Trees

- Linear access time of linked lists is prohibitive
  - Does there exist any simple data structure for which the running time of most operations (search, insert, delete) is O(\log N)?

Some Terminologies

- Child and parent
  - Every node except the root has one parent
  - A node can have an arbitrary number of children
- Leaves
  - Nodes with no children
- Sibling
  - Nodes with same parent

Trees

- A tree is a collection of nodes
  - The collection can be empty
  - (recursive definition) If not empty, a tree consists of a distinguished node \( r \) (the root), and zero or more nonempty subtrees

Some Terminologies

- Path
- Length
  - number of edges on the path
- Depth of a node
  - length of the unique path from the root to that node
  - The depth of a tree is equal to the depth of the deepest leaf
- Height of a node
  - length of the longest path from that node to a leaf
  - all leaves are at height 0
  - The height of a tree is equal to the height of the root
- Ancestor and descendant
  - Proper ancestor and proper descendant

Example: UNIX Directory
Binary Trees

- A tree in which no node can have more than two children
- The depth of an “average” binary tree is considerably smaller than N, even though in the worst case, the depth can be as large as N – 1.

Example: Expression Trees

- Leaves are operands (constants or variables)
- The other nodes (internal nodes) contain operators
- Will not be a binary tree if some operators are not binary

Tree traversal

- Used to print out the data in a tree in a certain order
- Pre-order traversal
  - Print the data at the root
  - Recursively print out all data in the left subtree
  - Recursively print out all data in the right subtree

Preorder, Postorder and Inorder

- Preorder traversal
  - node, left, right
  - prefix expression

  - Postorder traversal
  - left, right, node
  - postfix expression

  - Inorder traversal
  - left, node, right
  - infix expression

• Postorder

Preorder, Postorder and Inorder

Algorithm Preorder(x)
Input: x is the root of a subtree.
1. If x ≠ NULL
2. then output key(x);
3. Preorder(left(x));
4. Preorder(right(x));

Algorithm Postorder(x)
Input: x is the root of a subtree.
1. if x ≠ NULL
2. then Postorder(left(x));
3. Postorder(right(x));
4. output key(x);

Algorithm Inorder(x)
Input: x is the root of a subtree.
1. if x ≠ NULL
2. then Inorder(left(x));
3. output key(x);
4. Inorder(right(x));

Binary Search Trees

• Stores keys in the nodes in a way so that searching, insertion and deletion can be done efficiently.

Binary search tree property
- For every node X, all the keys in its left subtree are smaller than the key value in X, and all the keys in its right subtree are larger than the key value in X.

A binary search tree
Not a binary search tree

Binary Trees

• Possible operations on the Binary Tree ADT
  - parent
  - left_child, right_child
  - sibling
  - root, etc

• Implementation
  - Because a binary tree has at most two children, we can keep direct references in them.

```cpp
struct BinaryNode
{
    Object element;        // The data in the node
    BinaryNode *left;     // Left child
    BinaryNode *right;    // Right child
};
```
Binary search trees

- Average depth of a node is $O(\log N)$;

Two binary search trees representing the same set:

Implementation

```cpp
template <class Comparable>
class BinarySearchTree;

template <class Comparable>
class BinaryNode {
  Comparable *element;
  BinaryNode *left;
  BinaryNode *right;
  BinaryNode( const Comparable &theElement, BinaryNode *lt, BinaryNode *rt ) : element(theElement), left(lt), right(rt) {};
};

friend class BinarySearchTree<Comparable>;
```

Figure 4.16 The BinaryNode class

Searching (Find)

- Find $X$: return a pointer to the node that has key $X$, or NULL if there is no such node

```cpp
template <class Comparable>
BinaryNode<Comparable> *
BinarySearchTree<Comparable>::find( const Comparable &x, BinaryNode<Comparable> *t ) const {
  if ( t == NULL )
    return NULL;
  else if ( x < t->element )
    return find( x, t->left );
  else if ( x > t->element )
    return find( x, t->right );
  else
    return t;  // Match
}
```

- Time complexity

Searching BST

- If we are searching for 15, then we are done.
- If we are searching for a key < 15, then we should search in the left subtree.
- If we are searching for a key > 15, then we should search in the right subtree.

Inorder traversal of BST

- Print out all the keys in sorted order

Inorder: 2, 3, 4, 6, 7, 9, 13, 15, 17, 18, 20
findMin/ findMax

- Return the node containing the smallest element in the tree
- Start at the root and go left as long as there is a left child. The stopping point is the smallest element

```cpp
template <class Comparable>
BinaryTreeNode<Comparable>* findMin(BinaryTreeNode<Comparable>* t) const
{
    if (t == NULL)
        return NULL;
    if (t->left == NULL)
        return t;
    return findMin(t->left);
}
```

- Similarly for findMax
- Time complexity = O(height of the tree)

insert

- Proceed down the tree as you would with a find
- If X is found, do nothing (or update something)
- Otherwise, insert X at the last spot on the path traversed

![Tree traversal](image)

- Time complexity = O(height of the tree)

delete

- When we delete a node, we need to consider how we take care of the children of the deleted node.
  - This has to be done such that the property of the search tree is maintained.

Three cases:

1. the node is a leaf
   - Delete it immediately
2. the node has one child
3. the node has 2 children
   - replace the key of that node with the minimum element at the right subtree
   - delete the minimum element
     - Has either no child or only right child because if it has a left child, that left child would be smaller and would have been chosen. So invoke case 1 or 2.

![Delete with one child](image)

- Time complexity = O(height of the tree)

Priority Queues – Binary Heaps

homes.cs.washington.edu/~anderson/iucee/Slides_326.../Heaps.ppt
Recall Queues

- FIFO: First-In, First-Out
- Some contexts where this seems right?
- Some contexts where some things should be allowed to skip ahead in the line?

Applications of the Priority Queue

- Select print jobs in order of decreasing length
- Forward packets on routers in order of urgency
- Select most frequent symbols for compression
- Sort numbers, picking minimum first
- Anything greedy

Queues that Allow Line Jumping

- Need a new ADT
- Operations: Insert an Item, Remove the “Best” Item

Potential Implementations

<table>
<thead>
<tr>
<th></th>
<th>insert</th>
<th>deleteMin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unsorted list (Array)</td>
<td>O(1)</td>
<td>O(n)</td>
</tr>
<tr>
<td>Unsorted list (Linked-List)</td>
<td>O(1)</td>
<td>O(n)</td>
</tr>
<tr>
<td>Sorted list (Array)</td>
<td>O(n)</td>
<td>O(1)*</td>
</tr>
<tr>
<td>Sorted list (Linked-List)</td>
<td>O(n)</td>
<td>O(1)</td>
</tr>
</tbody>
</table>

Priority Queue ADT

1. **PQueue data**: collection of data with priority
2. **PQueue operations**
   - insert
   - deleteMin
3. **PQueue property**: for two elements in the queue, x and y, if x has a lower priority value than y, x will be deleted before y

Recall From Lists, Queues, Stacks

- Use an ADT that corresponds to your needs
- The right ADT is efficient, while an overly general ADT provides functionality you aren’t using, but are paying for anyways
- Heaps provide O(log n) worst case for both insert and deleteMin, O(1) average insert
Binary Heap Properties

1. Structure Property
2. Ordering Property

Tree Review

A Perfect binary tree – A binary tree with all leaf nodes at the same depth. All internal nodes have 2 children.

Heap Structure Property

- A binary heap is a complete binary tree.

Examples:

Tree Terminology

More Tree Terminology

Implicit (array) implementation:
Why this approach to storage?

Heap Order Property

Heap order property: For every non-root node X, the value in the parent of X is less than (or equal to) the value in X.

