

Tree Data Structures

261217 Data Structures for Computer Engineers

Patiwet Wuttisarnwattana, Ph.D.

patiwet@eng.cmu.ac.th

Computer Engineering, Chiang Mai University

Why do we need Tree?

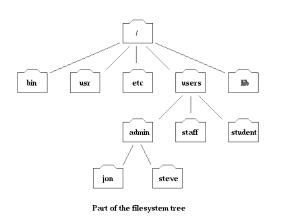
- Lists, Stacks, and Queues are linear relationships
 - Accessing time can be super fast O(1)
 - But searching for a key takes O(n)
 - Some say this is too slow
 - Can you find something that the search is faster than O(n)
- Well, Tree can be the answer

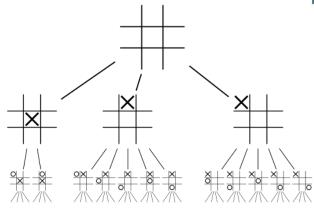
What is Tree?

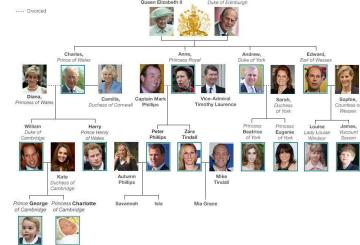
- ■What is Tree
 - Tree consists of nodes that contain data
 - ■Tree can contain no node (empty)
 - A node can point to one or more nodes
 - A node can point to null (leaf node)
 - A node that no other node point to it; is called **root** node
 - Tree contains no loop/cycle of nodes

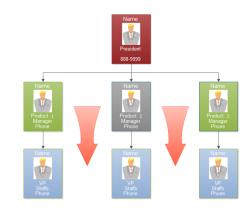
Trees

- Information often contains hierarchical
 - relationships
 - Family
 - File directories or folders
 - Moves in game
 - Hierarchies in organizations





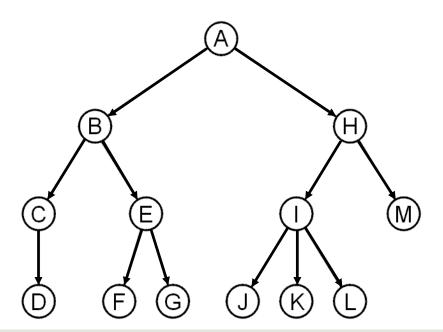




Tree Terminology

A tree data structure stores information in *nodes*

- Similar to linked lists:
 - There is a first node, or *root*
 - Each node has variable number of references to successors
 - Each node, other than the root, has exactly one node pointing to it
 - In the diagram, the line that connected between nodes is called edge
 - The edge should be directed (either be unidirected or bidirected)



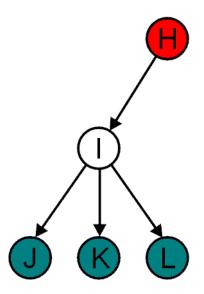
Parent and Child

All nodes will have zero or more child nodes or children

I has three children: J, K and L

For all nodes other than the root node, there is one parent node

- H is the parent I

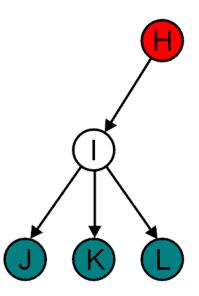


Node Degree

The *degree* of a node is defined as the number of its children: $deg(\mathbf{I}) = 3$

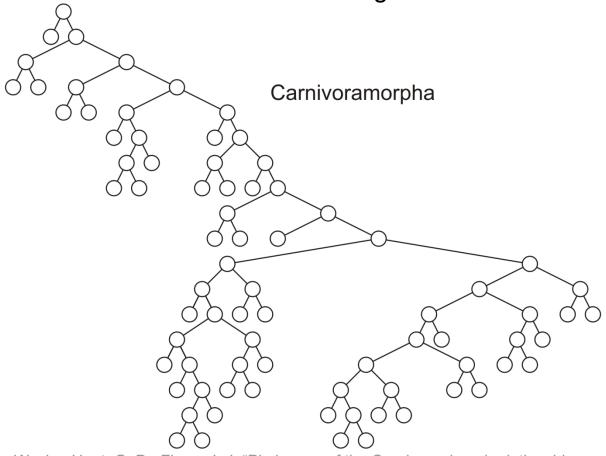
Nodes with the same parent are *siblings*

J, K, and L are siblings



Phylogenetic tree

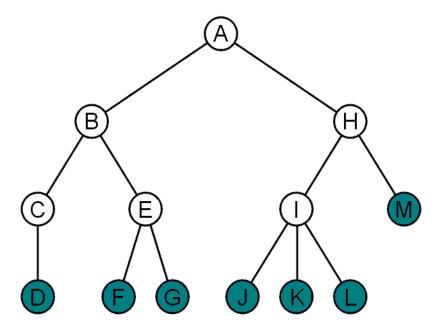
Phylogenetic trees have nodes with degree 2 or 0:



Leaf Nodes and Internal Nodes

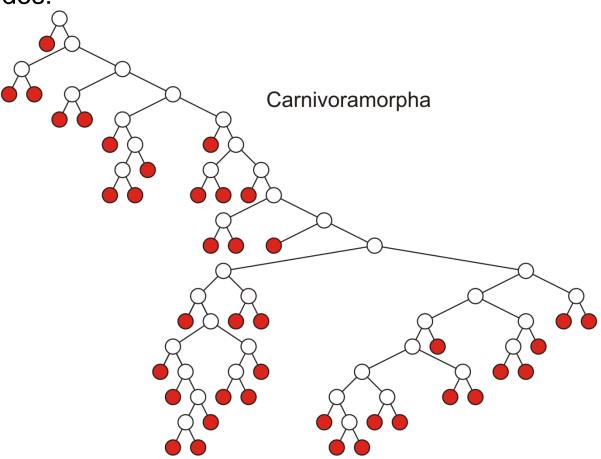
Nodes with degree zero are also called *leaf nodes*

All other nodes are said to be *internal nodes*, that is, they are internal to the tree



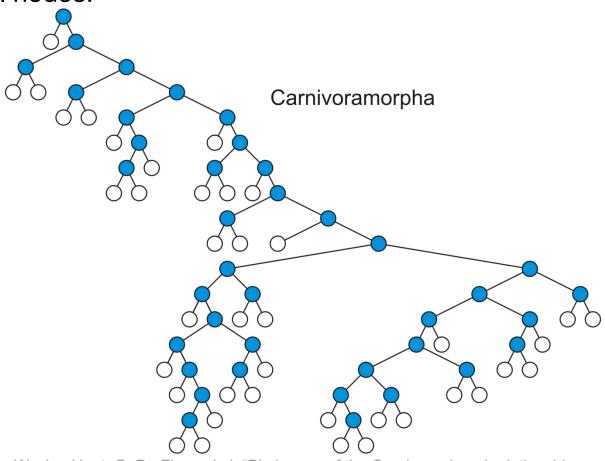
Leaf Nodes

Leaf nodes:



Internal Nodes

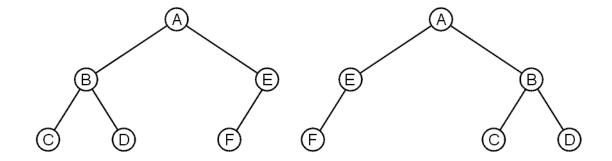
Internal nodes:



Unordered vs Ordered Trees

These trees are equal if the order of the children is ignored

unordered trees

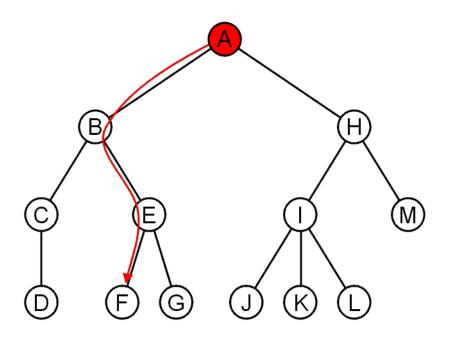


They are different if order is relevant (*ordered trees*)

- We will usually examine ordered trees (linear orders)
- In a hierarchical ordering, order is not relevant

Root node

- The shape of a rooted tree gives a natural flow from the root node, or just root
- Every node grows from the root
- Tree in Data Structure grows upside down



Path

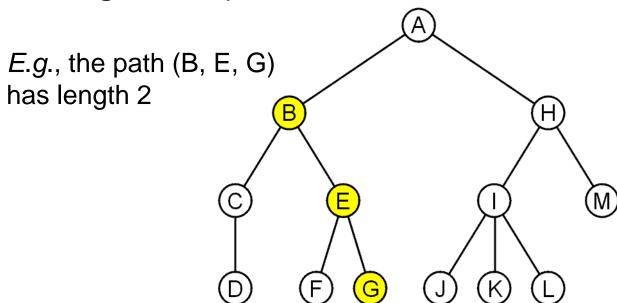
A **path** is a sequence of nodes

$$(a_0, a_1, ..., a_n)$$

where a_{k+1} is a child of a_k is

The **length** of this path is n

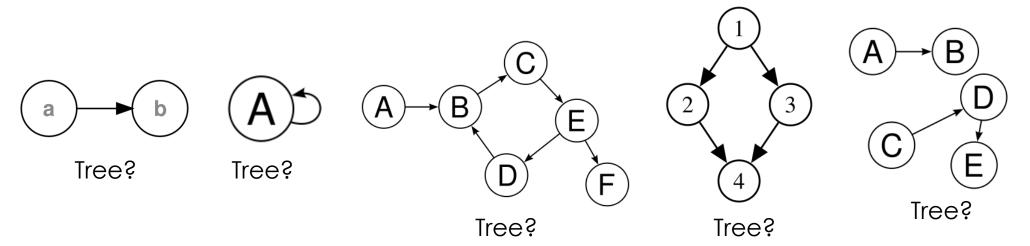
A tree with N nodes always has N-1 edges



Loops in Tree?

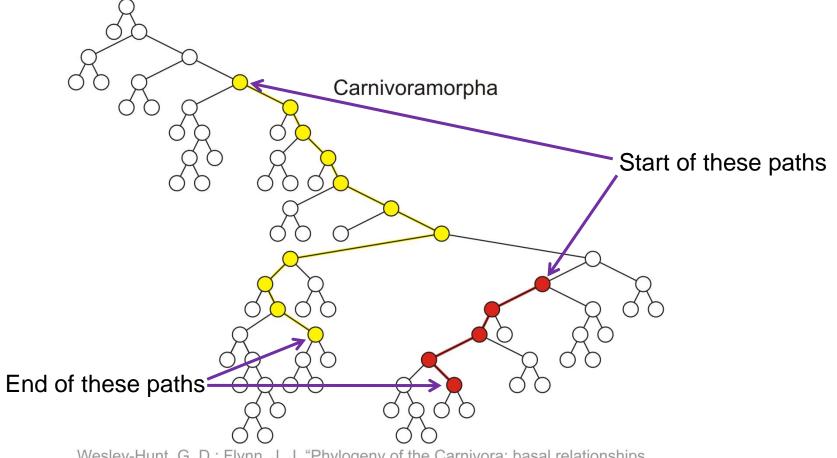
- Two nodes in a tree have at most one path between them
- Can a non-zero path from node N reach node N again?

No. Trees can never have cycles (loops)



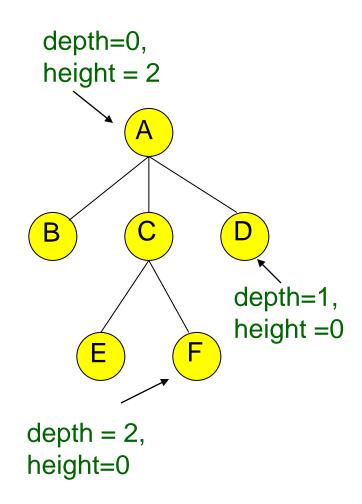
Path Example

Paths of length 10 (11 nodes) and 4 (5 nodes)



Distance Measurement

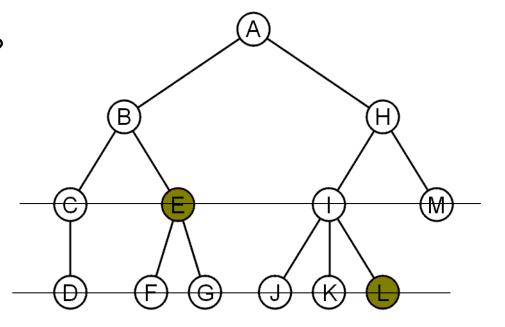
- Length of a path = number of edges
- Depth of a node N = length of path from root to N
- Height of node N = length of longest path from N to a leaf
- Depth of tree = depth of deepest node
- Height of tree = height of root



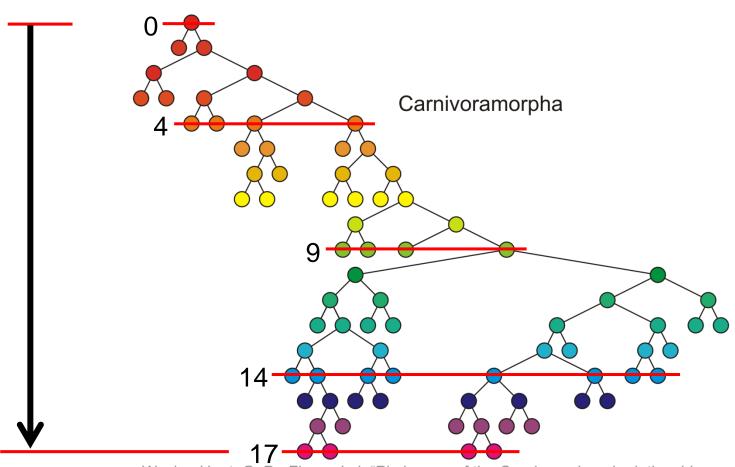
Node Depth

For each node in a tree, there exists a unique *path from the root* node to that node

- What is depth of the node E?
 - **2**
- What is depth of the node L?
 - **3**



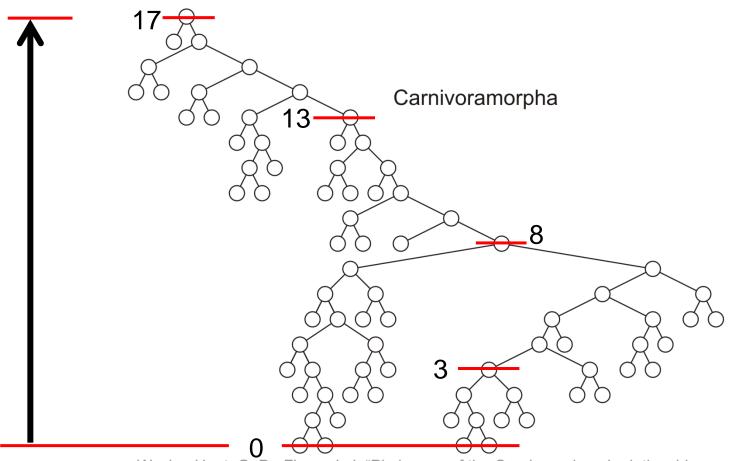
Node Depth and Tree Depth



Height of Node and Tree

- The height of a node is the path length from the node to the deepest descendant.
- The height of a tree is the path length from root node to the deepest leaf.
- The height of a tree with a single node is 0
- The height of a leaf node is 0
- The height of a tree is equal to the depth of the tree
- For convenience, we define the height of the empty tree to be -1

Node Height vs Tree Height



Ancestor and Descendent

If a path exists from node *a* to node *b*:

- a is an ancestor of b
- b is a descendent of a

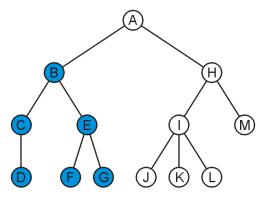
Thus, a node is both an ancestor and a descendant of itself

- We can add the adjective *strict* to exclude equality: a is a *strict* descendent of b if a is a descendant of b but $a \neq b$

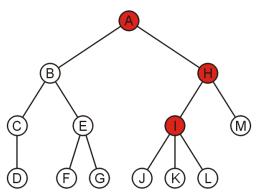
The root node is an ancestor of all nodes

Ancestor and Descendent

The descendants of node B are B, C, D, E, F, and G:

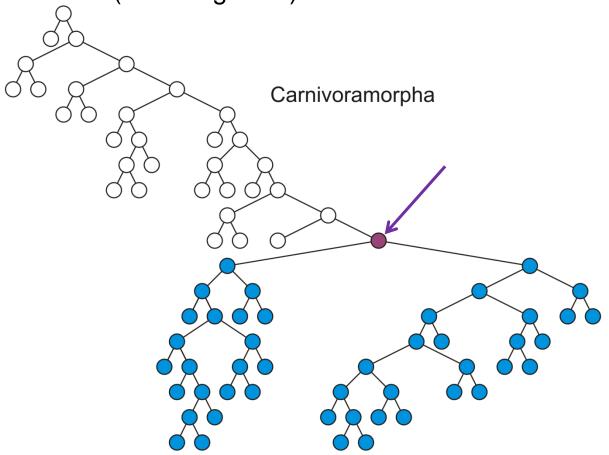


The ancestors of node I are I, H, and A:



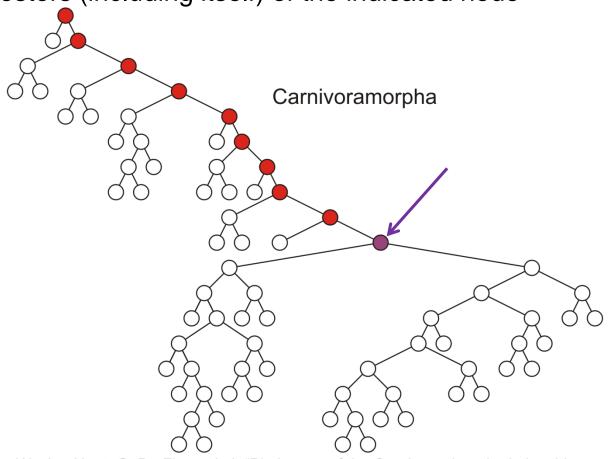
Descendants

All descendants (including itself) of the indicated node



Ancestors

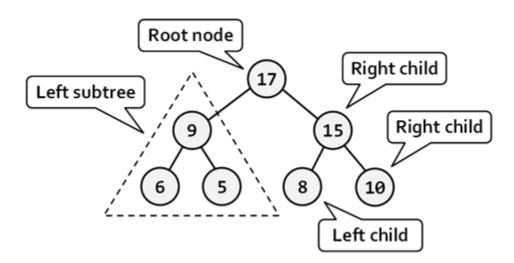
All ancestors (including itself) of the indicated node



Alternative Definition of Tree

Another approach to a tree is to define the tree recursively:

- A tree can be represented by a root note
- Any node can be a root node
- A child node of a node is also a root node of a tree (sub-tree)



Example: XHTML

- The XML of XHTML has a tree structure
- Consider the following XHTML document

```
<html>
<head>
<title>Hello World!</title>
</head>
<body>
<h1>This is a <u>Heading</u></h1>
This is a paragraph with some
<u>u>underlined</u> text.
</body>
</html>
```

Example: XHTML

- The XML of XHTML has a tree structure
- Consider the following XHTML document



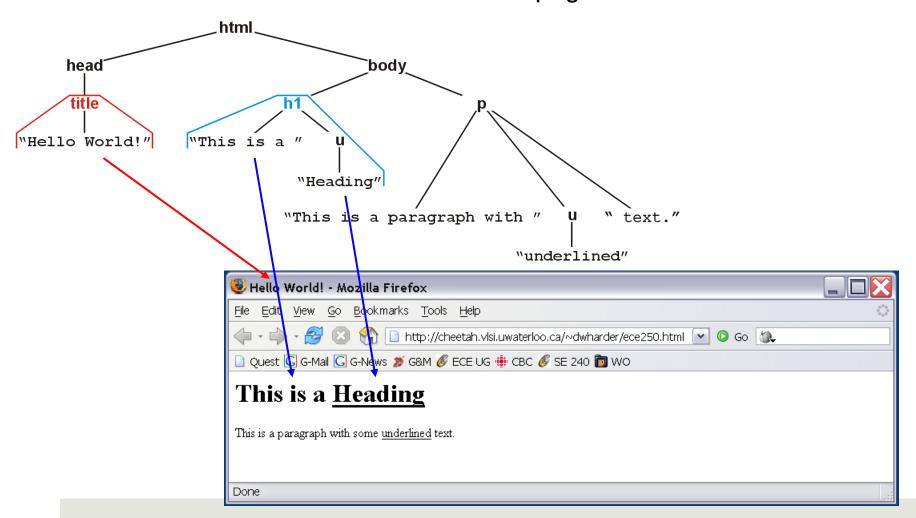
Converting XML to Tree Data Structure

The nested tags define a tree rooted at the HTML tag

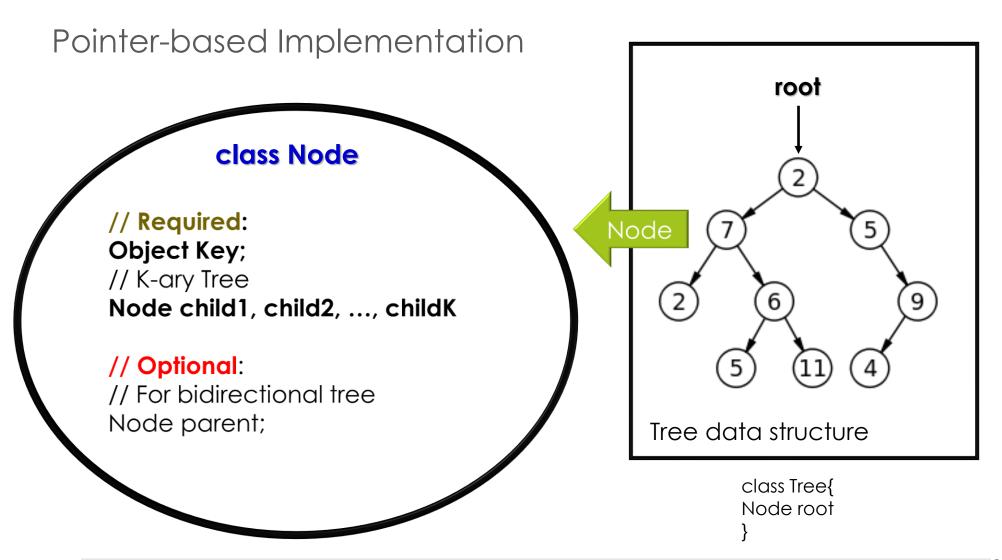
```
<html>
    <head>
         <title>Hello World!</title>
    </head>
    <body>
         <h1>This is a <u>Heading</u></h1>
        This is a paragraph with some
        <u>underlined</u> text.
                         html
    </body>
</html>
         head<sup>2</sup>
                                         body
         title
                      "This is a "
   "Hello World!"
                                  "Heading"
                                                                  " text."
                                "This is a paragraph with "
                                                         "underlined"
```

Example: XHTML

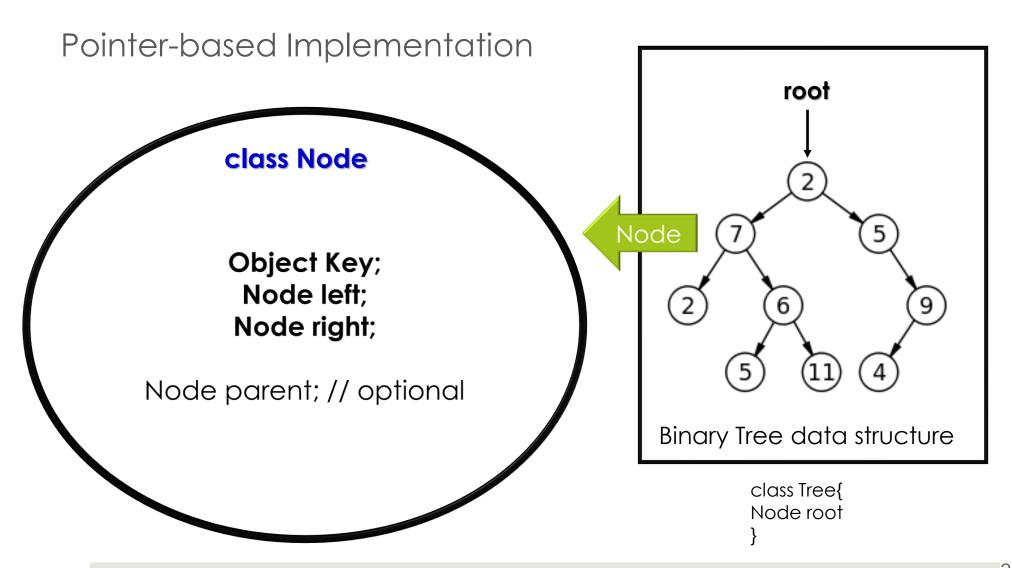
Web browsers render this tree as a web page



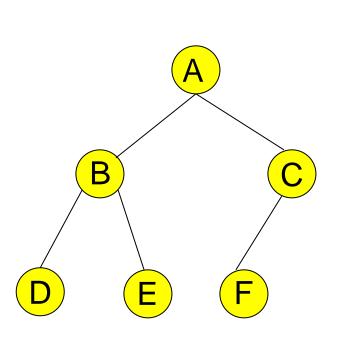
Implementation of Trees (K-ary Tree)

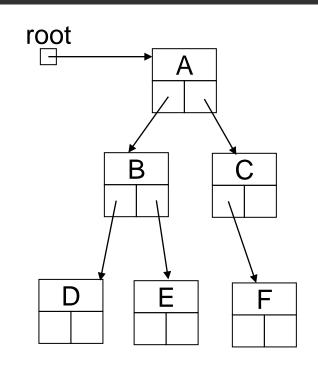


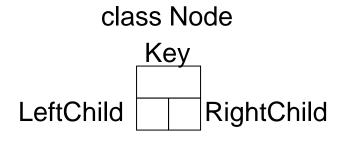
Implementation of Binary Trees



Binary Tree Structure (applicable to K-nary tree)





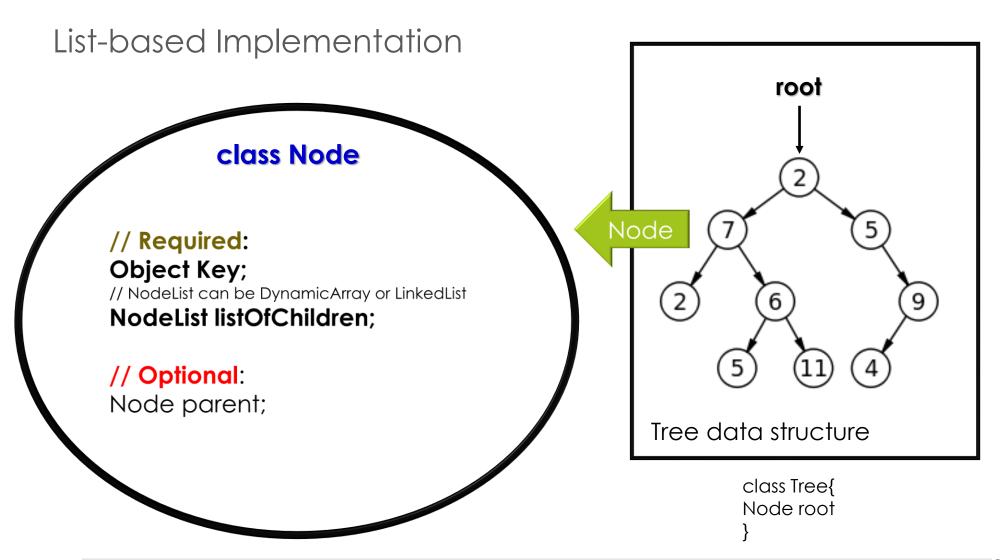


Potential issue?

