CSc 120
Introduction to Computer Programming II
Adapted from slides by Dr. Saumya Debray

14: Trees
trees: basic concepts
Hierarchies

- Hierarchically organized "stuff" are everywhere
Trees

Nodes
Trees

root node

nodes
Trees

X's parent

X: a node

X's children
Trees: terminology

• A tree is a collection of nodes
• Each node has:
  – ≥ 0 child nodes
  – 0 or 1 parent nodes

• A node with 0 children is called a leaf node
• A node with 0 parent nodes is called the root node
• A tree has:
  – ≥ 1 leaf nodes
  – exactly one root node
Trees: leaves and root

root node

leaf nodes
Tree: Example

- Arizona
  - Maricopa
    - Tempe
    - Glendale
  - Pima
  - Coconino
    - Flagstaff
Tree: Example

$$(7+3) \times (5-2)$$
Binary trees

• A tree where each node has at most two children is called a *binary tree*
Binary trees

left subtree

right subtree
Trees: node representation

• A node in a general tree:
  – value(s) at the node
  – references to child nodes:
    o an extensible data structure (e.g., a list, a linked list, or dictionary)
  – (infrequently) reference to parent

• A node in a binary tree:
  – value(s) at the node
  – a reference to the left subtree
  – a reference to the right subtree
  – (infrequently) reference to parent
class BinaryTree:
    def __init__(self, value):
        self._value = value  # the value at the node
        self._left = None    # left child
        self._right = None   # right child

...
binary search trees
Examine this binary tree:

What can we say about the values in the nodes to the left of 8?

What can we say about the values in the nodes to the right of 8?
Binary search tree (BST)

A *binary search tree* is a binary tree where every node satisfies the following:

- The values at these nodes are smaller than $a$.
- The values at these nodes are bigger than $a$.

The value at this node is $a$. 
Binary search tree: Searching
Searching a BST

Given a BST $T$ and a value $v$, is there a node in $T$ with value $v$?
Searching a BST

Given a BST $T$ and a value $v$, is there a node in $T$ with value $v$?

Idea: at each node with value $a$:
- if $a == v$: done
- if $v < a$: search left subtree
- if $v > a$: search right subtree
Searching a BST

def search(T, v):
    if T == None:
        return False
    if v == T._value:
        return True
    if v < T._value:
        return search(T._left, v)
    else:
        return search(T._right, v)
Searching a BST: Example 1

$v < T._value$:  
$\Rightarrow$ search left subtree
Searching a BST: Example 1

\( v > T._\text{value} \):
\[ \Rightarrow \text{search right subtree} \]
Searching a BST: Example 1

\[ v \approx T.\text{value}: \]  
⇒ return True
Searching a BST: Example 2

\( v > T._\text{value}: \Rightarrow \text{search right subtree} \)
Searching a BST: Example 2

\[ v < T._\text{value}: \Rightarrow \text{search left subtree} \]
Searching a BST: Example 2

\[ v > T.value: \]
\[ \Rightarrow \text{search right subtree} \]
Searching a BST: Example 2

\[ T == \text{None:} \Rightarrow \text{return False} \]
Exercise

Create a BST from this sequence: 7, -2, 10, 0, 13, 14, 3

Do prob. 2 in the handout.

Let’s add nodes to the tree up on the Elmo.

Do prob. 3 in the handout:
modify the code for search(T,v) to define insert(T,v)
Constructing a BST

Given a BST $T$ and a value $v$, return the tree $T'$ obtained by inserting $v$ into $T$

- if $T$ is empty: return a node with value $v$
- otherwise:
  - if $v < T._{value}$: insert into $T$'s left subtree
  - if $v == T._{value}$: done
  - if $v > T._{value}$: insert into $T$'s right subtree
def insert(T, v):
    if T == None:
        return Node(v)
    if v < T._value:
        T._left = insert(T._left, v)
    elif v > T._value:
        T._right = insert(T._right, v)
    return T
Constructing a BST: Example

Sequence of values: $8 \ 3 \ 11 \ 2 \ 9 \ 5$

```
def insert(T, v):
    # (v = 8, T = None)
    if T == None:
        return Node(v)
    if v < T.value:
        T.left = insert(T.left, v)
    elif v > T.value:
        T.right = insert(T.right, v)
    return T
```

T: None
Constructing a BST: Example

Sequence of values: \[8\ 3\ 11\ 2\ 9\ 5\]

def insert(T, v):
    if T == None:
        return Node(v)
    if v < T.value:
        T.left = insert(T.left, v)
    elif v > T.value:
        T.right = insert(T.right, v)
    return T
Constructing a BST: Example

Sequence of values: 8 3 11 2 9 5

def insert(T, v):
    if T == None:
        return Node(v)
    if v < T.value:
        T.left = insert(T.left, v)
    elif v > T.value:
        T.right = insert(T.right, v)
    return T
Constructing a BST: Example

Sequence of values: 8 3 11 2 9 5

```python
def insert(T, v):
    if T == None:
        return Node(v)
    if v < T.value:
        T.left = insert(T.left, v)
    elif v > T.value:
        T.right = insert(T.right, v)
    return T
```

T:

```
  8
  |
  |
```

35
Constructing a BST: Example

Sequence of values: 8 3 11 2 9 5

def insert(T, v):
    if T == None:
        return Node(v)
    if v < T.value:
        T.left = insert(T.left, v)
    elif v > T.value:
        T.right = insert(T.right, v)
    return T
Constructing a BST: Example

Sequence of values: 8 3 11 2 9 5

```python
def insert(T, v):
    if T == None:
        return Node(v)
    if v < T.value:
        T.left = insert(T.left, v)
    elif v > T.value:
        T.right = insert(T.right, v)
    return T
```
Constructing a BST: Example

Sequence of values: 8 3 11 2 9 5

```
def insert(T, v):
    if T == None:
        return Node(v)
    if v < T.value:
        T.left = insert(T.left, v)
    elif v > T.value:
        T.right = insert(T.right, v)
    return T
```
Constructing a BST: Example

Sequence of values: 8 3 11 2 9 5

def insert(T, v):
    if T == None:
        return Node(v)
    if v < T.value:
        T.left = insert(T.left, v)
    elif v > T.value:
        T.right = insert(T.right, v)
    return T
Constructing a BST: Example

Sequence of values: 8 3 11 2 9 5

def insert(T, v):
    if T == None:
        return Node(v)
    if v < T.value:
        T.left = insert(T.left, v)
    elif v > T.value:
        T.right = insert(T.right, v)
    return T
Constructing a BST: Example

Sequence of values: 8 3 11 2 9 5

def insert(T, v):
    if T == None:
        return Node(v)
    if v < T.value:
        T.left = insert(T.left, v)
    elif v > T.value:
        T.right = insert(T.right, v)
    return T
Constructing a BST: Example

Sequence of values: 8 3 11 2 9 5

```python
def insert(T, v):
    if T == None:
        return Node(v)
    if v < T.value:
        T.left = insert(T.left, v)
    elif v > T.value:
        T.right = insert(T.right, v)
    return T
```
Constructing a BST: Example

Sequence of values: 8 3 11 2 9 5

```python
def insert(T, v):
    if T == None:
        return Node(v)
    if v < T.value:
        T.left = insert(T.left, v)
    elif v > T.value:
        T.right = insert(T.right, v)
    return T
```
Constructing a BST: Example

Sequence of values: 8 3 11 2 9 5

def insert(T, v):
    if T == None:
        return Node(v)
    if v < T.value:
        T.left = insert(T.left, v)
    elif v > T.value:
        T.right = insert(T.right, v)
    return T
Constructing a BST: Example

Sequence of values: 8 3 11 2 9 5

```python
def insert(T, v):
    if T == None:
        return Node(v)
    if v < T.value:
        T.left = insert(T.left, v)
    elif v > T.value:
        T.right = insert(T.right, v)
    return T
```

![BST Diagram]
Constructing a BST: Example

Sequence of values: 8 3 11 2 9 5

def insert(T, v):
    if T == None:
        return Node(v)
    if v < T.value:
        T.left = insert(T.left, v)
    elif v > T.value:
        T.right = insert(T.right, v)
    return T
Constructing a BST: Example

Sequence of values:  8  3  11  2  9  5

def insert(T, v):
    if T == None:
        return Node(v)
    if v < T.value:
        T.left = insert(T.left, v)
    elif v > T.value:
        T.right = insert(T.right, v)
    return T
Constructing a BST: Example

Sequence of values: 8 3 11 2 9 5

def insert(T, v):
    if T == None:
        return Node(v)
    if v < T.value:
        T.left = insert(T.left, v)
    elif v > T.value:
        T.right = insert(T.right, v)
    return T
Constructing a BST: Example

Sequence of values: 8 3 11 2 9 5

def insert(T, v): (v = 5, T.value = 8)
    if T == None:
        return Node(v)
    if v < T.value:
        T.left = insert(T.left, v)
    elif v > T.value:
        T.right = insert(T.right, v)
    return T
Constructing a BST: Example

