Augmented Search Trees
(CLRS 14)

1 Red-Black Trees

• Last time we discussed red-black trees:
  – Balanced binary trees—all elements in left (right) subtree of node \( x \) are \(< x \) (\( > x \)).
  – Every node is colored RED or BLACK and we maintained red-blue invariant:
    - Root is BLACK.
    - A RED node can only have BLACK children.
    - Every path from the root to a leaf contains the same number of BLACK nodes.

• We saw how the red-blue invariant guaranteed \( O(\log n) \) height.
• We could reestablish the red-blue invariant after an insertion or deletion in \( O(\log n) \) time
  – \( O(\log n) \) node recolorings (no structural changes).
  – \( O(1) \) rotations:

• Red-black tree also supports SEARCH, SUCCESSOR, and PREDECESSOR in \( O(\log n) \) as in binary search trees.
• We will now discuss how to develop data structures supporting other operations by augmenting red-black tree.
2 Augmented Data Structures

- We want to add an operation Select\((i)\) to a red-black tree
  - We have previously seen how to select the \(i\)'th element among \(n\) elements in \(O(n)\) time.
  - Can we support it faster if we have the elements stored in a data structure?
  - We can of course support the operation in \(O(1)\) time if we have the elements sorted in an array but what is we also want to be able to insert and delete elements?

- We augment every node \(x\) in red-black tree with a field \(size(x)\) equal to the number of nodes in the subtree rooted in \(x\)
  - \(size(x) = size(left(x)) + size(right(x)) + 1\)

Example:

```
    20
   /  \
  17   41
 /     /  \
14 12 21 5
 /     /  /  \
10 7 16 221 30 5
 /  /  /  /  /  /  \
7 2 12 1 14 1 21 1 28 1
     /  \
    3 1
```

- We can use this field to implement Select\((i)\):

\[
Select(x, i) \\
= \begin{cases} 
  r = size(left(x)) + 1 \\
  \text{IF } i = r \text{ THEN Return } x \\
  \text{IF } i < r \text{ THEN Return } Select(left(x), i) \\
  \text{IF } i > r \text{ THEN Return } Select(right(x), i-r)
\end{cases}
\]

Example (Select(17)):
Since we only follow one root-leaf path the operation takes $O(\log n)$ time.

- Actually, we can also use the field to perform the “opposite” operation in $O(\log n)$ time—determining the rank of the element in node $x$:

$$\text{Rank}(x) = \begin{cases} r = \text{size}(\text{left}(x)) + 1 \\ y = x \\ \text{WHILE } y \neq \text{root of tree } \text{DO} \\ \quad \text{IF } y = \text{right}(\text{parent}(y)) \text{ THEN} \\ \quad \quad r = r + \text{size}(\text{left}(\text{parent}(y))) + 1 \\ \quad \quad y = \text{parent}(y) \\ \quad \text{FI} \\ \text{OD} \\ \text{Return } r \end{cases}$$

Example (Rank of element 38):

- We need to maintain the extra field during updates:
  - $\text{INSERT}(i)$:
Search down one root-leaf part as usual for position where \(i\) should be inserted.

* Increment \(\text{size}(x)\) for all nodes \(x\) on root-leaf path (these are the only nodes for which the size field change).

Example (Insertion of element 32)

* Rebalancing using Red-black tree rules—recall that we do \(O(\log n)\) recolorings and \(O(1)\) rotations:
  - Color change rules do not affect extra field
  - Rotations do affect size extra fields but we can still easily perform a rotation in \(O(1)\) time

\[
\begin{align*}
\text{size}(y') &= \text{size}(root(B)) + \text{size}(root(C)) + 1 \\
\text{size}(x') &= \text{size}(y) = \text{size}(root(A)) + \text{size}(root(B)) + \text{size}(root(C)) + 2
\end{align*}
\]

\(\Downarrow\)

INSERT performed in \(O(\log n)\) time.

\(-\text{DELETE}(i):\)
  * Find element to delete and decrement size field on one root-leaf path (recall that conceptually we always delete a node with at most one child).
  * Rebalance using rotations.

\(\Downarrow\)

DELETE performed in \(O(\log n)\) time.

- Note: The key to maintaining the size field during updates is that the field of node \(x\) only depend on the field of the children of \(x\) ⇒
  - Insertion or deletion only affect one root-leaf path.
Rotations can be handled in $O(1)$ time locally.

- In general we can easily prove the following:

<table>
<thead>
<tr>
<th>A field $f$ in a red-black tree can be maintained in $O(\log n)$ time during updates if $f(x)$ can be computed using only information in $x$, $\text{left}(x)$ and $\text{right}(x)$ (including $f(\text{left}(x))$ and $f(\text{right}(x))$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>- When changing field in a node $x$, $f$ can only change for the $O(\log n)$ ancestors of $x$ on the path to the root.</td>
</tr>
<tr>
<td>- Rotations can be handled in $O(1)$ time locally.</td>
</tr>
</tbody>
</table>

3 Interval Tree

- We now consider a slightly more complicated augmentation. We want to solve the following problem:
  - Maintain a set of $n$ intervals $[i_1, i_2]$ such that one of the intervals containing a query point $q$ (if any) can be found efficiently.


To solve the problem we use the so-called “Interval tree”:

- Red-black tree with intervals in nodes
  - Key is left endpoint
- Node $x$ augmented with maximal right endpoint in subtree rooted in $x$

Example: Interval tree on intervals from previous figure:
• We can maintain the interval tree dynamically during insertions and deletions in $O(\log n)$ time
  
  – because augmented field in $x$ only depends on augmented fields in the children of $x$ and the interval stored in $x$.

  – $\text{max}(x) = \text{max}(\text{rightendpoint}(x), \text{max}(\text{left}(x)), \text{max}(\text{right}(x)))$

• We can also answer a query in $O(\log n)$ time:

  1. We first check if $q$ is contained in interval stored in root $r$—if it is we are done.

  2. Next we check if $q$ is on left side of left endpoint of interval in $r$—if it is we recursively search in left subtree ($q$ cannot be contained in any interval in right subtree).

  3. If $q$ is to the right of left endpoint of interval in $r$ we have two cases:

     (a) If $\text{max}(\text{left}(r)) > q$ there must be a segment in left subtree containing $q$ and we recurse left.

     (b) If $\text{max}(\text{left}(r)) < q$ there is no segment in left subtree containing $q$ and we recurse right.

\[
\begin{array}{c}
\text{Case 1} & q & \text{Case 3a} \hline
\hline
\text{Case 2} & q & \text{Case 3b} \\
\end{array}
\]

\[
\text{QUERY}(x, q)
\]

\[
\begin{array}{l}
\text{IF } q \text{ contained in } x \text{ interval THEN Return } x \\
\text{IF } \text{max}(\text{left}(x)) \geq q \text{ THEN} \\
\quad \text{Return Query(left}(x), q) \\
\text{ELSE} \\
\quad \text{Return Query(right}(x), q) \\
\end{array}
\]

\[
\downarrow
\]

We search down one root-leaf path $\Rightarrow O(\log n)$ time.

Example: Query with $q = 23$: