CSC 212: Data Structures and Abstractions

Balanced trees (part 1)

Prof. Marco Alvarez

Department of Computer Science and Statistics
University of Rhode Island

Spring 2025

THINK BIG WE DO℠

From lab session

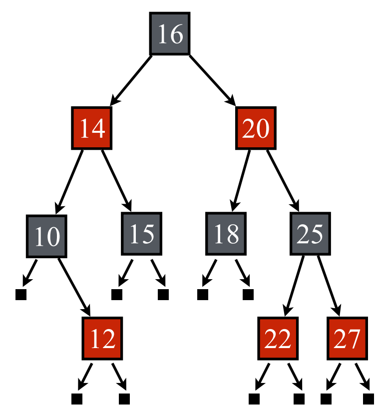
- Assume the dictionary has n keys, and the book has m words (tokens)
  - ✓ what is the computational cost of finding all words in the book that are not in the dictionary?
    - dictionary is represented as a BST and assume that $h = O(\log n)$

Balanced search trees

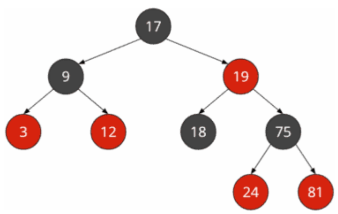
- **Balanced search trees** are a type of search trees that maintain a balanced structure to ensure that the height of the tree is logarithmic in the number of nodes
  - ✓ among the most useful data structures in computer science
  - ✓ many programming languages have built-in support: e.g. Java's `TreeSet` and `TreeMap`, C++'s `std::set` and `std::map`
- Examples of balanced trees:
  - ✓ AVL trees, **Red-Black trees**, B-trees, Splay trees, Treaps, etc.

Red-black trees

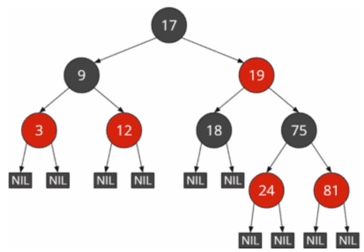
- Red-black trees are BSTs that maintain a balanced structure by enforcing the following properties on the nodes:
  - ✓ each node is colored either **red** or **black**
  - ✓ the root node is always **black**
  - ✓ **red** nodes cannot have **red** children (no two red nodes can be adjacent)
  - ✓ *null* nodes are considered **black**
  - ✓ every *root-to-null* path must have the same number of **black** nodes



Examples



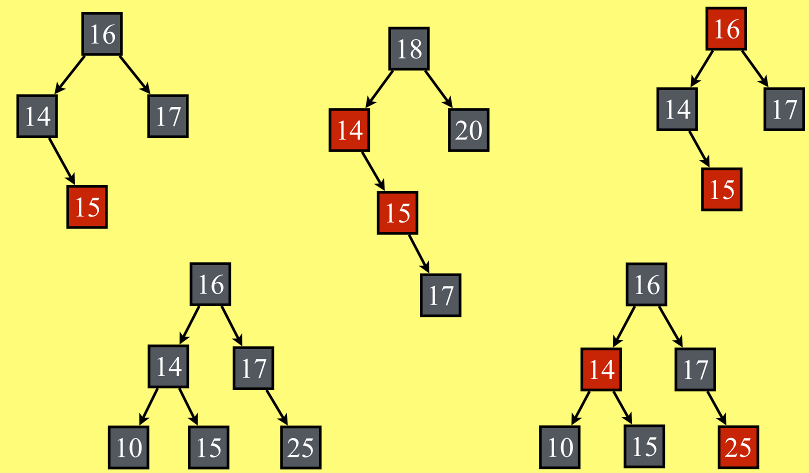
Red-black tree with implicit NIL leaves



Red-black tree with explicit NIL leaves

Practice

- Are these valid red-black trees? — (null nodes not shown)



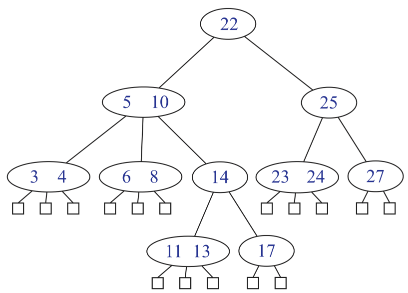
Analysis

- A red-black tree on n nodes has $h = O(\log n)$
  - ✓ after performing an insertion or deletion the tree may become unbalanced
  - ✓ to restore balance, we efficiently modify the tree to satisfy the red-black tree properties
    - done by performing a sequence of **rotations** and **recoloring** nodes
- Equivalence to **B-trees**
  - ✓ red-black trees are equivalent to B-trees of order 4
  - ✓ it is easier to understand the complexity analysis and rebalancing operations of red-black trees by thinking of them as B-trees

B-Trees (interlude)

Multi-way search trees

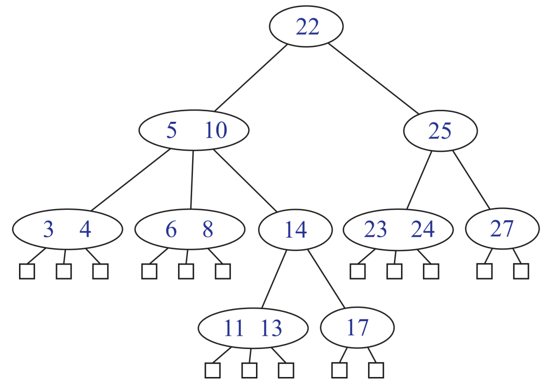
- A <u>multi-way search tree</u> is a generalization of a BST that allows nodes to have more keys and more than two children
  - ✓ the keys in each node are **sorted** in increasing order
  - ✓ the keys in the left subtree of a key k are less than k, and the keys in the right subtree are greater than k



note that null pointers are illustrated as external nodes

Search on a multi-way search tree

- Perform **search** for 12, 17, 24, and 50 on the following tree
  - ✓ note that null pointers are illustrated as external nodes



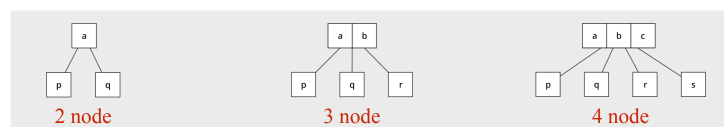
Assume d denotes the maximum number of keys of any node of T, and h denotes the height of T. What is the cost of search?

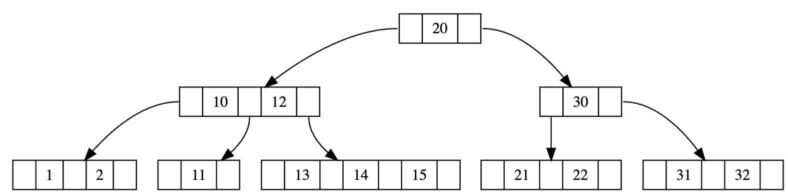
Balanced multi-way search trees

- Balanced multi-way search tree
  - ✓ cap the number of children to a fixed number and keep the leaf nodes at the same depth
  - ✓ the tree is <u>**always balanced**</u>
    - <u>all leave nodes have the same depth</u>
    - search, insertion, and deletion can be performed in $O(\log n)$ time
- <u>**B-tree:**</u> specific type of a balanced multi-way search tree
  - ✓ on a B-tree of order m, each node, except the root, must have between $\lceil b/2 \rceil$ and b children
    - note there are differences in terminology (multiple “order” definitions)
  - ✓ heavily used in databases and file systems to store large amounts of data (common orders: 1024, 2048, 4096, …)

2-3-4 tree

- A 2-3-4 tree (a.k.a. 2-4 tree) is a <u>B-tree of order 4</u>
  - ✓ each node can have 2, 3, or 4 children
  - ✓ i.e. all nodes must have at least 1 key and at most 3 keys, except the root node that can have 0 keys when the tree is empty

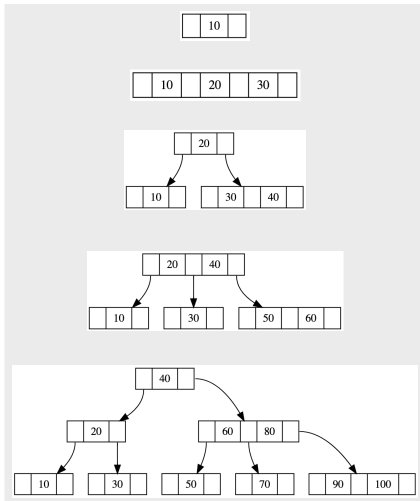




Insertion (2-3-4 tree)

- Steps
  - ✓ start at the root and traverse down the tree to find the appropriate leaf node
  - ✓ if the leaf node has less than 3 keys, insert the new key in sorted order
  - ✓ if the leaf node has 3 keys, split it into two nodes and promote the middle key to the parent node
    - insert the new key in the appropriate child node
    - if the parent node also has 3 keys, repeat the splitting process up to the root
- Tree remains balanced after each insertion
  - ✓ all leaf nodes are at the same level

Insert 10, 20, 30, 40, 50, 60, 70, 80, 90, 100



Practice

- Insert the following sequence into a 2-3-4 tree
  - ✓ 15, 10, 25, 5, 1, 30, 45, 60, 100, 70, 80, 40, 35, 90

Practice

- What is the max h of a 2-3-4 tree with n nodes?
  - ✓ to maximize the height, we want to minimize the number of keys per node (instance of a worst-case)
  - ✓ draw an example tree and express h in terms of n

Practice

- What is the cost of search and insert on a 2-3-4 tree?
  - ✓ worst-case scenario

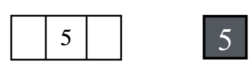
So far …

- The cost of operations in a B-tree of order b is $O(b \log_b n)$
  - ✓ insert, search, remove
  - ✓ small values of b make this cost optimal
- In practice …
  - ✓ B-trees are <u>widely used in databases and file systems</u> to manage large amounts of data efficiently
  - ✓ useful for systems that read and write large blocks of data
    - B-trees can minimize the number of disk accesses required (much larger order values)

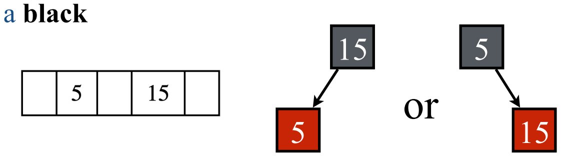
Red-black trees

Red-black trees <=> 2-3-4 trees

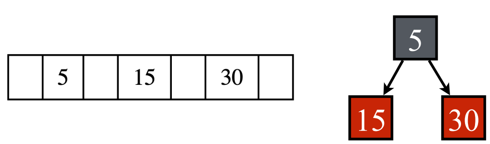
- A 2-node in a 2-3-4 tree corresponds to a **black** node in a red-black tree



- A 3-node corresponds to a **black** node with one **red** child

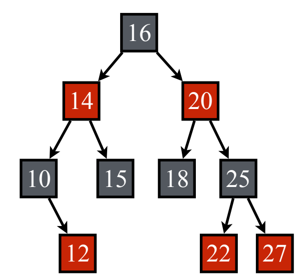


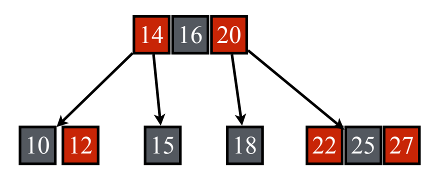
- A 4-node corresponds to a **black** node with two **red** children



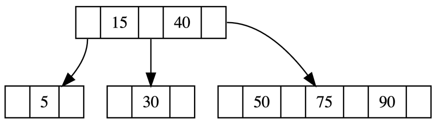
Red-black trees <=> 2-3-4 trees

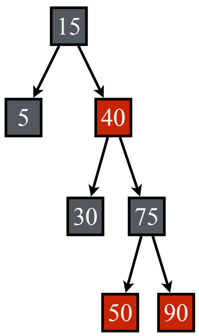
- Red-black trees are <u>**isometric**</u> to 2-3-4 trees
  - ✓ the <u>number of **black** nodes</u> on any *root-to-null* path corresponds to the <u>number of levels</u> of the 2-3-4 tree
- Every red-black tree can be transformed into an equivalent 2-3-4 tree and vice versa
  - ✓ the relationship between the trees is not bijective (1-1)





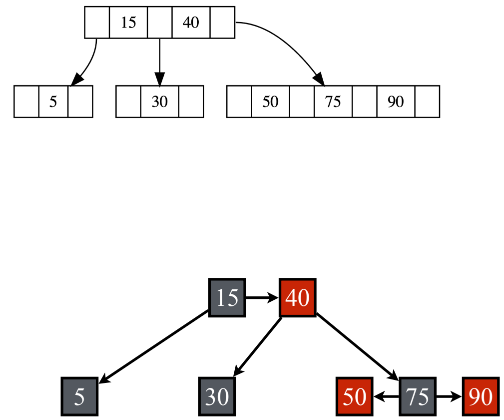
Example





- each node is colored either **red** or **black**
- the root node is always **black**
- **red** nodes cannot have **red** children (no two red nodes can be adjacent)
- *null* nodes are considered **black**
- every *root-to-null* path must have the same number of **black** nodes

Example



- each node is colored either **red** or **black**
- the root node is always **black**
- **red** nodes cannot have **red** children (no two red nodes can be adjacent)
- *null* nodes are considered **black**
- every *root-to-null* path must have the same number of **black** nodes

Practice

- Draw the red-black tree that corresponds to the following 2-3-4 tree