Lecture 8

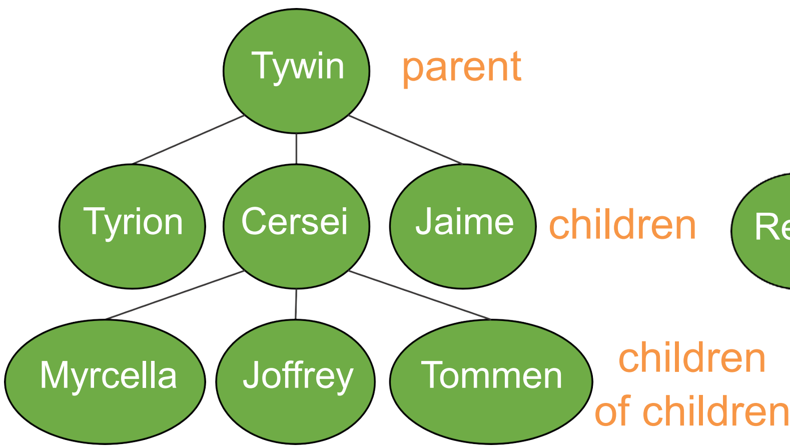
Binary Search Tree and Trie

Department of Computer Science
Hofstra University

Lecture Goals

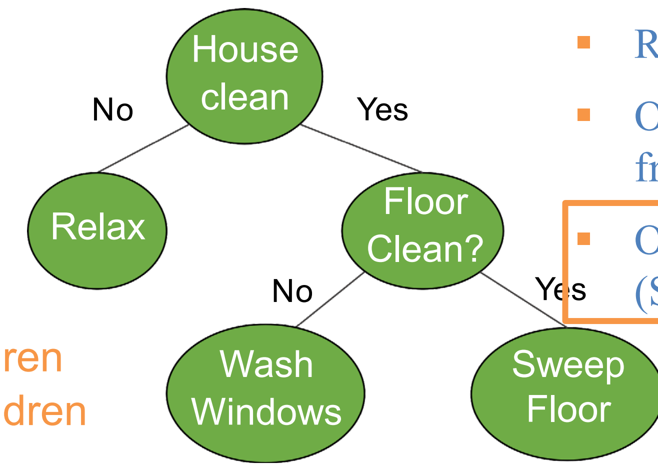
- Describe the value of trees and their data structure
- Explain the need to visit data in different orderings
- Perform pre-order, in-order, post-order and level-order traversals
- Define a Binary Search Tree
- Perform search, insert, delete in a Binary Search Tree
- Explain the running time performance to find an item in a BST
- Compare the performance of linked lists and BSTs
- Explain what a trie data structure is
- Describe the algorithm for finding keys in and adding keys to a trie
- Compare the time to find a key in a BST to a trie
- Implement a trie data structure in Java

Different Trees in Computer Science



Family Trees

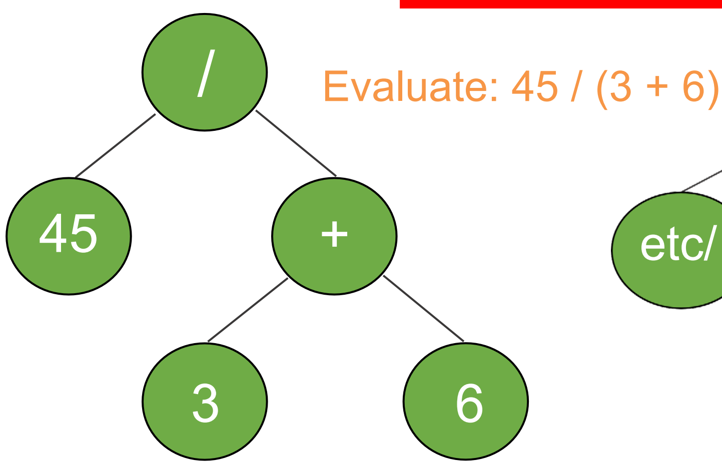
Why trees?



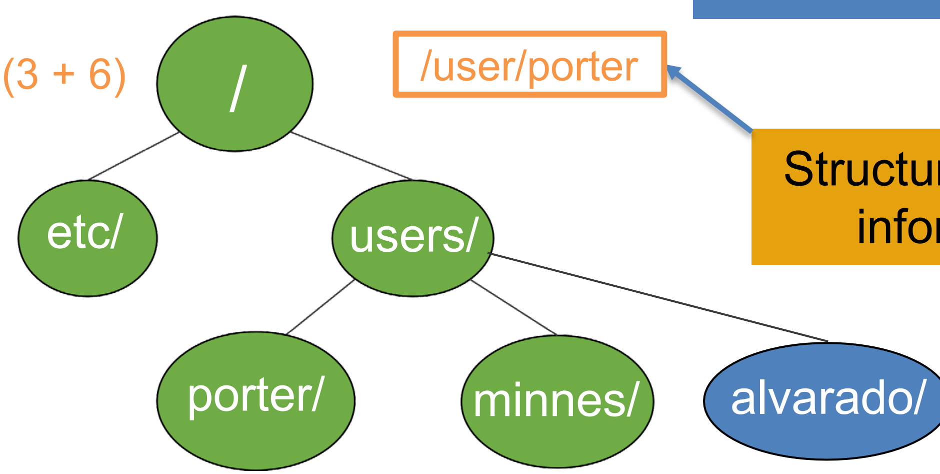
Decision Trees

- Root is most important (Heap)
- Organized by character frequency (Huffman Tree)
- Organized by node ordering (Search Trees)
- Etc…

