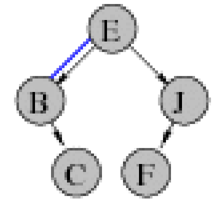
Tree Depth Average Case

Section 4.3.5

Ali Erkan
Ithaca College

Path Length

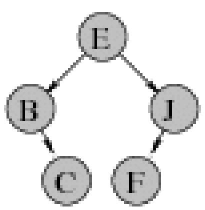
Path length of the following tree: 1



Internal Path Length Of A Tree

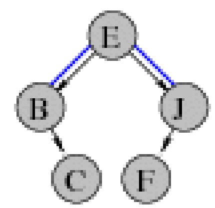
The sum of the depths of all nodes in a tree is known as the internal path length (AKA just path length in these slides).

What is the path length of the following tree:



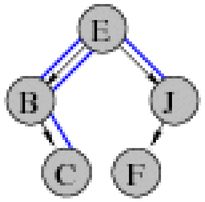
Path Length

Path length of the following tree: 1 + 1



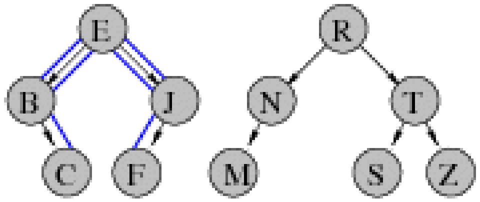
Path Length

Path length of the following tree: 1 + 1 + 2



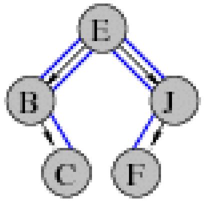
Path Length

Path length of the other tree:



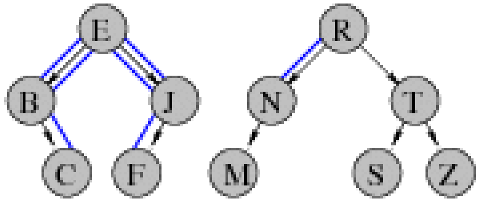
Path Length

Path length of the following tree: 1 + 1 + 2 + 2 = 6



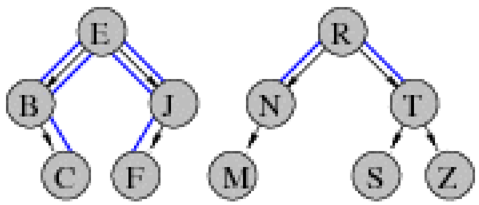
Path Length

Path length of the other tree: 1



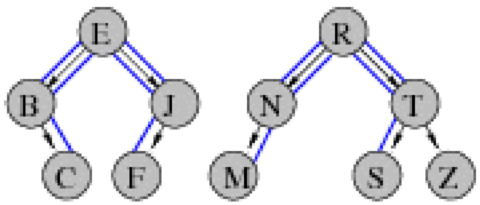
Path Length

Path length of the other tree: 1 + 1



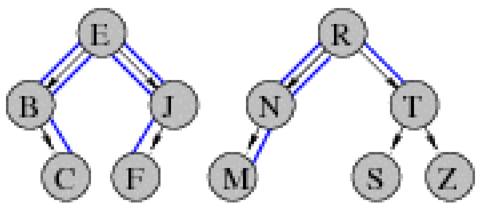
Path Length

Path length of the other tree: 1 + 1 + 2 + 2



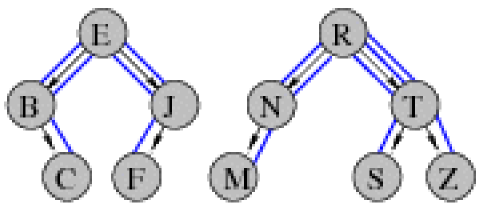
Path Length

Path length of the other tree: 1 + 1 + 2

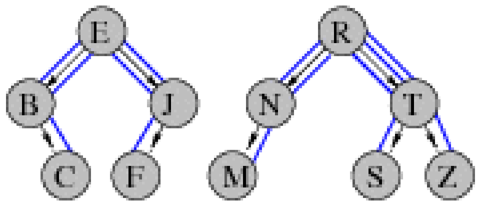


Path Length

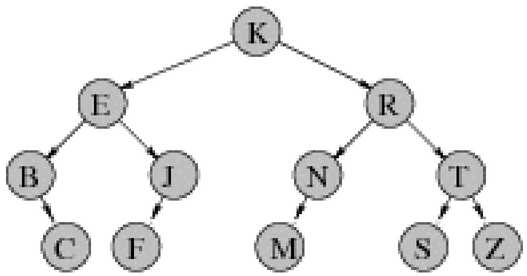
Path length of the other tree: 1 + 1 + 2 + 2 + 2 = 8



Path Length

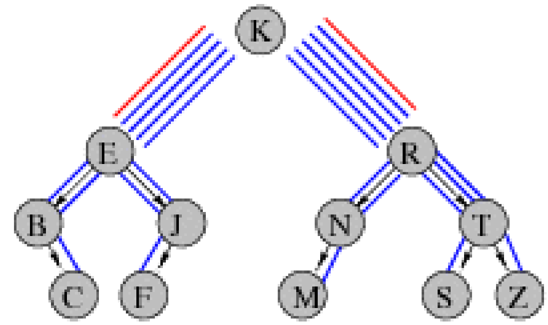


What is the path length of the following:



Path Length

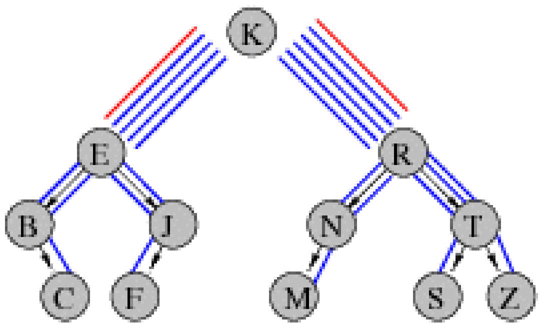
What is the path length of the following:



Path length of the left tree + Number of nodes in the left tree
Path length of the right tree + Number of nodes in the right tree

Path Length

What is the path length of the following:



Path length of the left tree + (Number of nodes in the left tree - 1)
Path length of the right tree + (Number of nodes in the right tree - 1) +
1 + 1

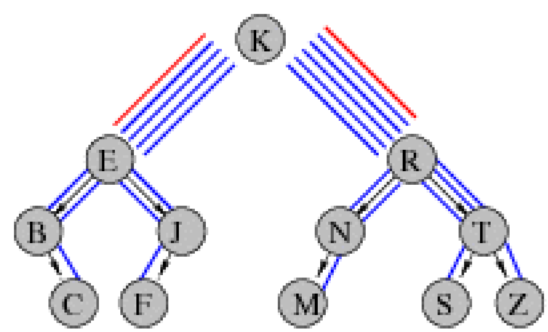
Average Case Analysis (This Is A First For Us)

$$D(N) = \text{Internal path length of a tree with } N \text{ nodes}$$

$$D(1) = 0$$

$$\text{For } 0 \leq i < N, D(N) = D(i) + D(N - i - 1) + (N - 1)$$

$$\text{e.g. } D(14) = D(5) + D(6) + (12)$$



Average Case Analysis (This Is A First For Us)

$$D(1) = 0$$

$$\text{For } 0 \leq i < N, D(N) = D(i) + D(N-i-1) + (N-1)$$

$$D(N) = \frac{2}{N}\left(\sum_{j=0}^{N-1} D(j)\right) + (N-1)$$

Average Case Analysis (This is a First For Us)

- Could we have solved the following:

$$D(N) = \frac{2}{N}\left(\sum_{j=0}^{N-1} D(j)\right) + (N-1)$$

  No! Did we thoroughly understand its derivation? Maybe...

- What is the missing piece? Understanding expected values.
- When will we study expected value? Right before hash tables (that is next week, boys and girls).
- What will we do in the mean time? See the coding question of the problem set due next Monday.

Average Case Analysis (This is a First For Us)

$$D(N) = \frac{2}{N}\left(\sum_{j=0}^{N-1} D(j)\right) + (N-1)$$

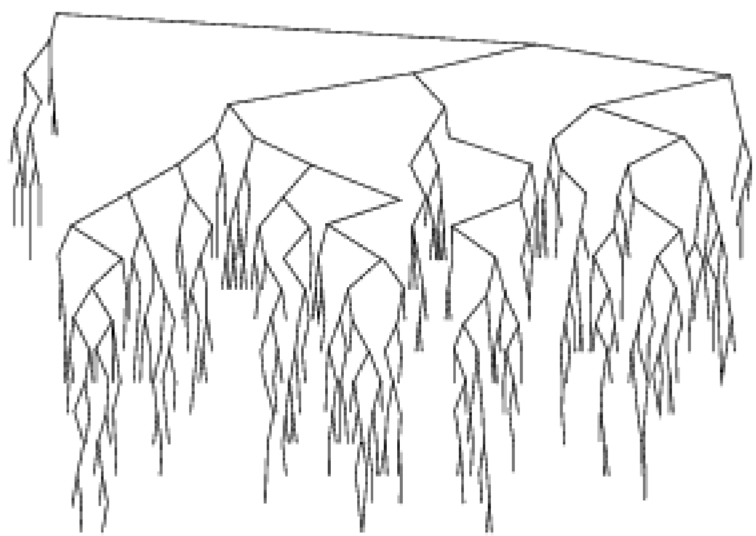
This beauty is a recurrence relation that tells us the expected internal path length of a binary search tree. Do we know its closed form?

$$D(N) = O(N \lg N)$$

Therefore, the expected height is

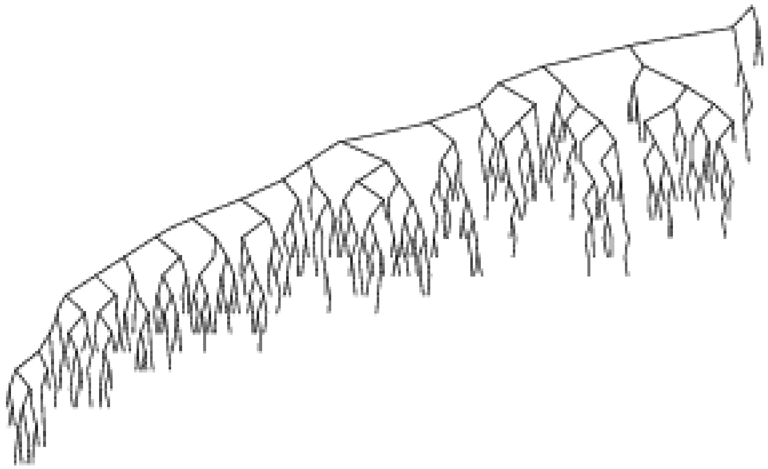
$$H(N) = \frac{D(N)}{N} = \frac{O(N \lg N)}{N} = O(\lg N)$$

Randomly Generated 500 Node BST



Previous result tells us depth of above should be around $\lg 500 \approxeq 9$

BSTTree After $\Theta(N^2)$ `insert()/remove()` Calls



Anything disturbing about the above? It is rumored $H(N) = O(\sqrt{N})$ after random sequences of `insert()/remove()` calls.
In Shakespearean terms, degeneration without ordered input.

What Do We Do?

Enter AVL Trees...