Heap Operations

• findMin:
• insert(val): percolate up.
• deleteMin: percolate down.

Heap – Insert(val)

Basic Idea:
1. Put val at “next” leaf position
2. Percolate up by repeatedly exchanging node until no longer needed

Insert Code (optimized)

```
void insert(Object o) {
    assert(!isFull());
    size++;  
    newPos = percolateUp(size,o);
    Heap[newPos] = o;
}
```

```
int percolateUp(int hole, Object val) {
    while (hole > 1 && val < Heap[hole/2]) {
        Heap[hole] = Heap[hole/2];
        hole /= 2;
    }
    return hole;
}
```

runtime:

(Code in book)
Heap – Deletemin

Basic Idea:
1. Remove root (that is always the min!)
2. Put “last” leaf node at root
3. Find smallest child of node
4. Swap node with its smallest child if needed.
5. Repeat steps 3 & 4 until no swaps needed.

DeleteMin: percolate down

Data Structures
Binary Heaps

DeleteMin Code (Optimized)

Building a Heap
Building a Heap

- Adding the items one at a time is $O(n \log n)$ in the worst case
- I promised $O(n)$ for today

Working on Heaps

- What are the two properties of a heap?
  - Structure Property
  - Order Property
- How do we work on heaps?
  - Fix the structure
  - Fix the order

BuildHeap: Floyd’s Method

Add elements arbitrarily to form a complete tree. Pretend it’s a heap and fix the heap-order property!

BuildHeap pseudocode

```java
private void buildHeap() {
    for (int i = currentSize/2; i > 0; i--)
        percolateDown(i);
}
```

BuildHeap: Floyd’s Method
Finally...

More Priority Queue Operations

• **decreaseKey**
  - given a pointer to an object in the queue, reduce its priority value

  Solution: change priority and __________________________

• **increaseKey**
  - given a pointer to an object in the queue, increase its priority value

  Why do we need a *pointer*? Why not simply data value?

  Solution: change priority and __________________________

More Priority Queue Operations

• **Remove(objPtr)**
  - given a pointer to an object in the queue, remove the object from the queue

  Solution: set priority to negative infinity, percolate up to root and deleteMin

• **FindMax**
Facts about Heaps

Observations:
• Finding a child/parent index is a multiply/divide by two
• Operations jump widely through the heap
• Each percolate step looks at only two new nodes
• Inserts are at least as common as deleteMins

Realities:
• Division/multiplication by powers of two are equally fast
• Looking at only two new pieces of data: bad for cache!
• With huge data sets, disk accesses dominate

A Solution: $d$-Heaps

• Each node has $d$ children
• Still representable by array
• Good choices for $d$:
  – (choose a power of two for efficiency)
  – fit one set of children in a cache line
  – fit one set of children on a memory page/disk block

Tries

Preprocessing Strings

• Preprocessing the pattern speeds up pattern matching queries
  – After preprocessing the pattern, KMP’s algorithm performs pattern matching in time proportional to the text size
• If the text is large, immutable and searched for often (e.g., works by Shakespeare), we may want to preprocess the text instead of the pattern
• A trie is a compact data structure for representing a set of strings, such as all the words in a text
  – A trie supports pattern matching queries in time proportional to the pattern size

Standard Tries

• The standard trie for a set of strings $S$ is an ordered tree such that:
  – Each node but the root is labeled with a character
  – The children of a node are alphabetically ordered
  – The paths from the external nodes to the root yield the strings of $S$
• Example: standard trie for the set of strings $S = \{\text{bear, bell, bid, bull, buy, sell, stock, stop} \}$
Analysis of Standard Tries
- A standard trie uses $O(n)$ space and supports searches, insertions and deletions in time $O(dm)$, where:
  - $n$ total size of the strings in $S$
  - $m$ size of the string parameter of the operation
  - $d$ size of the alphabet

Word Matching with a Trie
- insert the words of the text into trie
- Each leaf is associated with one particular word
- leaf stores indices where associated word begins ("see" starts at index 0 & "see" leaf stores those indices)

Compact Representation
- Compact representation of a compressed trie for an array of strings:
  - Stores at the nodes ranges of indices instead of substrings
  - Uses $O(n)$ space, where $n$ is the number of strings in the array
  - Serves as an auxiliary index structure

Suffix Trie
- The suffix trie of a string $X$ is the compressed trie of all the suffixes of $X$
- Uses $O(nw)$ space
- Supports arbitrary pattern matching queries in $X$ in $O(dw)$ time, where $w$ is the size of the pattern
- Can be constructed in $O(nw)$ time

Compressed Tries
- A compressed trie has internal nodes of degree at least two
- It is obtained from standard trie by compressing chains of "redundant" nodes
- ex. the "i" and "d" in "bid" are "redundant" because they signify the same word
1/9/2014

Encoding Trie (1)

- A code is a mapping of each character of an alphabet to a binary code-word
- A prefix code is a binary code such that no code-word is the prefix of another code-word
- An encoding trie represents a prefix code
  - Each leaf stores a character
  - The code word of a character is given by the path from the root to the leaf storing the character (0 for a left child and 1 for a right child)

Encoding Trie (2)

- Given a text string $X$, we want to find a prefix code for the characters of $X$ that yields a small encoding for $X$
  - Frequent characters should have short code-words
  - Rare characters should have long code-words
- Example
  - $X = \text{abracadabra}$
  - $T_1$ encodes $X$ into 29 bits
  - $T_2$ encodes $X$ into 24 bits

Huffman’s Algorithm

- Given a string $X$, Huffman’s algorithm construct a prefix code the minimizes the size of the encoding of $X$
- It runs in time $O(n + d \log d)$, where $n$ is the size of $X$ and $d$ is the number of distinct characters of $X$
- A heap-based priority queue is used as an auxiliary structure

Algorithm HuffmanEncoding($X$)

Input: string $X$ of size $n$
Output: optimal encoding trie for $X$

$C \leftarrow$ distinctCharacters($X$)
computeFrequencies($C$, $X$)
$Q \leftarrow$ new empty heap
for all $c \in C$
  $T \leftarrow$ new single-node tree storing $c$
  $Q.insert$($getFrequency(c), T$)
while $Q.size() > 1$
  $f_1 \leftarrow Q.min()$
  $T_1 \leftarrow Q.removeMin()$
  $f_2 \leftarrow Q.min()$
  $T_2 \leftarrow Q.removeMin()$
  $T \leftarrow$ join($T_1, T_2$)
  $Q.insert(f_1 + f_2, T)$
return $Q.removeMin()$

Example

$X = \text{abracadabra}$
Frequencies

- $a$: 5
- $b$: 2
- $c$: 1
- $d$: 2
- $r$: 2

The End

One More Operation

- Merge two heaps
  - $O(n \log n)$
- Add the items from one into another?
  - $O(n \log n)$
- Start over and build it from scratch?
  - $O(n)$
CSE 326: Data Structures

Priority Queues
Leftist Heaps & Skew Heaps

New Heap Operation: Merge
Given two heaps, merge them into one heap
– first attempt: insert each element of the smaller heap into the larger.
  \textit{runtime}:
– second attempt: concatenate binary heaps’ arrays and run buildHeap.
  \textit{runtime}:

Leftist Heaps
Idea:
Focus all heap maintenance work in one small part of the heap

Leftist heaps:
1. Most nodes are on the left
2. All the merging work is done on the right

Definition: Null Path Length
\text{null path length (npl)} of a node \( x \) = the number of nodes between \( x \) and a null in its subtree
\text{OR}
\text{npl}(x) = \text{min} \text{ distance to a descendant with 0 or 1 children}

\begin{itemize}
\item \text{npl}(\text{null}) = -1
\item \text{npl}(\text{leaf}) = 0
\item \text{npl}(\text{single-child node}) = 0
\end{itemize}

