Data Structure Evolution

- Standard operations on data structures
  - Search
  - Insert
  - Delete
- Linear Lists
  - Implementation: Arrays (Unsorted and Sorted)
- Dynamic Linear Lists
  - Implementation: Linked Lists
- Dynamic Trees
  - Implementation: Binary Search Trees
## Common Data Structures

<table>
<thead>
<tr>
<th></th>
<th>Search</th>
<th>Insert</th>
<th>Delete</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Unsorted Arrays</strong></td>
<td>O(N)</td>
<td>O(1)</td>
<td>O(N)</td>
<td></td>
</tr>
<tr>
<td><strong>Sorted Arrays</strong></td>
<td>O(log N)</td>
<td>O(N)</td>
<td>O(N)</td>
<td></td>
</tr>
<tr>
<td><strong>Unsorted Linked Lists</strong></td>
<td>O(N)</td>
<td>O(1)</td>
<td>O(N)</td>
<td></td>
</tr>
<tr>
<td><strong>Sorted Linked Lists</strong></td>
<td>O(N)</td>
<td>O(N)</td>
<td>O(N)</td>
<td></td>
</tr>
<tr>
<td><strong>Binary Search Trees</strong></td>
<td>O(H)</td>
<td>O(H)</td>
<td>O(H)</td>
<td>H = O(N)</td>
</tr>
<tr>
<td><strong>Balanced BSTs</strong></td>
<td>O(log N)</td>
<td>O(log N)</td>
<td>O(log N)</td>
<td>As H = O(log N)</td>
</tr>
</tbody>
</table>
Animations

- https://visualgo.net/
- http://www.cs.jhu.edu/~goodrich/dsa/trees/
- https://www.youtube.com/watch?v=Y-5ZodPvhmM
- http://www.algoanim.ide.sk/
Red-Black (RB) Trees

- Every node in a red-black tree is colored either red or black.
  - The root is always black.
  - Every path on the tree, from the root down to the leaf, has the same number of black nodes.
  - No red node has a red child.
  - Every NIL pointer points to a special node called NIL[T] and is colored black.
- Every RB-Tree with $n$ nodes has black height at most $\log n$
- Every RB-Tree with $n$ nodes has height at most $2\log n$
Red-Black Tree Insert

RB-Insert (T,z)  // pg 315
   // Insert node z in tree T
   y = NIL[T]
x = root[T]
while (x ≠ NIL[T]) do
   y = x
   if (key[z] < key[x])
      x = left[x]
x = right[x]
p[z] = y
if (y == NIL[T])
   root[T] = z
else if (key[z] < key[y])
   left[y] = z
else right[y] = z
   // new stuff
   left[z] = NIL[T]
   right[z] = NIL[T]
color[z] = RED
   RB-Insert-Fixup (T,z)

RB-Insert-Fixup (T,z)
   while (color[p[z]] == RED) do
      if (p[z] = left[p[p[z]]]) then
         y = right[p[p[z]]]
      if (color[y] == RED) then  // C-1
         color[p[z]] = BLACK
         color[y] = BLACK
         z = p[p[z]]
color[z] = RED
      else if (z = right[p[z]]) then  // C-2
         z = p[z]
         LeftRotate(T,z)
color[p[z]] = BLACK  // C-3
color[p[p[z]]] = RED
         RightRotate(T,p[p[z]])
else
   // Symmetric code: “right” ↔ “left”
   . . .
color[root[T]] = BLACK
Case 1: Non-elbow; sibling of parent (y) red
Case 2: Elbow case
Case 3: Non-elbow; sibling of parent black
Rotations

\texttt{LeftRotate(T,x)} \quad // \texttt{pg 278}

\begin{verbatim}
    // right child of x becomes x's parent.
    // Subtrees need to be readjusted.
    y = right[x]
    right[x] = left[y] \quad // y's left subtree becomes x's right
    p[left[y]] = x
    p[y] = p[x]
    if (p[x] == NIL[T]) then
        root[T] = y
    else if (x == left[p[x]]) then
        left[p[x]] = y
    else right[p[x]] = y
    left[y] = x
    p[x] = y
\end{verbatim}
## More Dynamic Operations

