Data Structure I: Tutorial 3

First, we discuss expectations of solutions for the first experimental assignment.

- Create proper plots of all experimental results.
- Write all theoretical knowledge related to the assignment. Cite all sources.
- Discuss whether the theory correspond to experimental results.
- Submit only the report in PDF format.

Exercise 1 *The size $s(u)$ of a node u in a tree is the number of nodes in the subtree of u including u. The potential Φ of a tree is the sum of $\log_2(s(u))$ over all nodes u. Determine the potential of a path and a perfectly balanced binary tree on n nodes. Proof that $0 \leq \Phi \leq n \log n$ for every binary tree.*

Why Splay trees are so interesting?

Theorem 2 (Balance) *The cost of performing a sequence of m accesses is $O((m+n)\log n)$.*

This theorem implies that Splay trees perform as well as static balanced binary search trees on sequences of at least n accesses.

Exercise 3 *Consider a binary tree and iterate all elements in the increasing order using the operation Successor from the first assignment. What is the total time complexity?*

Exercise 4 *Now, use operation Find in binary trees to search every element once. For which binary tree we obtain the worst possible total time complexity of all searches?*

Theorem 5 (Scanning) *Accessing all n elements in the increasing order takes $O(n)$ time, regardless of the initial structure of the splay tree.*

Theorem 6 (Working set) *For a sequence S of m accesses, let $t(u)$ be the number of distinct elements accessed before the previous time element u was accessed. Then, the cost of performing S is $O(m + n \log n + \sum_{u \in S} \log t(u))$. Hence, if S accesses only z distinct elements, the cost is $O(m + n \log n + m \log(z))$.*

If S is a sequence of $z \geq 2$ distinct elements of length $m \geq n \log n$, then the cost of S is $O(m \log z)$ and the amortized cost of one access is $O(\log z)$.

Theorem 7 (Static optimality) *Let M be a set of n elements and S be a sequence of accesses to M in which every element of M is accessed at least once. Let T be an arbitrary static binary search tree on M and f be the total time to access M. Then, the total time to access M in Splay tree is $O(f)$, regardless of the initial structure of the splay tree.*

Exercise 8 *Let S be a sequence of accesses to elements of M and T be a static binary search tree on M. The cost of accessing S in T is $\sum_{u \in S} d_T(u)$ where $d_T(u)$ is the depth of u in T. Static optimal search tree for S is a binary search tree on M minimizing the cost. For a given sequence S, write an algorithm which finds a static optimal tree.*

Exercise 9 *We that that Splay trees cannot be asymptotically slower than the static optimal tree for any sufficiently long sequence of accesses. Is there a sequence S of accesses for which Splay tree is asymptotically faster than the static optimal tree (built for S)?*

Are the rules for rotations used in Splay trees the best possible? I.e. is it possible to modify rotations to asymptotically improve Splay trees (for some sufficiently long sequence of accesses)?