Sequence of values: 8 3 11 2 9 5

def insert(T, v):
    if T == None:
        return Node(v)
    if v < T.value:
        T.left = insert(T.left, v)
    elif v > T.value:
        T.right = insert(T.right, v)
    return T
Constructing a BST: Example

Sequence of values: 8 3 11 2 9 5

```python
def insert(T, v):
    if T == None:
        return Node(v)
    if v < T.value:
        T.left = insert(T.left, v)
    elif v > T.value:
        T.right = insert(T.right, v)
    return T
```
Constructing a BST: Example

Sequence of values: 8 3 11 2 9 5

def insert(T, v):
    if T == None:
        return Node(v)
    if v < T.value:
        T.left = insert(T.left, v)
    elif v > T.value:
        T.right = insert(T.right, v)
    return T
Constructing a BST: Example

Sequence of values: 8 3 11 2 9 5
Does Maricopa come before Coconino? If so, why?
Are the leaves more important than the nodes with children?
What does "order" mean here?
tree traversals
Tree traversals

• A *traversal* of a tree is a systematic way of visiting and processing the nodes of the tree

This usually comes down to the relative order between:
- traversing the subtrees of the node's children; and
- processing the node

"Doing something with the value at the node"
  – e.g., printing it out
Tree traversals (n-ary)

There are three widely used traversals:

• **Preorder traversal**
  – process the node first
  – then traverse (and process) its children

• **Inorder traversal**
  – traverse left subtree children
  – then process the node
  – then traverse right subtree

• **Postorder traversal**
  – traverse (and process) the children
  – then process the node
Tree traversals (binary)

There are three widely used traversals:

• **Preorder traversal**
  – process the node first
  – then traverse (and process) its children

• **Inorder traversal**
  – traverse left subtree children
  – then process the node
  – then traverse right subtree

• **Postorder traversal**
  – traverse (and process) the children
  – then process the node

"pre" – visit node first
"in" – visit node in between
"post" – visit node last
## BinaryTree Traversals

### 3 Traversals

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>preorder:</td>
<td>Visit</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>inorder:</td>
<td>-----</td>
<td>Visit</td>
<td>-----</td>
</tr>
<tr>
<td>postorder:</td>
<td>-----</td>
<td>-----</td>
<td>Visit</td>
</tr>
</tbody>
</table>
Preorder traversal

Algorithm:
- Visit the node
- Recurse on node's Left subtree
- Recurse on node's Right subtree

Ex:

```
(where's the base case?)
```

A

B C D

A B C D
Trace of preorder traversal

<table>
<thead>
<tr>
<th>Output</th>
<th>Trace</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Call preorder(A)</td>
</tr>
<tr>
<td></td>
<td>Visit A</td>
</tr>
<tr>
<td></td>
<td>Left (call preorder(B))</td>
</tr>
<tr>
<td>B</td>
<td>Visit B</td>
</tr>
<tr>
<td></td>
<td>Left – return immediately</td>
</tr>
<tr>
<td></td>
<td>Right (call preorder(C))</td>
</tr>
<tr>
<td>C</td>
<td>Visit C</td>
</tr>
<tr>
<td></td>
<td>Left – return</td>
</tr>
<tr>
<td></td>
<td>Right – return</td>
</tr>
<tr>
<td>D</td>
<td>Right(call preorder(D))</td>
</tr>
<tr>
<td></td>
<td>Visit D</td>
</tr>
<tr>
<td></td>
<td>Left – return</td>
</tr>
<tr>
<td></td>
<td>Right – return</td>
</tr>
</tbody>
</table>
Preorder traversal (n-ary)

def preorder(T):
    process(T._value)
    for i in range(len(T._children)):
        preorder(T._children[i])

(where's the base case?)
Preorder traversal: Example
Preorder traversal: Example

Title
Abstract
...
Sec 1
...
Sec 1.1
...
Sec 1.2
...
Sec 2
...
Sec 2.1
...
Sec 2.2
...
Sec 3
...
References
Inorder traversal

Algorithm:

- Recurse on node's Left subtree
- Visit node
- Recurse on node's Right subtree

Ex:

B C A D
Inorder traversal (binary trees)

def inorder(T):
    if T == None:
        return
    else:
        inorder(T.left())
        process(T.value)
        inorder(T.right())
Inorder traversal: Example

Print out the values in a BST in sorted order

1

2

3

4

5

6

i

print order
Postorder traversal

Algorithm:
- Recurse on node's Left subtree
- Recurse on node's Right subtree
- Visit node

