**Aleksandr Itskovich (ai1221)**

**Collaborator:** Elsie Kenyon

**Assignment 2**

**Question 1: Binary Search Trees**

a)

* Insert: 30, 15, 40, 10, 17, 50, 8



* Explain why this tree is an AVL tree.

An AVL tree is characterized by the following invariant: that the difference in heights between the subtrees rooted at any node is at most 1.

We see that this invariant is met in the tree drawn above. Take, for example, the root node at 30. The height of its left subtree, LH, (rooted at 15) is 2, since the longest path from 15 to a leaf has length 2. The height of the right subtree, meanwhile is 1. Repeating the same analysis for each node yields:

|  |  |  |  |
| --- | --- | --- | --- |
| **Node** | **Left Subtree Height** | **Right Subtree Height** | **Difference** |
| 30 | 2 | 1 | 1 |
| 15 | 1 | 0 | 1 |
| 40 | -1 | 0 | 1 |
| 10 | 0 | -1 | 1 |
| 17 | -1 | -1 | 0 |
| 50 | -1 | -1 | 0 |
| 8 | -1 | -1 | 0 |

We can see that the AVL height invariant is maintained at every node in the binary tree, therefore the tree is an AVL tree.

* Next, insert the key 9. Explain why the result of the insert no longer satisfies the AVL property.

Diagram

Description automatically generated

We appended an additional node to the left subtree of 30. Now we see that for node 30, the height of its left subtree rooted at 15 is 3, while the height of the right subtree remains unchanged at 1. Because the difference in height between the two subtrees is greater than 1. Therefore the tree is no longer an AVL tree.

* Show how you could perform at most 2 rotations in order to restore the AVL tree property.

We could rotate twice to restore the tree heights back to having a difference of at most 1.

Diagram, schematic

Description automatically generatedCalendar

Description automatically generated

In general, when an AVL tree is out of balance there are 4 types of rotations that are similar to the ones discussed in the RB Tree case: two bent rotations (handling left-right and right-left cases) and two straight rotations (handling left-left and right-right cases). After performing either straight rotation, the tree is balanced. Moreover, we can always convert a bent case into a straight case by performing one rotation. That means we are never more than 2 rotations away from having a balanced AVL tree.

We either have

1. Bent Case rotation, followed by a straight case rotation (2 rotations)
2. Or just a straight case (1 rotation)

These cases are summarized below:

A picture containing text, flying, map, group

Description automatically generated

b) Suppose T1 and T2 and T3 each reference the root nodes of a binary search tree on n

nodes. We wish to combine these three trees into one binary search tree. Explain how this can be done in time, in such a way that the resulting BST has height . You do not need to write pseudo-code for your procedure, but you must properly describe your steps.

We are given that we have 3 nodes that point at the roots of different binary search trees. Let’s assume that the sizes of the trees are n1, n2 and n3, respectively.

We can carry out the following procedure to build a minimal binary search tree in linear time.

1. Perform an in-order traversal of each tree, storing the values in an array as they are encountered. This produces 3 separate, *sorted* arrays, with sizes n1, n2 and n3, respectively.
   1. We know from class that in-order traversal is linear for each tree.
   2. That means the total work done for all 3 trees is
2. Merge the 3 sorted arrays produced in step one into a single array. The logic behind this step is identical to the way arrays are merged in merge sort, only now we have a 3-way merge instead of a 2 way merge. This produces a single, *sorted* array.
   1. Overall, we have to combine elements, so the total time is again .
3. Now that we have a single, sorted array, we know from class that we can produce a minimal BST from a sorted array by recursively selecting the median element from the left and right subarrays, and inserting that median element into the tree.
   1. Getting the median element recursively is given by , which we know to have a solution of by the Master Method.
   2. Inserting into the tree takes constant time per element because we don’t have to traverse the tree searching for the position of the newly inserted element; we know exactly where to insert elements since we’re selecting medians from subsequent sub-arrays, so in this case insertion is constant is constant. ]
4. Overall, we have 3 steps, each of which is shown to take time. That means the overall procedure also runs in time.

c)

LazyDelete(T, z):

while T != NIL

if T.deleted = false and T.key = z.key:

T.deleted = true

else

if T.key < z

T = T.right

else

T = T.left

* TreeSearch

We would have to update the TreeSearch method only slightly: if we find a node that has a key equal to the target search key, we also need to check if the node was deleted before returning the match; in other words, we just want to make sure we are only matching on non-deleted nodes. Otherwise the algorithm is the same. This change is shown below:

TreeSearch(T, z):

while T != NIL

**if T.deleted = false** and z = T.key

return T

else

if T.key < z

T = T.right

else

T = T.left

* TreeInsert

The TreeInsert algorithm doesn’t change. When inserting a child, we don’t care whether or not the parent is deleted or not, we just want the tree to maintain the necessary invariant for non-deleted nodes, which the algorithm in its current form already does. This also takes for granted that nodes being inserted are marked as non-deleted by default.

* Disadvantages:

Implementing lazy deletion makes the implementation of other methods more complicated. More importantly, it increases the overall space-complexity of the tree. Since deleted nodes are never removed from the tree, they continue to take up space. For example, if we delete half the nodes in a tree containing n nodes, we are effectively wasting half of the nodes. It also means that space only increases, never shrinks, since nodes are never truly removed. All of this also means that the time required to carry out an operation such as search or insert also increases by some constant factor as the number of deleted nodes increases.

**Question 2: Red-Black Trees**

a)

* For height b, what is the maximum number of nodes in the tree (excluding NIL nodes)? Describe the shape of the tree and its coloring.

To achieve the maximum number of nodes in a RB tree, we will want to have a full tree (that is, all possible nodes occupied) with the maximum height. We know from class that to achieve maximum height in a RB tree, we will want nodes levels to be alternating between red and black. That way we have half black nodes and half red nodes:

Shape

Description automatically generated with medium confidence

For a given black height, b, we know that for every black level, we are going to have an additional red level (because we want the tree to be as full as possible for a given black height, with alternation). That means that the number of levels is going to be given by 2b.

The number of nodes in each level is given by the level number raised to the 2. That means, the number of nodes in the first level (level is) is 20 = 1, the number of nodes in the second level is given by 21 = 2, etc.

We need to total sum of all internal nodes. Adding up all the nodes in all the levels gives us:

However, because we don’t want to count the last level of nil nodes (and only the internal nodes) we want to be just .

Therefore the total number of internal nodes in the maximum case is given by.

* Repeat for minimum number of nodes:

To get the minimum number of nodes, we want to get the smallest possible BT tree. This corresponds to the case when all leaves are black.

The total number of levels of this tree will be just b. In this base, we have

As above, we don’t want to count the nil nodes so becomes just .

Therefore the total number of internal nodes in the minimum case is given by.

b)

We can only increase the black height when we insert a node that causes a recoloring. Rotations never increase the black height.

c) It is impossible to transform this tree into a proper red-black tree.

Consider the right subtree first. The right height is 2. We can either have all nodes be black, or alternate red and black nodes. Therefore the right subtree has . More specifically, it can either be 2 or 3:

Chart, line chart

Description automatically generated

Now, we can look at the right subtree.

In the left subtree we have multiple paths down to NIL nodes. These paths have sizes 3, 4, 5 and 6 (including the NIL nodes). The shortest path has height 3, and the longest has height 6.

Let’s examine the left subtree with respect to the right subtree.

* Case 1: the right subtree has BH 2:

If the right subtree has a BH of 2, then it is impossible for the left subtree to maintain a BH of 2 along its longest path of length 6. Because we cannot have adjacent red nodes, they must be alternated with black nodes. The maximum number of red nodes we could have along that height would be 3, which means we would need to have *at least* 3 black nodes (including NIL). That means it is impossible for the right subtree to have a BH of 2 and still create a valid RB tree with the left subtree.

* Case 2: the right subtree has BH 3:

If the right subtree has a BH of 3, the only way to maintain a BH of 3 along its shortest path of length 3 is to have all nodes along that path to be black. This is depicted in the diagram below:

Diagram

Description automatically generated

There is no other way to achieve a path of 3 along nodes **a** and **b**, and the NIL node on **b**.

However, nodes **a** and **b** also form the backbone of paths **root-a-b-c** (length 4) and **root-a-b-c-d** (length 5).