Different Organizations → Different Trees



Expression Trees

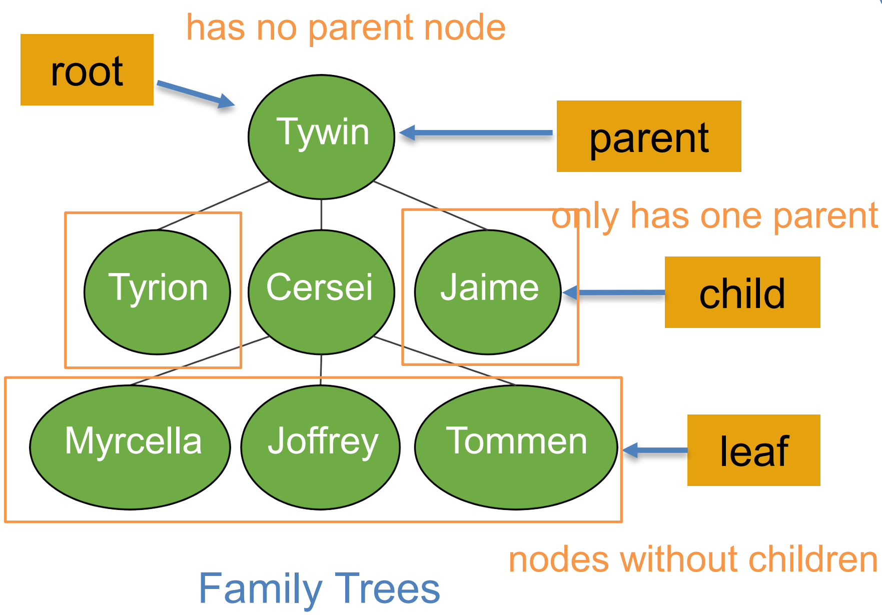


File System

Structure conveys information

Dynamic Data Structure

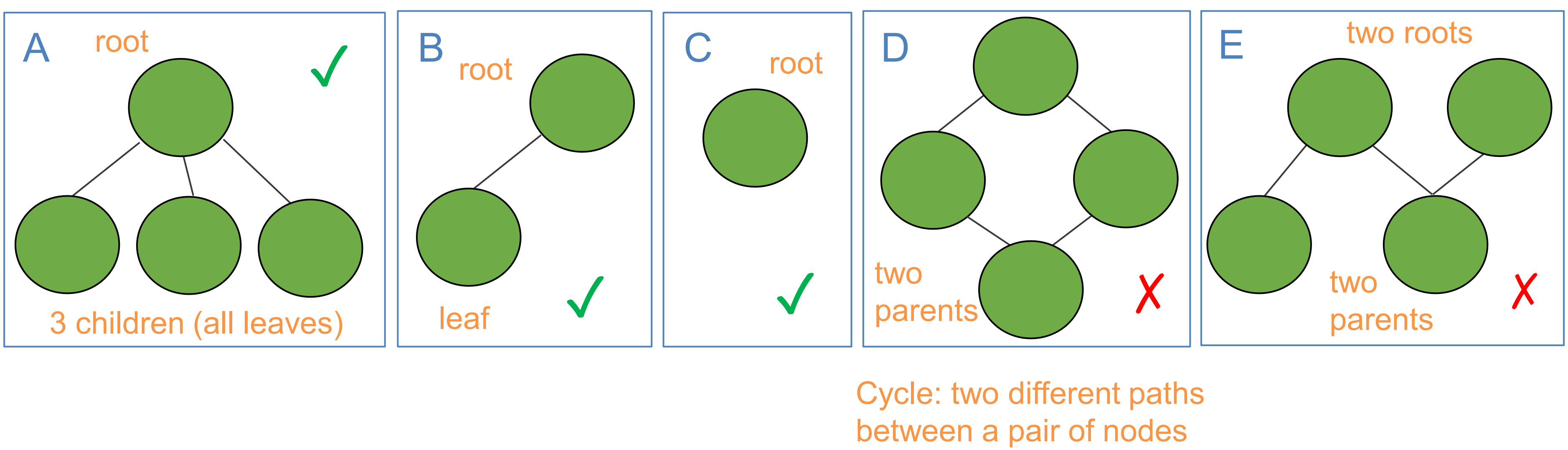
Defining Trees



What defines a tree?

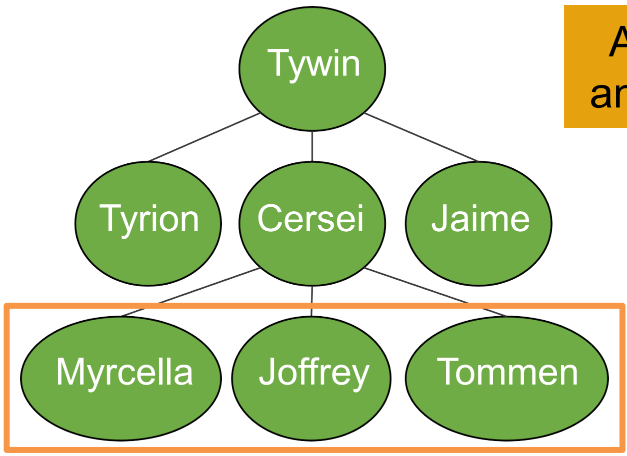
- Single root
- Each node can have only one parent (except for root)
- No cycles in a tree

Which are trees?



Binary Trees

Generic Tree



Any Parent can have any number of children

How would a general tree node differ?

A general tree would just have a list for children

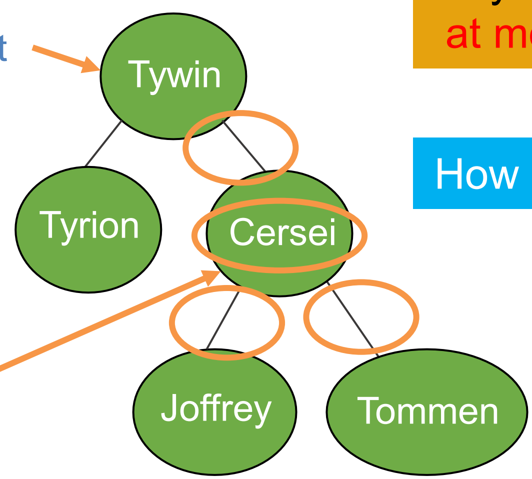
Binary Tree

Any Parent can have at most two children

A tree just needs a root node

like the head and tail for linked list

root



How do we construct a tree?

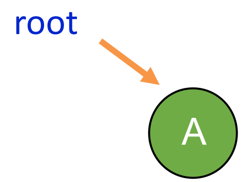
Each node needs:
1. A value
2. ~~A parent~~
3. A left child
4. A right child

Like Linked Lists, Trees have a "Linked Structure"

nodes are connected by references

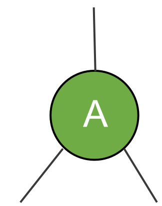
Write Code for Binary Tree

```
public class BinaryTree<E> {
    TreeNode<E> root;
    // more methods
}
```




```
public class TreeNode<E> {
        private E value;
        private TreeNode<E> parent;
        private TreeNode<E> left;
        private TreeNode<E> right;
        public TreeNode(E val, TreeNode<E> par) {
                this.value = val;
                this.parent = par;
                this.left = null;
                this.right = null;
        }
        public TreeNode<E> addLeftChild(E val) {
                this.left = new TreeNode<E>(val, this);
                return this.left;
        }
}
```

For root: TreeNode(val, null)



Let's write a constructor together

Next Step is to able to set/get children

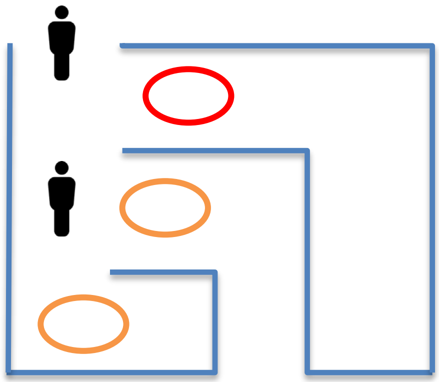
Fill in the blank:

A. this.parent
B. this.left
C. this.right
D. this

Tree Traversal - Motivation

Warning: These first examples are really graphs. We'll visit graphs in detail in the next course. Here they are used as motivating examples

start



finish

Strategy: go until hit a dead end, then retrace steps and try again

Maze Traversal

Imagine this is a hedge maze

What's my next step?

Mazes benefit from "Depth First Traversals"

Bottom line: Order we visit matters and we'll make choices based on our needs

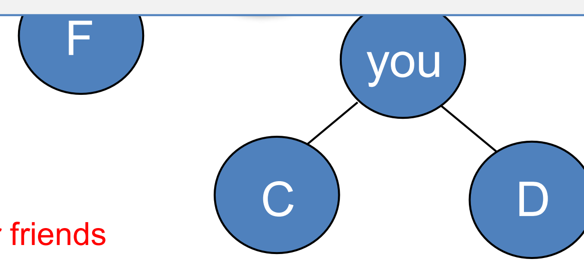
Suppose you have a list of your friends and each of your friends have lists

How closely are you connected with D?

What's my next step?

Strategy: look at all of your friends first, and then branch out.