Ex:

(C B D A)
Postorder traversal (n-ary)

def postorder(T):
    for i in range(len(T.children)):
        postorder(T.children[i])  # visit all children first
    process(T.value)
Postorder traversal - Example

Problem: evaluate this expression

\[(x + y * 3) / (n - 1)\]

suppose that: \(x = 3, y = 2, n = 4\)

Solution:

1) Convert the expression to post-fix notation
   • use an auxiliary stack
   • use the algorithm in the handout

2) Represent the expression as a binary tree
   • use a postorder traversal of the tree
Postorder traversal: Example

Evaluate: \((x + y \times 3) / (n - 1)\)

suppose that: \(x = 3, y = 2, n = 4\)
Evaluate: \((x + y \times 3) / (n - 1)\)
suppose that: \(x = 3, y = 2, n = 4\)
Trees ↔ ___str___

How should we print a tree?
Trees ↔ _ _str_ _

Use parenthesis to indicate the structure:

( A ( B None None ) (D None None ) )
Write a function `traverse(t)` for a BST that prints the values in this order: 2, 3, 5, 8, 9, 11.

class BinarySearchTree:
    def __init__(self):
        self._value = None
        self._left = None
        self._right = None
Exercise (ICA-9, p. 2)

Given a binary search tree, write a function `count_leaves(t)` that counts the number of leaf nodes.

- What is the base case? (Is there more than one?)

- What is the smaller amount of computation for the next round of recursion?
Trees ↔ traversals

What is the preorder traversal of this tree?
Inorder?
Postorder?
Trees $\iff$ traversals

- Given a tree, we can figure out its traversals
- Does the converse hold?
  I.e., given a traversal, can we figure out the tree?
  preorder: 1  2  3  4
Trees ↔ traversals

• The two trees below have the same preorder traversal.

Preorder traversal = 1 2 3 4
Exercise (ICA-9, p. 3)

Draw 2 trees that have the in-order traversal: 1 2 3 4
Trees ↔ traversals

• We cannot derive a *unique* tree from a single traversal

• What if we have two traversals?
  – Inorder: 3 5 7 9 4
    o hard to tell where the root is
  – Preorder: 5 3 9 7 4
    o now we know

• Let's draw the tree & figure out an algo to do it
Trees ↔ traversals

Preorder:
- $a$
- $\text{Pre}(T_{\text{left}})$
- $\text{Pre}(T_{\text{right}})$

Inorder:
- $\text{In}(T_{\text{left}})$
- $a$
- $\text{In}(T_{\text{right}})$

$T_{\text{left}}$  $T_{\text{right}}$
Trees ↔ traversals

• Given a preorder \textit{and} an inorder traversal, create the tree

• Given:
  - preorder\_list
  - inorder\_list \quad \text{node sequences from traversals of a tree}

Need to do: build a function:

traversals\_to\_tree(preorder\_list, inorder\_list)

that will return the tree for the given traversals.
Trees ↔ traversals

• Given:
  - preorder_list + inorder_list

• Suppose we can figure out:
  - root
  - preorder_left + preorder_right
  - inorder_left + inorder_right
Trees ↔ traversals

• Given:
  ‒ preorder_list + inorder_list

• Suppose we can figure out:
  ‒ root
  ‒ preorder_left + preorder_right
  ‒ inorder_left + inorder_right

• Then:
  traversals_to_tree(preorder_left, inorder_left)
  traversals_to_tree(preorder_right, inorder_right)
Trees ↔ traversals

• Given:
  – preorder_list + inorder_list

• Suppose we can figure out:
  – root
  – preorder_left + preorder_right
  – inorder_left + inorder_right

• Then:
  \[
  \text{traversals\_to\_tree}(\text{preorder\_left}, \text{inorder\_left}) \rightarrow T_{\text{left}} \\
  \text{traversals\_to\_tree}(\text{preorder\_right}, \text{inorder\_right}) \rightarrow T_{\text{right}}
  \]
  \[\text{recursion}\]
Trees ↔ traversals

- **Given:**
  - preorder_list + inorder_list

- **Suppose we can figure out:**
  - root
  - preorder_left + preorder_right
  - inorder_left + inorder_right

- **Then:**
  - traversals_to_tree(preorder_left, inorder_left) → \( T_{left} \)
  - traversals_to_tree(preorder_right, inorder_right) → \( T_{right} \)
Trees ↔ traversals

Preorder:
\[ \text{Pre}(T_{\text{left}}) \quad \text{Pre}(T_{\text{right}}) \]
\[ a \]

Inorder:
\[ \text{In}(T_{\text{left}}) \quad \text{In}(T_{\text{right}}) \]
\[ a \]
more traversals
Consider: game playing

Goal: to write a program to play a 2-person game (e.g., tic-tac-toe, chess, go, ...)

How does this work?
Consider: game playing

Goal: to write a program to play a 2-person game (e.g., tic-tac-toe, chess, go, ...)

Generate successive levels of board positions
- At each level, pick best move for the player at that level
- Work backwards to find the move that will lead to the best position $n$ moves later

*best position for the computer among all positions at this level*
Consider: game playing

- For a nontrivial game (e.g., chess, go) the tree is usually too large to build or explore fully
  - also, usually there are time constraints on play
  - our previous tree traversal algorithms don't work
- Game-playing algorithms typically explore the tree level by level
  - consider the nodes at depth 1, then depth 2, etc.
Level-by-level tree traversal
Level-by-level tree traversal
Level-by-level tree traversal

This order of traversal is called *breadth-first traversal*
Breadth-first tree traversal

Breadth-first traversal order:

1 2 3 4 5 6 7 8 9 10
Breadth-first tree traversal

Data structure: use a queue $q$

Algorithm:
• Create a queue $q$
• Put the root in $q$
• While $q$ not empty
  o node = $q$.dequeue()
  o process node
  o enqueue its children
Breadth-first vs. Depth-first

• Stacks and queues are closely related structures
• What if we use a stack in our tree traversal?
  – the deeper levels of the tree are explored first
  – this is referred to as depth-first traversal
BST / Complexity
Binary Search tree: complexity

Searching: $O(\log n)$, where $n$ is the number of elements in the tree

Note: this tree is balanced

What if the tree is not balanced?
BST / Complexity

• Unbalanced BST
  – How many comparisons does it take to find 2?
  – Worst case, complexity can be $O(n)$
  – Skewed trees can result from sorted input

• Balanced trees
  – AVL Trees: trees are kept balanced on insertion, deletion, etc.
Trees: summary

• An n-ary tree represents a hierarchy

• A binary tree represents a sequence

• They show up in all kinds of contexts
  – including many in computer science

• Various kinds of tree traversals reflect different ways of processing the information and structure of trees

• Recursion is often the simplest way to process trees
**Assgn 12 - Huffman Coding**

- ASCII uses fixed length encodings:

<table>
<thead>
<tr>
<th>char</th>
<th>ASCII</th>
<th>binary</th>
</tr>
</thead>
<tbody>
<tr>
<td>g</td>
<td>103</td>
<td>1100111</td>
</tr>
<tr>
<td>o</td>
<td>111</td>
<td>1101111</td>
</tr>
<tr>
<td>p</td>
<td>112</td>
<td>1110000</td>
</tr>
<tr>
<td>h</td>
<td>104</td>
<td>1101000</td>
</tr>
<tr>
<td>e</td>
<td>101</td>
<td>1100101</td>
</tr>
<tr>
<td>r</td>
<td>114</td>
<td>1110010</td>
</tr>
<tr>
<td>s</td>
<td>115</td>
<td>1110011</td>
</tr>
</tbody>
</table>

What is this word?
1101000110111111001011100111100101

binary: 1101000 1101111 1110010 1110011 1100101 (group by 7's)
ASCI: 104 111 114 115 101
Assign 12 – Huffman Coding

• Huffman coding:
• Use fewer bits (not 7) for more frequently occurring characters
• Do this by using a tree that stores characters at the leaves
• root-to-leaf paths provide the bit sequence used to encode the characters
Assign 12 – Huffman Coding

Source: Astrachan

path: right-right-left-left-left-left-right using the 0/1 convention
Assign 12 – Huffman Coding

The structure of the tree can be used to determine the coding of any leaf by using the 0/1 edge convention. A different tree gives a different coding.

The tree below gives the coding on the right.

What would the encoding be for hope?

Source: Astrachan
Prefix codes/Huffman codes

*prefix property:* no bit-sequence encoding of a character is the prefix of any other bit-sequence encoding.

When all characters are stored in leaves and every non-leaf node has two children, the coding produced by the 0/1 convention has the prefix property

*invented by Huffman 1952*
Exercise (Answer)

# count the leaves of a tree

def count_leaves(t):
    if t == None:
        return 0
    if t._left == None and t._right == None:
        return 1
    else:
        return count_leaves(t._left) + \ count_leaves(t._right)
Exercise

State three things about this tree.