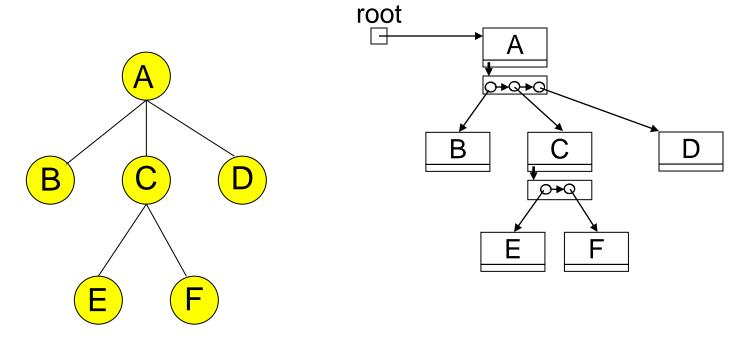
```
class Node
// Required:
Object Key;
// K-ary Tree
Node child1, child2, ..., childK
// Optional:
// For bidirectional tree
Node parent;
```

- The implementation is not flexible (It is not a generic tree)
- What should be k? Magic!
- If the tree has different number of possible children (> k), the class will need to be re-implemented (useless code; add more variables)

Version 2: Use List of Nodes instead of a fixed number of pointers to Nodes



List-based Implementation



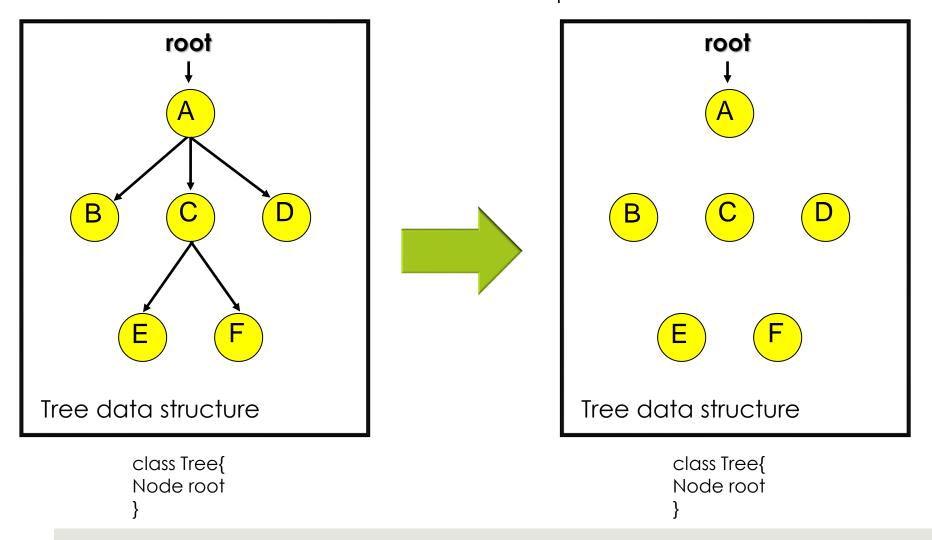
- Key
 ListOfChildren
- Now that each node can have a flexible number of children
- Slightly confusing for implementing
 - Need loop for accessing the children
 - Tree traversal/search implementation might be complicated

Potential issues with Version 2?

- It's quite complicated
 - Accessing the children requires implementation of loop
 - Tree traversal implementation could be complicated
- □ In general sense, siblings should know each other
 - Version 2: A node knows only its children; or its parent (if you implement the parent pointer)
 - It does not directly know its siblings; need to go up one step to the parent and then go down to the sibling nodes.
 - Solution is to implement sibling pointer <u>but</u> node cycle may form and tree data structure will be violated
 - Version 3 is introduced

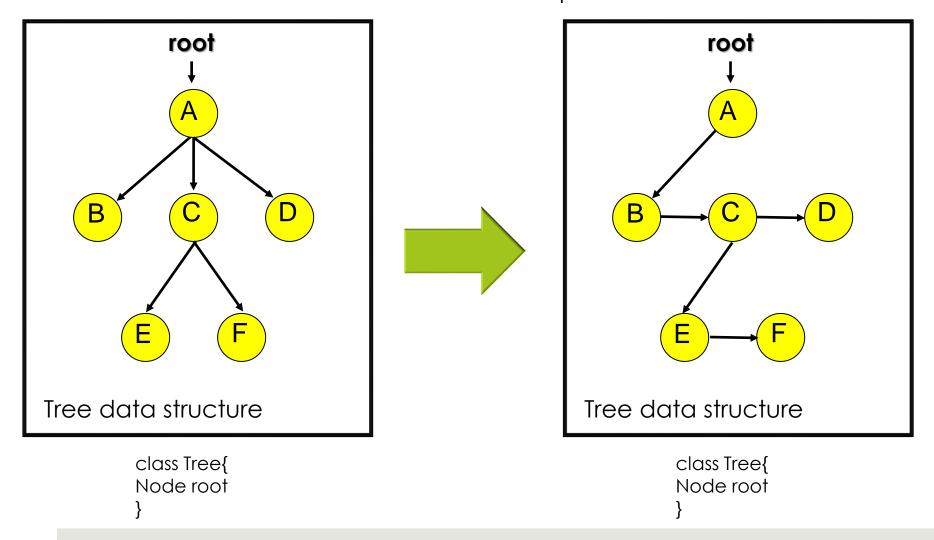
Instead of multiple pointers/list, let's fix to two pointer variables

How to re-define a node so that it can point to at most two nodes

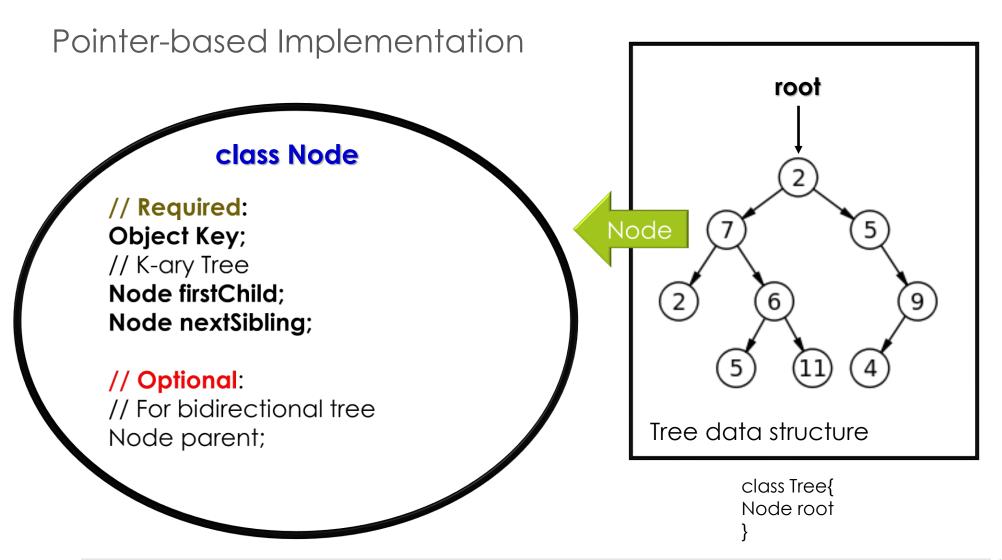


Instead of multiple pointers/list, let's fix to two pointer variables

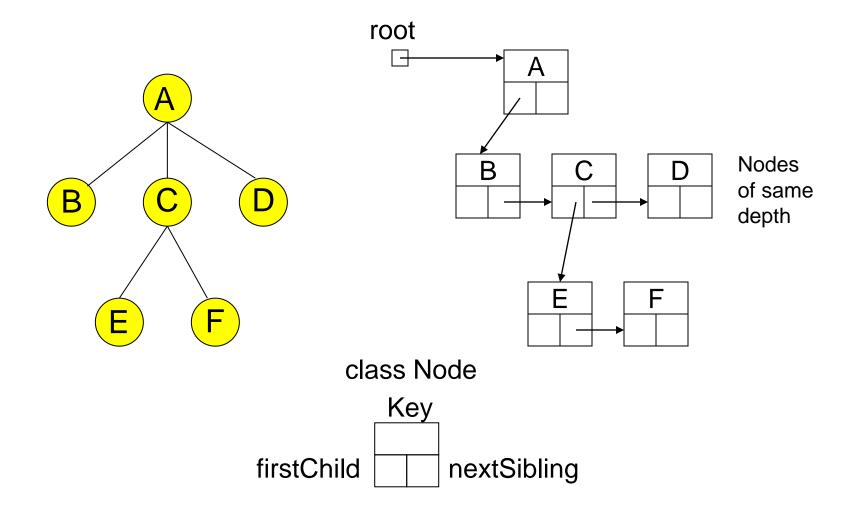
How to re-define a node so that it can point to at most two nodes



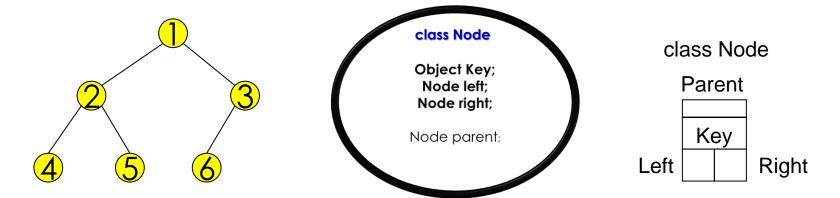
Version 3: Generic Tree Implementation



Generic Tree Implementation



Binary Tree



- Every node has at most two children
- Most popular tree in computer engineering field
- ☐ If I do not say otherwise, we will assume that we are talking about Binary Trees for the rest of the course.
- \square Given a binary tree with depth **d**, what is the possible number of nodes (**N**)?
- Given a number of nodes N, what is the minimum depth of a binary tree?

Minimum depth vs Node count

 At depth d, you can have N = 2^d to 2^{d+1}-1 nodes

minimum depth d is ⊕(log N)

```
T(n) = \Theta(f(n)) means

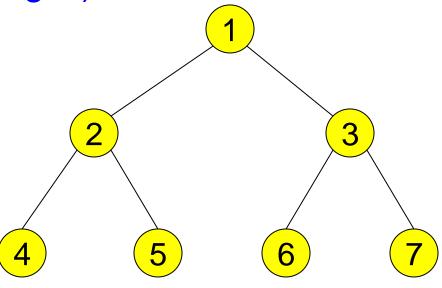
T(n) = O(f(n)) and f(n) = O(T(n)),

i.e. T(n) and f(n) have the same

growth rate

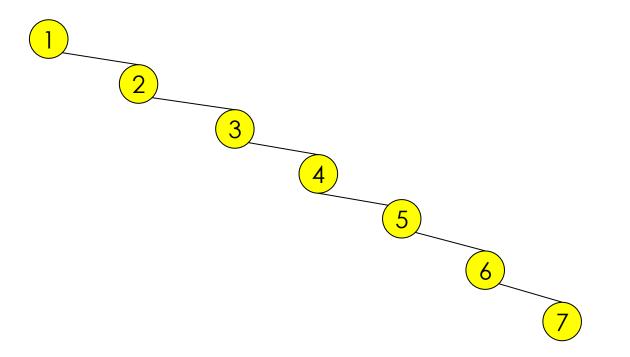
d=2
```

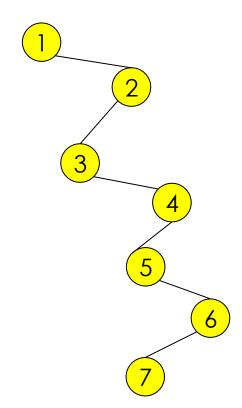
 $N=2^2$ to 2^3-1 (i.e, 4 to 7 nodes)



Minimum depth vs Node count

- What is the maximum depth of a binary tree?
 - Degenerate case: Tree is a linked list!
 - \blacksquare Maximum depth = N 1

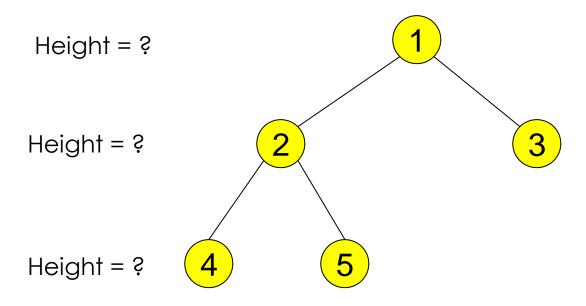




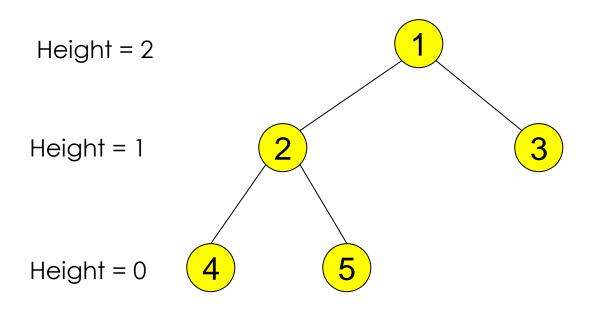
Thought Questions

- □ If a target key is often stored at the leaf nodes
 - What is the complexity of the searching for a key in a tree with the maximum depth (degenerate tree)?
 - What is the complexity of the searching for a key in a tree with the minimum depth (complete tree)?

Height Measurement



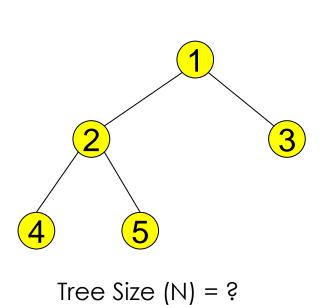
Height Measurement Algorithm

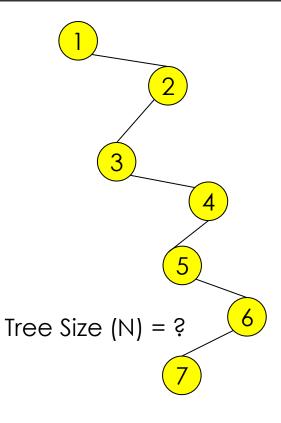


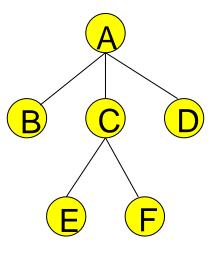
int Height(Node node)

```
if node == null
    return -1
else
    return 1 + max( Height(node.left), Height(node.right) )
```

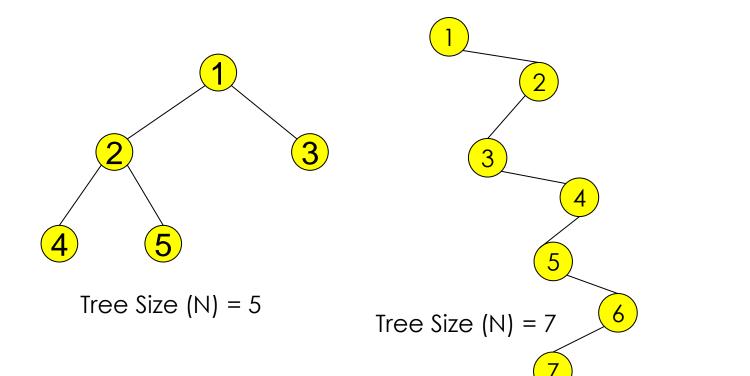
Tree Size Measurement

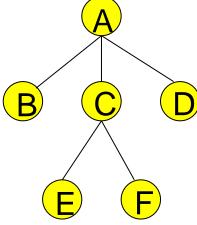






Tree Size Measurement Algorithm





Tree Size (N) = 6

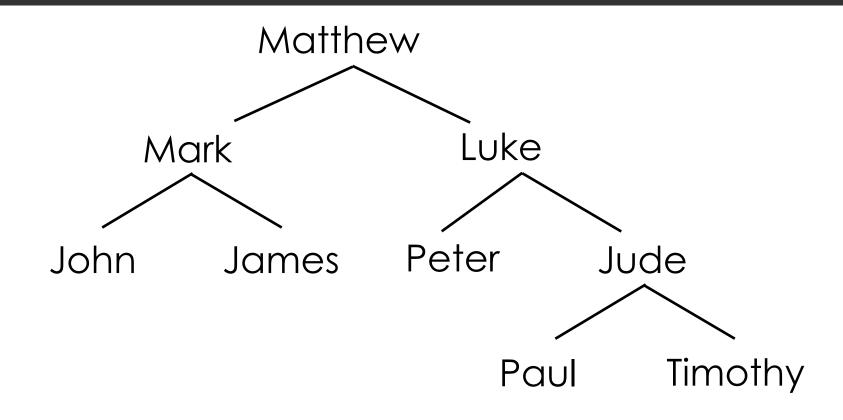
int Size(Node node)

```
if node == null
    return 0
else
    return 1 + Size(node.left) + Size(node.right)
```

Tree Traversal (Walking a Tree)

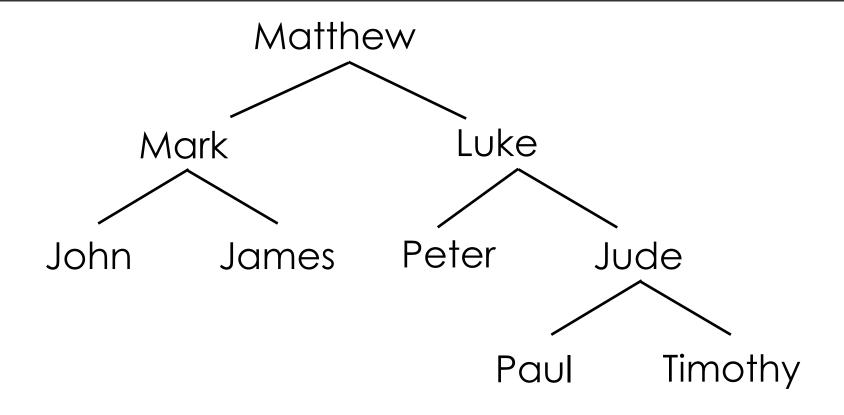
- We want to visit all nodes of the tree
 - Searching for a key in an unordered tree (O(n))
 - Re-evaluate values of one or more nodes
 - \square Print all the values contains within the tree $(\Theta(n))$
- Traversal order: Breadth-first and Depth-first
 - **Breadth-first**: We traverse all nodes at one level before progressing to the next level
 - e.g. Users tend to place files are usually places at the earlier folders rather than deeper subfolder
 - Depth-first: We completely traverse one sub-tree to before exploring a sibling sub-tree
 - e.g. Important data often stored at the leaves of the tree (such as operands) which should be read first before the internal nodes (operators)

Breadth-first Traversal (Level Traversal)



Output:

Breadth-first Traversal (Level Traversal)

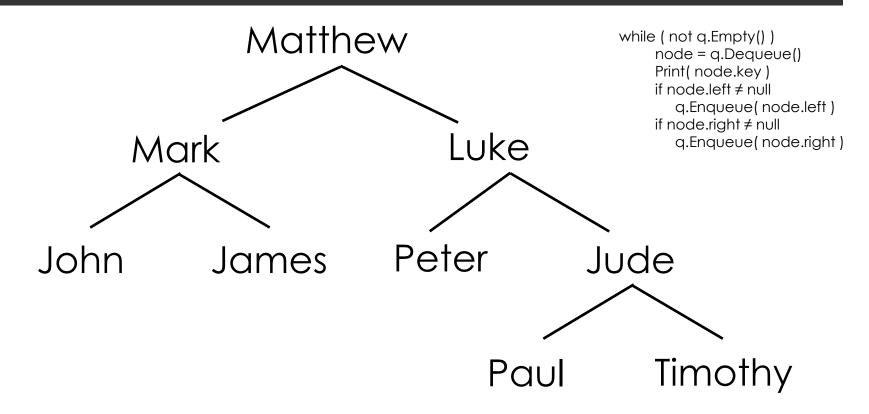


Output: Matthew, Mark, Luke, John, James, Peter, Jude, Paul, Timothy

Breadth-first Traversal implementation using Queue

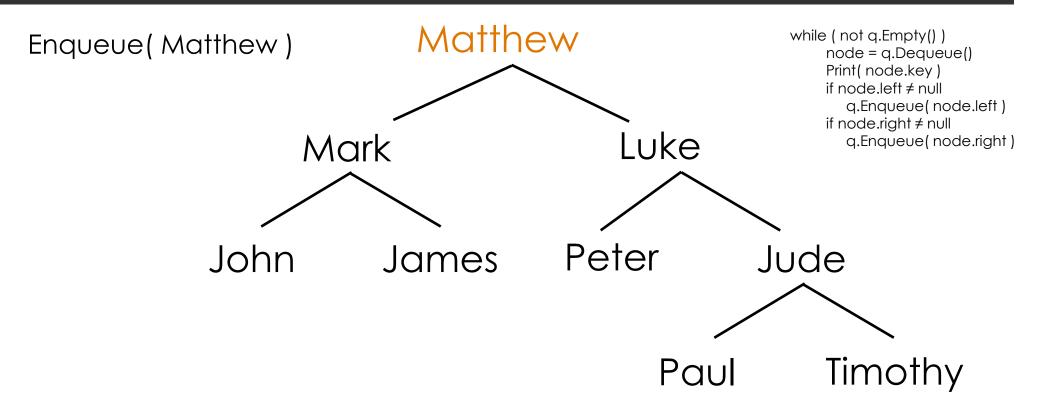
LevelTraversal (Node node)

```
if node == null
   return
else
   Queue q
   a.Enqueue(node)
   while (not q.Empty())
      node = q.Dequeue()
      Print( node.key ) // Do something to the node
      if node.left ≠ null
         q.Enqueue( node.left )
      if node.right ≠ null
         q.Enqueue( node.right )
```



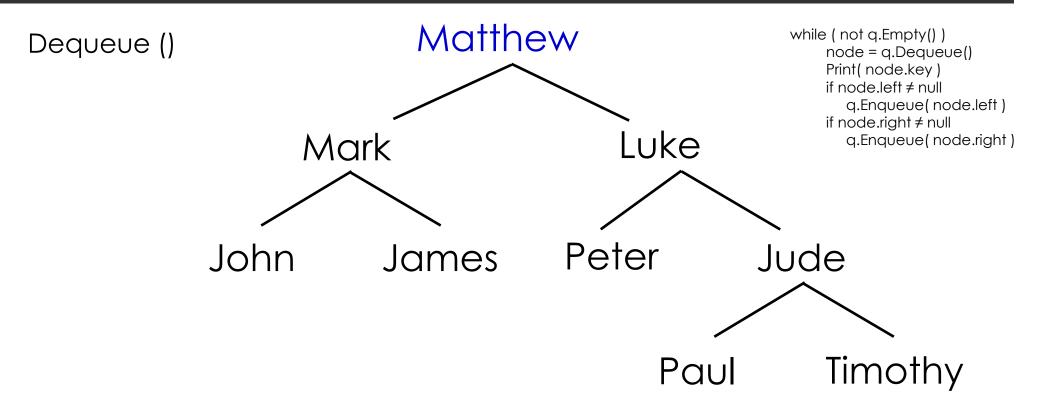
Output:

Queue:



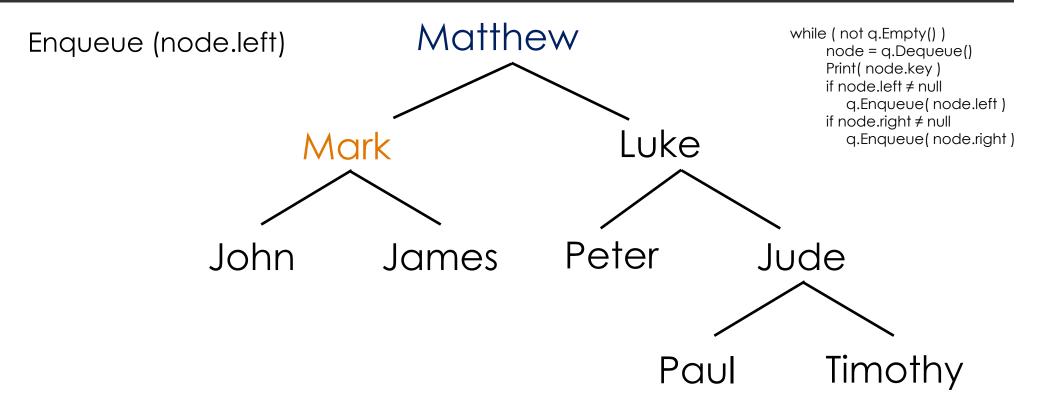
Output:

Queue: Matthew



Output: Matthew

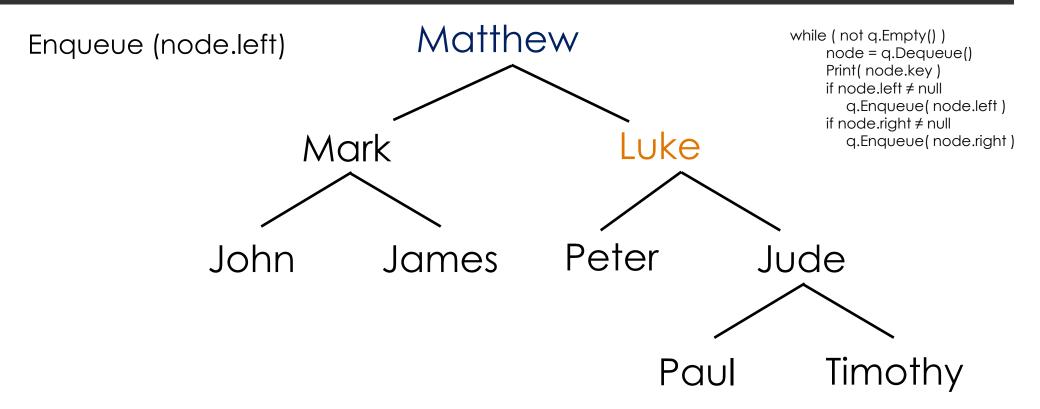
Queue:



Output: Matthew

node

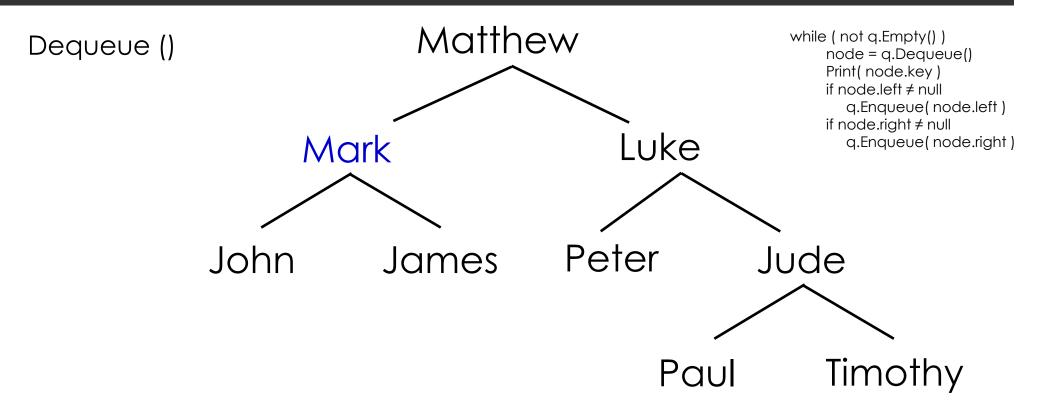
Queue: Mark



Output: Matthew

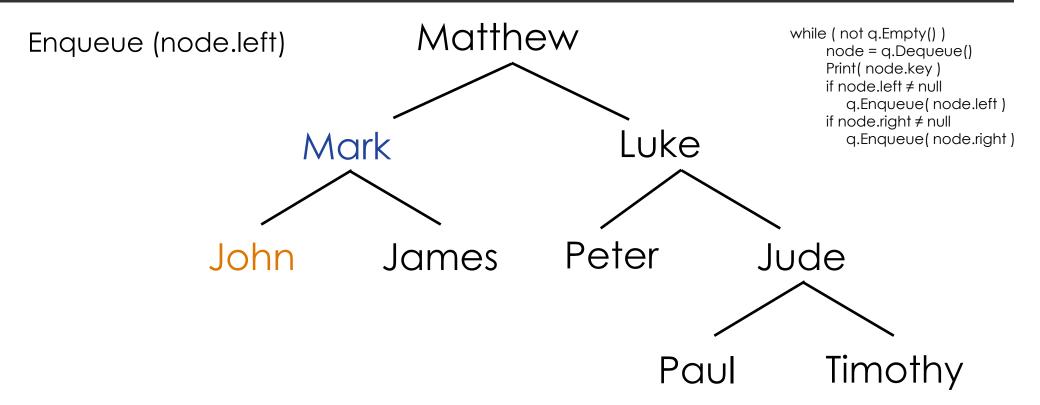
node

Queue: Mark, Luke



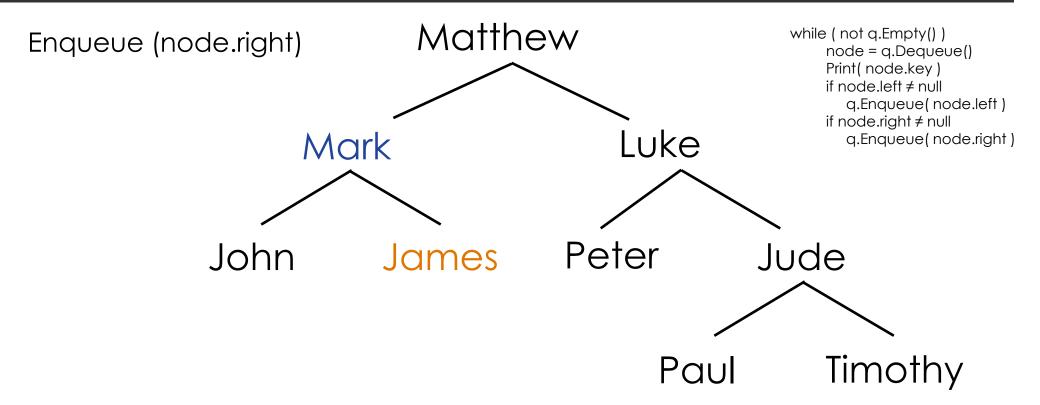
Output: Matthew, Mark

Queue: Luke



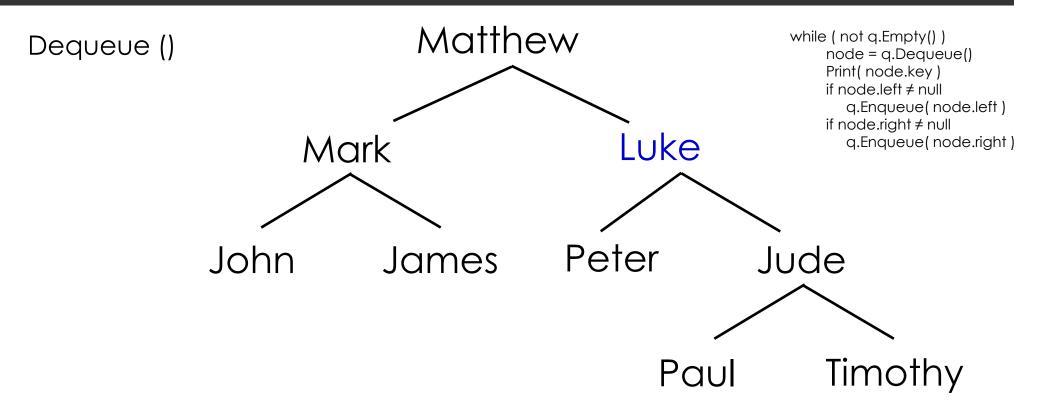
Output: Matthew, Mark

Queue: Luke, John



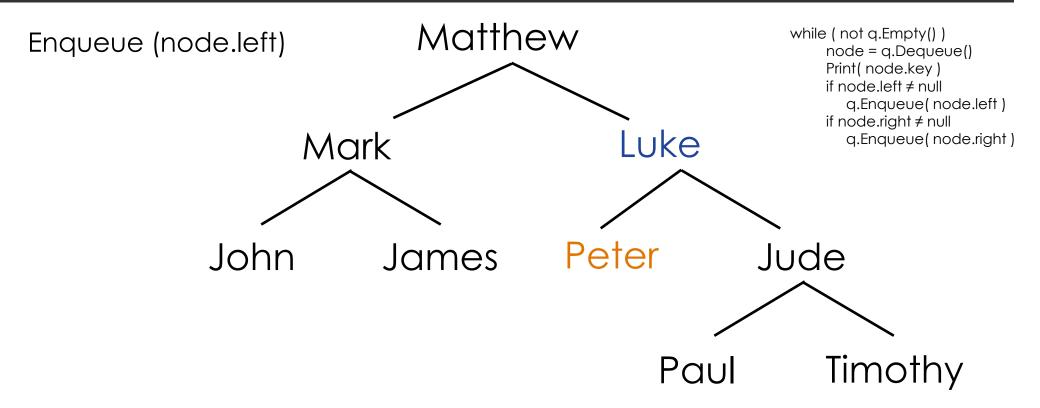
Output: Matthew, Mark

Queue: Luke, John, James



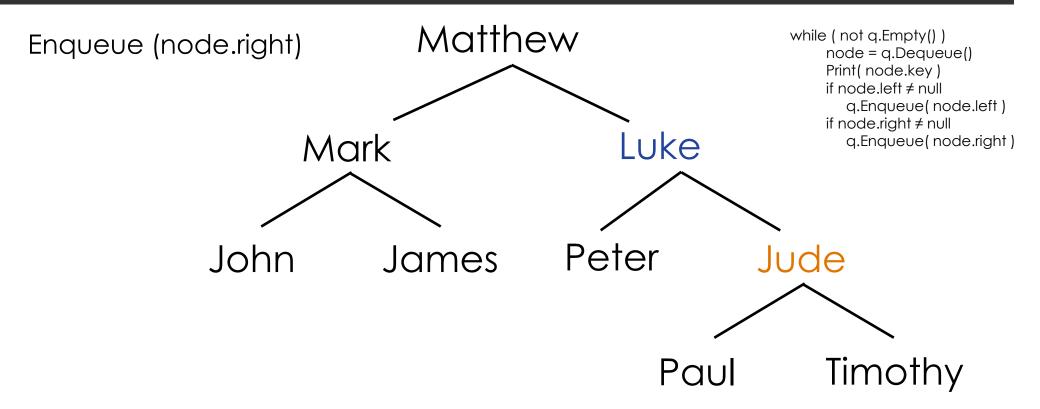
Output: Matthew, Mark, Luke

Queue: John, James



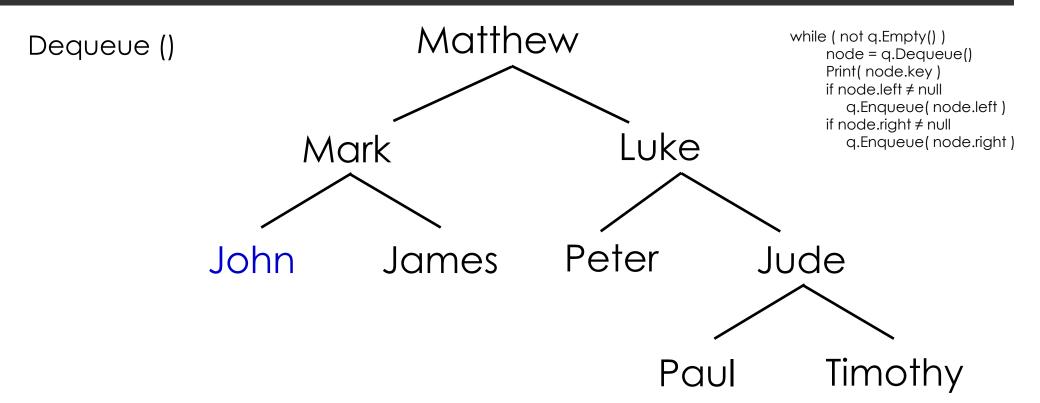
Output: Matthew, Mark, Luke

Queue: John, James, Peter



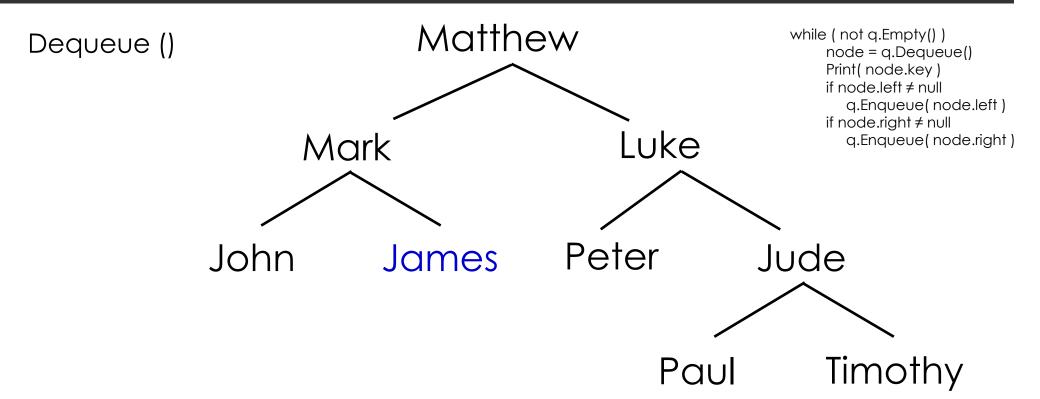
Output: Matthew, Mark, Luke

Queue: John, James, Peter, Jude



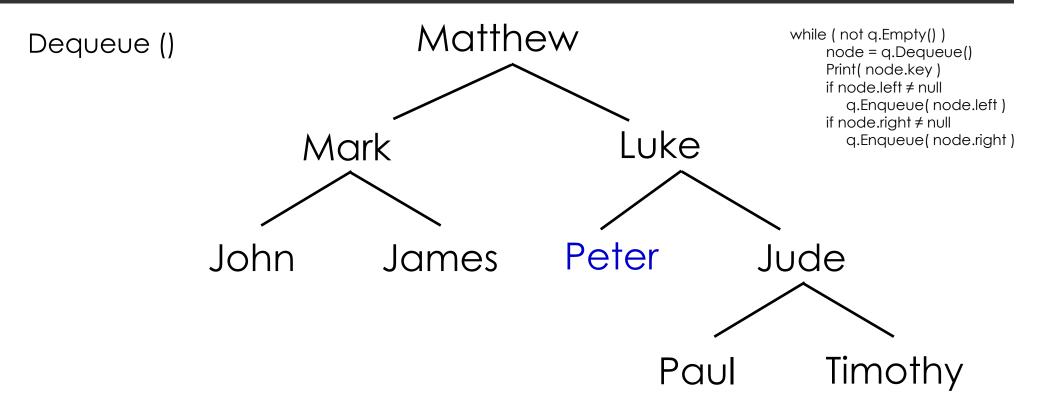
Output: Matthew, Mark, Luke, John

Queue: James, Peter, Jude



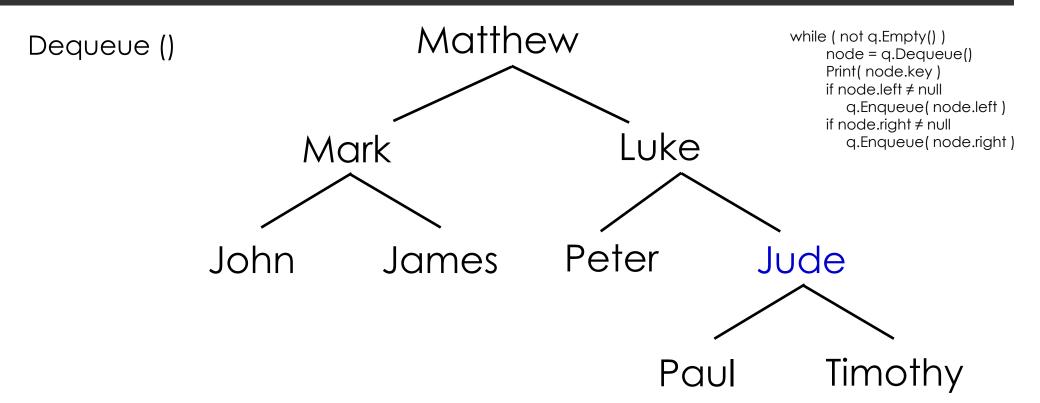
Output: Matthew, Mark, Luke, John, James

Queue: Peter, Jude



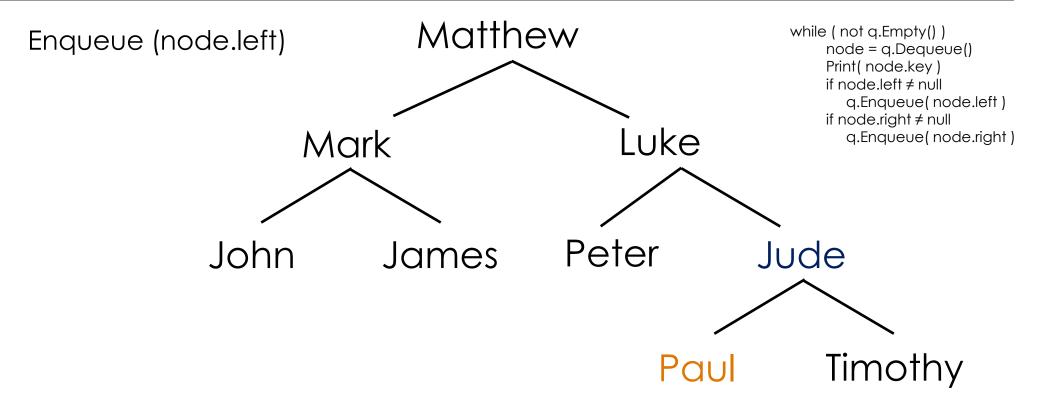
Output: Matthew, Mark, Luke, John, James, Peter

Queue: Jude



Output: Matthew, Mark, Luke, John, James, Peter, Jude

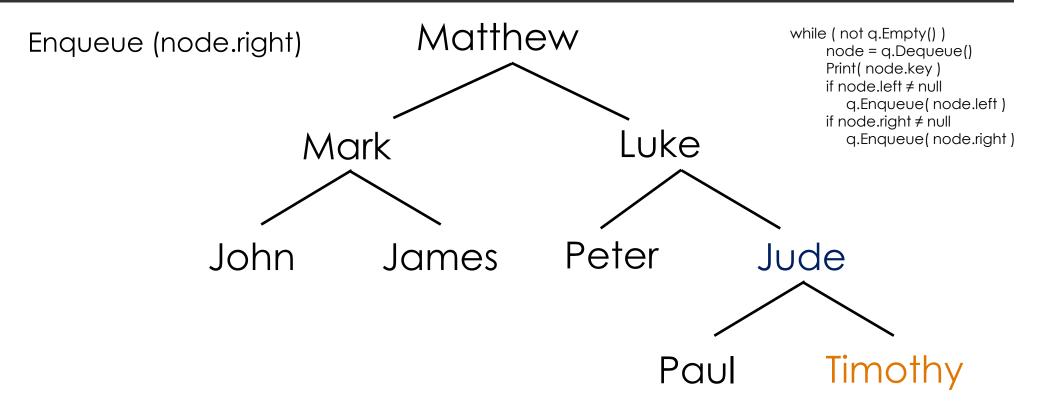
Queue:



Output: Matthew, Mark, Luke, John, James, Peter,

Jude

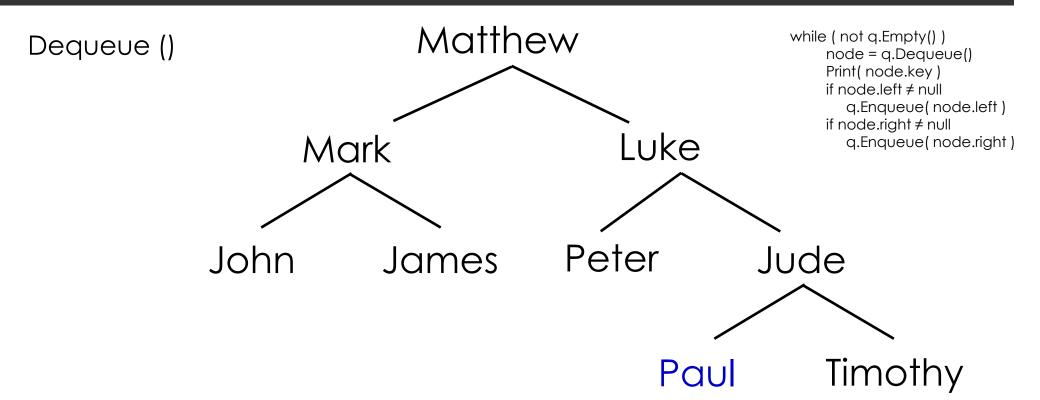
Queue: Paul



Output: Matthew, Mark, Luke, John, James, Peter,

Jude

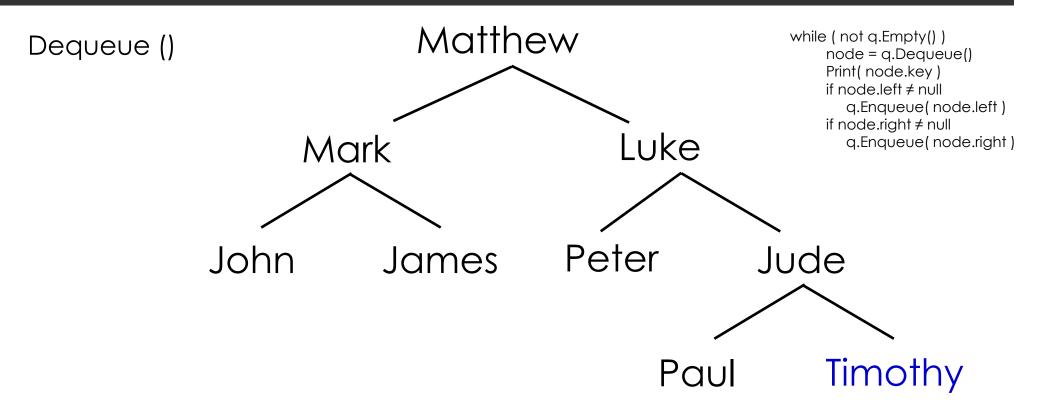
Queue: Paul, Timothy



Output: Matthew, Mark, Luke, John, James, Peter,

Jude, Paul

Queue: Timothy

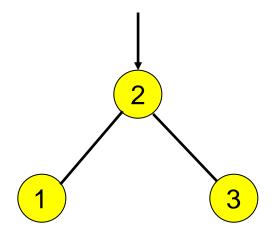


Output: Matthew, Mark, Luke, John, James, Peter, Jude, Paul, Timothy

Queue:

Depth-first Traversal Implementation using Recursion

- We completely traverse one sub-tree to before exploring a sibling sub-tree
- We can implement DFT using
 - Stack
 - Recursion
- Three modes of traversal
 - PreOrder, InOrder, PostOrder



PreOrder: 2, 1, 3

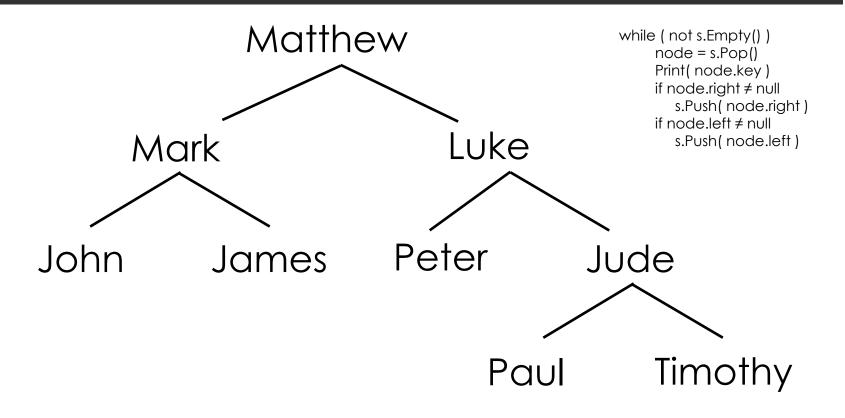
InOrder: 1, 2, 3

PostOrder: 1, 3, 2

Non-recursive Implementation of Depth-first Traversal

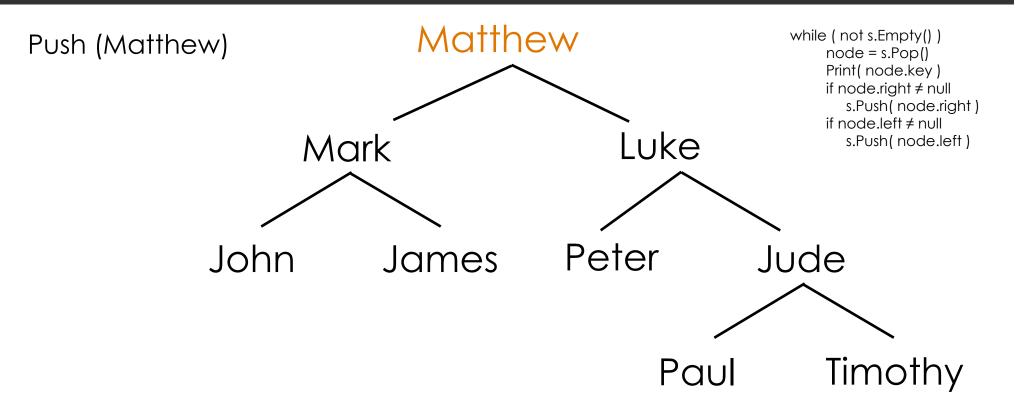
nonRecursiveDFT (Node node)

```
if node == null
   return
else
   Stack s
   s.push( node )
   while (not s.Empty())
      node = s.Pop()
      Print (node.key) // Do something to the node
      if node.right ≠ null
         s.push(node.right)
      if node.left ≠ null
         s.push(node.left)
```



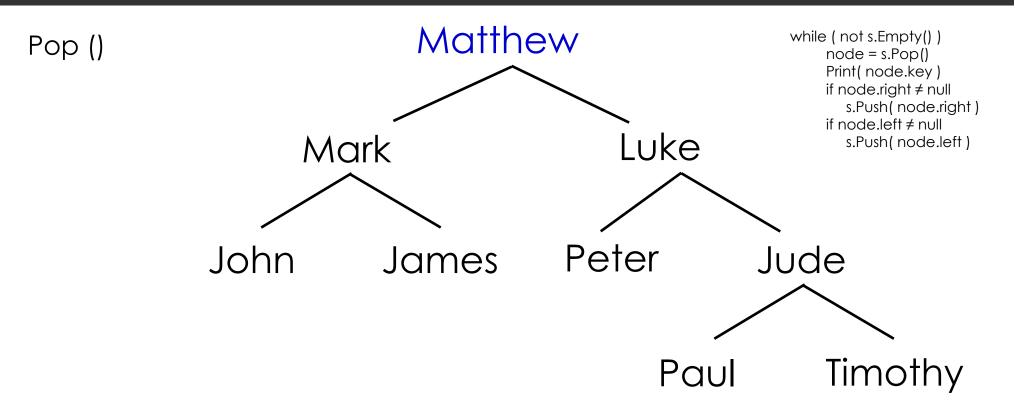
Output:

Stack:



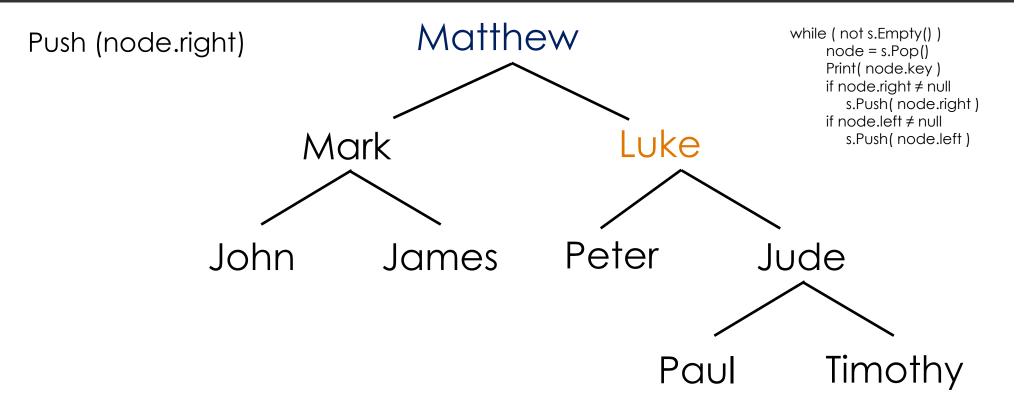
Output:

Stack: Matthew



Output: Matthew

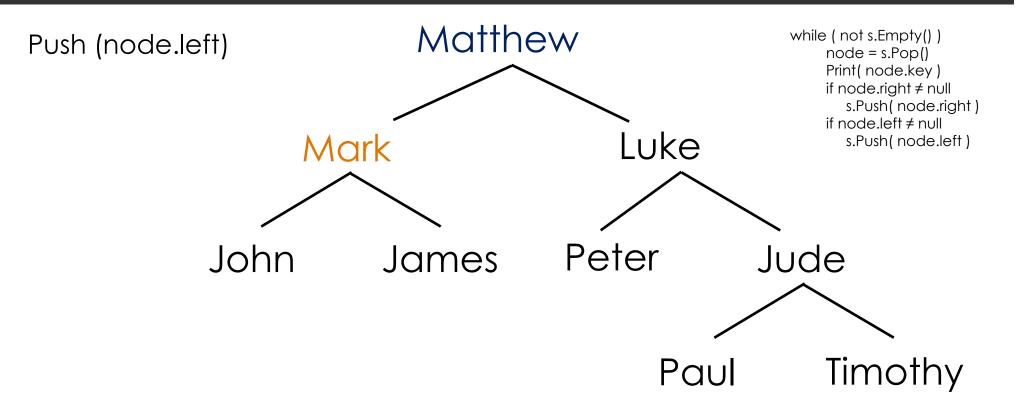
Stack:



Output: Matthew

node

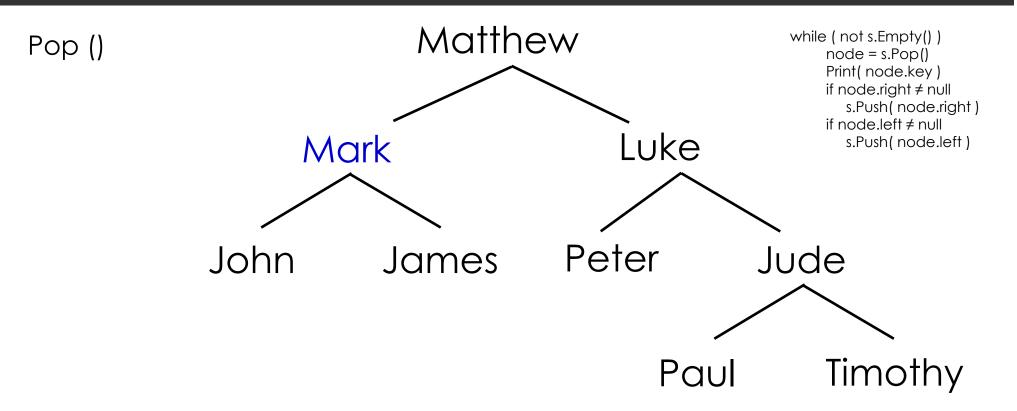
Stack: Luke



Output: Matthew

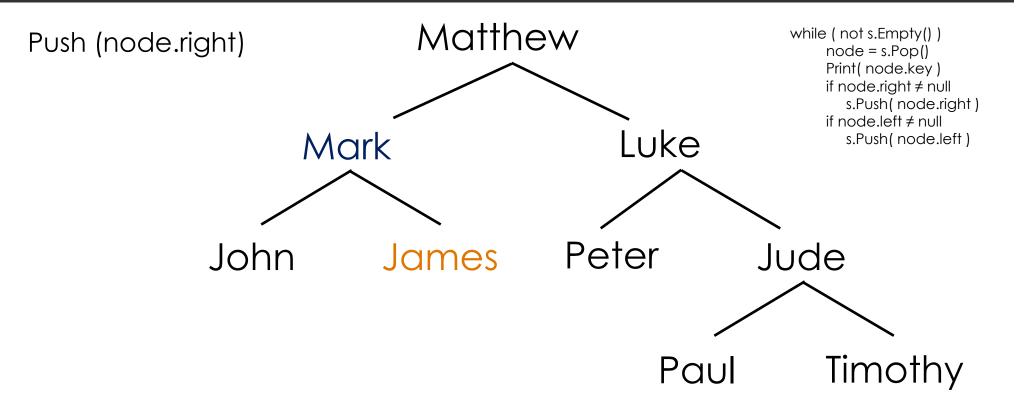
node

Stack: Luke, Mark



Output: Matthew, Mark

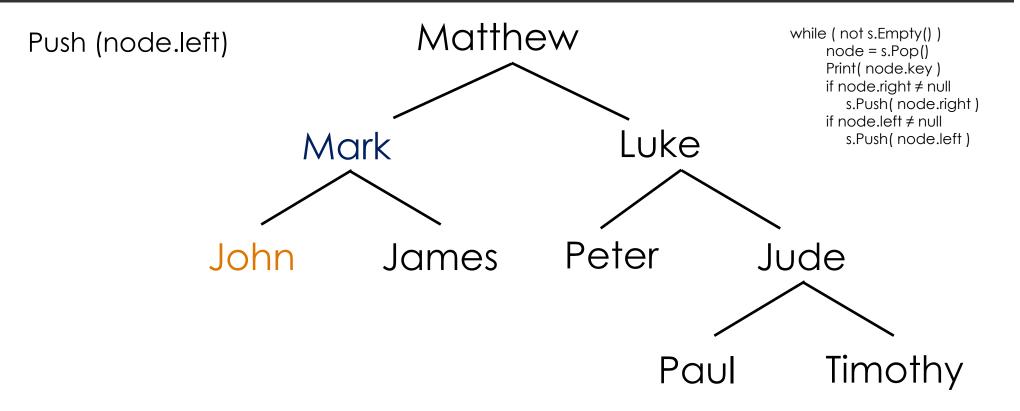
Stack: Luke



Output: Matthew, Mark

node

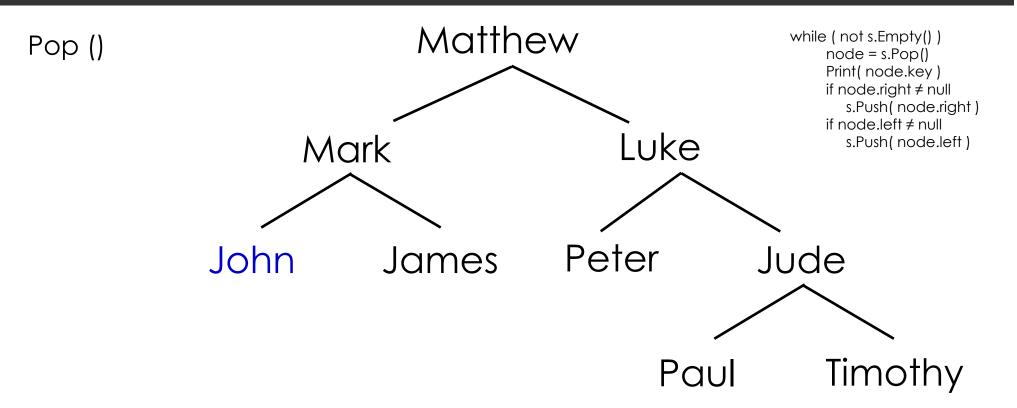
Stack: Luke, James



Output: Matthew, Mark

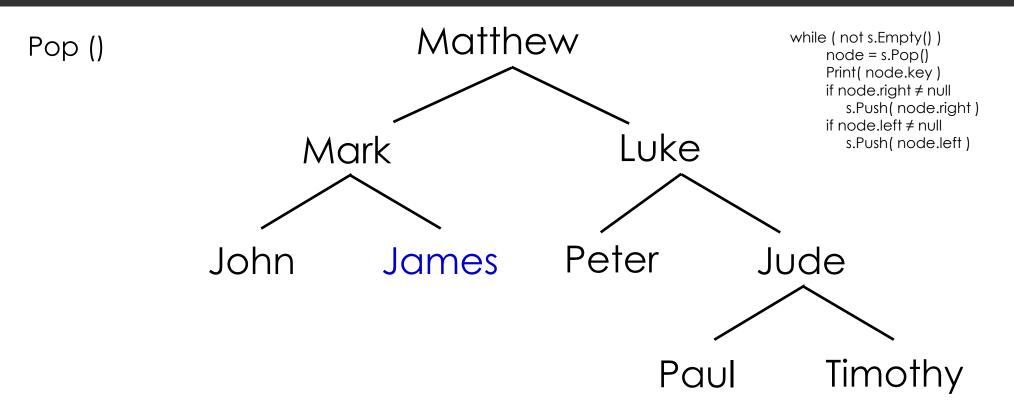
node

Stack: Luke, James, John



Output: Matthew, Mark, John

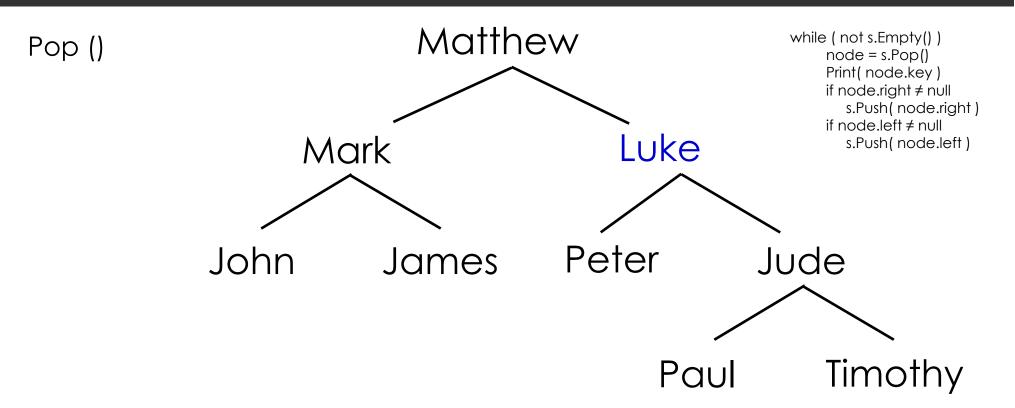
Stack: Luke, James



Output: Matthew, Mark, John, James

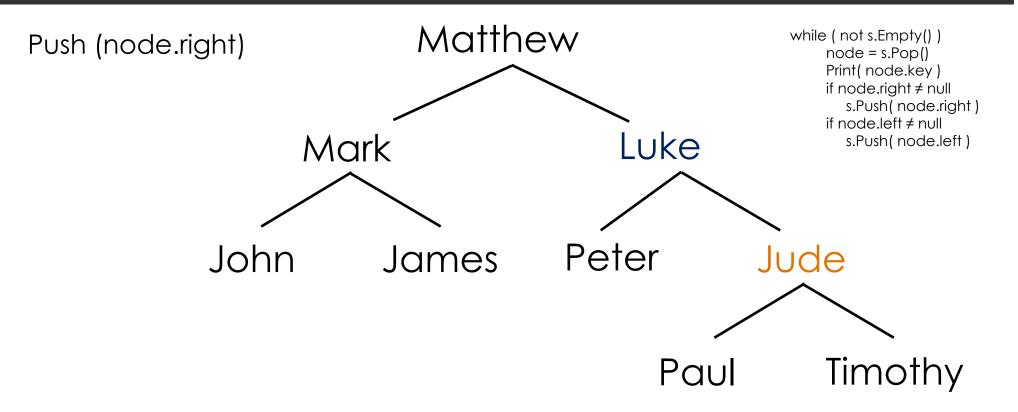
node

Stack: Luke



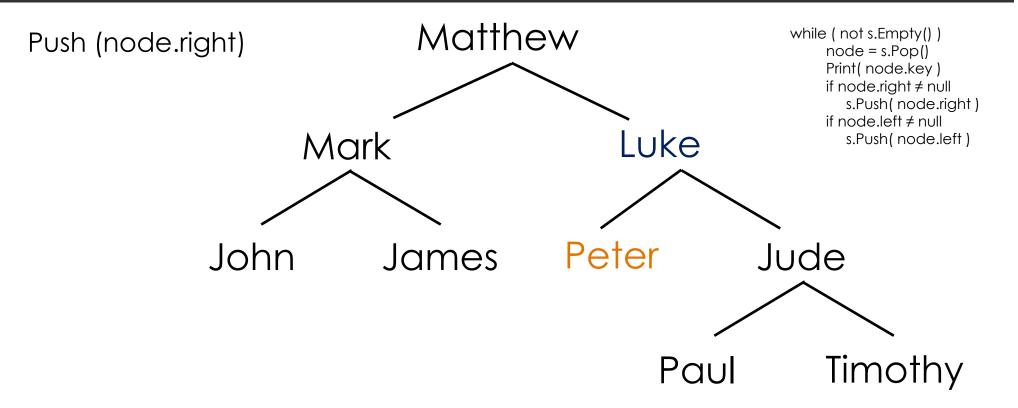
Output: Matthew, Mark, John, James, Luke

Stack:



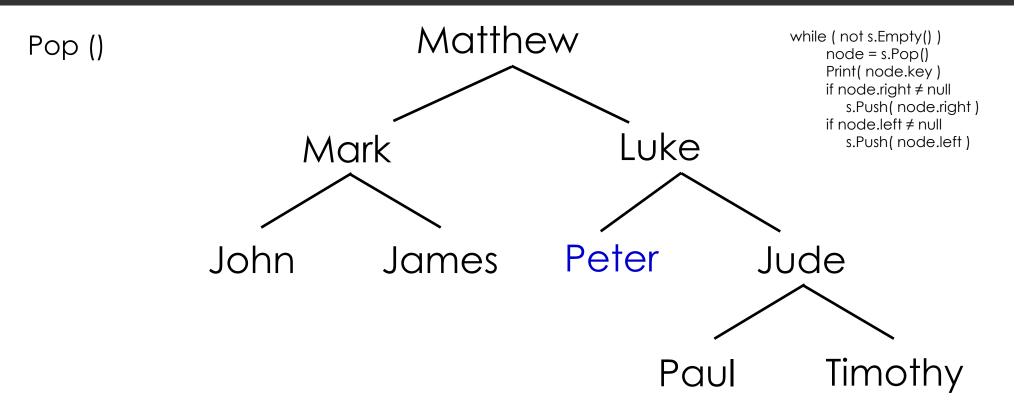
Output: Matthew, Mark, John, James, Luke

Stack: Jude



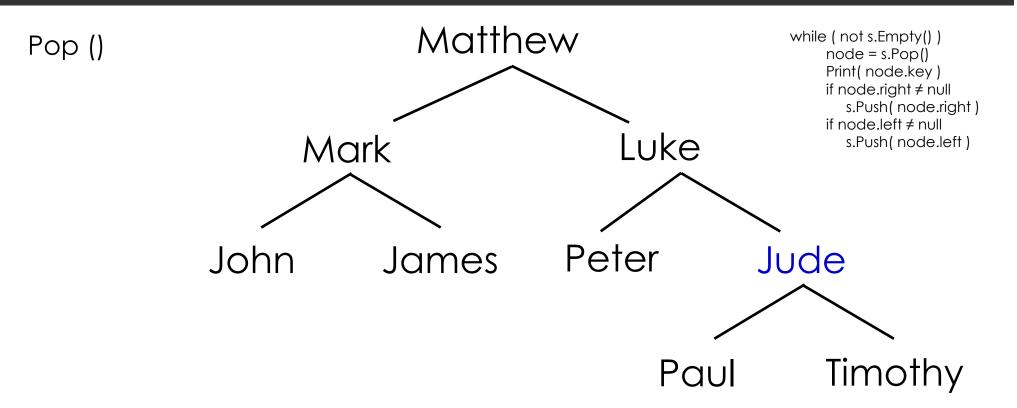
Output: Matthew, Mark, John, James, Luke

Stack: Jude, Peter



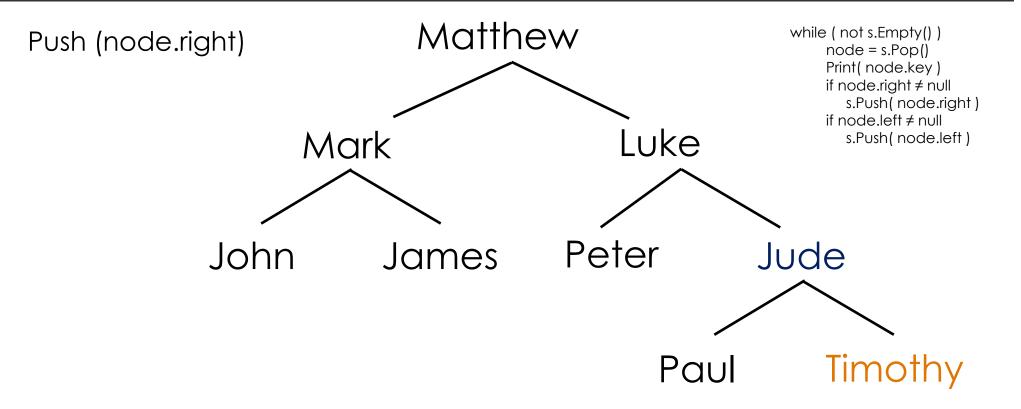
Output: Matthew, Mark, John, James, Luke, Peter

Stack: Jude



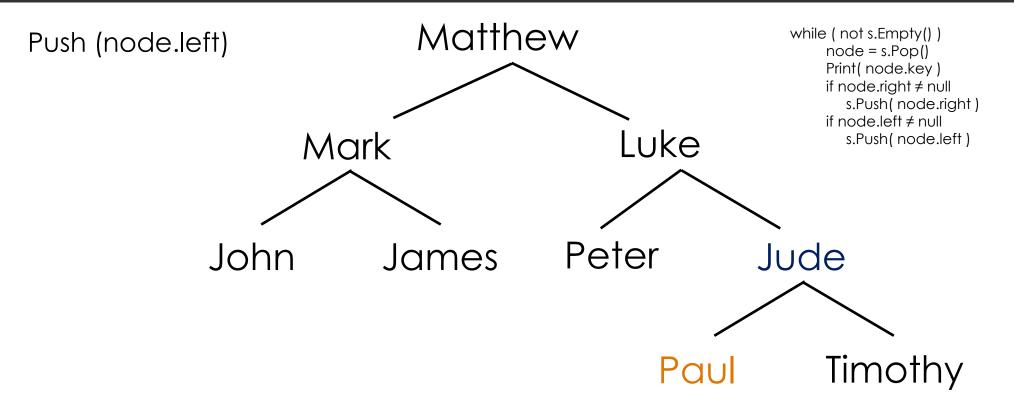
Output: Matthew, Mark, John, James, Luke, Peter, Jude

Stack:



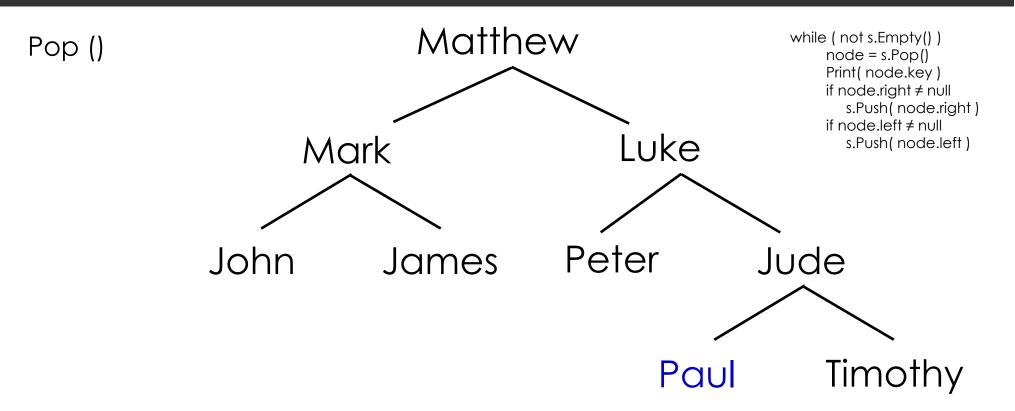
Output: Matthew, Mark, John, James, Luke, Peter, Jude

Stack: Timothy



Output: Matthew, Mark, John, James, Luke, Peter, Jude

Stack: Timothy, Paul

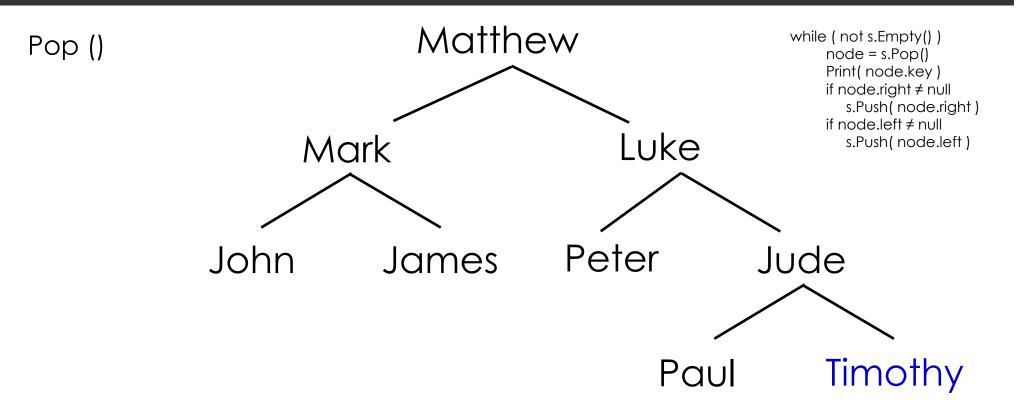


Output: Matthew, Mark, John, James, Luke, Peter, Jude,

Stack:

Paul

Timothy



Output:

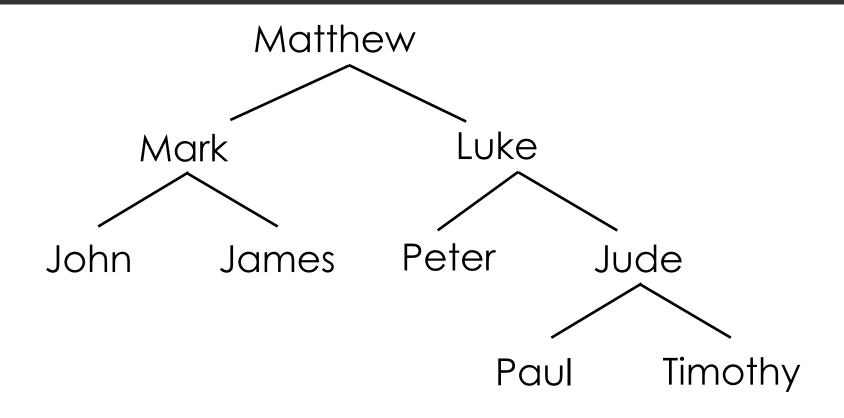
Matthew, Mark, John, James, Luke, Peter, Jude,

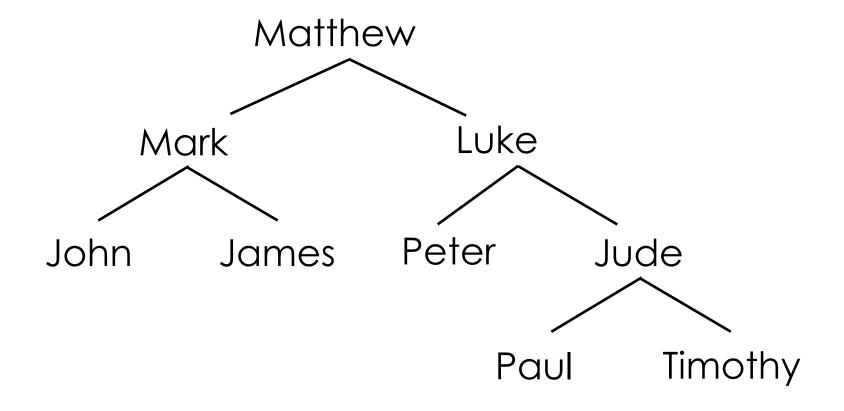
Paul, Timothy

Stack:

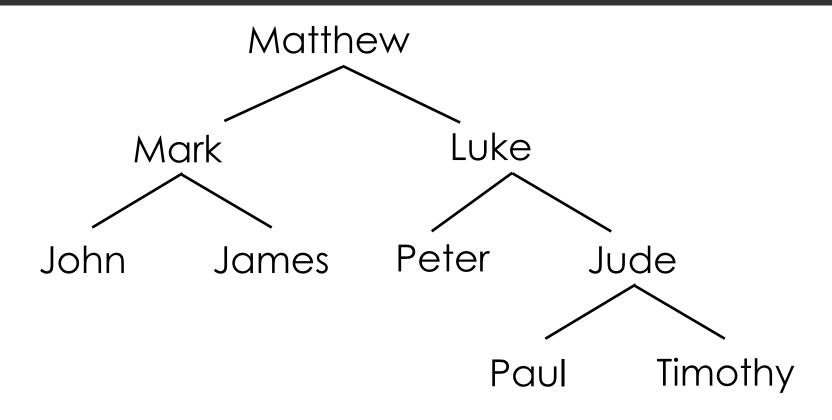
93

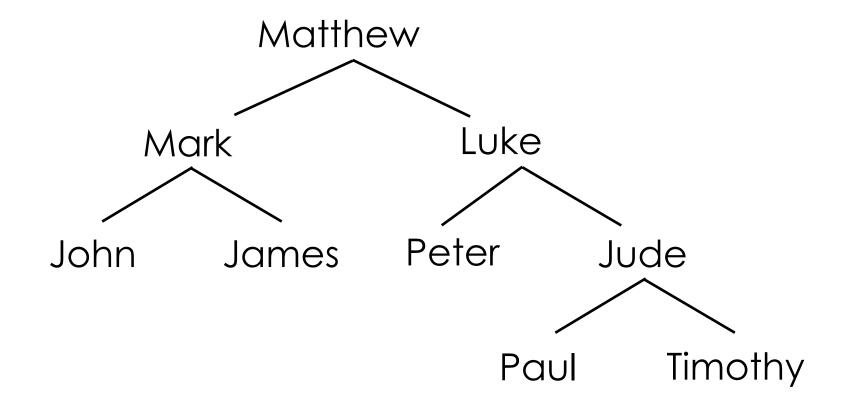
Depth-first Traversal (PreOrder) Implementation using **Recursion**



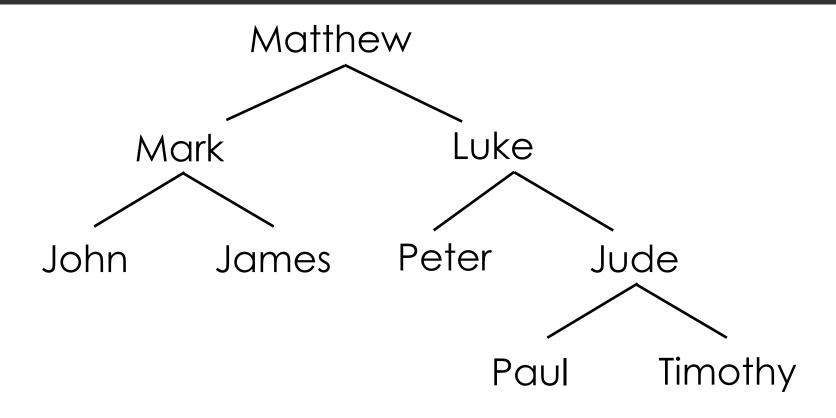


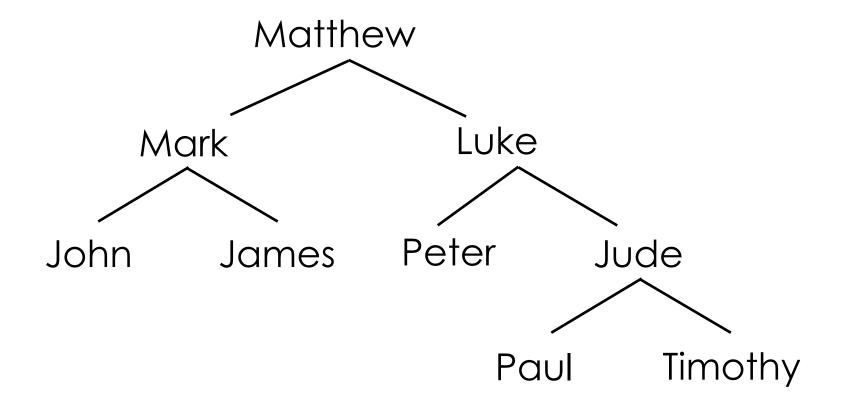
Output: Matthew, Mark, John, James, Luke, Peter, Jude, Paul, Timothy





Output: John, Mark, James, Matthew, Peter, Luke, Paul, Jude, Timothy

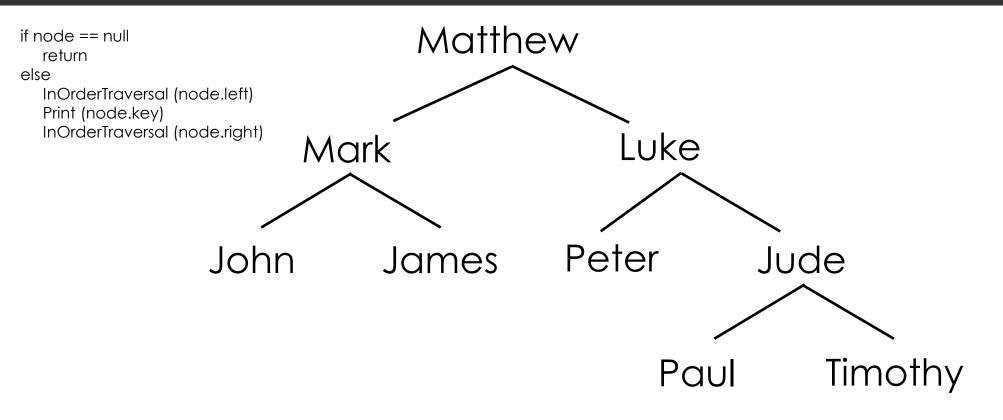


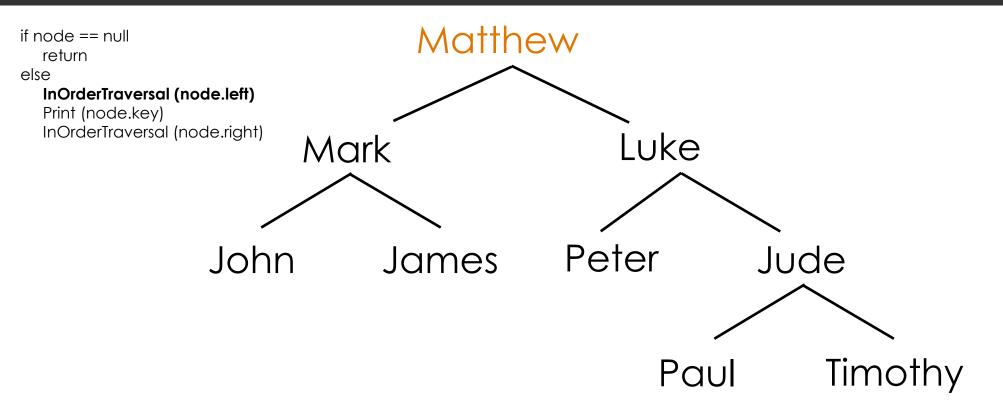


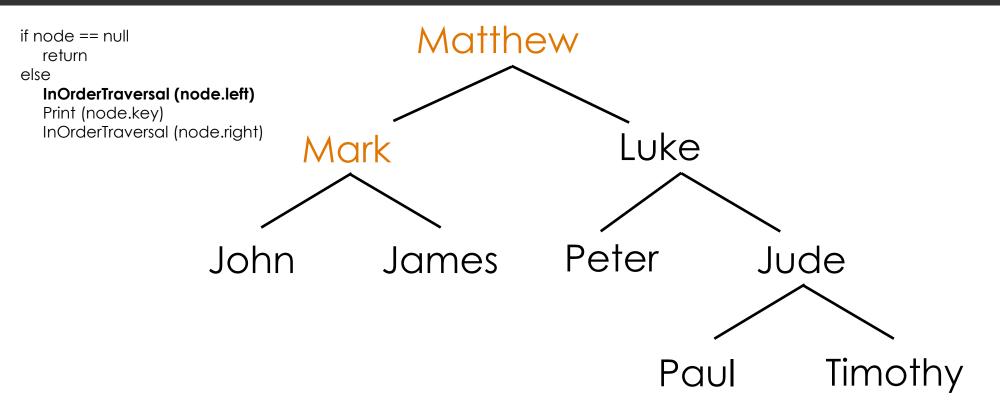
Output: John, James, Mark, Peter, Paul, Timothy, Jude, Luke, Matthew

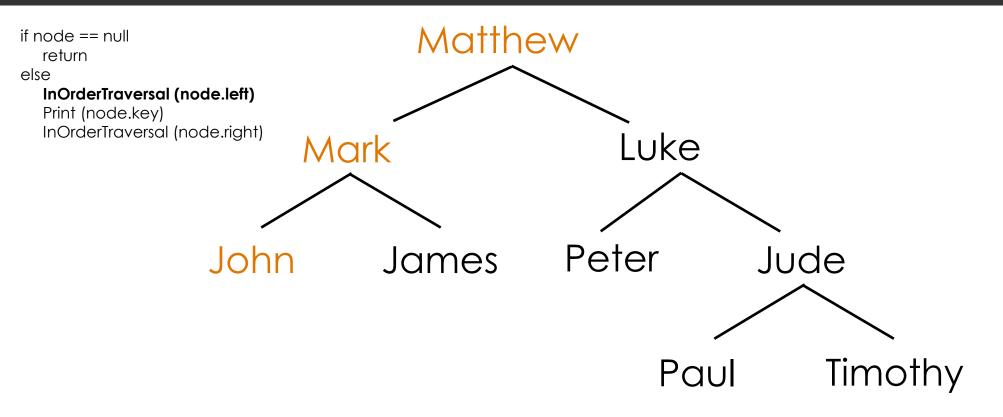
InOrderTraversal (Node node)

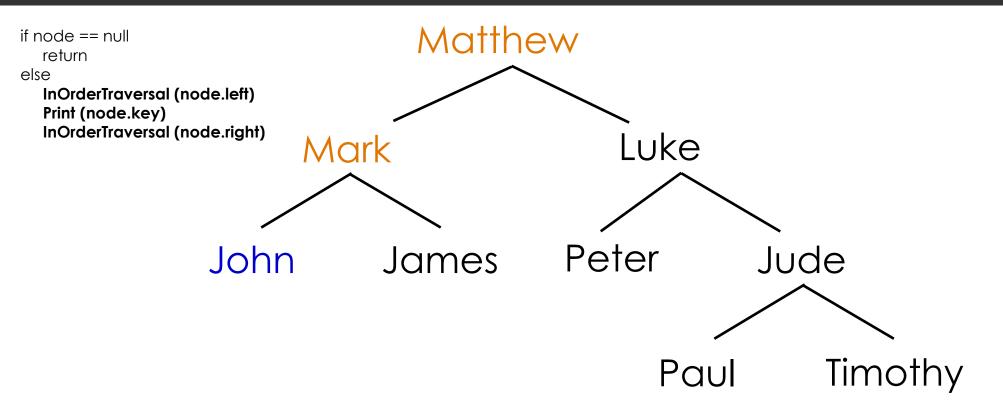
```
if node == null
    return
else
    InOrderTraversal (node.left)
    Print (node.key)
    InOrderTraversal (node.right)
```



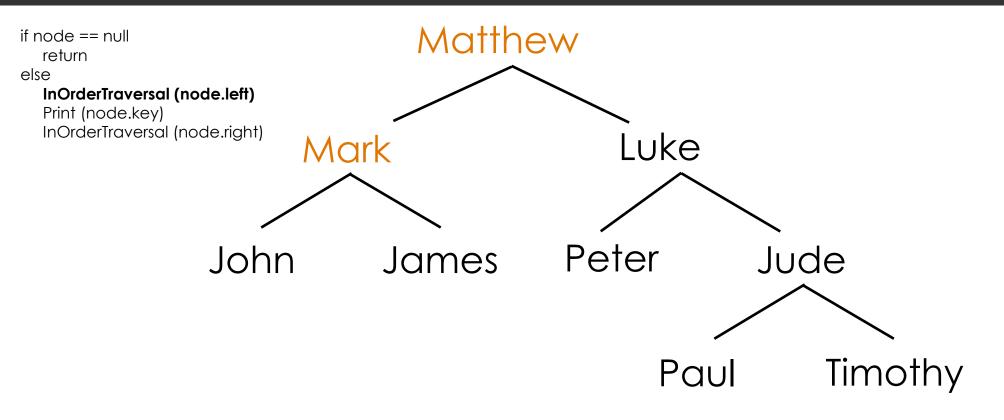




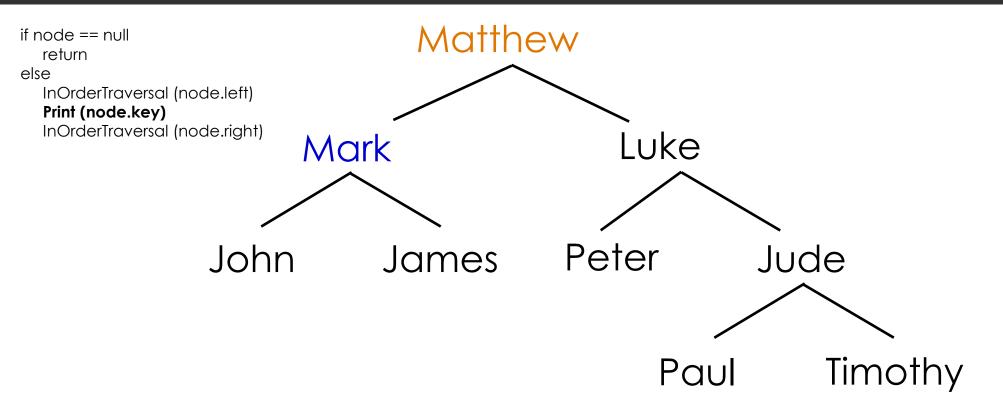




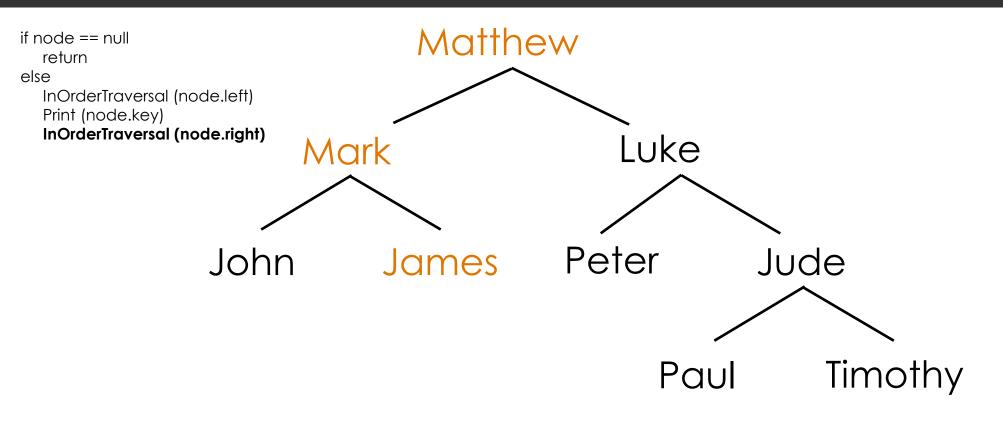
Output: John



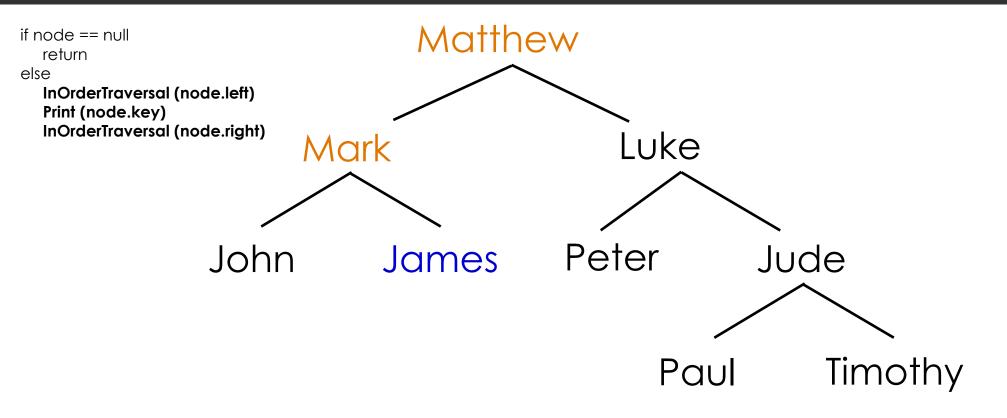
Output: John



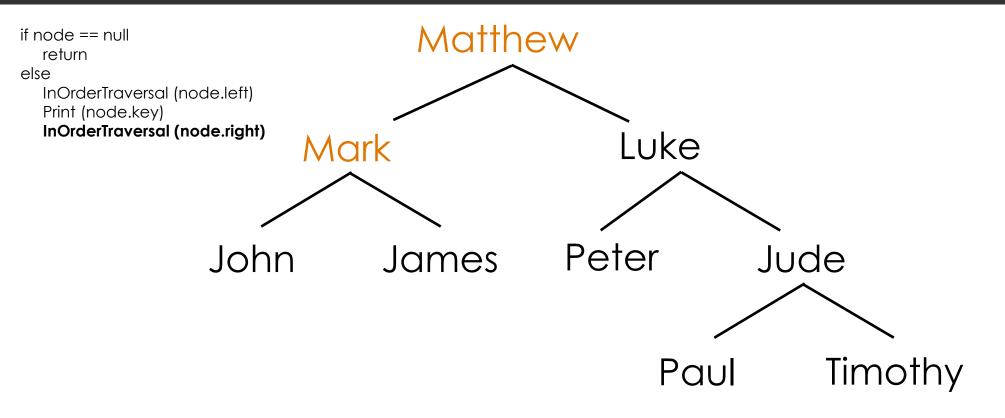
Output: John, Mark



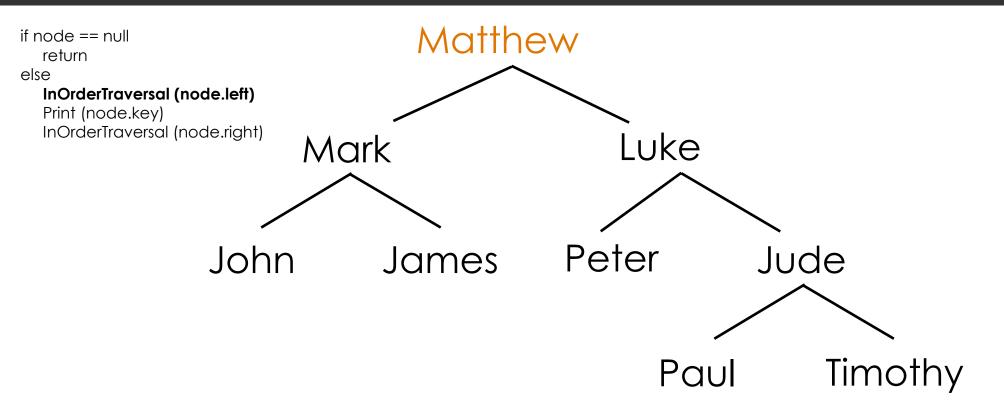
Output: John, Mark, James



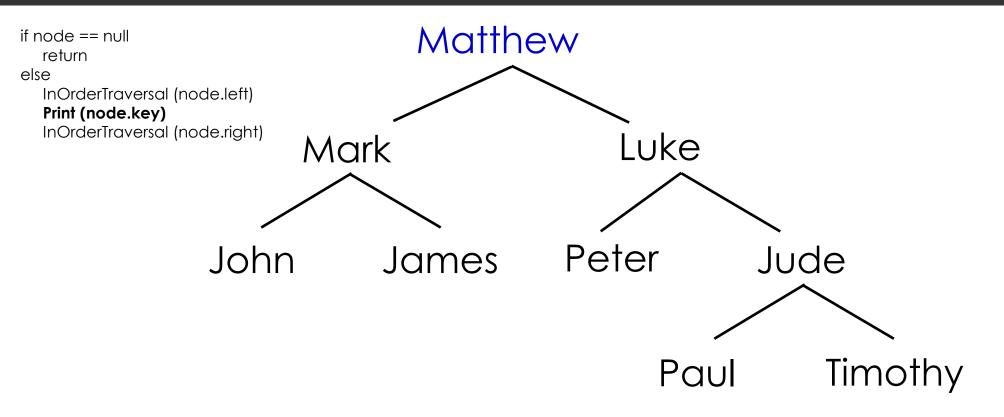
Output: John, Mark, James



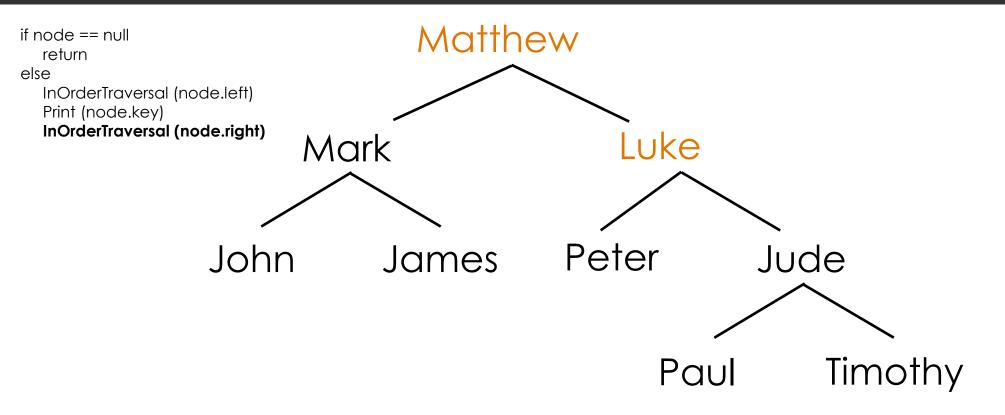
Output: John, Mark, James



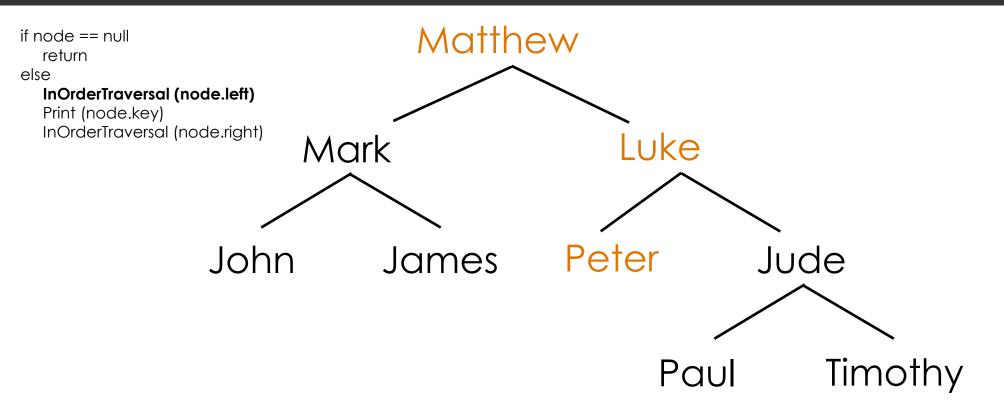
Output: John, Mark, James



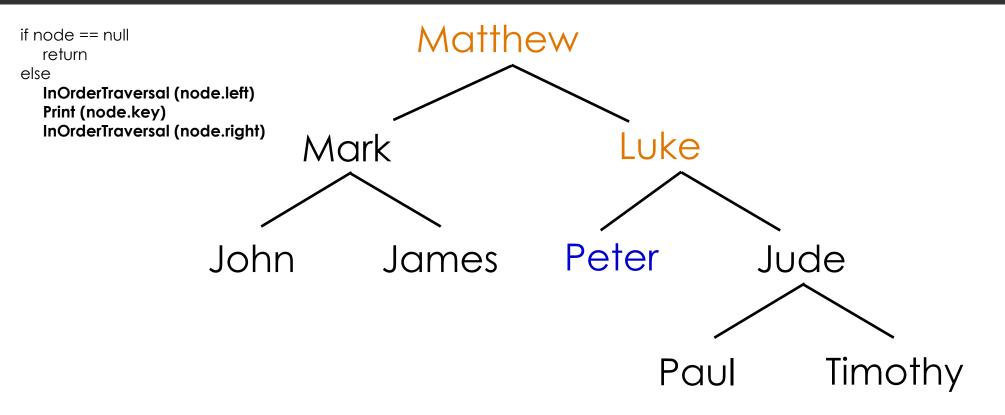
Output: John, Mark, James, Matthew



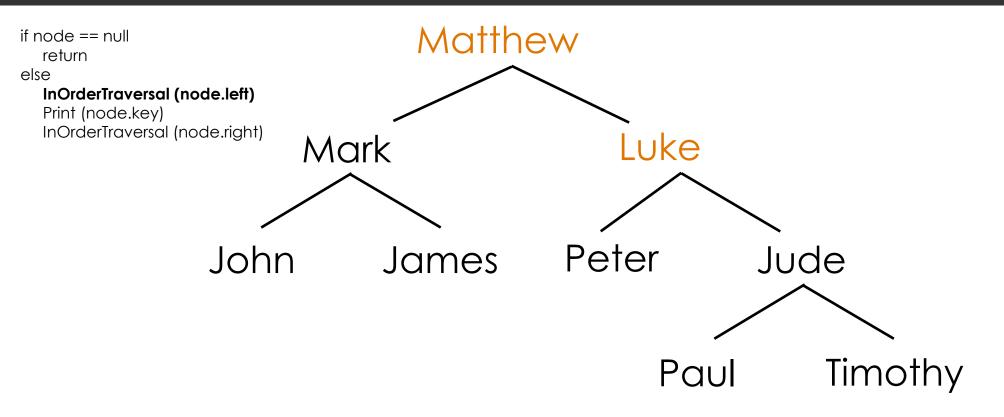
Output: John, Mark, James, Matthew



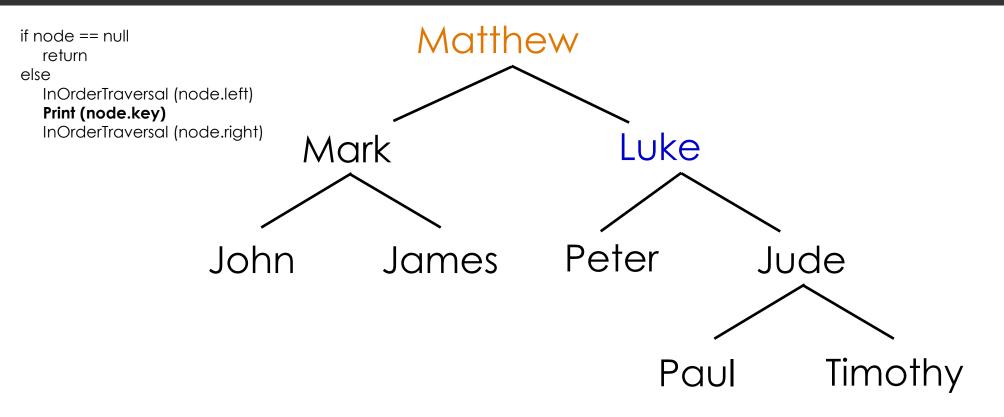
Output: John, Mark, James, Matthew



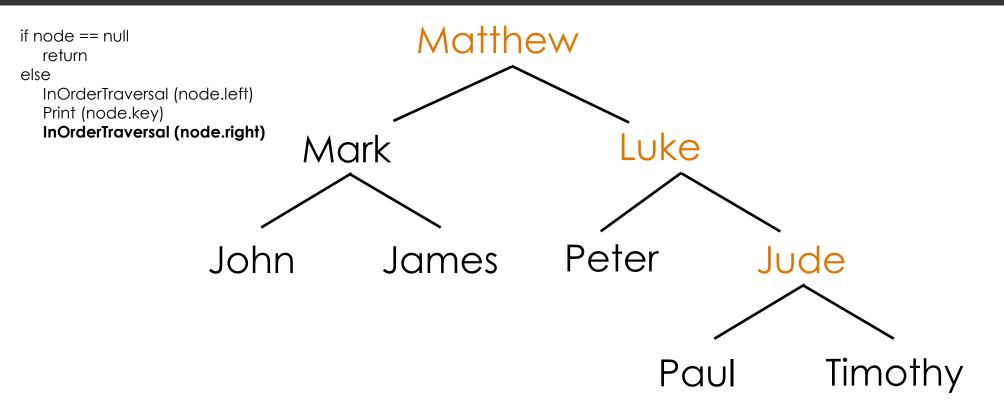
Output: John, Mark, James, Matthew, Peter