This problem benefits from "Breadth First Traversals"



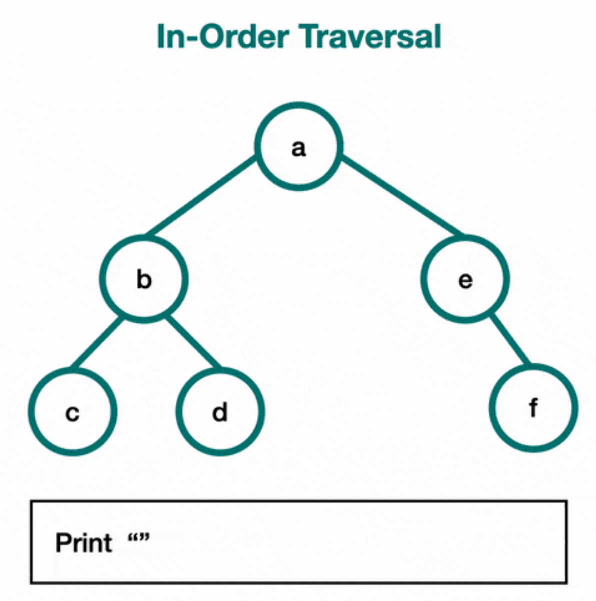
Social Network

Graph traversal with DFS: in-order, pre-order, post-order

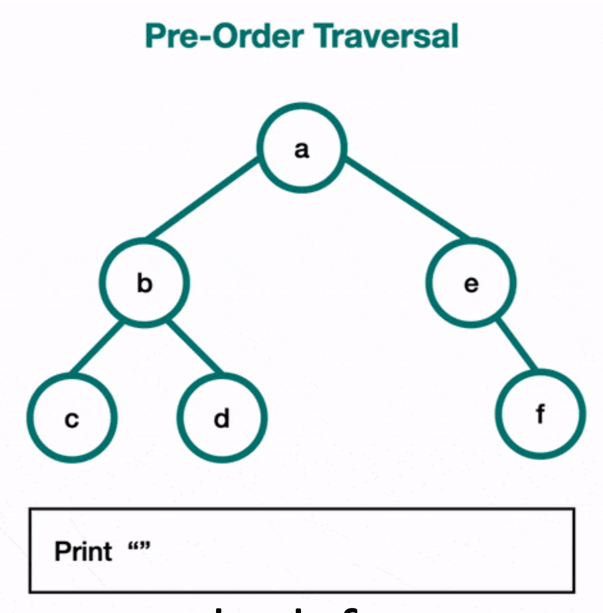
```
function inOrderTraversal(node) {
  if (node !== null) {
    inOrderTraversal(node.left);
    visitNode(node);
    inOrderTraversal(node.right);
  }
}
```

```
function preOrderTraversal(node) {
  if (node !== null) {
    visitNode(node);
    preOrderTraversal(node.left);
    preOrderTraversal(node.right);
  }
}
```

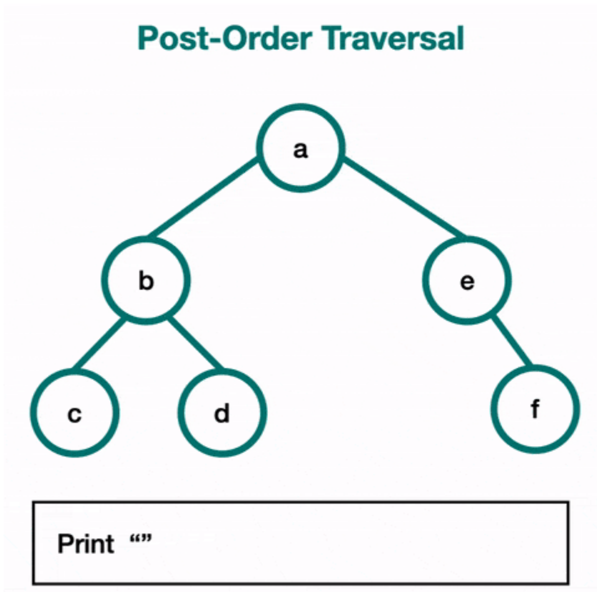
```
function postOrderTraversal(node) {
  if (node !== null) {
    postOrderTraversal(node.left);
    postOrderTraversal(node.right);
    visitNode(node);
  }
}
```