<table>
<thead>
<tr>
<th></th>
<th>Search</th>
<th>Insert</th>
<th>Delete</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unsorted Arrays</td>
<td>$O(N)$</td>
<td>$O(1)$</td>
<td>$O(N)$</td>
<td></td>
</tr>
<tr>
<td>Sorted Arrays</td>
<td>$O(\log N)$</td>
<td>$O(N)$</td>
<td>$O(N)$</td>
<td></td>
</tr>
<tr>
<td>Unsorted Linked Lists</td>
<td>$O(N)$</td>
<td>$O(1)$</td>
<td>$O(N)$</td>
<td></td>
</tr>
<tr>
<td>Sorted Linked Lists</td>
<td>$O(N)$</td>
<td>$O(N)$</td>
<td>$O(N)$</td>
<td></td>
</tr>
<tr>
<td>Binary Search Trees</td>
<td>$O(H)$</td>
<td>$O(H)$</td>
<td>$O(H)$</td>
<td>$H = O(N)$</td>
</tr>
<tr>
<td>Balanced BSTs</td>
<td>$O(\log N)$</td>
<td>$O(\log N)$</td>
<td>$O(\log N)$</td>
<td>As $H = O(\log N)$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Se/In/De</th>
<th>Rank</th>
<th>Select</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Balanced BSTs</td>
<td>$O(\log N)$</td>
<td>$O(N)$</td>
<td>$O(N)$</td>
<td></td>
</tr>
<tr>
<td>Augmented BBSTs</td>
<td>$O(\log N)$</td>
<td>$O(\log N)$</td>
<td>$O(\log N)$</td>
<td></td>
</tr>
</tbody>
</table>
Operations on Dynamic RB Trees

- **K-Selection**
  - Select an item with a specified rank
  - “Efficient” solution not possible without preprocessing
- Preprocessing - store additional information at nodes
- Inverse of K-Selection
  - Find rank of an item in the tree
- What information should be stored?
  - Rank
  - ??
OS-Rank

OS-RANK(x,y)

// Different from text (recursive version)
// Find the rank of x in the subtree rooted at y

1. \( r = \text{size}[\text{left}[y]] + 1 \)
2. if \( x = y \) then return \( r \)
3. else if ( key[\( x \)] < key[y] ) then
4. return OS-RANK(x, left[y])
5. else return \( r + \text{OS-RANK}(x, \text{right}[y]) \)

Time Complexity \( O(\log n) \)
OS-Select

OS-SELECT(x,i) //page 304
// Select the node with rank i
// in the subtree rooted at x
1. \( r = \text{size}[\text{left}[x]] + 1 \)
2. if \( i = r \) then
3. \( \text{return x} \)
4. elseif \( i < r \) then
5. \( \text{return OS-SELECT (left}[x], i) \)
6. else \( \text{return OS-SELECT (right}[x], i-r) \)

Time Complexity \( O(\log n) \)
RB-Tree Augmentation

- Augment x with \textbf{Size}(x), where
  - \textbf{Size}(x) = size of subtree rooted at x
  - \textbf{Size}(NIL) = 0
Augmented Data Structures

- Why is it needed?
  - Because basic data structures not enough for all operations
  - Storing extra information helps execute special operations more efficiently.

- Can any data structure be augmented?
  - Yes. Any data structure can be augmented.

- Can a data structure be augmented with any additional information?
  - Theoretically, yes.

- How to choose which additional information to store.
  - Only if we can maintain the additional information efficiently under all operations. That means, with additional information, we need to perform old and new operations efficiently maintain the additional information efficiently.
How to augment data structures

1. choose an underlying data structure
2. determine additional information to be maintained in the underlying data structure,
3. develop new operations,
4. verify that the additional information can be maintained for the modifying operations on the underlying data structure.
Augmenting RB-Trees

Theorem 14.1, page 309

Let $f$ be a field that augments a red-black tree $T$ with $n$ nodes, and $f(x)$ can be computed using only the information in nodes $x$, $\text{left}[x]$, and $\text{right}[x]$, including $f[\text{left}[x]]$ and $f[\text{right}[x]]$. Then, we can maintain $f(x)$ during insertion and deletion without asymptotically affecting the $O(\log n)$ performance of these operations.

For example,

$$\text{size}[x] = \text{size}[\text{left}[x]] + \text{size}[\text{right}[x]] + 1$$

$$\text{rank}[x] = ?$$
Augmenting information for RB-Trees

- Parent
- Height
- Any associative function on all previous values or all succeeding values.
- Next
- Previous
Reading for next class

- Red Black Trees
  - Properties
  - Invariants
  - Insert and Delete
- Mathematical Induction