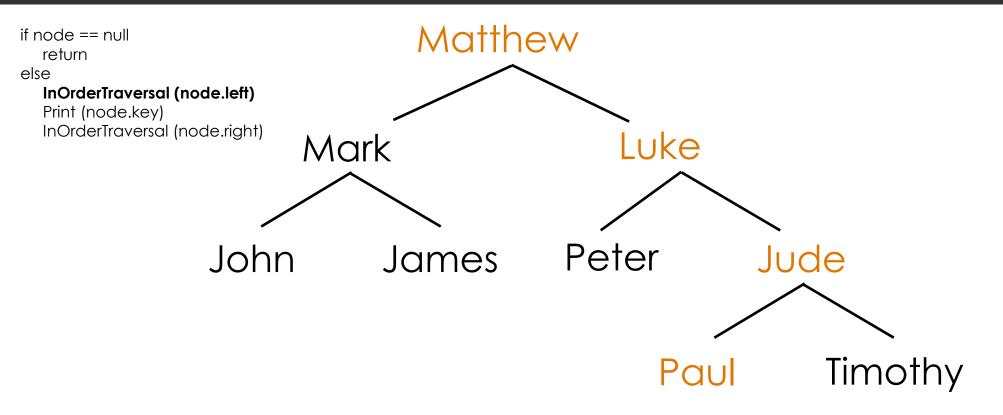
Output: John, Mark, James, Matthew, Peter



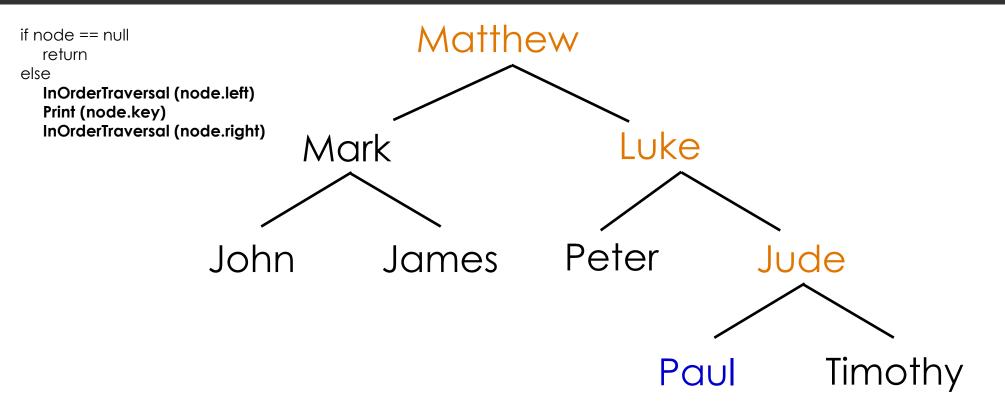
Output: John, Mark, James, Matthew, Peter, Luke



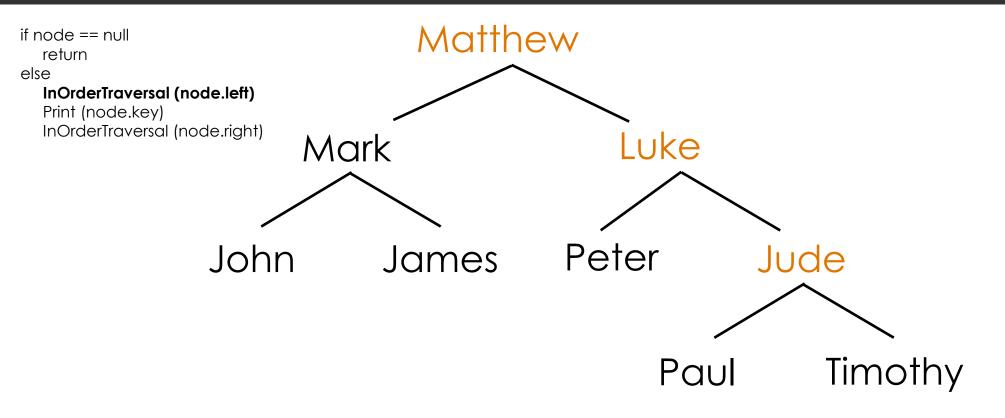
Output: John, Mark, James, Matthew, Peter, Luke



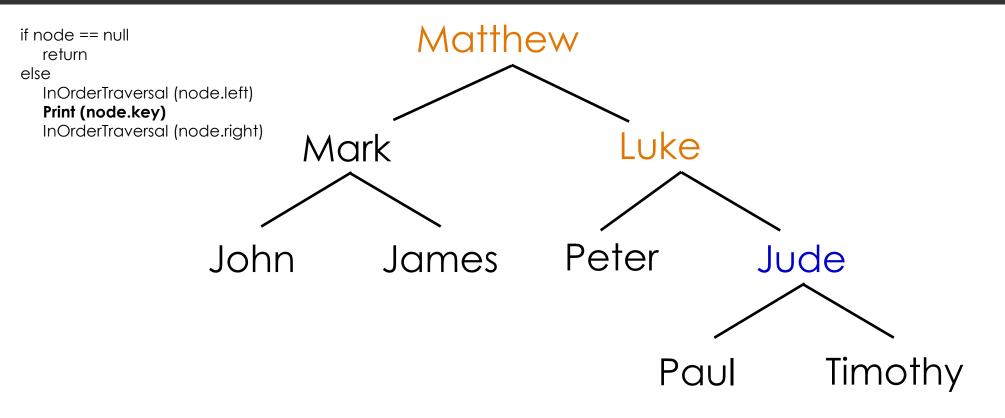
Output: John, Mark, James, Matthew, Peter, Luke

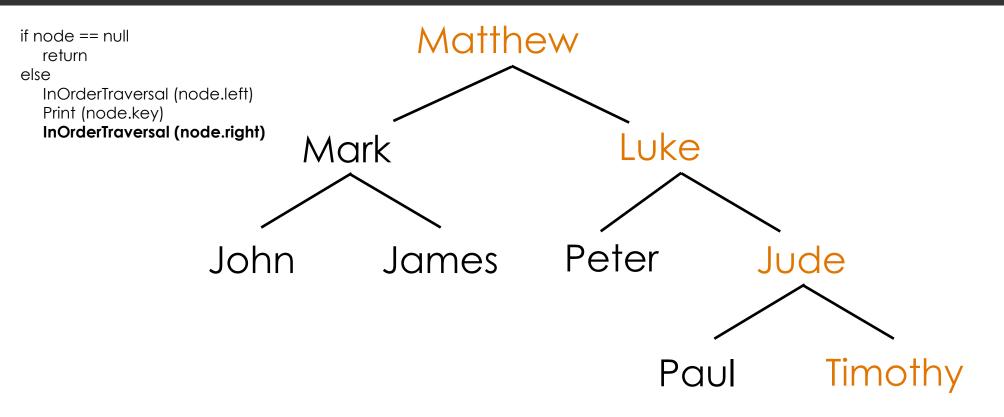


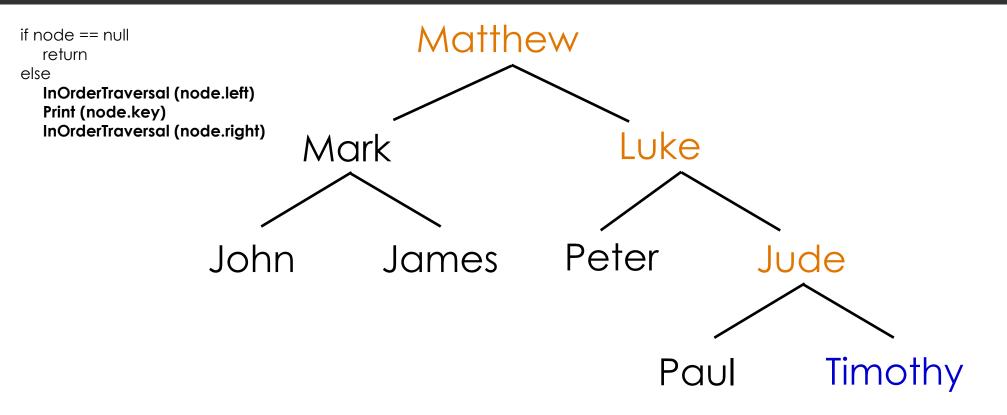
Output: John, Mark, James, Matthew, Peter, Luke, Paul

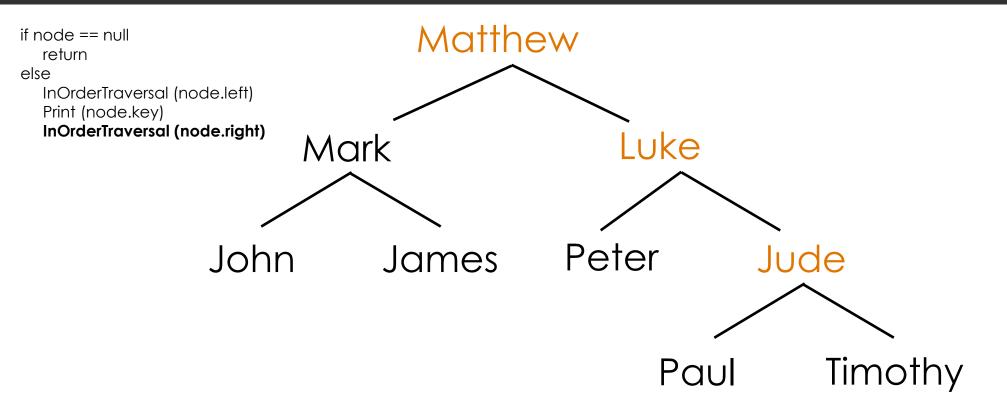


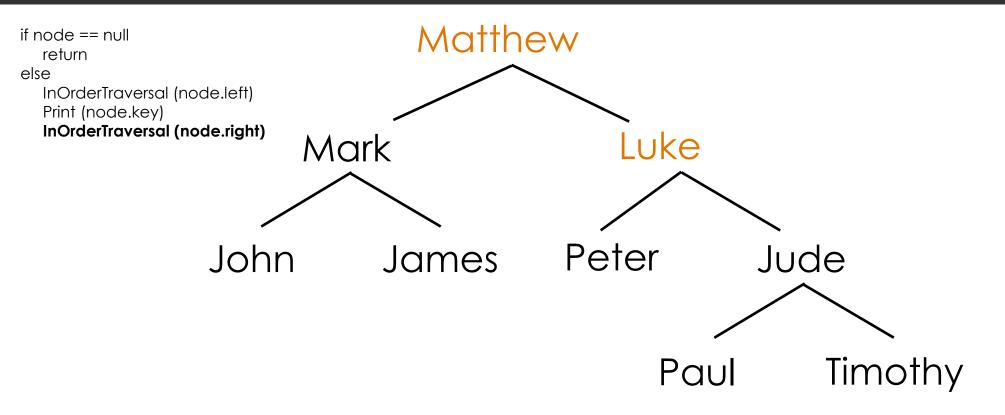
Output: John, Mark, James, Matthew, Peter, Luke, Paul

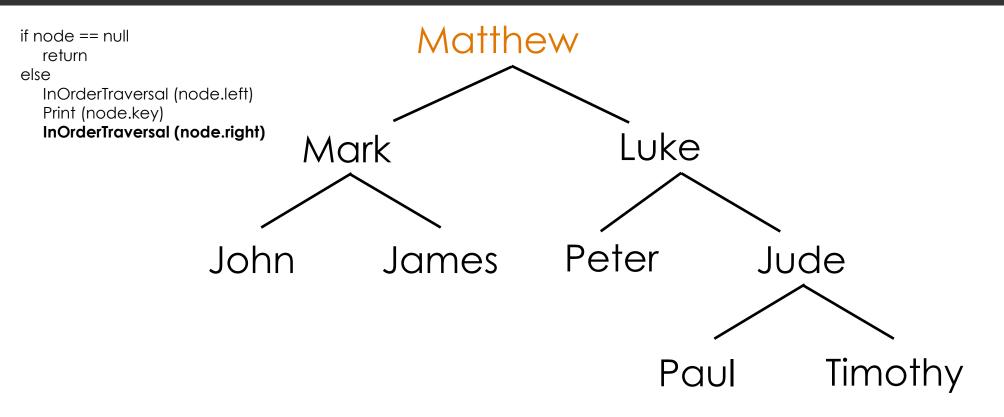


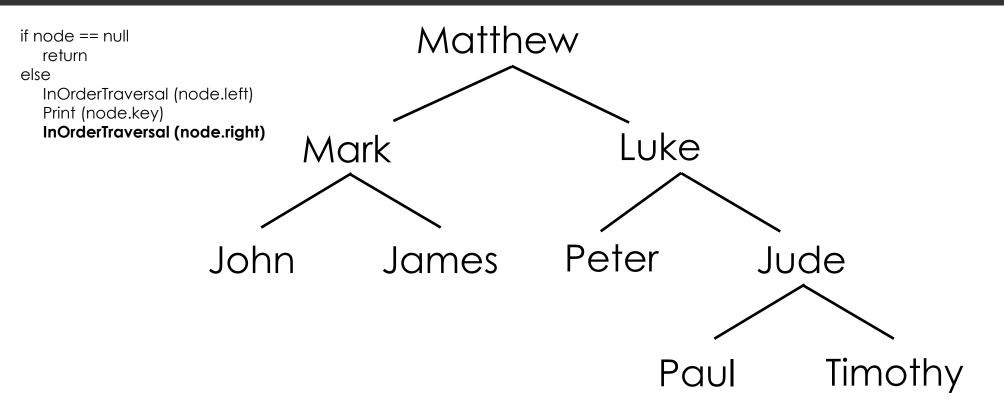






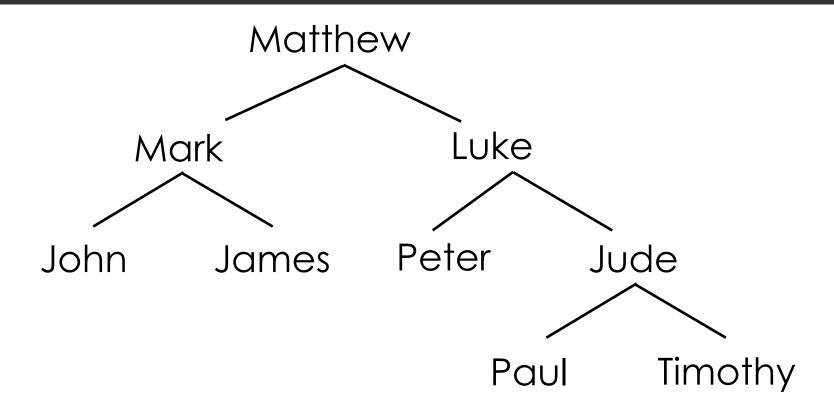




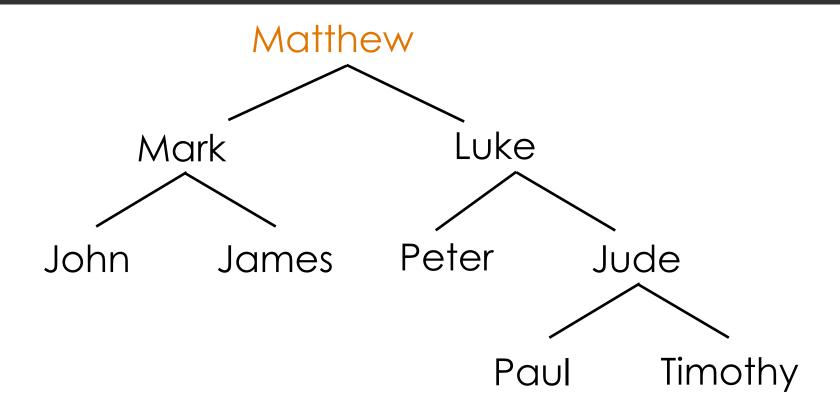


PreOrderTraversal (Node node)

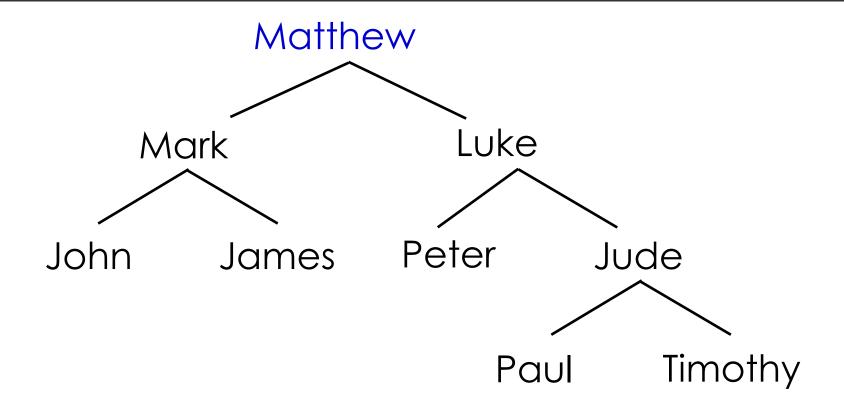
```
if node == null
    return
else
    Print (node.key)
    PreOrderTraversal (node.left)
    PreOrderTraversal (node.right)
```



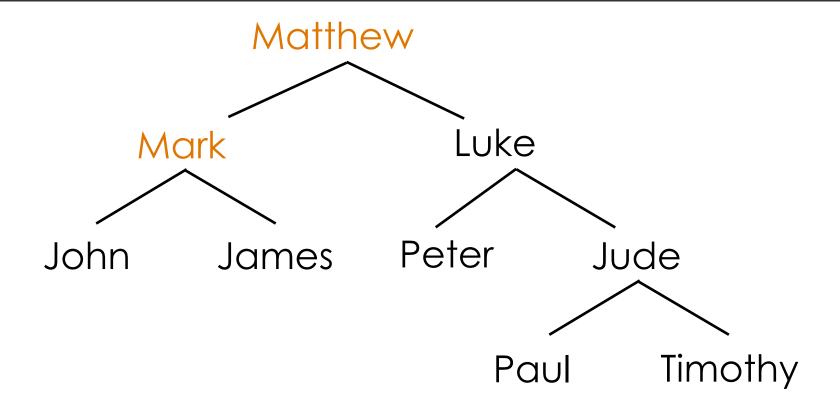
Output:



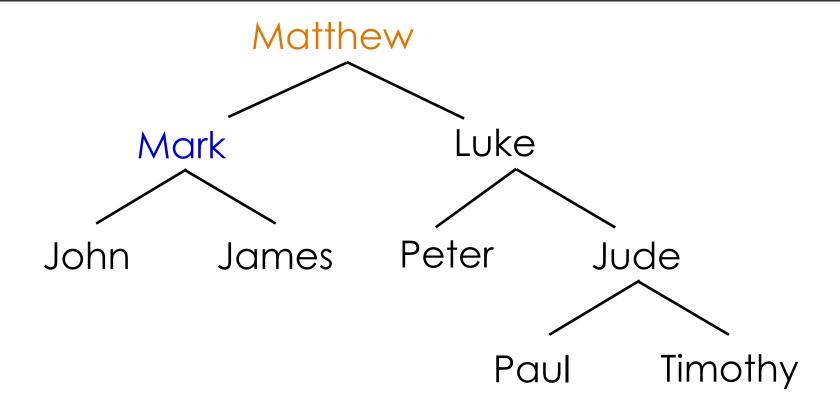
Output:



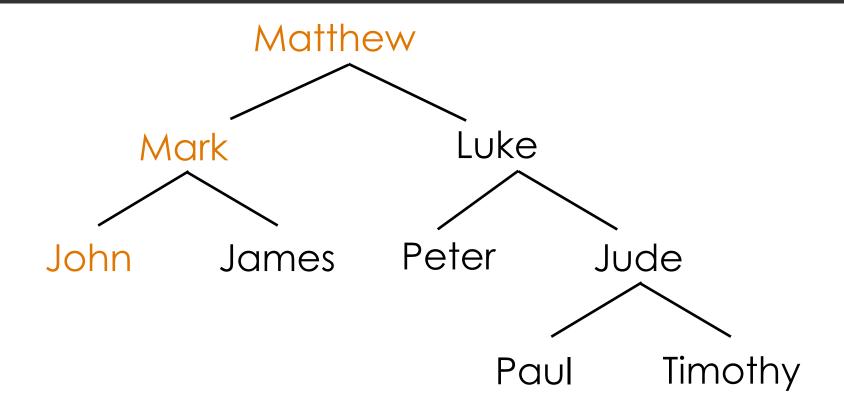
Output: Matthew



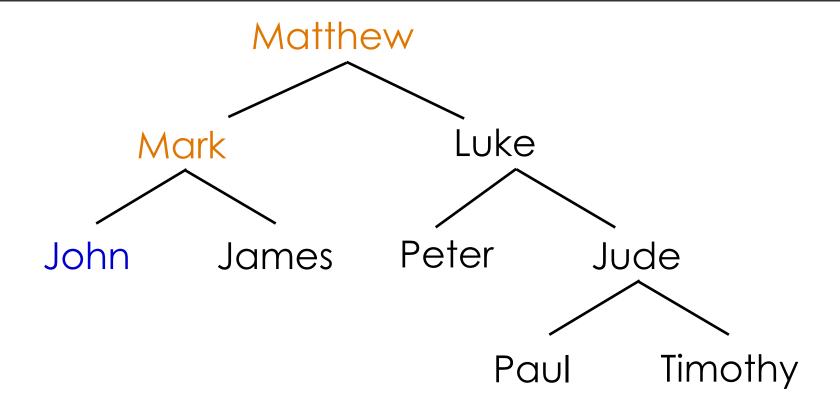
Output: Matthew



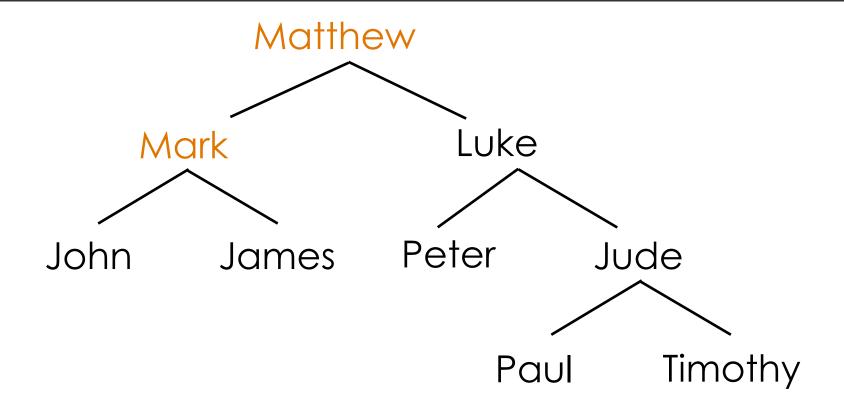
Output: Matthew, Mark



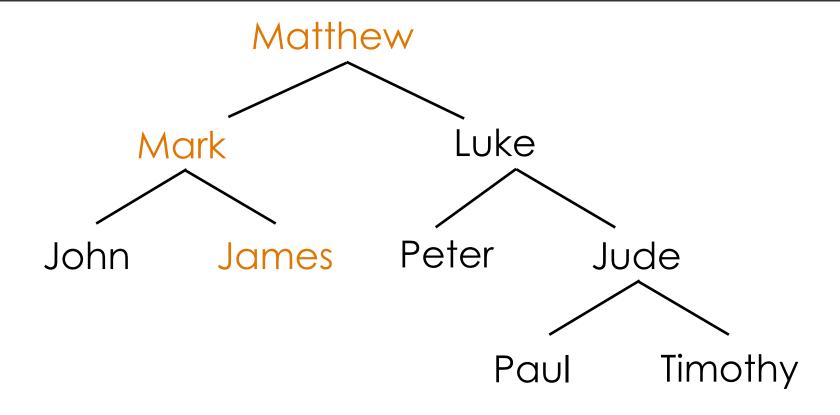
Output: Matthew, Mark



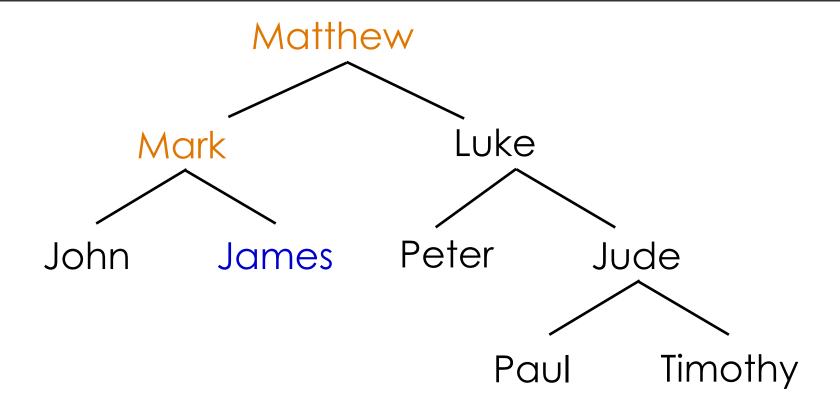
Output: Matthew, Mark, John

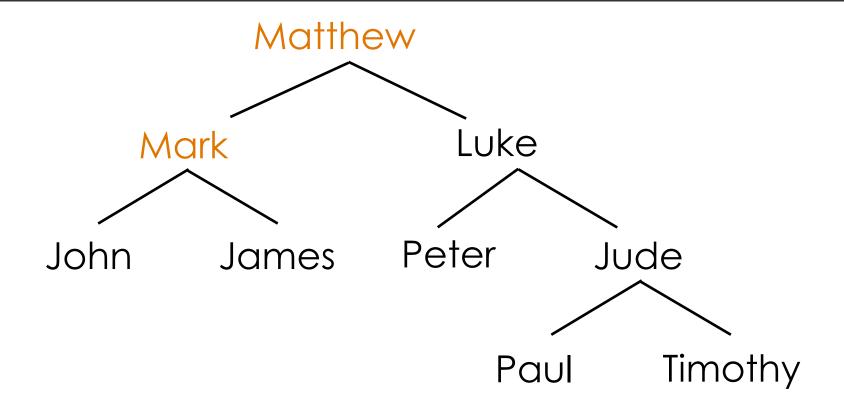


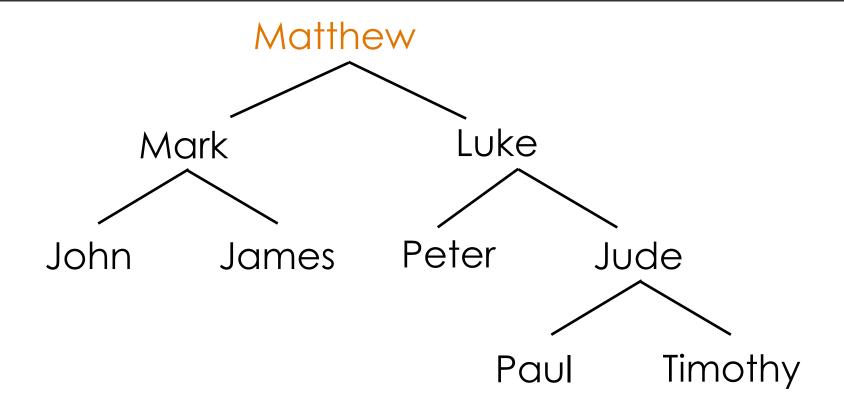
Output: Matthew, Mark, John

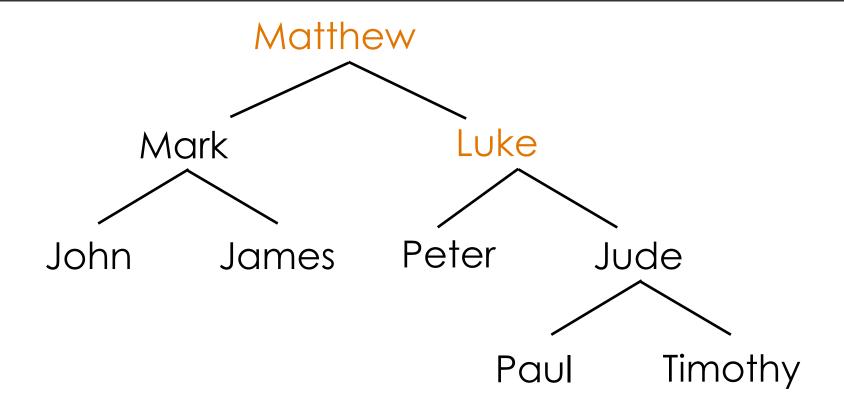


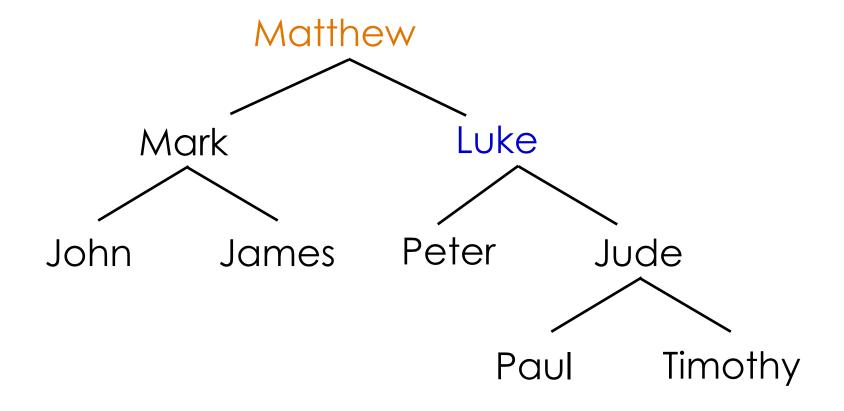
Output: Matthew, Mark, John

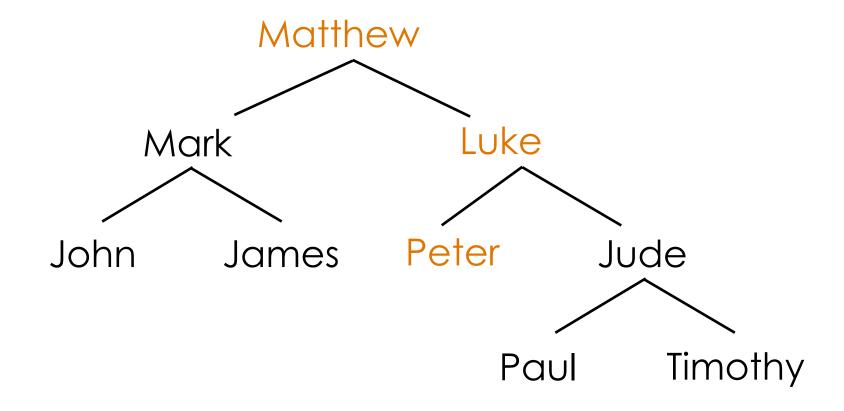


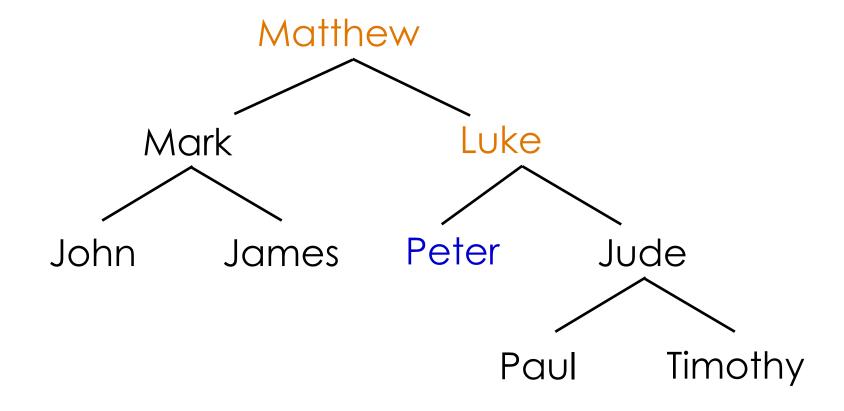




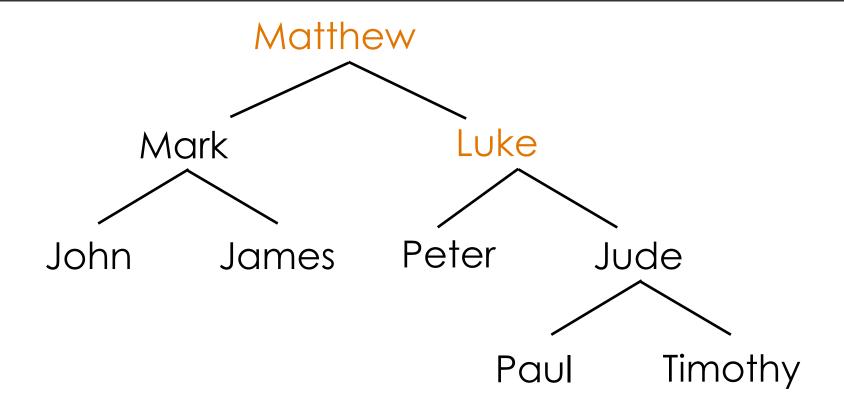




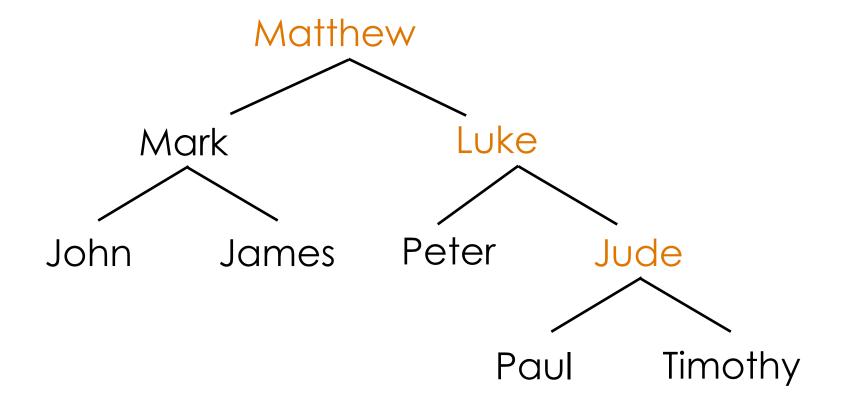




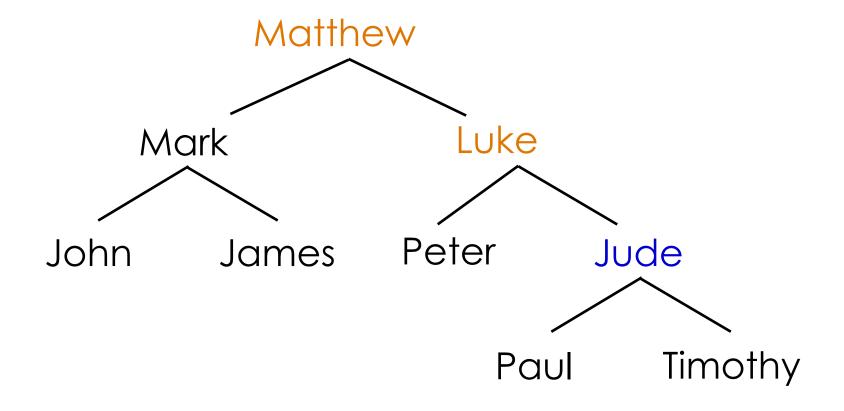
Output: Matthew, Mark, John, James, Luke, Peter



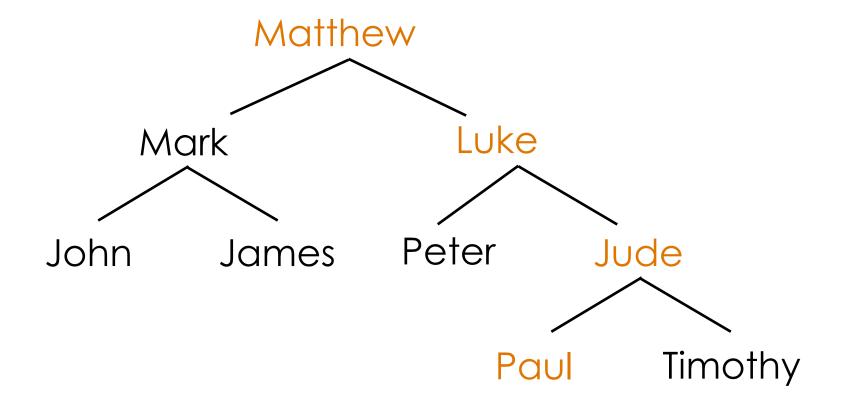
Output: Matthew, Mark, John, James, Luke, Peter



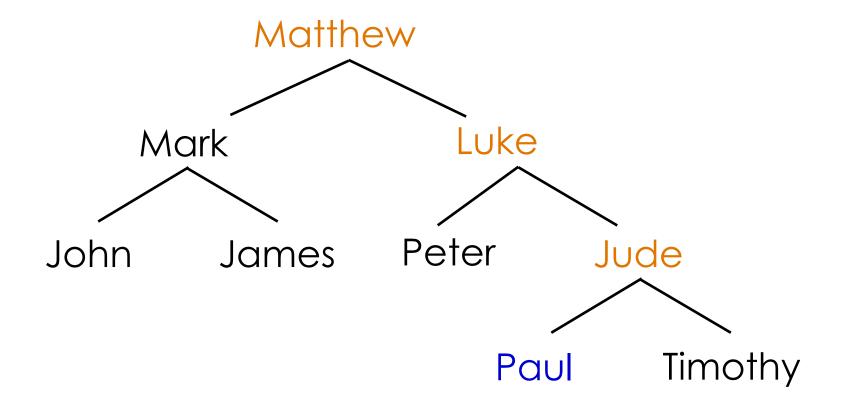
Output: Matthew, Mark, John, James, Luke, Peter

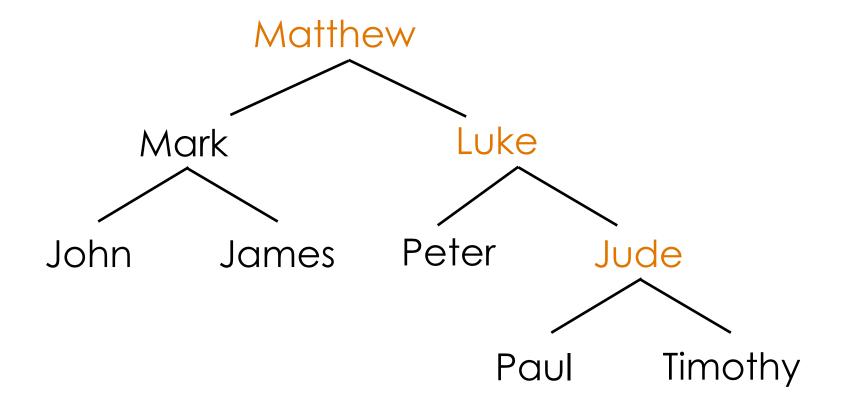


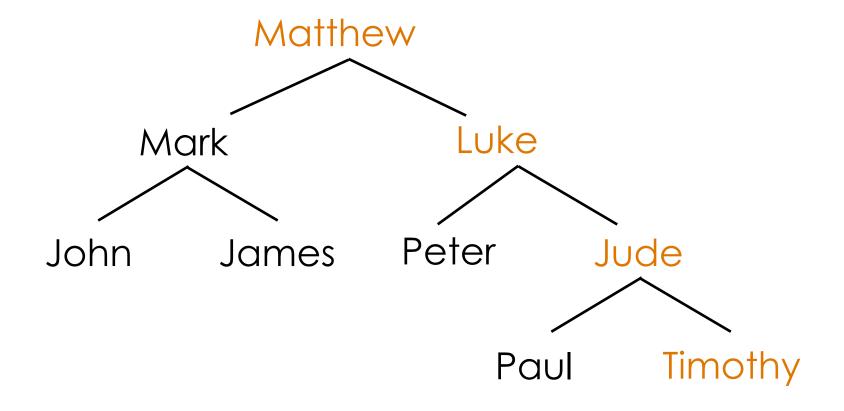
Output: Matthew, Mark, John, James, Luke, Peter, Jude

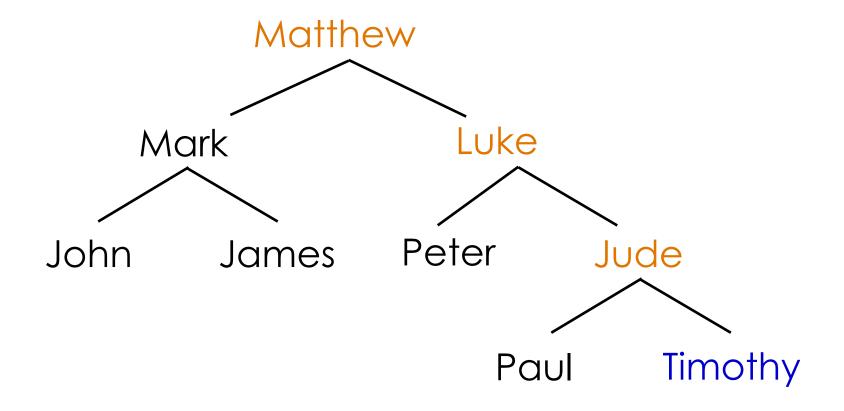


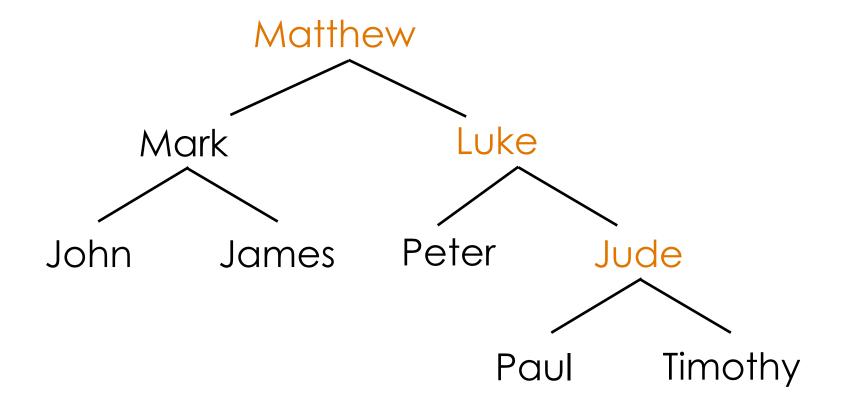
Output: Matthew, Mark, John, James, Luke, Peter, Jude

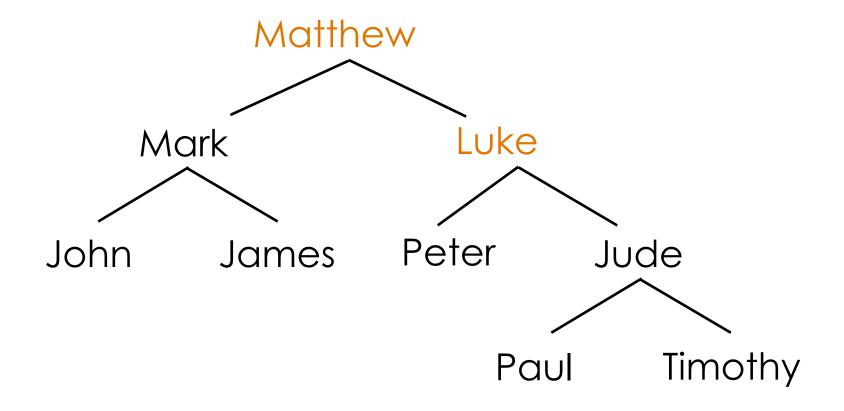


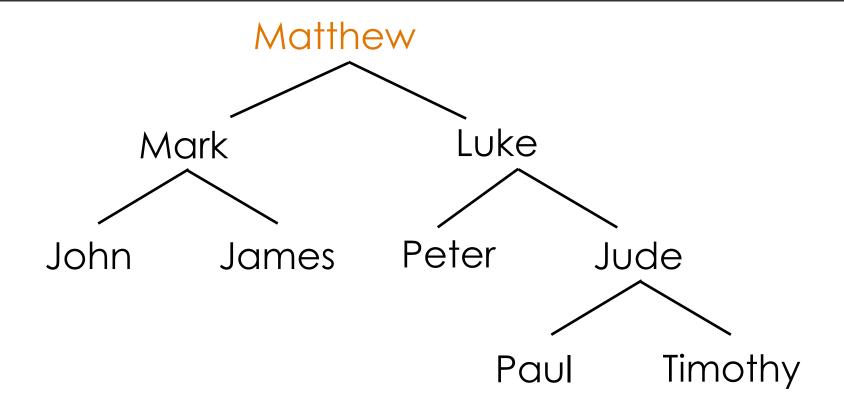


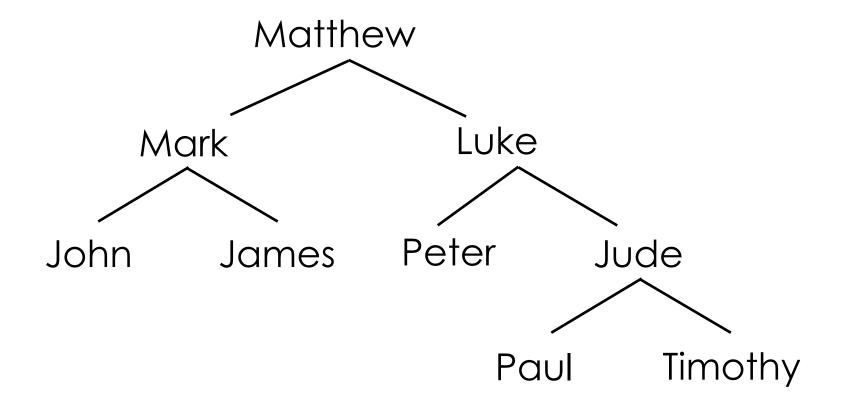






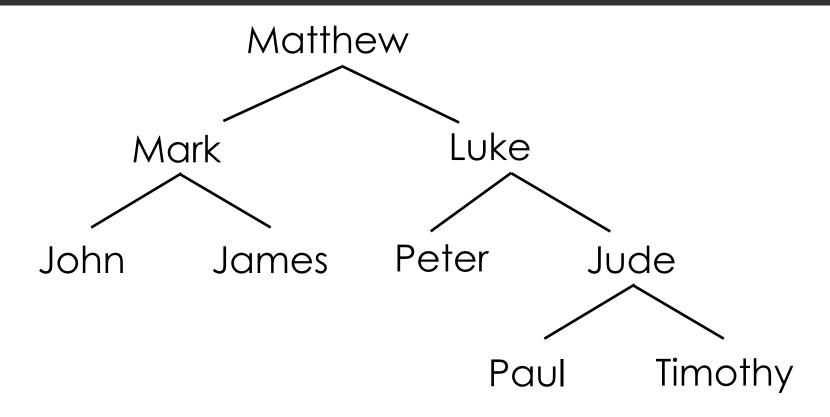


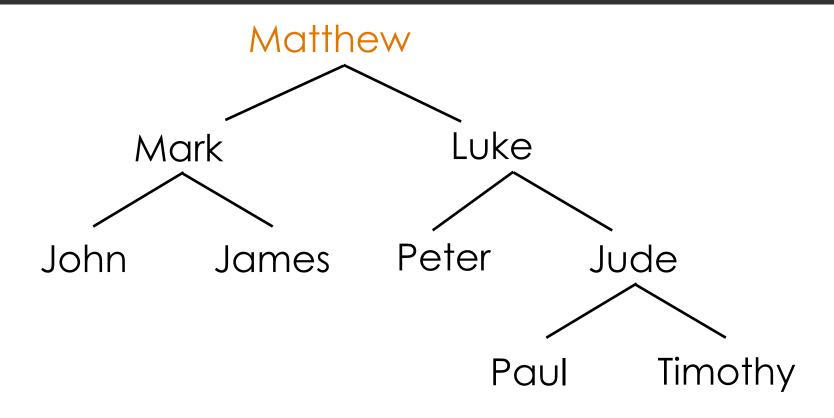


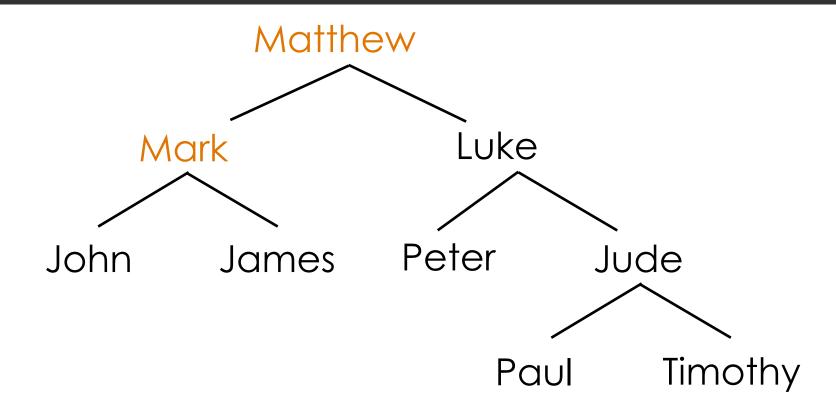


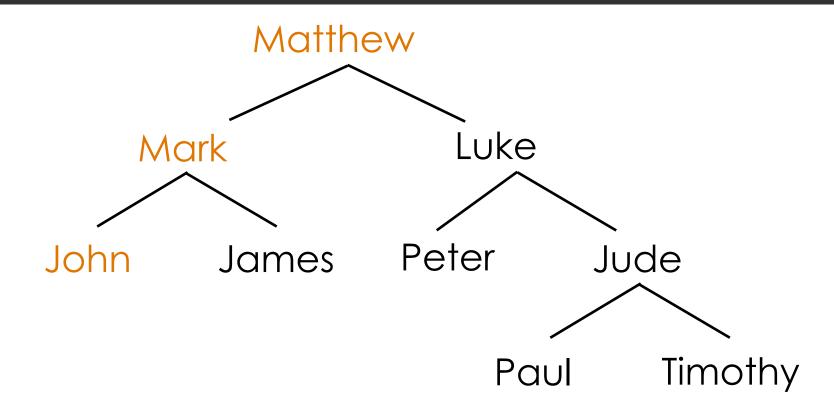
PostOrderTraversal (Node node)

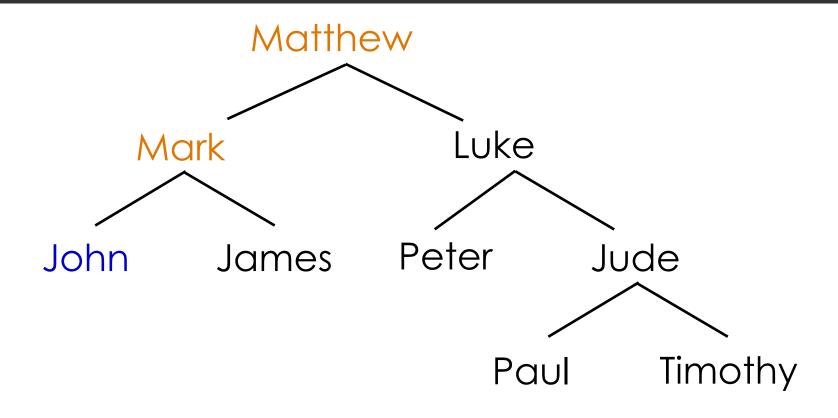
```
if node == null
    return
else
    PostOrderTraversal (node.left)
    PostOrderTraversal (node.right)
    Print (node.key)
```



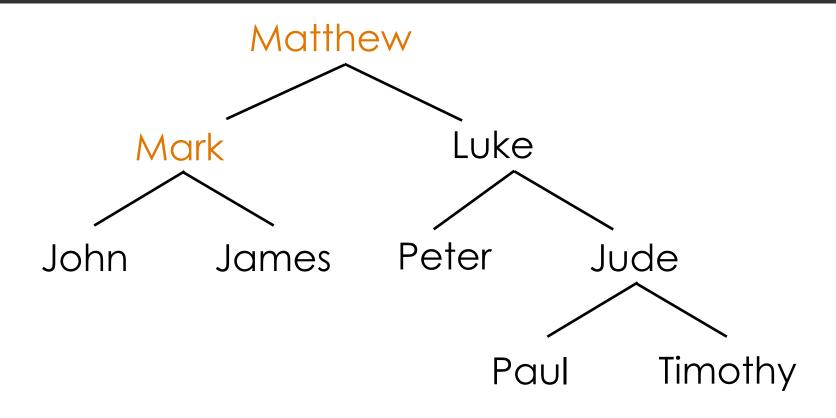




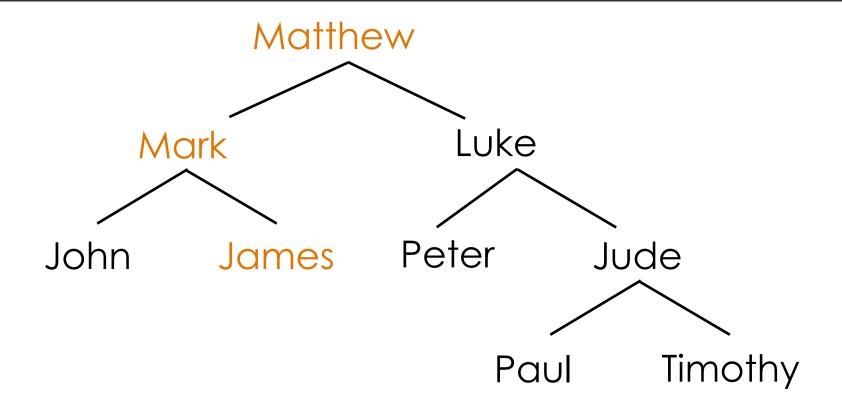




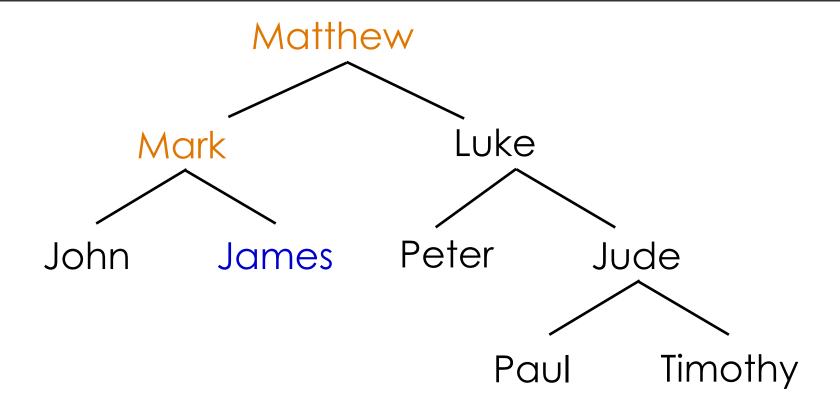
Output: John



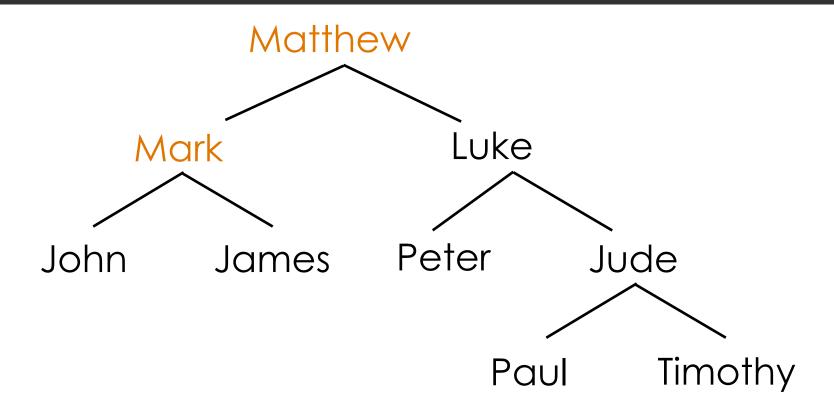
Output: John



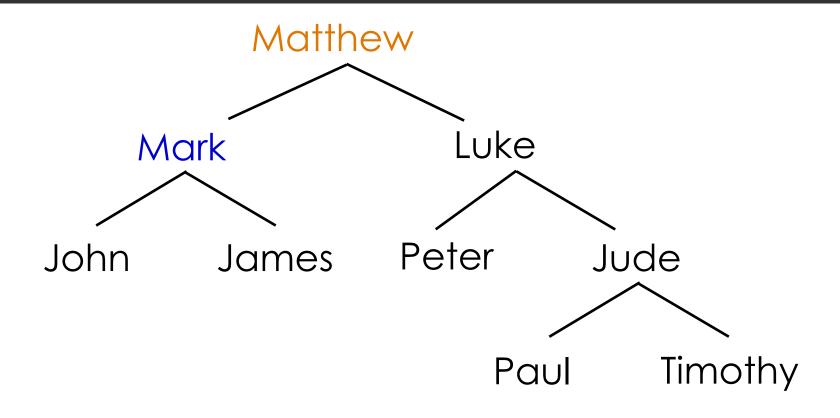
Output: John

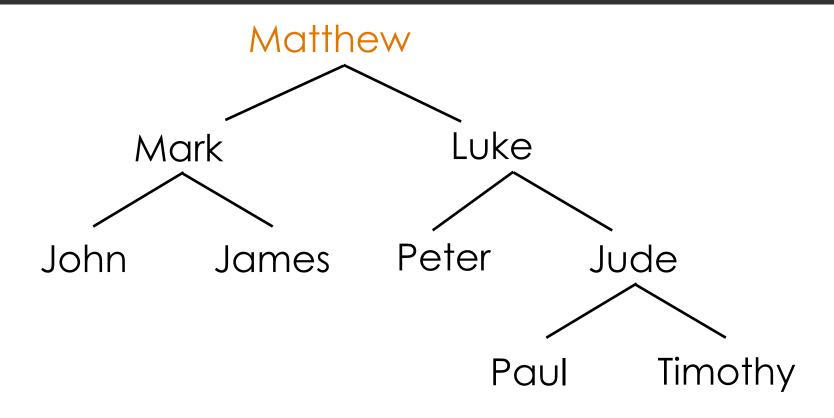


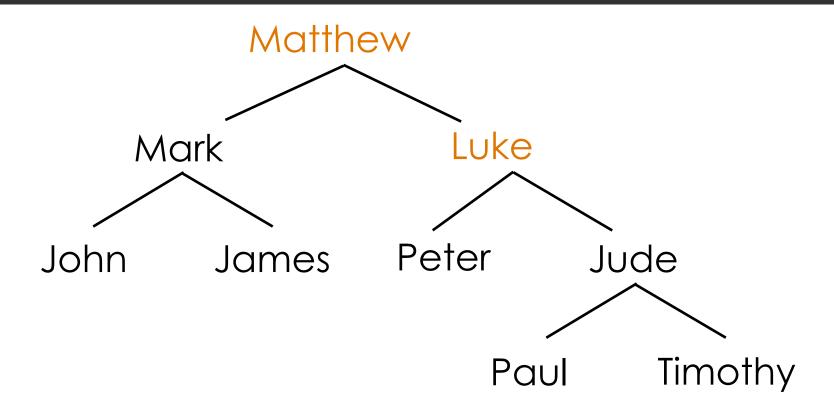
Output: John, James

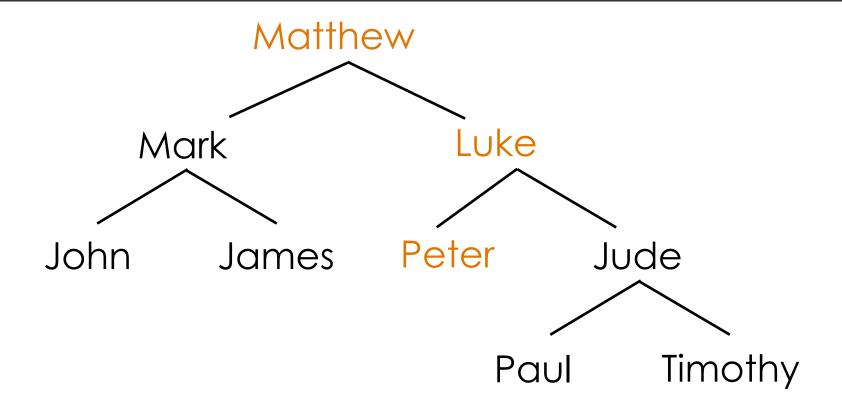


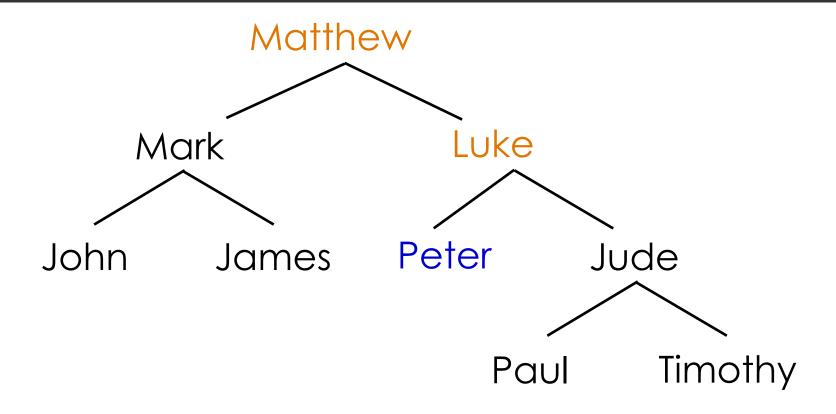
Output: John, James

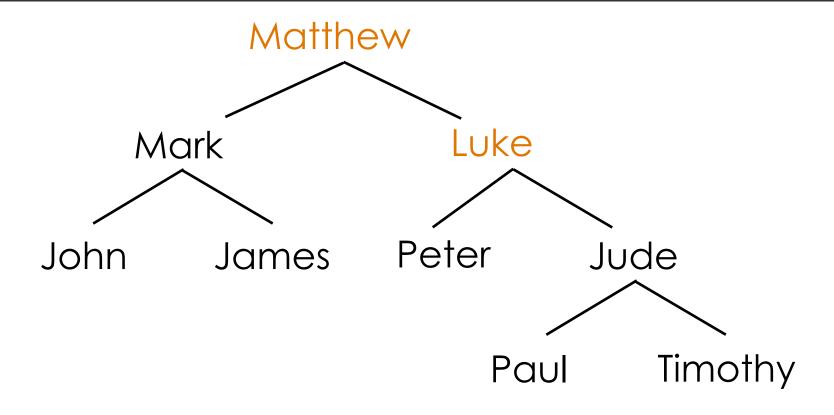


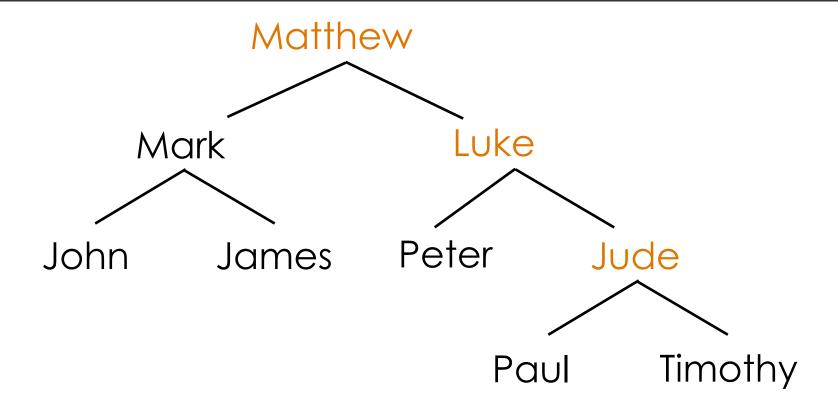


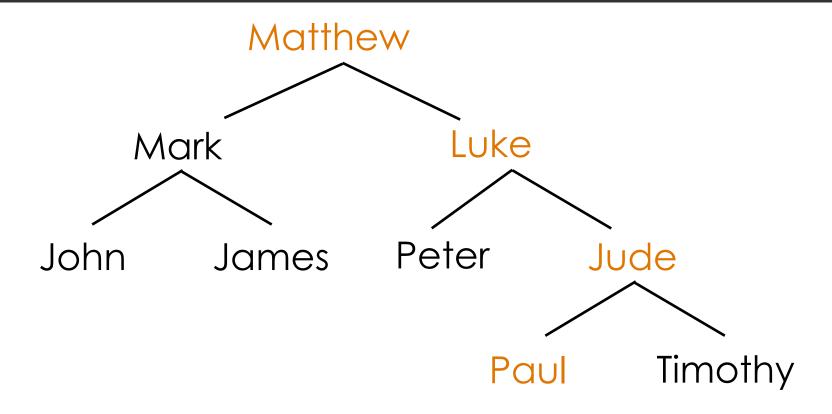


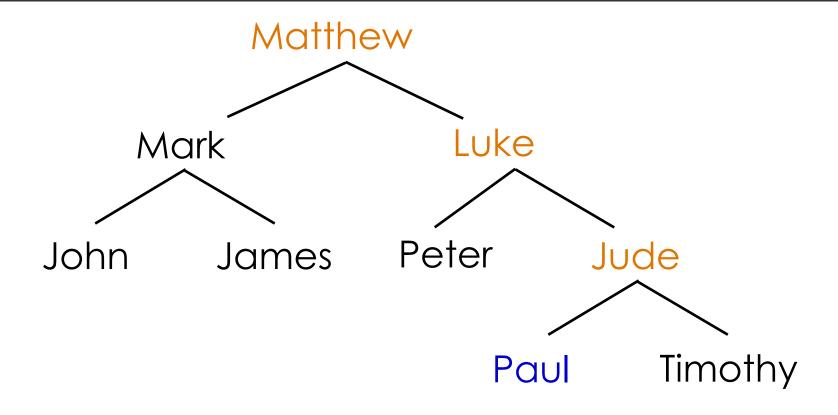


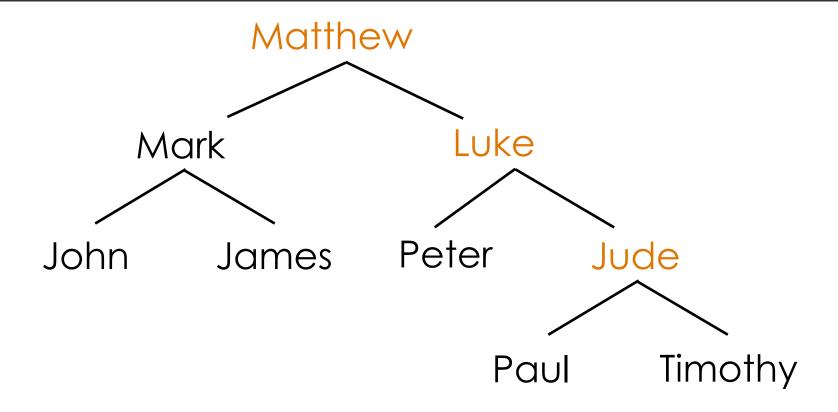


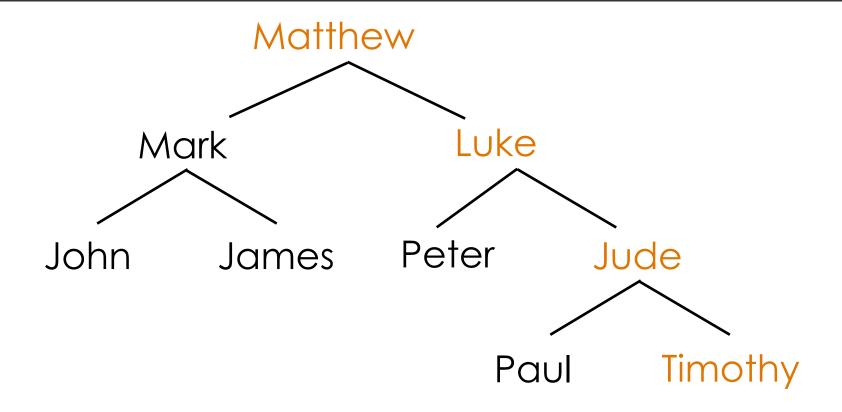


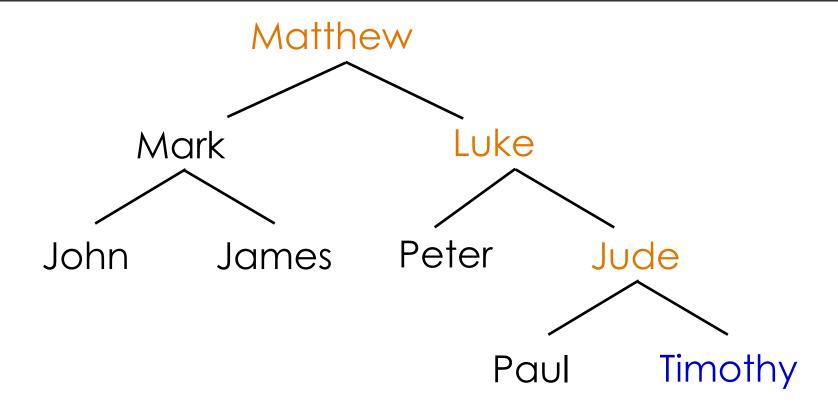




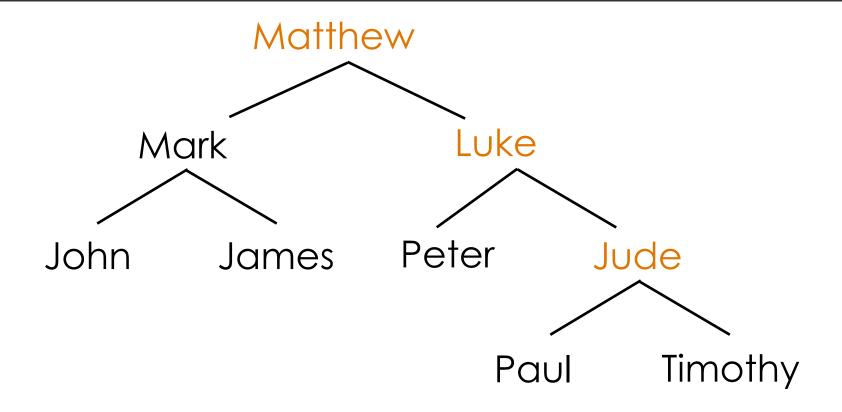




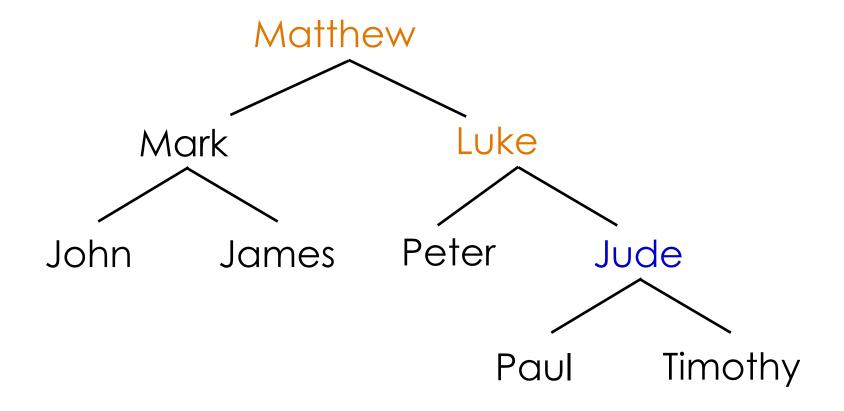




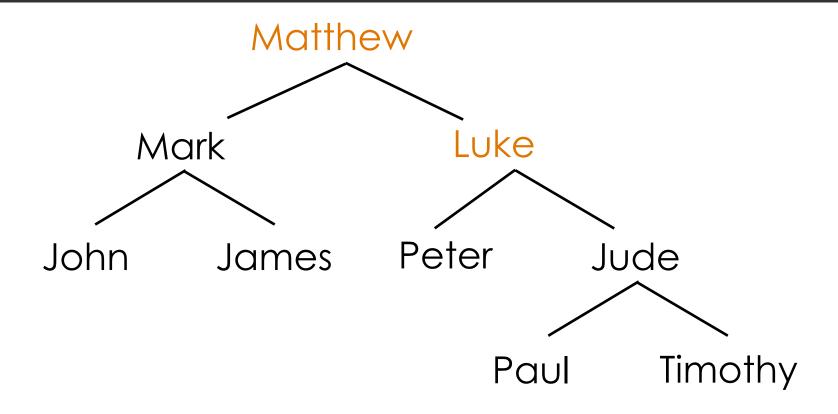
Output: John, James, Mark, Peter, Paul, Timothy



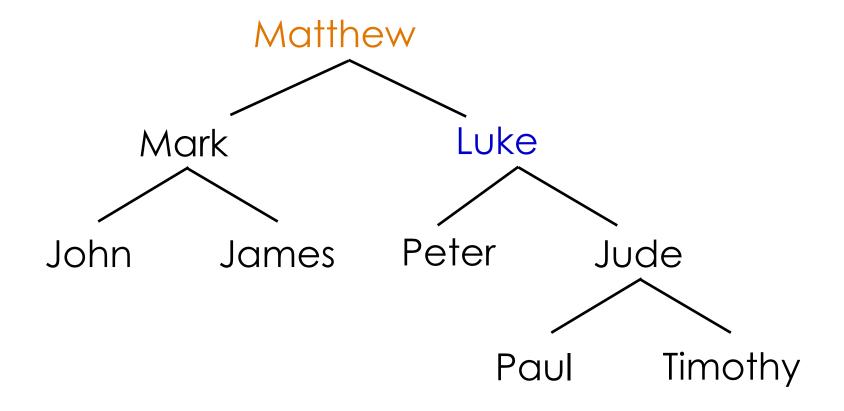
Output: John, James, Mark, Peter, Paul, Timothy



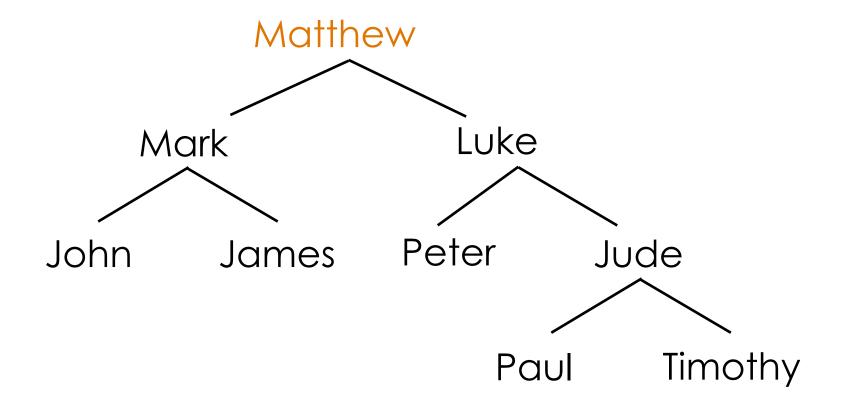
Output: John, James, Mark, Peter, Paul, Timothy, Jude



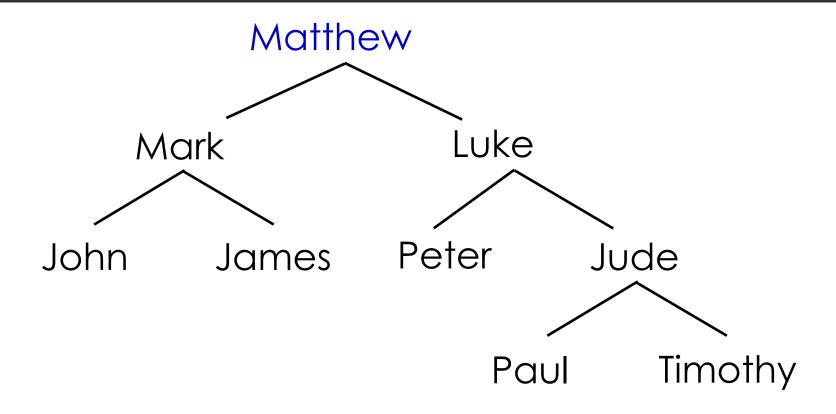
Output: John, James, Mark, Peter, Paul, Timothy, Jude



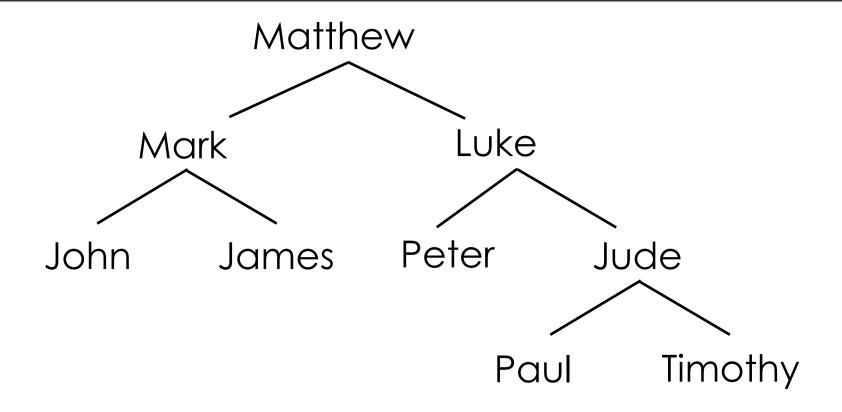
Output: John, James, Mark, Peter, Paul, Timothy, Jude, Luke



Output: John, James, Mark, Peter, Paul, Timothy, Jude, Luke



Output: John, James, Mark, Peter, Paul, Timothy, Jude, Luke, Matthew



Output: John, James, Mark, Peter, Paul, Timothy, Jude, Luke, Matthew

Unordered Tree Summary

- Tree is Data Structure with a node that can point to multiple nodes
- Searching a key in unordered tree takes O(n)
- □ Degenerate trees look like Linked list which takes Θ(n) to build
- If the target key tends to be in the earlier nodes, Breadth-first search is the best strategy
- If the target key tends to be in the deeper nodes or leaf nodes, Depth-first search is the best strategy
- Traversal can be: PreOrder, InOrder, or PostOrder
- Traversal implementation can be either recursive or non-recursive
- Non-recursive implementation of BFT is to use Q whileas DFT is to use Stack