cbdaef



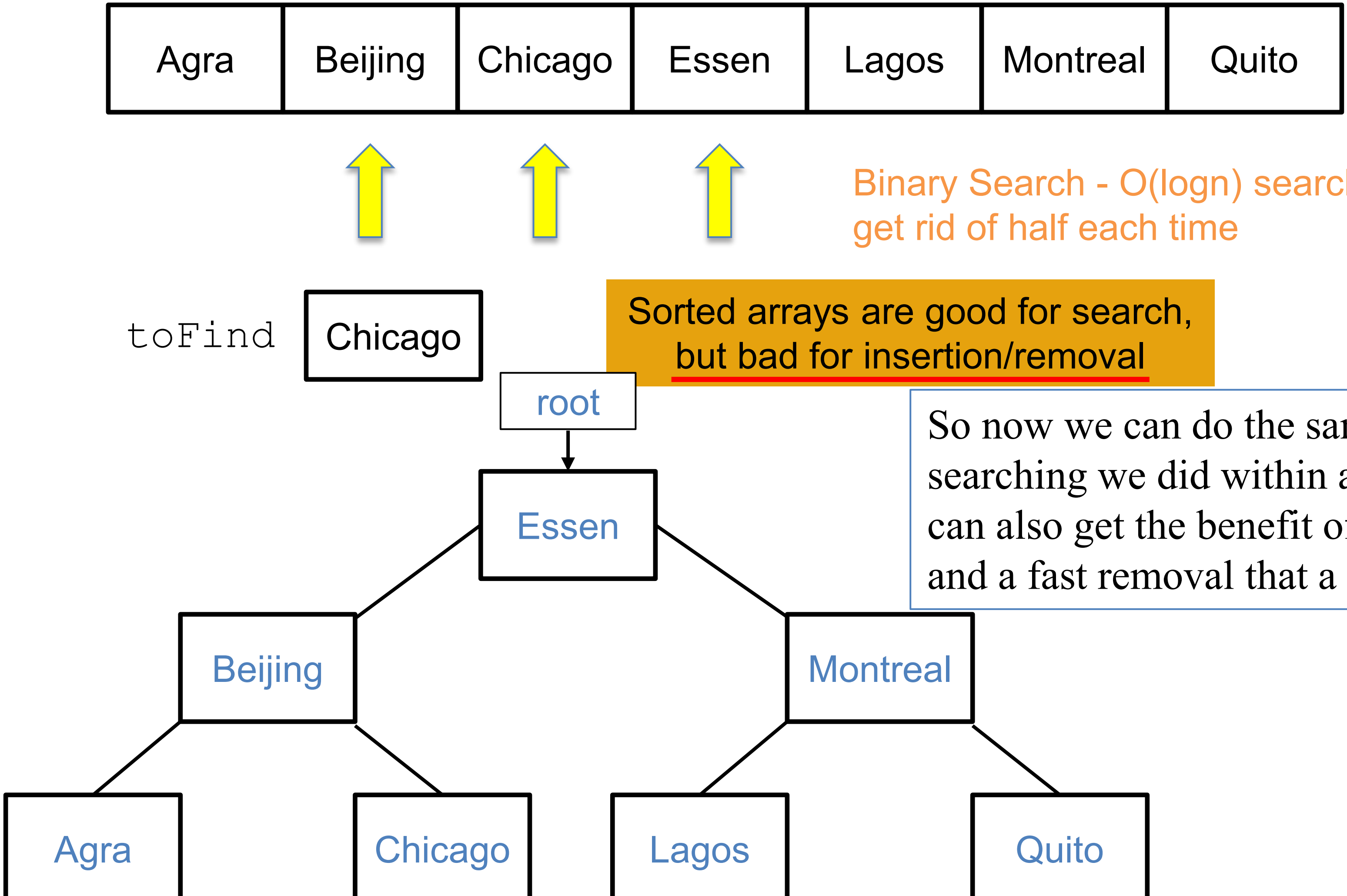
abcdef



cdbfea

https://skilled.dev/course/tree-traversal-in-order-pre-order-post-order

Motivation for Binary Search Tree



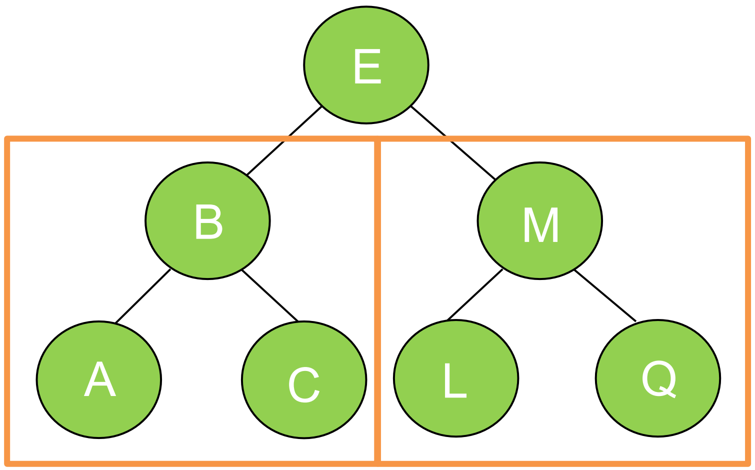
Binary Search - O(logn) search:
get rid of half each time

Sorted arrays are good for search, but bad for insertion/removal

So now we can do the same kind of fast searching we did within an array, but we can also get the benefit of a fast insert and a fast removal that a tree provides.

Binary Search Trees (BST) Explained in Animated Demo
https://www.youtube.com/watch?v=mtvbVLK5xDQ

Binary Search Trees

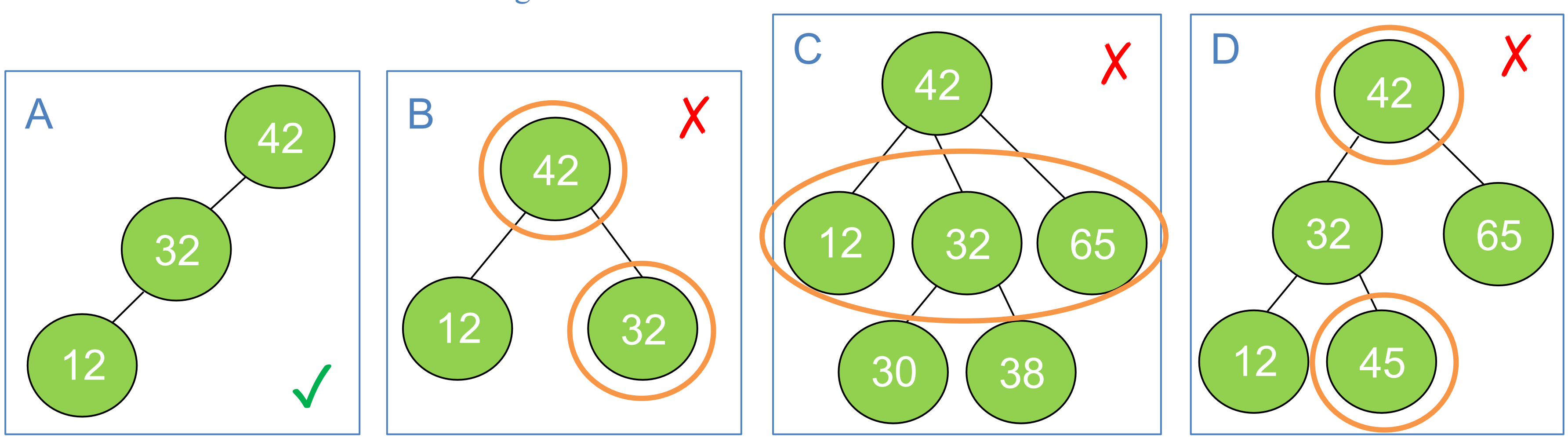


Left subtree's values must be lesser

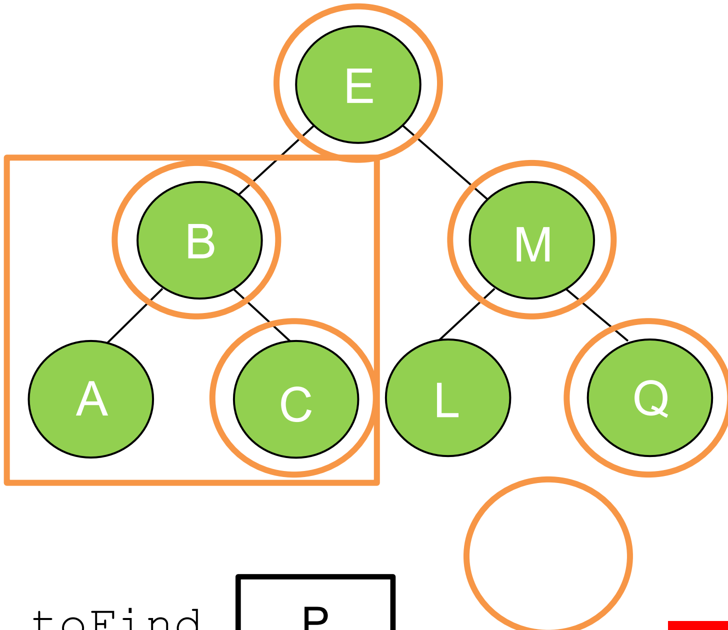
Right subtree's values must be greater

- Ordered, or sorted, binary trees.
- Each node can have 2 subtrees.
- Items to the left of a given node are smaller.
- Items to the right of a given node are larger.

Which of these are binary search trees?



Searching a BST



Same fundamental idea as binary search of an array

toFind C

Compare: E and C

Compare: B and C

Compare: C and C

Found it!

toFind P

Compare: E and P

Compare: M and P

Compare: Q and P

Node is null

Not Found!

How to implement this?

You could solve this with recursion.

You could also solve it with iteration by keeping track of your current node.

Searching a BST Iteratively

```
public class BinaryTree<E> {          <E extends Comparable<? super E>> {
        TreeNode<E> root;
        public boolean search(E toSearch) {
                TreeNode<E> curr = root;
                while (curr != null) {
                        int comp = toSearch.compareTo(curr.getValue());
                                if (comp < 0)
                                curr = curr.getLeftChild();
                                else if (comp > 0)
                                curr = curr.getRightChild();
                                else // comp = 0
                                return true;
                }
                return false;
        }
}
```

It means that either the class E itself or one of its super classes implements Comparable

Doesn't work with objects

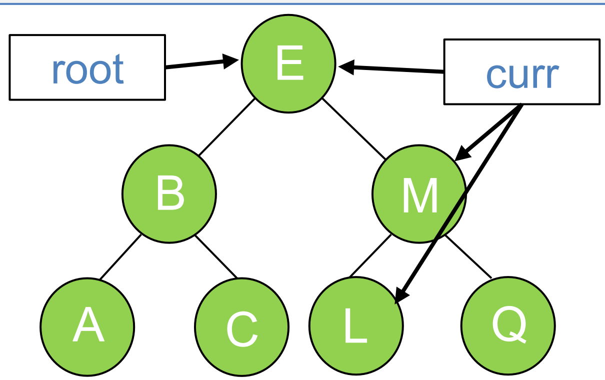
Do NOT change root pointer!