Because node **c** has a NIL node, it cannot be colored black. If we made it black, we would have a BH of 4 along nodes **a-b-c-NIL**. This violates the RB tree invariant since we know the right subtree has a BH of 3.

Therefore, we have to color node **c** red. However, when node **c** is red, node **d** must then be black because we cannot have two adjacent red nodes. Now, however, the path along **a-b-c-d-NIL** has a BH of 4, violating the RB tree invariant.

There is therefore no way to make this tree a valid red-black tree, according to the above analysis.

**Question 3: Augmented BSTs**

**a)** Given an unsorted set of *n* distinct numbers, suppose we are interested in carrying out the following operations:

* Selecting the element of rank *k*. This operation will be carries out times.
* Inserting a new element into the set. This operation will be carried out times.

The order of the above operations is not known.

In class we saw 3 different ways of solving this problem: one that uses an augmented binary search tree, one that uses the original select algorithm, and a third method which simply sorts the input. For each approach, you must describe how to solve the above two operations and provide the overall runtime.

* **Augmented Binary Search Tree**

Elements are stored in a BST, which may be self-balancing, such as a Red-Black tree.

*Select k:*

We know from class that selecting an element of rank *k* requires time . When using a self-balancing tree, such as an AVL or Red-Black tree, the height is guaranteed to be in the worst-case.

For *n* select operations, the overall runtime for selection would be .

For *m* select operations, the overall runtime for selection would be where *m* is a constant, which simplifies to just .

For select operations the overall runtime would be .

*Insert element:*

Similar to select, insertion in a binary search tree also requires time , which is for a self-balancing tree.

Inserting *n* elements would have runtime .

Inserting *p* elements would have runtime where *p* is a constant, which simplifies to just .

Inserting elements would have a runtime of .

* **Select algorithm**

Elements are stored in an array.

*Select k*:

The select algorithm takes an array of numbers and a target rank, *k*, and is able to provide the number of the element corresponding to rank *k* in a worst-case runtime of .

For *n* select operations, the runtime would be .

For *m* select operations, the runtime would be where *m* is a constant, which simplifies to .

For select operations the runtime would be .

*Insert element*:

Insertion into an array is amortized constant, with worst-case linear insertion time. The expected time per insertion would be .

For *n* elements, the overall time for all insertion operations would be expected to be .

For *m* elements, the overall time for all insertion operations would be where *m* is a constant, which simplifies to .

For elements, the overall time for all insertion operations would be expected to be

* **Sorting the input**

Elements are stored in an array.

*Select k:*

We can use different sorting algorithms to sort the input, then return the element stored in the slot corresponding to rank *k*. Using *merge-sort* or *quicksort* this would require time (If we could use a linear sort such a counting or radix sorts, this would only require time . However, since the prompt only says “numbers”, not specifying integers, or their range, or a constant amount of decimal precision, or specifying that the values are uniformly distributed, I am assuming that the available linear sorts are out of scope.)

Because the array is already sorted, we can get the element correspond to any rank in constant time. In other words, if we want to carry out multipleselect operations one after the other, it is necessary to only sort the array once.

For *n* operations, the runtime would be the time taken to sort the values once, followed by the time necessary to get the rank of each value. Sorting using quicksort takes time , while retrieving each values takes a constant amount of time. For *n* operations this would be which simplifies to .

For *m* operations, the runtime would be the time taken to sort the values, followed by the time necessary to get *m* ranks. Following the logic above, this would be where *m* is a constant, which simplifies to just .

For operations, the runtime would be the time taken to sort the values, followed by the time necessary to get ranks. This would be which simplifies to .

*Insert element*:

Insertion into an array takes amortized constant time, with worst-case linear time. However, the expected time per insertion would be .

For *n* elements, the overall time for all insertion operations would be expected to be .

For *m* elements, the overall time for all insertion operations would be where *m* is a constant, which simplifies to .

For elements, the overall time for all insertion operations would be .

