University of Freiburg
Dept. of Computer Science
Prof. Dr. F. Kuhn
P. Bamberger, M. Fuchs,
A. Lehmann, P. Schneider

Algorithms and Datastructures
Summer Term 2022
Sample Solution Exercise Sheet 6

Due: Wednesday, June 15th, 4pm

Exercise 1: Binary Search Tree - Range Queries *(10 Points)*

(a) Implement the binary search tree (BST) data structure and the `insert` operation. You can use the template `BST.py`. *(4 Points)*

(b) Implement the operation `getrange`(x_{min}, x_{max}) efficiently on binary search trees which returns all keys x in the tree with $x_{min} \leq x < x_{max}$ (cf. lectue notes week 6 slide 21). *(4 Points)*

(c) Use your implementation of BST and your `insert` function to insert all words from the file `inputs.txt` into a BST with respect to the lexicographic ordering on words over the alphabet $\{a, \dots, z\}$[1]. Use your data structure to output all words from the BST beginning with a certain prefix.[2] Output all words with prefix "qw". Copy the result into your `experiences.txt` file. *(2 Points)*

Sample Solution

Cf. `BST.py` for part (a) and (b). For part (c) it was sufficient to run `getrange('qw','qx')` on the BST filled with the words from `input.py`. The correct output is `['qwb', 'qwdjbcsm', 'qweli', 'qwgconj', 'qwgzykg', 'qwivkay', 'qwlybcn', 'qwmwwi', 'qwo', 'qwohudf', 'qwpoh', 'qwqrn', 'qwrmxd', 'qwtq', 'qwxpyjl', 'qwxrm', 'qwyiwh']`.

Exercise 2: Binary Search Tree - Operations *(10 Points)*

(a) Describe a function which returns the depth of a binary search tree and analyze the runtime. *(2 Points)*

(b) Describe a function that for a given binary search tree with n nodes and a given $k \leq n$ returns a list with the k smallest keys from the tree. Analyze the runtime. *(4 Points)*

(c) Describe a function that takes a binary search tree B and a key x as input and generates the following output:

- If there is an element v in B with $v.key = x$, return v.
- Otherwise, return the pair (u, w) where u is the tree element with the next smaller key and w is the element with the next larger key. It should be $u =$ None if x is smaller than any key in the tree and $w =$ None if x is larger than any key in the tree.

For your description you can use pseudo code or a sufficiently detailed description in English.

Analyze the runtime of your function. *(4 Points)*

[1] Python supports the comparison of strings with respect to the lexicographic ordering, i.e., you can use "<", "<=".

[2] If you enter `Python3` and `from BST import BST` into the command prompt you can use the class `BST` from the command line. We provided a method for inserting the content of `inputs.txt`.

Sample Solution

(a) We can do a recursive traversal of the tree where we keep track of the current recursion depth. Then a call of `depth`(r) on the root r ot the BST returns its depth.

Algorithm 1 `depth`(v, R)

> **if** $v =$ None **then**
>   **return** -1 ▷ *depth of a childless node must be 0, hence we define the depth of None as -1*
> **else return** max $\big($`depth`$(v.$`left`$)+1$, `depth`$(v.$`right`$)+1\big)$

The runtime corresponds to the runtime of the traversal of the whole tree which is $\mathcal{O}(n)$ as we have just one recursive call for each node and each recursive call costs $\mathcal{O}(1)$ (c.f., pre-, in-, post-order traversal algorithms given in the lecture).

As an alternative solution, we can run a BFS which takes $\mathcal{O}(n)$. If v is the node visited last by the BFS, do

Algorithm 2 `traverse-up`(v)

> $d \leftarrow 0$
> **while** $v.$`parent` $\neq$ None **do**
> $d \leftarrow d + 1$
> $v \leftarrow v.$`parent`
> **return** d

This takes $\mathcal{O}(d)$ where d is the depth of the tree. Since $d \leq n$ the overall runtime is $\mathcal{O}(n{+}d) = \mathcal{O}(n)$.

(b) Initialize an empty list K. We roughly do the following. Make an in-order traversal of the tree and each time visiting a node, add it to K. Stop if $|K| \geq k$. The following pseudocode formalizes this.

Algorithm 3 `inorder_variant`(node) ▷ Assume list K is given globally, initially empty

> **if** node $\neq$ `None` **then**
>   `inorder_variant`(node.`left`)
>   **if** $|K| \geq k$ **then**
>   **return**
> K.append$(v.key)$
>   `inorder_variant`(node.`right`)

The runtime is $\mathcal{O}(d + k)$ where d is the depth of the tree. We prove this in the following.

Let K be the set of k nodes representing the k smallest keys in the BST. Obviously, the in-order traversal must visit all nodes in K once. In accorddance with the lecture a call of `inorder_variant`(root) adds all keys in ascending order to K.

Let A be the set of nodes in the BST on which are not in K but in which a recursive call will be made. Since the recursion is aborted (with the `return` statement) after reporting k nodes, the set A contains exactly the nodes which are ancestors of a node in K, but are not in K themselves. Since the runtime of a single recursive call (neglecting subcalls) is $\big(1\big)$ the total runtime is $\mathcal{O}(|A| + |K|)$.

By definition we have $|K| = k$, so it remains to determine the size of A. We claim that all nodes in a A are on a path from the root to a leaf, that is, $|A| \leq d$. This is the case if there do not exist two nodes in A so that neither is an ancestor of the other.

For a contradiction, suppose that two such nodes u, v exist so that neither u is ancestor of v nor vice versa. Assume (without loss of generality) that `key`$(u) \leq$ `key`(v). That means u is in the left and v is in the right subtree of some common ancestor a of u and v.

By definition v has a node $w \in K$ in its subtree. Since v is in the right subtree and u is in the left subtree of a, we have `key`$(w) \geq$ `key`(u) and w has a higher in-order-position. But then we would have $u \in K$ as well, a contradiction to $u \in A$.

(c) **Algorithm 4** `return-closest`(x)

$v \leftarrow$ `find`(x)
if $v \neq$ None **then**
    return v
else
    `insert`(x)
    $(p, s) \leftarrow$ (`pred`(x), `succ`(x))
    `delete`(x)
    return (p, s)

All subprocedures that we call (`find`, `insert`, `pred`, `succ`) are known from the lecture and take $\mathcal{O}(d)$ with d being the depth of the tree. So the overall runtime is $\mathcal{O}(d)$.