Equivalent definitions:
1. \text{npl}(x) is the height of largest complete subtree rooted at \( x \)
2. \text{npl}(x) = 1 + \min\{\text{npl}(\text{left}(x)), \text{npl}(\text{right}(x))\}

Leftist Heap Properties
\begin{itemize}
\item \text{Heap-order property}
   – parent’s priority value is \( \leq \) to children’s priority values
   – \textit{result}: minimum element is at the root
\item \text{Leftist property}
   – For every node \( x \), \( \text{npl(\text{left}(x))} \geq \text{npl(\text{right}(x))} \)
   – \textit{result}: tree is at least as “heavy” on the left as the right
\end{itemize}

Are These Leftist?
Every subtree of a leftist tree is leftist!
Right Path in a Leftist Tree is Short (#1)

**Claim:** The right path is as short as any in the tree.

**Proof:** (By contradiction)
- Pick a shorter path: \( D_1 < D_2 \)
- Say it diverges from right path at \( x \)

\[ \text{np}(L) \leq D_1 - 1 \] because of the path of length \( D_1 - 1 \) to null

\[ \text{np}(R) \geq D_2 - 1 \] because every node on right path is leftist

Leftist property at \( x \) violated!

Merge two heaps (basic idea)
- Put the smaller root as the new root,
- Hang its left subtree on the left.
- Recursively merge its right subtree and the other tree.

Right Path in a Leftist Tree is Short (#2)

**Claim:** If the right path has \( r \) nodes, then the tree has at least \( 2^{r-1} \) nodes.

**Proof:** (By induction)

**Base case:** \( r = 1 \). Tree has at least \( 2^{1-1} = 1 \) node

**Inductive step:** assume true for \( r' < r \). Prove for tree with right path at least \( r \).

1. Right subtree: right path of \( r-1 \) nodes
   \[ \Rightarrow 2^{r-1-1} \text{ right subtree nodes (by induction)} \]

2. Left subtree: also right path of length at least \( r-1 \) \((by \ previous \ slide)\) \[ \Rightarrow 2^{r-1-1} \text{ left subtree nodes (by induction)} \]

Total tree size: \( 2^{r-1-1} + (2^{r-1-1} + 1) = 2^{r-1} \)

Merging Two Leftist Heaps

- \( \text{merge}(T_1, T_2) \) returns one leftist heap containing all elements of the two (distinct) leftist heaps \( T_1 \) and \( T_2 \)

Why do we have the leftist property?

Because it guarantees that:
- the **right path is really short** compared to the number of nodes in the tree
- A leftist tree of \( N \) nodes, has a right path of at most \( \log(N+1) \) nodes

**Idea** – perform all work on the right path

Merge Continued

- If \( \text{np}(R') > \text{np}(L_i) \)
- \( R' = \text{Merge}(R_i, T_3) \)

runtime:
Let's do an example, but first…

Other Heap Operations

• insert ?
• deleteMin ?

Operations on Leftist Heaps

• merge with two trees of total size $n$: $O(\log n)$
• insert with heap size $n$: $O(\log n)$
  – pretend node is a size 1 leftist heap
  – insert by merging original heap with one node heap

• deleteMin with heap size $n$: $O(\log n)$
  – remove and return root
  – merge left and right subtrees

Leftest Merge Example

Leftist Heaps: Summary

Good

Bad
Random Definition: Amortized Time

**Amortized time:**
- Running time limit resulting from “writing off” expensive runs of an algorithm over multiple cheap runs of the algorithm, usually resulting in a lower overall running time than indicated by the worst possible case.
- Amortized time per operation is $O(\log N)$

Difference from average time:

Skew Heaps

Problems with leftist heaps
- Extra storage for npl
- Extra complexity/logic to maintain and check npl
- Right side is “often” heavy and requires a switch

