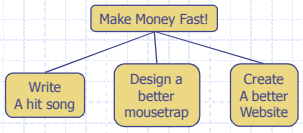
Lecture 3: Trees and Amortized Analysis

Sequential Unfoldment of Natural Law

Wholeness Statement

Trees are data structures that provide wide ranging capabilities and a highly flexible perspective on a set of element objects. The whole range of space and time is open to individuals with fully developed awareness.

Trees

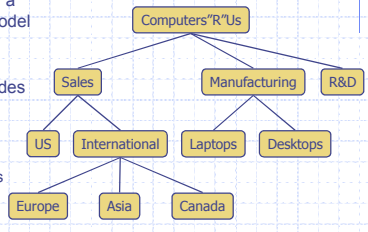


Outline and Reading

- Tree ADT (§2.3.1)
- Preorder and postorder traversals (§2.3.2)
- BinaryTree ADT (§2.3.3)
- Inorder traversal (§2.3.3)
- Euler Tour traversal (§2.3.3)
- Template method pattern
- Data structures for trees (§2.3.4)

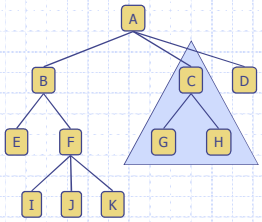
What is a Tree

- In computer science, a tree is an abstract model of a hierarchical structure
- A tree consists of nodes with a parent-child relation
- Applications:
  - Organization charts
  - File systems
  - Programming environments



Tree Terminology

- ***Root***: only node without parent (A)
- ***Internal node***: node with at least one child (A, B, C, F)
- ***External node*** (a.k.a. leaf): node without children (E, I, J, K, G, H, D)
- ***Ancestors of a node***: parent, grandparent, grand-grandparent, etc.
- ***Depth of a node***: number of ancestors
- ***Height of a tree***: maximum depth of any node (3 in tree to right)
- ***Descendant of a node***: child, grandchild, grand-grandchild, etc.
- ***Subtree***: tree consisting of a node and its descendants



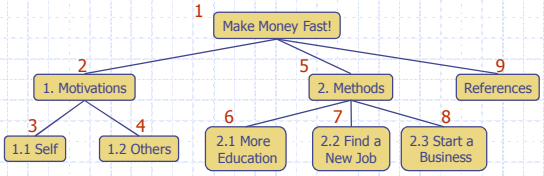
Tree ADT

- We use positions to abstract nodes
- Generic methods:
  - integer size()
  - boolean isEmpty()
  - objectIterator elements()
  - positionIterator positions()
- Accessor methods:
  - position root()
  - position parent(p)
  - positionIterator children(p)
- Query methods:
  - boolean isInternal(p)
  - boolean isExternal(p)
  - boolean isRoot(p)
- Update methods:
  - swapElements(p, q)
  - object replaceElement(p, o)
- Additional update methods may be defined by data structures implementing the Tree ADT

Preorder Traversal

- A traversal visits the nodes of a tree in a systematic manner
- In a preorder traversal, a node is visited before its descendants
- Application: print a structured document

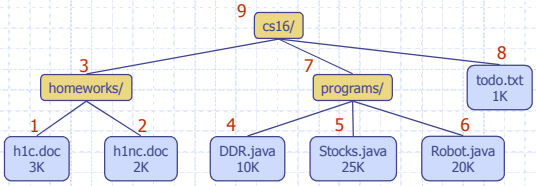
Algorithm *preOrder(v)*
  visit(v)
  for each child *w* of *v*
    preOrder (w)



Postorder Traversal

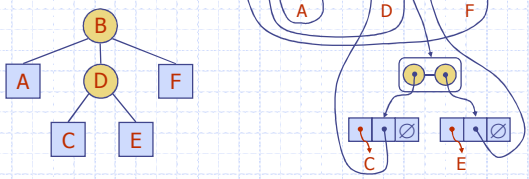
- In a postorder traversal, a node is visited after its descendants
- Application: compute space used by files in a directory and its subdirectories

Algorithm *postOrder(v)*
  for each child *w* of *v*
    postOrder (w)
  visit(v)



Data Structure for Trees

- A node is represented by an object storing
  - Element
  - Parent node
  - Sequence of children nodes
- Node objects implement the Position ADT



Performance of Tree ADT Linked Implementation

- Generic methods:
  - integer size()
  - boolean isEmpty()
  - objectIterator elements()
  - positionIterator positions()
- Accessor methods:
  - position root()
  - position parent(p)
  - positionIterator children(p)
- Query methods:
  - boolean isInternal(p)
  - boolean isExternal(p)
  - boolean isRoot(p)
- Update methods:
  - swapElements(p, q)
  - object replaceElement(p, o)

Linked Implementation of the Tree ADT

Operation	Time
size, isEmpty	
positions, elements	
swapElements(p, q), replaceElement(p, o)	
root, parent(p)	
children(v)	
isInternal(p), isExternal(p), isRoot(p)	

Linked Implementation of the Tree ADT

Operation	Time
size, isEmpty	1
positions, elements	n
swapElements(p, q), replaceElement(p, o)	1
root, parent(p)	1
children(v)	c_v
isInternal(p), isExternal(p), isRoot(p)	1

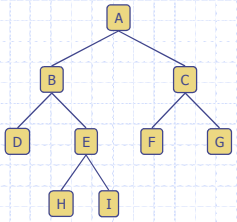
Main Point

1. The Tree ADT models a hierarchical structure between objects simplified to a parent-child relation. Nodes store arbitrary objects and connect to other nodes in the tree. A rooted tree has a root node without a parent; all other nodes have parents.
   Pure consciousness is the root of the tree of life.

Binary Tree

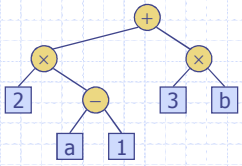
- A (proper) binary tree is a tree with the following properties:
  - Each internal node has two children
  - The children of a node are an ordered pair
- We assume that all binary trees are proper
- We call the children of an internal node left child and right child
- A binary tree is either
  - a tree consisting of a single node, or
  - a tree whose root has an ordered pair of children, each of which is a binary tree

- Applications:
  - arithmetic expressions
  - decision processes
  - searching



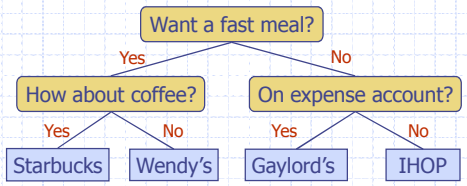
Arithmetic Expression Tree

- Binary tree associated with an arithmetic expression
  - internal nodes: operators
  - external nodes: operands
- Example: arithmetic expression tree for the expression $(2 \times (a - 1) + (3 \times b))$



Decision Tree

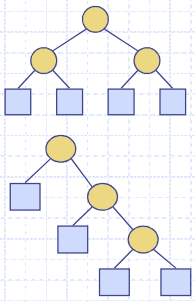
- Binary tree associated with a decision process
  - internal nodes: questions with yes/no answer
  - external nodes: decisions
- Example: dining decision



Binary Tree Theorem 2.8 (page 84)

Properties of Binary Trees
0. $e = i + 1$
1. $h \leq i \leq 2^h - 1$
2. $h+1 \leq e \leq 2^h$
3. $2h+1 \leq n \leq 2^{h+1} - 1$
4. $\log (n+1) - 1 \leq h \leq (n-1)/2$

where
- n number of nodes
- h height of the tree
- e number of external nodes
- i number of internal nodes



Proof of Theorem 2.8

Properties 2, 3, and 4 can be proven based on Property 1

Property 2:

	$h \le i \le 2^h - 1$	(Property 1)
<==>	$h \le e - 1 \le 2^h - 1$	(Property 0, e = i+1)
<==>	$h+1 \le e \le 2^h$	(arithmetic) ■

Property 3:

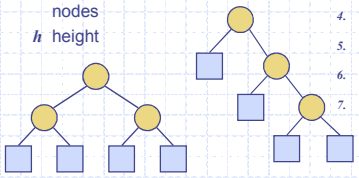
	$h+1 \le e \le 2^h \ \wedge \ h \le i \le 2^h - 1$	(Properties 1 and 2)
<==>	$2h+1 \le e+i \le 2^h + 2^h - 1$	(arithmetic)
<==>	$2h+1 \le n \le 2^{h+1} - 1$	(since n = i+e and $2(2^h) = 2^{h+1}$) ■

Property 4:

	$n = 2e - 1$	(since i = e – 1 and n = e+i)
<==>	$(n+1)/2 = e$	(algebra)
<==>	$(n+1)/2 = e \ \wedge \ h+1 \le (n+1)/2$	(since h+1 ≤ e, from Property 1)
<==>	$(n+1)/2 = e \ \wedge \ h \le (n-1)/2$	(algebra)
==>	$(n+1)/2 \le 2^h \ \wedge \ h \le (n-1)/2$	($e \le 2^h$, from Property 1)
<==>	$\log_2 (n+1)/2 \le \log_2 2^h \ \wedge \ h \le (n-1)/2$	(logarithms)
<==>	$\log_2 (n+1) - 1 \le h \ \wedge \ h \le (n-1)/2$	(logarithms) ■

Properties of Binary Trees

- Notation
  - n number of nodes
  - e number of external nodes
  - i number of internal nodes
  - h height
- Properties:
  1. $e = i + 1$
  2. $n = 2e - 1$
  3. $h \le i$
  4. $h \le (n-1)/2$
  5. $e \le 2^h$
  6. $h \ge \log_2 e$
  7. $h \ge \log_2 (n+1) - 1$



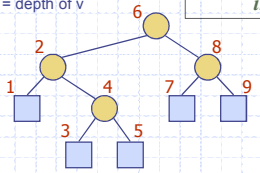
BinaryTree ADT

- The BinaryTree ADT extends the Tree ADT, i.e., it inherits all the methods of the Tree ADT
- Additional methods:
  - position leftChild(p)
  - position rightChild(p)
  - position sibling(p)
- Update methods may be defined by data structures implementing the BinaryTree ADT

Inorder Traversal

- In an inorder traversal a node is visited after its left subtree and before its right subtree
- Application: draw a binary tree
  - x(v) = inorder rank of v
  - y(v) = depth of v

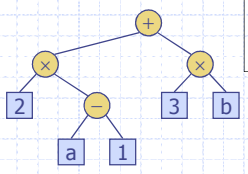
```
Algorithm inOrder(v)
    if isInternal (v) then
        inOrder (leftChild (v))
    visit(v)
    if isInternal (v) then
        inOrder (rightChild (v))
```



Print Arithmetic Expressions

- Specialization of an inorder traversal
  - print operand or operator when visiting node
  - print "(" before traversing left subtree
  - print ")" after traversing right subtree

```
Algorithm printExpression(v)
    if isInternal (v) then
        print("(")
        printExpression(leftChild (v))
    print(v.element ())
    if isInternal (v) then
        printExpression(rightChild (v))
        print (")")
```

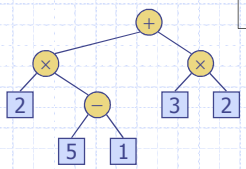


$((2 \times (a - 1)) + (3 \times b))$

Evaluate Arithmetic Expressions

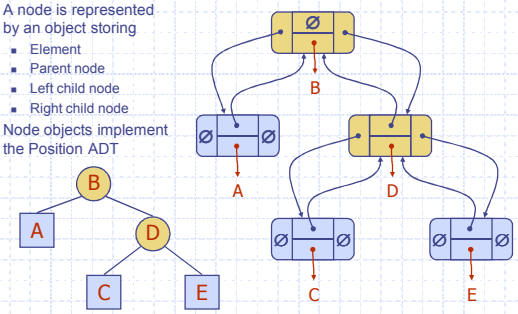
- Specialization of a postorder traversal
  - recursive method returning the value of a subtree
  - when visiting an internal node, combine the values of the subtrees

```
Algorithm evalExpr(v)
    if isExternal (v) then
        return v.element ()
    else
        x ← evalExpr(leftChild (v))
        y ← evalExpr(rightChild (v))
        ◊ ← operator stored at v
        return x ◊ y
```



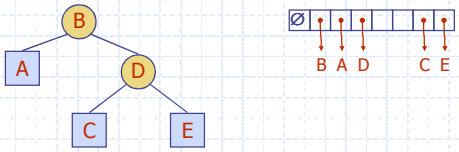
Data Structure for Binary Trees

- A node is represented by an object storing
  - Element
  - Parent node
  - Left child node
  - Right child node
- Node objects implement the Position ADT



Data Structure for Binary Trees

- Another alternative: use an array to store the binary tree.
- Node objects are referenced by index:
  - Index 0 is empty and not used.
  - Root node is at index 1
  - Left child is at 2*index
  - Right child is at 2*index+1



Array-Based Implementation of Binary Tree

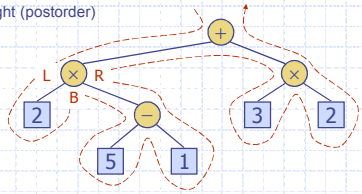
Operation	Time
size, isEmpty	
positions, elements	
swapElements(p, q), replaceElement(p, e)	
root, parent(p), children(p)	
isInternal(p), isExternal(p), isRoot(p)	

Array-Based Implementation of Binary Tree

Operation	Time
size, isEmpty	1
positions, elements	n
swapElements(p, q), replaceElement(p, e)	1
root, parent(p), children(p)	1
isInternal(p), isExternal(p), isRoot(p)	1

Euler Tour Traversal

- Generic traversal of a binary tree
- Includes as special cases the preorder, postorder, and inorder traversals
- Walk around the tree and visit each node three times:
  - on the left (preorder)
  - from below (inorder)
  - on the right (postorder)



Template Method Pattern

- Generic algorithm that can be specialized by redefining certain steps
- Implemented by means of an abstract Java class
- Visit methods that can be redefined by subclasses
- Template method eulerTour
  - Recursively called on the left and right children
  - A Result object with fields leftResult, rightResult and finalResult keeps track of the output of the recursive calls to eulerTour

```
public abstract class EulerTour {
    protected BinaryTree tree;
    protected void visitExternal(Position p, Result r) { }
    protected void visitLeft(Position p, Result r) { }
    protected void visitBelow(Position p, Result r) { }
    protected void visitRight(Position p, Result r) { }
    protected Object eulerTour(Position p) {
      Result r = new Result();
      if tree.isExternal(p) {
        visitExternal(p, r);
      } else {
        visitLeft(p, r);
        r.leftResult = eulerTour(tree.leftChild(p));
        visitBelow(p, r);
        r.rightResult = eulerTour(tree.rightChild(p));
        visitRight(p, r);
      }
      return r.finalResult;  …
```

Specializations of EulerTour

- We show how to specialize class EulerTour to evaluate an arithmetic expression
- Assumptions
  - External nodes store Integer objects
  - Internal nodes store Operator objects supporting method operation (Integer, Integer)

```
public class EvaluateExpression
                extends EulerTour {
  protected void visitExternal(Position p, Result r) {
    r.finalResult = (Integer) p.element();
  }
  protected void visitRight(Position p, Result r) {
    Operator op = (Operator) p.element();
    r.finalResult = op.operation(
                (Integer) r.leftResult,
                (Integer) r.rightResult
                );
  }
  …
}
```


Main Point

2. Each internal node of a Binary Tree has two children and each external node has no children. Thus the height, h, of a binary tree ranges as follows: $i \geq h \geq \log_2 e$, that is, $O(n) \geq h \geq O(\log_2 n)$. Pure consciousness spans the full range of life, from smaller than the smallest to larger than the largest.

Amortization (§1.5)

Analysis of growable array-based stacks and queues

Amortization (§1.5)

- Comes from the field of accounting
  - Provides a monetary metaphor for algorithm analysis
- Useful for understanding the running time of algorithms that have steps with widely varying performance
  - i.e., each step performs a widely varying amount of work
  - Rather than focusing on individual operations, we study the interactions of a series of operations

Aggregate Amortized Analysis

- The average time required to perform an operation within a sequence of operations
  - The worst-case total running time of a series of operations divided by the number of operations
- Guarantees the average performance of each operation in the worst case

Aggregate Analysis

- Determine an upper bound, T(n),
  - the total cost of a sequence of n operations
- The average cost per operation is then T(n)/n
- The average cost becomes the amortized cost of each operation
- Thus all operations have the same amortized cost
  - Even though the cost of each individual operation varies widely

Growable Array-based Stack

- In a push operation, when the array is full, instead of throwing an exception, we can replace the array with a larger one
- How large should the new array be?
  - incremental strategy: increase the size by a constant c
  - doubling strategy: double the size

```
Algorithm push(o)
  if t = S.length − 1 then
    A ← new array of
        size …
    for i ← 0 to t do
      A[i] ← S[i]
    S ← A
  t ← t + 1
  S[t] ← o
```


Comparison of the Strategies

- We compare the incremental strategy and the doubling strategy by analyzing the total time $T(n)$ needed to perform a series of n push operations
- We start with an empty stack represented by an array of size 1
- We call amortized time of a push operation the average time taken by a push over the series of operations, i.e., $T(n)/n$

```
Algorithm push(o)
  if t = S.length − 1 then
    A ← new array of
        size …
    for i ← 0 to t do
      A[i] ← S[i]
    S ← A
  t ← t + 1
  S[t] ← o
```


Incremental Strategy Analysis

- We replace the array $k = n/c$ times
- The total time $T(n)$ of a series of n push operations is proportional to

$$n + c + 2c + 3c + 4c + \ldots + kc =$$
$$n + c(1 + 2 + 3 + \ldots + k) =$$
$$n + ck(k + 1)/2$$

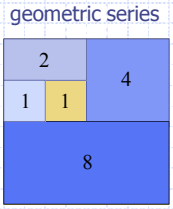
- Since c is a constant, $T(n)$ is $O(n + k^2)$, i.e., $O(n^2)$
- Thus the amortized time of a push operation is $O(n)$

Doubling Strategy Analysis

- We replace the array $k = \log_2 n$ times
- The total time $T(n)$ of a series of n push operations is proportional to

$$n + 1 + 2 + 4 + 8 + \ldots + 2^k =$$
$$n + 2^{k+1} - 1 = 3n - 1$$

- $T(n)$ is $O(n)$
- The amortized time of a push operation is $O(1)$



Array Based Stack

- Advantage
  - Avoids the usual cost of copying array elements because there is no inserting or deleting of elements in the middle of the array
- Disadvantage
  - If many more pushes than pops, the array has to be resized often, which is costly

Quiz: Growable Array-based Queue

- In an enqueue operation, when the array is full, instead of throwing an exception, we can replace the array with a larger one
- What is the amortized running time of the enqueue operation for incremental and doubling strategies?
  - Hint: Similar to what we did for a growable array-based stack

Growable Array-based Queue

- The enqueue operation has amortized running time
  - $O(n)$ with the incremental strategy
  - $O(1)$ with the doubling strategy

Other Amortization Techniques

The Accounting Method

- Uses a scheme of debits and credits to keep track of the running time of a series of operations
- Some operations are overcharged, others are undercharged
- The amount charged is called its *amortized cost*
- When amortized cost exceeds actual cost, the difference is assigned to specific objects within the data structure as credit
- Credits are used to pay for other operations that are charged less than they actually cost
- Amortized costs must be chosen carefully
- The total amortized cost of a sequence of operations must be an upper bound on the actual cost

Accounting Method Example:

push(o) – actual cost 1
pop() – actual cost 1
multipop(k) – actual cost *min*(k, n)

Accounting method:
push(o) – amortized cost 2
pop() – amortized cost 0
multipop(k) – amortized cost 0

When we do a push, we charge the actual cost (1 unit) and associate a credit of 1 unit with each element on the stack
When we do a pop or multipop, we charge 0 but use the credit associated with each element popped to pay for the operation

The Potential Method

- Determine the amortized cost of each operation
- Overcharge operations early to compensate for undercharges later
- Maintains the credit as the "potential energy" of the data structure as a whole instead of associating the credit with individual objects within the data structure

Main Point

3. The idea of "borrowing" and later "repaying" a data structure or program can be useful for determining the worst case time complexity of algorithms that have operations with widely varying running times. The basic idea of amortized analysis is that, even though a few operations are very costly, they do not occur often enough to dominate the entire algorithm; that is, the number of less costly operations far outnumber the costly ones over a large number of executions. Natural law (physics) says that for every action there is an equal and opposite reaction. To avoid mistakes, it is important to perform action from the silent, orderly level of our own consciousness.

Connecting the Parts of Knowledge with the Wholeness of Knowledge

1. The tree ADT is a generalization of the linked-list in which each tree node can have any number of children instead of just one. A binary tree is a special case of the generic tree ADT in which each node has either 0 or 2 children (a left and right child).
2. Any ADT will have a variety of implementations of its operations with varying efficiencies, e.g., the binary tree can be implemented as either a set of recursively defined nodes or as an array of elements.

3. **Transcendental Consciousness** is pure intelligence, the abstract substance out of which the universe is made.
4. **Impulses within Transcendental Consciousness**: Within this field, the laws of nature continuously organize and govern all activities and processes in creation.
5. **Wholeness moving within itself** : In Unity Consciousness, awareness is awake to its own value, the full value of the intelligence of nature. One's consciousness supports the knowledge that outer is the expression of inner, creation is the play and display of the Self.