We need to do this over and over if not found

if calling object is less than parameter, compareTo returns a value < 0

if calling object is greater than parameter, compareTo returns a value > 0

Are we done?




```
t.search('L')
```

if calling object is equal to parameter, compareTo returns 0

Traverse down tree until:

a) end is reached
b) element is found

Searching a BST Recursively

```
public class BinaryTree<E extends Comparable<? super E>> {
  TreeNode<E> root;

    private boolean search(TreeNode<E> p, E toSearch) {
        if (p == null)
            return false;
        int comp = toSearch.compareTo(p.getValue());
        if (comp == 0)
            return true;
        else if (comp < 0)
            return search(p.left, toSearch);
        else // comp > 0
            return search(p.right, toSearch);
    }
    public boolean search(E toSearch) {
        return search(root, toSearch);
    }
}
```

Root of the tree we look at

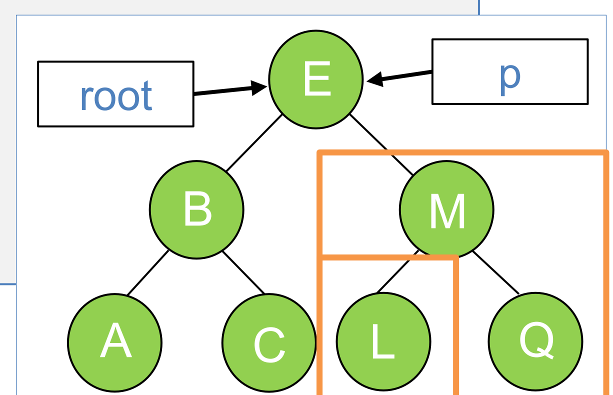
Tree is empty

Found it!

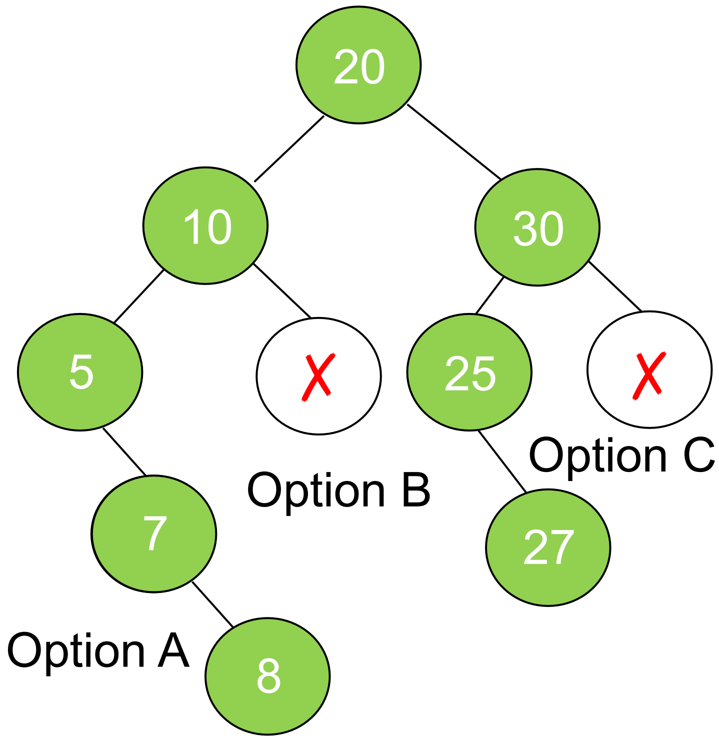
look left

look right

```
t.search(’L’)
```



Inserting into a BST



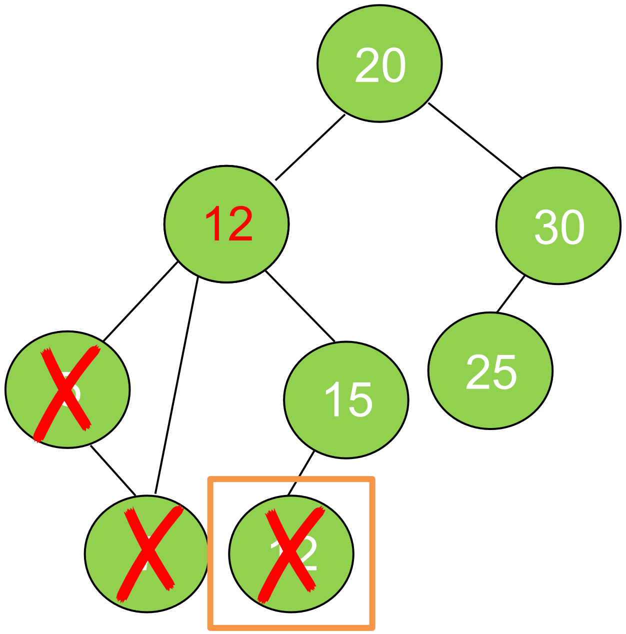
Where should we insert 7?

Insert 27?

Insert 8?

Option D: Either Option A or Option B are fine.

Deleting from a BST



Which of the following is true about the smallest element in a node's right subtree?

A. Its left child is null

B. Its right child is null

C. Both of its children are null

Delete 7

If leaf node: Delete parent's link 7

Delete 5

If only one child, hoist child

Delete 10

When a deleted node has two children, this gets tricky.

Find smallest value in right subtree

Replace deleted element with smallest right subtree value

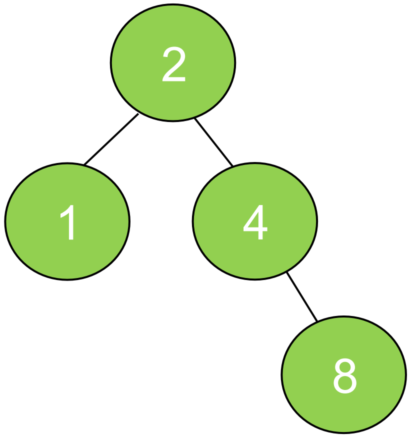
Then delete right subtree duplicate (12)

Binary Search Tree Shape

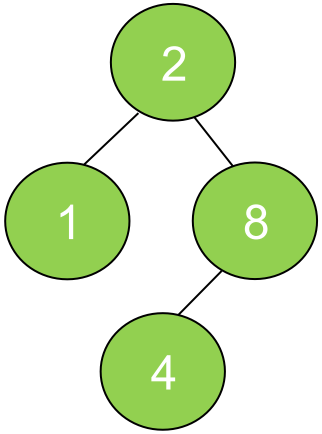
Which of the following Binary Search Trees could be the result of adding elements: 1, 2, 4, and 8 in some order.

These are all valid binary search trees!

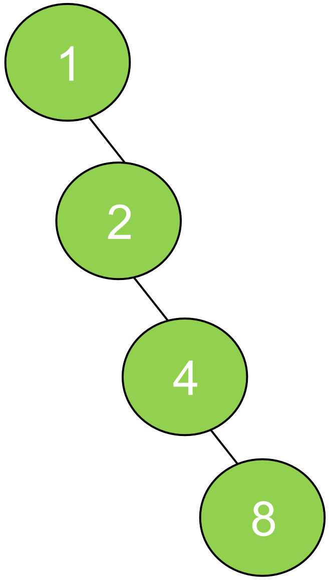
A



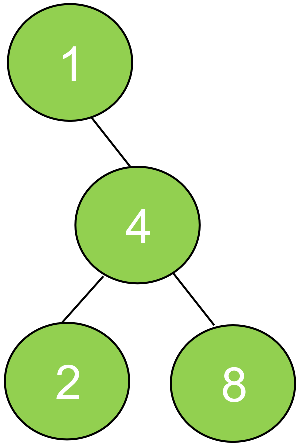
B



C



D

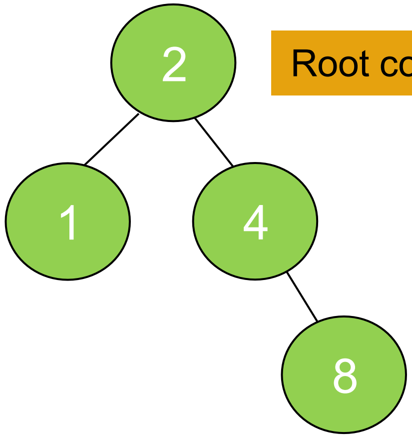


Binary Search Tree Shape (Contd.)

Inserting a node means making it a child of an existing node

Insert nodes as leaves

A ✓



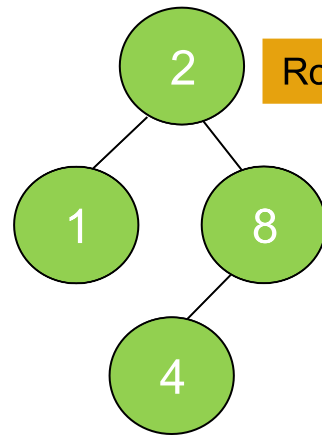
Root comes first

8 needs to be inserted AFTER 4

2			
2	4	1	8
2	1	4	8
2	4	8	1

2			
2	8	1	4
2	1	8	4
2	8	4	1

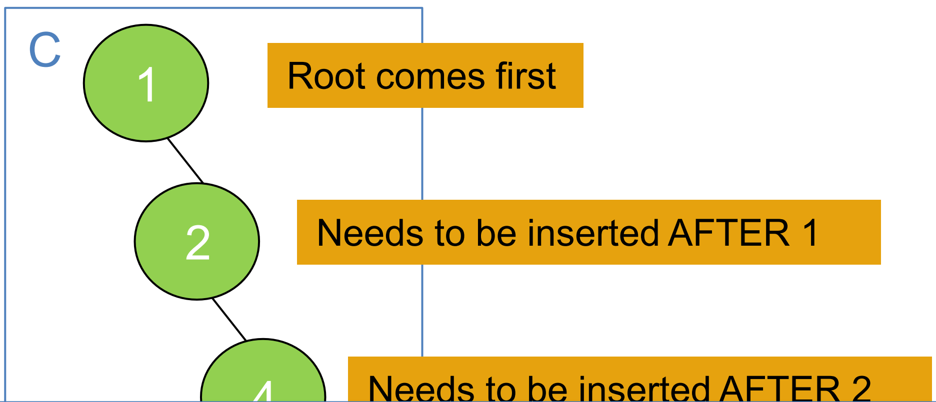
B ✓

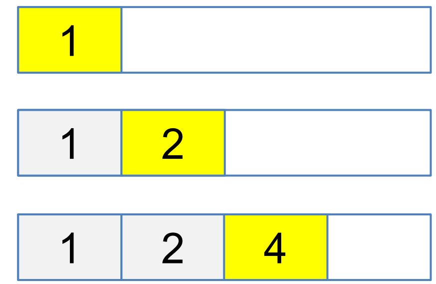


Root comes first

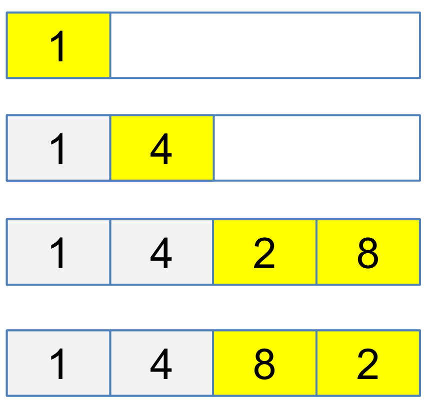
4 needs to be inserted AFTER 8

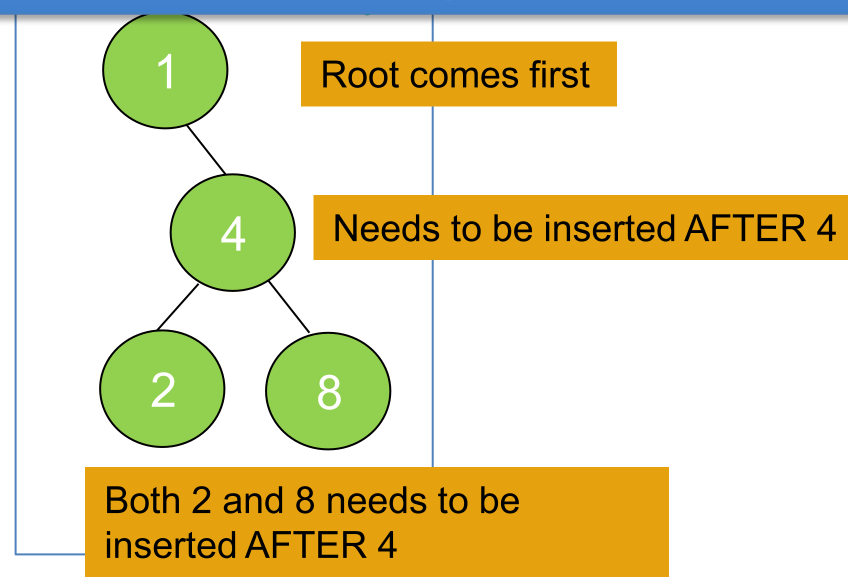
Binary Search Tree Shape (Contd.)





The order in which we put elements into a BST impacts the shape, and what you‘ll see is that the shape of BST will have a huge impact on the performance of operations.



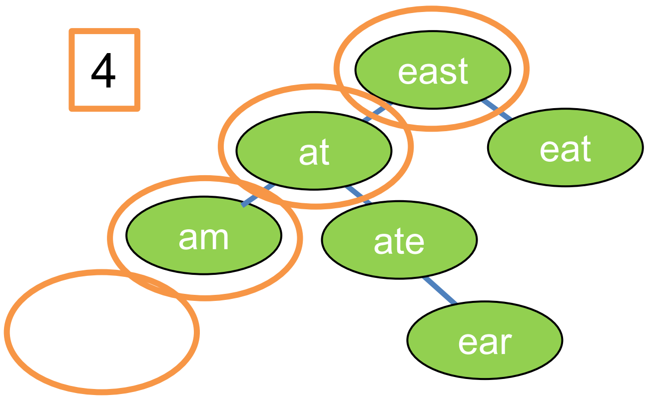


Performance Analysis of BST

Storing a dictionary as a BST

{ am, at, ate, ear, eat, east }

Structure of a BST depends on the order of insertion

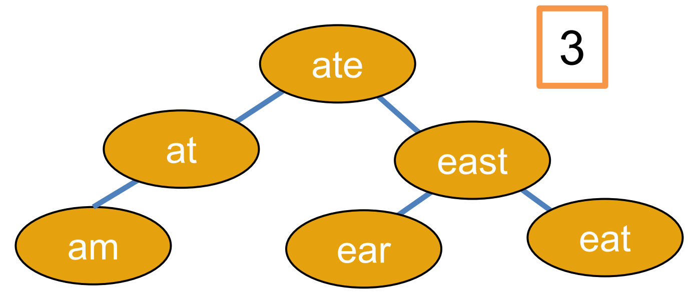


`isWord(east)`

Best case: O(1)

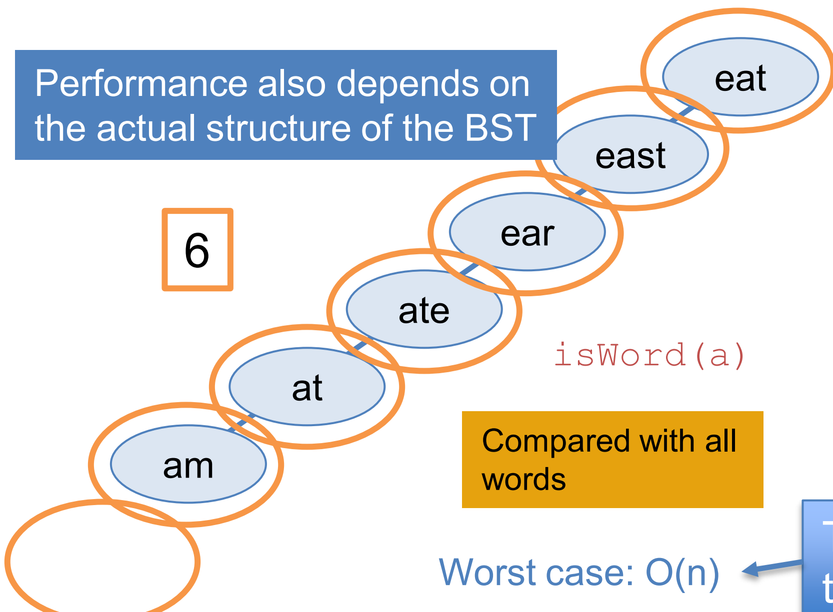
`isWord(a)`

Compared with 3 out of 7 words



How does the performance of isWord relate to input size n?

Performance also depends on the actual structure of the BST



`isWord(a)`

Compared with all words

`isWord(String wordToFind)`

1. Start at root
2. Compare word to current node
   1. If current node is null, return false
   2. If wordToFind is less than word at current node, continue searching in left subtree
   3. If wordToFind is greater than word at current node, continue searching in right subtree
   4. If wordToFind is equal to word at current node, return true

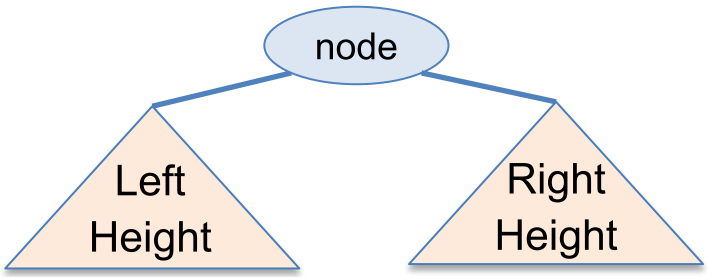
Worst case: O(n)

To optimize the worst case, we can modify the tree to control the max distance until leaf

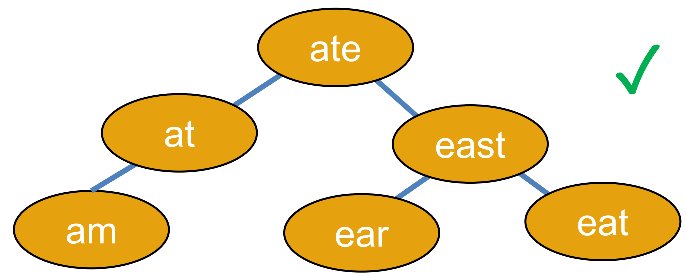
height

Balanced BST

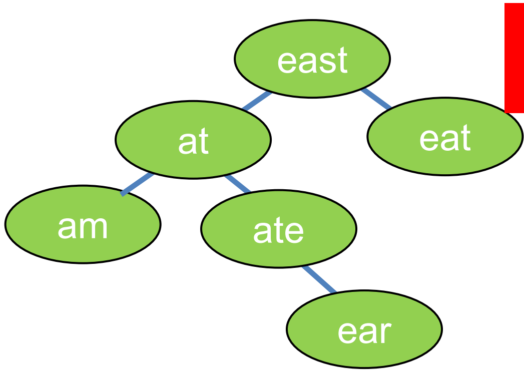
We want to keep the height down as much as we can while still maintaining the same number of nodes.

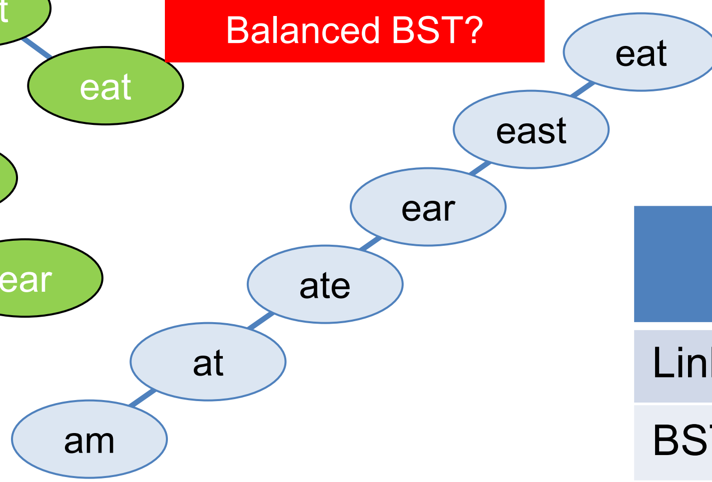


| LeftHeight – RightHeight | <=1



Which is the Balanced BST?





height ≈ log(n)

Especially if insert to BST in order!