Using the results, decide which method you should use when:

* *p = m =*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | **BST-Select** | **Select** | **Sort** |
| *insert* | *p = O(log n)* |  |  |  |
| *select* | *m = O(log n)* |  |  |  |
|  | *Final runtime* |  |  |  |

**Winner: BST-Select**

* *p = 4, m = n*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | **BST-Select** | **Select** | **Sort** |
| *insert* | *p = 4* |  |  |  |
| *select* | *m = n* |  |  |  |
|  | *Final runtime* |  |  |  |

**Winner: BST-Select** (could be considered a tie with Sort since both simplify to )

* *m = 2, p = n*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | **BST-Select** | **Select** | **Sort** |
| *select* | *m = 2* |  |  |  |
| *insert* | *p = n* |  |  |  |
|  | *Final runtime* |  |  |  |

**Winner: Select**

* *m = p = n*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | **BST-Select** | **Select** | **Sort** |
| *select* | *m = n* |  |  |  |
| *insert* | *p = n* |  |  |  |
|  | *Final runtime* |  |  |  |

**Winner: BST-Select**

**b)** Suppose we augment a binary search tree so that each node has an additional attribute

called *x.leaves* which represents the number of leaves in the subtree rooted at x. Let T be the root of such a BST. Your job is to update the BST-Delete(T, z) algorithm from class so that it correctly deletes node z from the tree T and updates the values of *x.leaves* for all necessary nodes in the tree. You must provide the pseudo-code for the updated version and explain why the runtime is still *O(h)*.

**BST-Delete(T, z)**

if z = T and z.left = NIL

return z.right

else if z = T and z.right = NIL

return z.left

else

if z.left = NIL and z.right = NIL

if z = z.parent.left

z.parent.left = NIL

else

z.parent.right = NIL

z.parent.leaves = z.parent.leaves - 1;

if z.parent.right != NIL and z.parent.left != NIL

y = z.parent;

while y != NIL

y = y.parent

y.leaves = y.leaves – 1

else if z.left = NIL

Replace(z, z.right)

else if z.right = NIL

Replace(z, z.left)

else

y = Find-Min(z.right)

z.key = y.key

if y.right = NIL # y is a leaf

y.parent.leaves = y.parent.leaves - 1

Replace(y, y.right)

return T

**c)** (8 points) In our practice problems we defined AVL trees. Each node of the AVL tree is augmented with an attribute called the *x.balance-factor* defined as follows: if both subtrees of x have the same height, the balance-factor is 0. Otherwise the balance-factor is 1 (left side is higher) or -1 (right side is higher). Given a tree T that is augmented with this information, update the Tree-Insert(T; z) algorithm from class so that the balance-factors are correctly updated after the insertion. If the resulting tree is not a proper AVL tree after the insertion, you do not need to repair it.

*\* The height of an empty tree is defined as -1*

Tree-Insert(T, z)

if T = NIL # no root, z becomes the root

z.balance-factor = 0

return z

else x = T

while x != NIL

y = x

if (z.key < x.key)

x = x.left

else

x = x.right

z.parent = y

if (z.key < y.key)

y.left = z;

else y.right = z;

**# fix balance factors**

**z.balance-factor = 0**

**p = z.parent**

**while (p != NIL)**

**lh = getHeight(p.left)**

**rh = getHeight(p.right)**

**if (lh = rh)**

**p.balance-factor = 0**

**else if (lh > rh)**

**p.balance-factor = 1**

**else**

**p.balance-factor = -1**

**p = z.parent**

return T

**d)** A project manager would like to store a set of n project intervals. Each interval consists

of a start and end time (over a year-long period). The manager would like a data structure that organizes the project intervals in such a way that she can carry out the following operations:

**Data Structure**: use a red-black interval tree augmented with subtree sizes.

* Because it is a RB tree, we are guaranteed a tree height of .
* Because it is an interval tree, we can store project intervals, i.e. the start and end-times, as well as the max end-time for a subtree.
* The interval tree’s BST invariant is maintained by the *x.int.low* property, as we’ve seen in class.
* Construction time is . We know from class that the time it takes to insert into a RB tree is . Because we have *n* elements to insert, the total build time is multiplied by *n* inserts, so overall.
* Because we have subtree sizes, we can quickly perform operations like range queries.

1. Given a **new project interval**, *i*, determine if project interval i overlaps the project that starts *last* among all the project. Time .

* The project that starts last has the greatest *x.int.low* value from all values in the tree.
* Because the tree uses *x.int.low* as the key to satisfy the BST property, the project the starts last is the *max* (rightmost) element in the tree.
* To find the maximum element in the tree, we have to traverse the entire height of the tree, from root to leaf, in the worst-case. Because we are maintaining a RB tree, we know that this height is guaranteed to be . Therefore, finding the maximum element in the tree takes time .
* Once we have the maximum element (the project starting last among all projects), we can compare it directly to the new project interval, *i*, to see whether there is an overlap. This requires constant time.
* Therefore, overall, the runtime for determining if project interval i overlaps the project that starts *last* among all the project is .

2. Given a project start time *t*, determine if there is a project that starts at exactly time t.

* The project start time is the property *x.int.low* in the tree.
* Given a project with start time *t*, to determine if there is already a project in the tree with the same start time, we have to perform a tree search for an element where *t* = *x.int.low*.
* We search the tree using the fact that the BST property is maintained by *x.int.low*. If *t* < *x.int.low*, we search left. Otherwise, we search right.
* Once we find the element where *x.int.low* = t, we can return true. Otherwise, if we run out of nodes to search, we return false.
* In the worst-case we have to search all nodes from root to leaf that correspond to the longest path in the tree. We know that with a RB tree in the worst-case this is . Therefore, the time required to perform this search operation is

3. Given a start time *t*, where *t* is the start time of some project in the set, the goal is to determine *the next project* to start after time t. Time .

* Given some start time, we want to know the *next project* to start after time *t.* That is to say, we want to find the *next closest project start time* to *t*.
* Because we know that that *t* corresponds to an interval (project), *i*, already in our set (tree), this is the same thing as saying that we want to find a *successor* to *i*.
* Before we find a successor, though, we need to find the location of interval *i*. We conduct a search for *i* using the steps outlined in above in 2. This takes time .
* Once we have interval *i* we can conduct a successor search:
  + If *i* has a right subtree, we look for an interval *i2* that is the minimum element of the right subtree. In other words, we would do *findMin(i.right)*.
  + If *i* does not have a right subtree, we need to look for the successor from the root. We look from the root to interval *i* keeping track of intervals where *i2.int.low > i.int.low*. We update to make sure we take the smallest value of *i2.int.low* that is still larger than *i.int.low* as we encounter it.
* Because in the worst case, as before, our search would need to be performed from root to leaf along the longest path of the tree, we know that the worst-case run time of performing a successor search is , given that we have a RB tree.

4. Given a time *t*, output all projects that starts after time *t*. Time: where *k* is the number of projects that start after time *t*.

* To output all projects starting after time t, we need a modification of the range query.
* The idea here is to first find the first element that has a start time (*x.int.low)* greater than time *t*. Call this node *v*. This search takes time
* Next, we perform a search for all elements on *v*’s left subtree for all elements where *x.int.low* is greater than *t*. If we encounter a node where *x.int.low* is greater than *t*, we output that element and search left. Otherwise, we search right. We continue searching until we either encounter an element where *x.int.low = t* or reach a leaf node. Because in the worst-case this search is along the longest path, and we perform a constant amount of work at each encountered node, this takes time in the worst-case for a self-balancing tree.
* Finally, we need to output all element’s to *v*’s right, all of which have a start time greater (*x.int.low)* greater than *t*. We can do this using an *in-order traversal*. Because the number of elements in-scope for the in-order traversal is at most *k* elements, this takes time .
* Ultimately, we conduct two search operations both taking time and one traversal operation taking at most time . Therefore, the entire runtime is therefore .

5. Output the first 5 projects to start. Time: .

* The 5 first projects to start correspond in our data structure to the 5 intervals with the smallest *x.int.low* values.
* We can call *findMin* and *delete* 5 times to output the 5 smallest elements. For a RB tree, both steps take time .
* After finishing this procedure, we can reinsert the 5 deleted elements to restore the original data structure. Each insertion also takes time .
* Altogether, each iteration – findMin, delete, insert – takes time . Because this is happening 5 times, the new runtime is then , which simplifies to just .

6. Output the project with the latest finish time: .

* The max value of the root (*x.max* in our data structure) comes from the element with the greatest *x.int.high* value.
* This value corresponds to the project that has the latest finish time.
* Knowing the value of *x.int.high* that we are interested in, we have to conduct a search for the interval that has that *x.int.high* value. This is the interval (project) we want to output.
* To search for this project, we need to search using *k =* *x.max* and interval *i*:
  + if *i.int.high* = *k*, return I (we have found the project with the latest start date)
  + else if *x.left != nil* and *x.left.max* = k, search left
  + otherwise, search right
* Because this search, in the worst-case, has to search through the longest path of the tree, we know that the runtime is at worst because we know the height of the tree is at most .
* Therefore, the runtime of the entire algorithm is .

7. Output the pair of projects whose start times are closest together. Time: .

* To output the pair of projects whose start times are closest together, we need to find the *gap* between successive pairs of projects.
  + I.e. if P1 has start time s1, and P2 has start time s2, the gap is given by:
* First, we perform an *inorder* traversal of the tree, outputting elements into an array. An *inorder* traversal outputs elements in a BST in sorted order. Because our tree is keyed by *x.int.low*, the projects are ordered from smallest to greatest start time. We have to visit each element once performing constant work, so this takes time .
* Next, we loop through the sorted elements in the array to find the pair of start times with the smallest gap. This corresponds to the start times that are closest together.
  + As we loop through the array, we keep track of the *minimum* gap between two adjacent projects. If we encounter a pair of adjacent projects that has a smaller gap than the current minimum, we update the current minimum and the pair of projects. Because we loop through all elements of the array performing constant work, this step takes time .
* Finally, we can return the pair of elements that correspond to the minimum. These are the projects whose start times are closest together.
* Because both steps take time the overall runtime of this process is .

8. Output the project that starts immediately before project *x*. Time:.

* Because our RB interval tree is keyed using *x.int.low*, i.e. the project start times, the project that starts immediately before project *x* is project *x’s* immediate *predecessor.*
* Finding the predecessor is analogous to searching for the successor; both operations run in time .
* If *x* has a left subtree, the predecessor is the maximum value of the left subtree. In other words, findMax(x.left).
* If *x* does not have a left subtree, we have to search from the root. If *i.int.low < x.int.low*, we keep track of *i* and search right. Otherwise, we search left.
* We keep track of all nodes we encounter where *i.int.low < x.int.low*, keeping track of the maximum value of *i.int.low* we encounter that is still less than smaller than *x.int.low*.
* We stop when *i = x*.
* Because in the worst-case we are searching the longest path of the tree, because we have a RB tree we know this is at most .
* Because both the search for

9. Output the number of projects that start between project *x*'s start time, and project *y*'s start time.

Time: .

* Because our data structure is keyed on project start times and maintains subtree sizes at each node, we can use a *range* operation to select all projects that start between project *x*’s start time and project *y*’s start time.
* If we say that project *x*’s start time is *a* and project *y*’s start time is *b*, we want the number of projects (intervals) where . This is exactly what the range query gives us.
* Because the problem does not explicitly state otherwise, assume that the start time bounds are inclusive.
* Using the standard range approach, we first locate the first interval where *x.int.low* is between *a* and *b*. Call this node *v*. The worst-case time for this search, as we’ve described above, is
* Once we have node *v*, we carry out two searches. First, we search the left subtree for all nodes where *x.int.low* is *greater than or equal to* *a*. We maintain a running sum, *left-total* that counts each node in the path that is greater than or equal to a, as well as that node’s right subtree (if it exists). We continue searching the path until we either find the project with *x.int.low = a* or reach a leaf node. Because the search in worst-case is along the longest path in the tree, this takes time for a self-balancing tree.
* Next, we search the right subtree for all nodes where *x.int.low* is smaller than or equal to *b*. We maintain a running sum, *right-total* that counts each node in the path that is less than or equal to b, as well as that node’s left subtree (if it exists). We continue searching the path until we find the node with *x.int.low = b* or reach a leaf. Because the search in worst-case is along the longest path in the tree, this takes time for a self-balancing tree.
* Finally, we return *left-total* + *right-total* + 1 as the number of projects that start between project *x*’s start time and project *y*’s start time.
* Because all 3 search operations take time and all other operations are constant, the overall runtime for this algorithm takes time