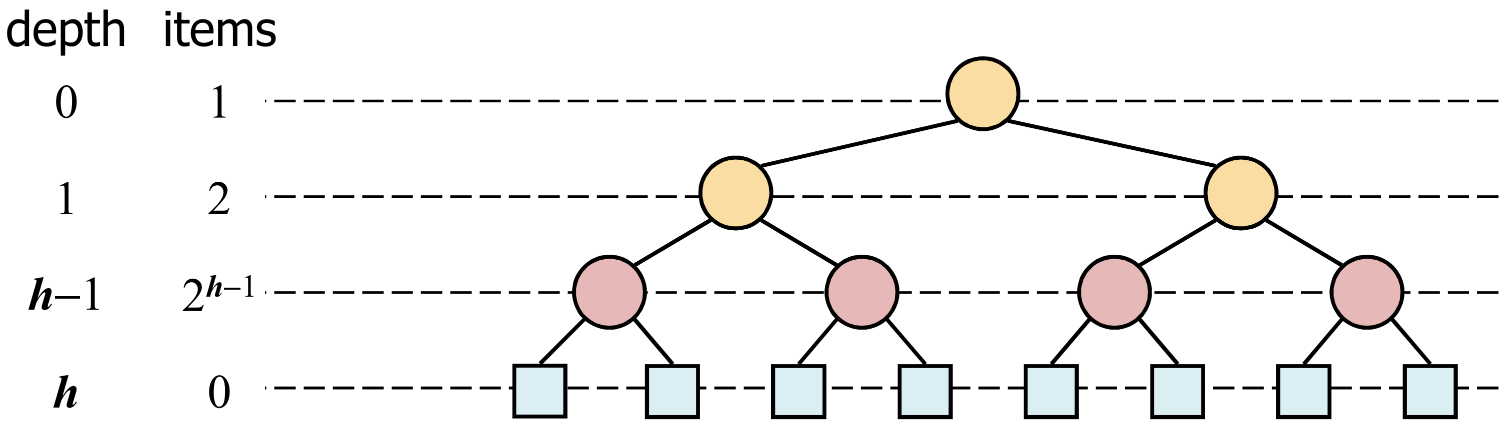
UNIST
ULSAN NATIONAL INSTITUTE OF SCIENCE AND TECHNOLOGY

Height of 2-3-4 Tree

- 2-3-4 tree storing $\boldsymbol{n}$ items has height $\boldsymbol{O}(\log \boldsymbol{n})$
  - Let $\boldsymbol{h}$ be the height
  - Since there are at least 2^i items at depth $\boldsymbol{i} = 0, \ldots, \boldsymbol{h}$ and no items at depth $\boldsymbol{h}$, we have

$$\boldsymbol{n} \geq 1 + 2 + 4 + \ldots + 2^{\boldsymbol{h}-1} = 2^{\boldsymbol{h}} - 1$$

  - Thus, $\boldsymbol{h} \leq \log(\boldsymbol{n} + 1)$



ULSAN NATIONAL INSTITUTE OF SCIENCE AND TECHNOLOGY

Choice of m

- Worst-case search time
  - (time to fetch a node + time to search node) x height
- Search time ↑ if m is too small or too large
- Pick m so that single node fits to a single memory fetch unit (e.g., cache line)
  - Time to fetch a node is constant
  - Time to search node is fast with spatial locality
  - Height of the tree is smaller (fewer non-spatial accesses)