Solution: skew heaps
- “blindly” adjusting version of leftist heaps
- Merge always switches children when fixing right path
- Amortized time for: merge, insert, deleteMin = $O(\log n)$
- However, worst case time for all three = $O(n)$

Merging Two Skew Heaps

**Example**

**Skew Heap Code**

```c
void merge(heap1, heap2) {
    case {
        heap1 == NULL: return heap2;
        heap2 == NULL: return heap1;
        heap1.findMin() < heap2.findMin():
            temp = heap1.right;
            heap1.right = heap1.left;
            heap1.left = merge(heap2, temp);
            return heap1;
        otherwise:
            return merge(heap2, heap1);
    }
    }
}
```

Runtime Analysis: Worst-case and Amortized

- No worst case guarantee on right path length!
- All operations rely on merge

⇒ worst case complexity of all ops =
- Probably won’t get to amortized analysis in this course, but see Chapter 11 if curious.
- Result: $M$ merges take time $M \log n$

⇒ amortized complexity of all ops =
Comparing Heaps
- Binary Heaps
- d-Heaps
- Leftist Heaps
- Skew Heaps

Still room for improvement! (Where?)

The Binomial Tree, $B_h$
- $B_h$ has height $h$ and exactly $2^h$ nodes
- $B_h$ is formed by making $B_{h-1}$ a child of another $B_{h-1}$
- Root has exactly $h$ children
- Number of nodes at depth $d$ is binomial coeff.
  - Hence the name; we will not use this last property

Data Structures
Binomial Queues

Yet Another Data Structure: Binomial Queues
- Structural property
  - Forest of binomial trees with at most one tree of any height
- Order property
  - Each binomial tree has the heap-order property

Binomial Queue with $n$ elements
Binomial Q with $n$ elements has a unique structural representation in terms of binomial trees!
Write $n$ in binary: $n = 1101_{(base 2)} = 13_{(base 10)}$

Properties of Binomial Queue
- At most one binomial tree of any height
- $n$ nodes $\Rightarrow$ binary representation is of size $\Theta$?
  $\Rightarrow$ deepest tree has height $\Theta$?
  $\Rightarrow$ number of trees is $\Theta$?

Define: height(forest $F$) = $\max_{\text{tree } T \in F} \{\text{height}(T)\}$

Binomial Q with $n$ nodes has height $\Theta(\log n)$
Operations on Binomial Queue

- Will again define `merge` as the base operation
  - insert, deleteMin, buildBinomialQ will use `merge`
- Can we do `increaseKey` efficiently?
- `decreaseKey`?
- What about `findMin`?

Merging Two Binomial Queues

Essentially like adding two binary numbers!

1. Combine the two forests
2. For $k$ from 0 to maxheight
   a. $m \leftarrow$ total number of $B_k$'s in the two BQs
   b. if $m=0$: continue;
   c. if $m=1$: continue;
   d. if $m=2$: combine the two $B_k$'s to form $B_{k+1}$
   e. if $m=3$: retain one $B_k$ and combine the other two to form a $B_{k+1}$

Claim: When this process ends, the forest has at most one tree of any height
Example: Binomial Queue
Merge

H1: H2:

Example: Binomial Queue
Merge

H1: H2:

Example: Binomial Queue
Merge

H1: H2:

Example: Binomial Queue
Merge

H1: H2:

Insert in a Binomial Queue

Insert(x): Similar to leftist or skew heap

runtime
Worst case complexity: same as merge
\( O(\ ) \)

Average case complexity: \( O(1) \)
Why??  Hint: Think of adding 1 to 1101

Complexity of Merge

Constant time for each height
Max number of heights is: \( \log n \)

\( \Rightarrow \) worst case running time = \( \Theta(\ ) \)

deleteMin in Binomial Queue

Similar to leftist and skew heaps….

deleteMin: Example

BQ

\( \Rightarrow \) find and delete smallest root

merge BQ (without the shaded part) and BQ’
deleteMin: Example

Result:

runtime: