Programming Abstractions

CS106B

Cynthia Lee

Topics:

- **Map** implemented as a Binary Search Tree (BST)
  - Starting with a dream: binary search in a linked list?
  - How our dream provided the inspiration for the BST
  - BST insert
  - Big-O analysis of BST
  - BST balance issues
- Traversals
  - Pre-order
  - In-order
  - Post-order
  - Breadth-first
- Applications of Traversals

Question about our put() algorithm:

Pretty simple!

- If key > node's key
  - Go right!
- If key < node's key
  - Go left!

Q. What do we do if the key is equal to the node's key?

Stanford Map example:

```
Map<int, string> mymap;
mymap.put(5, "five");
mymap.put(5, "cinco");          // what should happen?
cout << mymap.get(5) << endl; // what should print?
```

BST put()

Insert: 22, 9, 34, 18, 3

How many of these result in the same tree structure as above?

22, 34, 9, 18, 3

22, 18, 9, 3, 34

22, 9, 3, 18, 34

A. None of these
B. 1 of these
C. 2 of these
D. All of these

Performance analysis:
What is the WORST CASE cost for doing containsKey() in BST?

A. O(1)
B. O(log n)
C. O(n)
D. O(n log n)
E. $O(n^2)$

What is the worst case cost for doing containsKey() in BST ***if the BST is balanced?***

O(logN)—awesome!

BSTs are great when balanced

BSTs are bad when unbalanced

…and Balance depends on order of insert of elements…

…but user controls this, not “us” (author of the Map class)

…no way for “us” (author of Map class) to ensure our Map doesn’t perform terribly! ☹ ☹

Ok, so, long-chain BSTs are bad, should we worry about it? [math puzzle time]

One way to create a bad BST is to insert the elements in *decreasing* order: 34, 22, 9, 3

That's not the only way…

How many **distinctly structured** BSTs are there that exhibit the worst case height (height equals number of nodes) for a tree with the 4 nodes listed above?

A. 2-3
B. 4-5
C. 6-7
D. 8-9
E. More than 9

Bonus question: general formula for any BST of size n?

Extra bonus question (CS109): what is this as a fraction of all trees (i.e., probability of worst-case tree).

BST Balance Strategies

We need to balance the tree (keep O(logN) instead of O(N)), how can we do that if the tree structure is decided by key insert order?

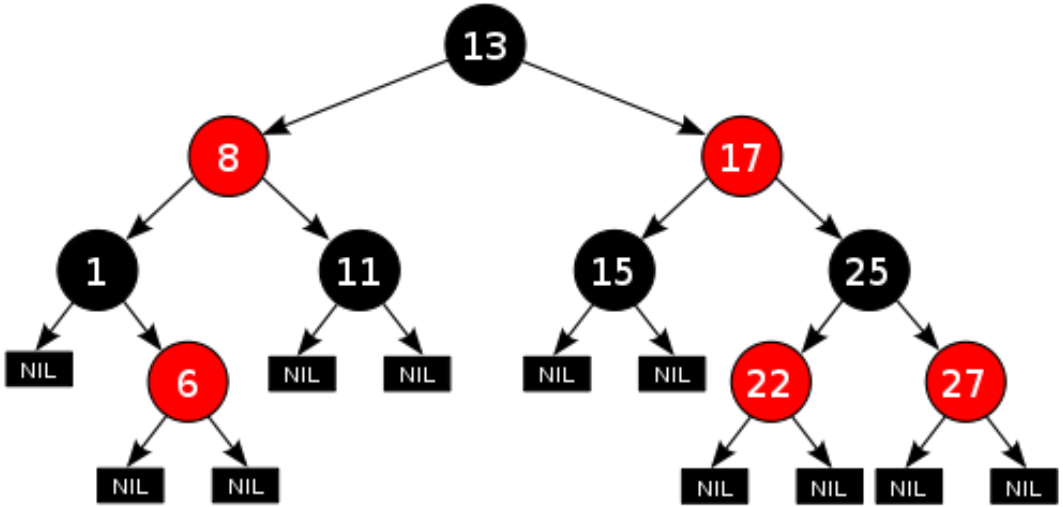
Red-Black trees

One of the most famous (and most tricky) strategies for keeping a BST balanced

Not guaranteed to be perfectly balanced, but “close enough” to keep O(log n) *guarantee* on operations

Video: http://www.youtube.com/watch?v=vDHFF4wjWYU

Red-Black trees



Every simple path from a given node to any of its descendant leaves contains the same number of black nodes.

- (This is what guarantees “close” to balance)

A few BST balance strategies

- AVL tree
- Red-Black tree
- Treap (BST + heap in one tree! What could be cooler than that, amirite? ♥ 😍 ♥)

Other fun types of **BST**

Splay tree

- Rather than only worrying about balance, Splay Tree dynamically readjusts based on **how often users search for an item**. Most commonly-searched items move to the top, saving time
  - Example: if Google did this, "Bieber" would be near the root, and "splay tree" would be further down by the leaves

B-Tree

- Like BST, but a node can have many children, not just two
- More branching means an even "flatter" (smaller height) tree
- Used for huge databases

BST and Heap quick recap/cheat sheet

BST and Heap Facts (cheat sheet)

Heap (**Priority Queue**)

- Structure: must be “complete”
- Order: parent priority must be <= both children
  - This is for min-heap, opposite is true for max-heap
  - No rule about whether left child is > or < the right child
- Big-O: guaranteed log(n) enqueue and dequeue
- Operations: always add to end of array and then “bubble up”; for dequeue do “trickle down”

BST (**Map**)

- Structure: any valid binary tree
- Order: leftchild.key < self.key < rightchild.key
  - No duplicate keys
  - Because it’s a Map, values go along for the ride w/keys
- Big-O: log(n) if balanced, but might not be balanced, then O(n)
- Operations: recursively repeat: start at root and go left if key < root, go right if key > root

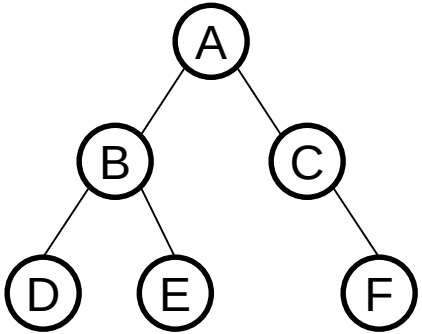
Tree Traversals!

These are for any binary trees, but we often do them on BSTs

What does this print?
(assume we call traverse on the root node to start)

```
void traverse(Node *node) {
  if (node != NULL) {
     cout << node->key << " ";
     traverse(node->left);
     traverse(node->right);
  }
}
```

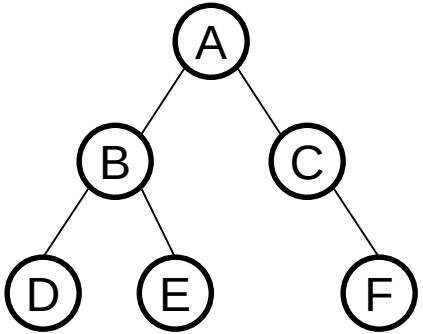
A. A B C D E F
B. A B D E C F
C. D B E F C A
D. D E B F C A
E. Other/none/more



What does this print?
(assume we call traverse on the root node to start)

```
void traverse(Node *node) {
  if (node != NULL) {
     traverse(node->left);
     traverse(node->right);
     cout << node->key << " ";
  }
}
```

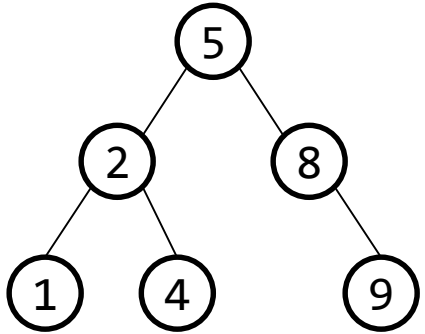
A. A B C D E F
B. A B D E C F
C. D B E F C A
D. D E B F C A
E. Other/none/more



What does this print?
(assume we call traverse on the root node to start)

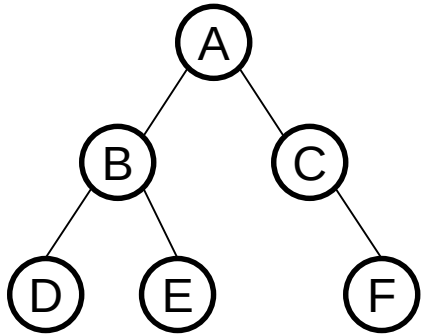
```
void traverse(Node *node) {
  if (node != NULL) {
     traverse(node->left);
     cout << node->key << " ";
     traverse(node->right);
  }
}
```

A. 1 2 4 5 8 9
B. 1 4 2 9 8 5
C. 5 2 1 4 8 9
D. 5 2 8 1 4 9
E. Other/none/more



How can we get code to print our ABCs in order as shown? (note: not BST order)

```
void traverse(Node *node) {
  if (node != NULL) {
 ??  cout << node->key << " ";
     traverse(node->left);
     traverse(node->right);
  }
}
```



You can't do it by using this code and moving around the cout—we already tried moving the `cout` to all 3 possible places and it didn't print in order

- You can but you use a **queue** instead of recursion
- **"Breadth-first"** search
- *Again we see this key theme of BFS vs DFS!*

Applications of Tree Traversals

Beautiful little things from an algorithms/theory standpoint, but they have a practical side too!

Traversals a very commonly-used tool in your CS toolkit

```
void traverse(Node *node) {
  if (node != NULL) {
     traverse(node->left);
     // "do something”
     traverse(node->right);
  }
}
```

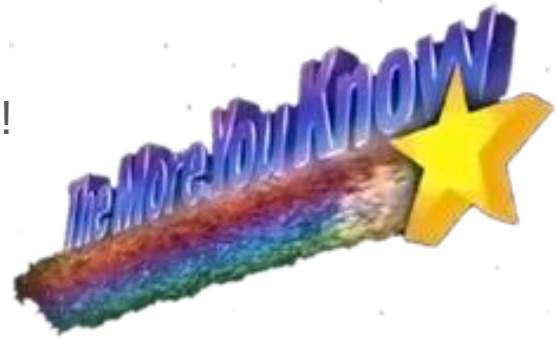
- Customize and move the “do something,” and that’s the basis for dozens of algorithms and applications

Map interface implemented with BST

- Remember how when you iterate over the Stanford library Map you get the keys in sorted order?
  - (we used this for the word occurrence counting code example in class)

```
void printMap(const Map<string, int>& theMap) {
    for (string s : theMap) {
        cout << s << endl; // printed in sorted order
    }
}
```

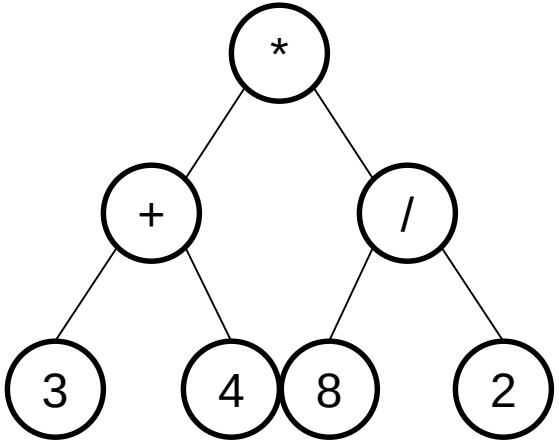
- Now you know why it can do that in O(N) time!
  - "In-order" traversal



Applications of the traversals

- You have a tree that represents **evaluation of an arithmetic expression**. Which traversal would form the foundation of your evaluation algorithm?

A. Pre-order
B. In-order
C. Post-order
D. Breadth-first



Applications of the traversals

- You are writing the **destructor** for a BST class. Given a pointer to the root, it needs to free each node. Which traversal would form the foundation of your destructor algorithm?
  - A. Pre-order
  - B. In-order
  - C. Post-order
  - D. Breadth-first