	Best case	Average case	Worst case
Linked List	O(1)	O(n)	O(n)
BST	O(1)	O(log n)	O(n)
Balanced BST	O(1)	O(log n)	O(log n)

How to keep balanced? TreeSet and TreeMap in Java API

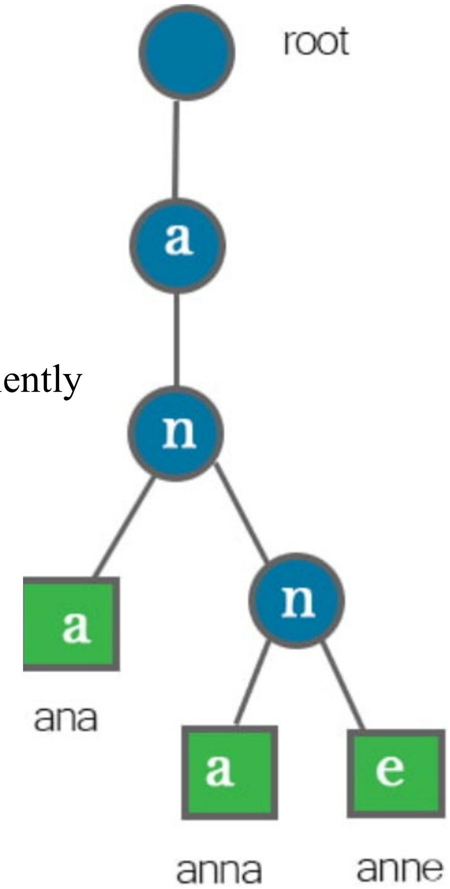
`isWord(String wordToFind)`

BST vs. Hash Table

- Time Complexity
  - Average case:
    - Hash Tables generally offer O(1) average time complexity for insertion, deletion, and search operations.
    - BSTs provide O(log n) time complexity for these operations, assuming the tree is balanced.
  - Worst case
    - Hash Tables can degrade to O(n) performance in cases of poor hash function design or many collisions.
    - BSTs maintain O(log n) performance even in the worst-case for self-balancing BST.
- Ordered Operations
  - BSTs excel at operations requiring ordered data
    - In-order traversal yields sorted elements.
    - Efficient range searches and finding closest elements.
  - Hash Tables do not inherently maintain order, making these operations more difficult.

Tree vs. Trie

- Structure and Purpose
  - Trees:
    - General-purpose data structure for representing hierarchical relationships
    - Each node can contain any type of data
    - Nodes typically have a value and references to child nodes
  - Tries:
    - Specialized tree structure for storing and retrieving strings efficiently
    - Also known as a prefix tree
    - Optimized for operations on strings or sequences
- Node Content
  - Trees:
    - Each node stores a value directly
  - Tries:
    - Nodes typically do not store complete strings
    - The path from the root to a node represents a string or prefix
    - Characters are stored along the edges between nodes



https://romankurnovskii.com/en/posts/tree-vs-trie-data-structures/

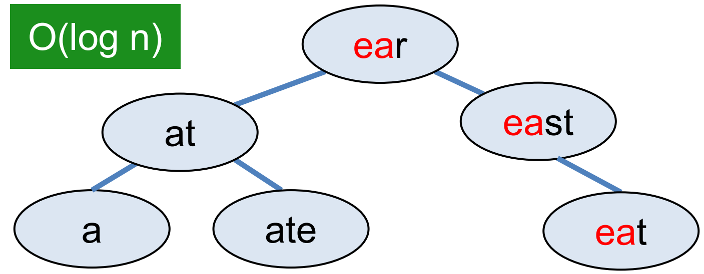
Trie Data Structure

re(TRIE)ve

Storing a dictionary as a (balanced) BST

BSTs don't take advantage of shared structure

O(log n)



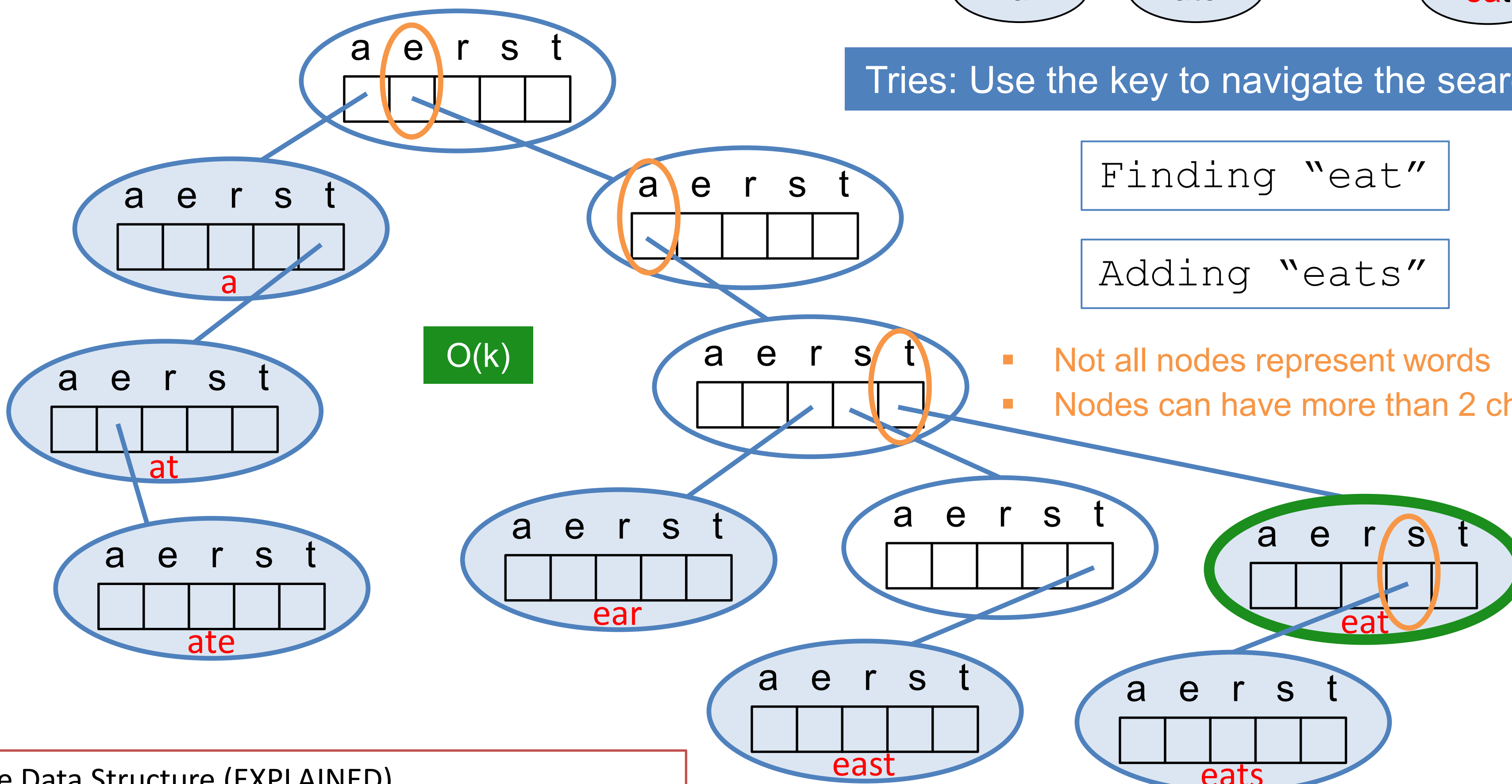
Tries: Use the key to navigate the search

```
Finding "eat"
```

```
Adding "eats"
```

- Not all nodes represent words
- Nodes can have more than 2 children

O(k)



$\log_2(250000) \approx 18$

Trie Data Structure (EXPLAINED)
https://www.youtube.com/watch?v=-urNrIAQnNo

Additional Resources

- Trees and Binary Search Trees
  - http://www.openbookproject.net/thinkcs/archive/java/english/chap17.htm -- explains trees, how to build and traverse it
  - http://algs4.cs.princeton.edu/32bst/ -- about binary search trees
  - Data structures: Binary Search Tree
    - https://www.youtube.com/watch?v=pYT9F8_LFTM
- Tries
  - https://www.toptal.com/java/the-trie-a-neglected-data-structure -- explains with solid example
  - https://www.topcoder.com/community/data-science/data-science-tutorials/using-tries/ -- explains as well as providing code