University of Waterloo
CS240 - Spring 2023
Assignment 4

Due Date: Wednesday July 12, 5pm

You should have submitted AID02 before the due date of this assignment. The agreement will indicate what you must do to ensure the integrity of your grade. If you are having difficulties with the assignment, course staff are there to help (provided it isn't last minute).

The Academic Integrity Declaration must be signed and submitted on time or the assessment will not be marked.

Please read `https://student.cs.uwaterloo.ca/~cs240/s23/assignments.phtml#guidelines` for guidelines on submission. **Each question must be submitted individually to MarkUs as a PDF** with the corresponding file names: a4q1.pdf, a4q2.pdf, ... It is a good idea to submit questions as you go so you aren't trying to create several PDF files at the last minute.

Late Policy: Assignments are due at 5:00pm, with the grace period until 11:59pm. Assignments submitted after 11:59pm on the due date will not be accepted but may be reviewed (by request) for feedback purposes only.

Question 1 Selection in an AVL tree [7 marks]

We work with a slightly modified version of AVL trees, where each node x also stores the number of keys in the subtree rooted at x (this includes x itself); you do not have to explain how this is done. Give an algorithm $\texttt{Selection}(T, k)$ that takes as input an AVL tree T with n keys and an index $k \in \{0, \ldots, n-1\}$, and returns the kth key in T (for the usual in-ordering, so $k = 0$ is the smallest key and $k = n-1$ the largest). Give the complexity of your algorithm; worst-case runtimes $\Theta(n)$ will not get full credit.

Question 2 Leaf levels in AVL trees [1+2+4+4 marks]

Consider an AVL tree T with n nodes, and let v be an arbitrary leaf in T. We want to give a lower bound on the *level* ℓ of v, that is, the length of the path from the root of T to v.

a) Explain why knowing the bound $\Theta(\log(n))$ on the height of an AVL tree is not enough to answer this question.

b) Let $v_0, \ldots, v_\ell$ be the path from the root of T to v (where v_0 is the root of T and $v_\ell = v$), and let $T_0, \ldots, T_\ell$ be the subtrees of T rooted at respectively $v_0, \ldots, v_\ell$. What is T_0 and what is T_ℓ?

c) Prove by induction that for $i = 0, \ldots, \ell$, $T_{\ell-i}$ has height at most $2i$.

d) Using the previous question, deduce a lower bound of the form $\ell \in \Omega(g(n))$, for a certain function $g(n)$.

Question 3 Self-organizing search [4+4 marks]

Consider the list of keys:

$$[1\ 2\ 3\ 4\ 5\ 6\ 7\ 8\ 9\ 10]$$

and assume we perform the following searches:

$$10,\ 7,\ 2,\ 2,\ 4^\star,\ 1,\ 2,\ 1,\ 2,\ 1^\star,\ 1,\ 7,\ 1,\ 9^\star$$

a) Using the move-to-front heuristic, give the list ordering after the starred ($\star$) searches are performed.

b) Same question, using the transpose heuristic instead of the move-to-front heuristic.

Question 4 Skip lists [2+2+3+4 marks]

In class, we saw that skip lists have expected $O(n)$ space requirements and expected height $O(\log(n))$. The proof involves a few steps, which you will generalize for this question. The main difference is that in this problem, we assume that the probability of adding a level to a tower is a fixed number p (with $0 < p < 1$), instead of $\frac{1}{2}$.

You will probably need somewhere the equality $\sum_{i \geq 0} p^i = 1+p+p^2+\cdots+p^i+\cdots = 1/(1-p)$.

a) Show that the probability that a given tower in the skip list has height at least i is p^i.

b) Assuming there are n distinct keys (that is, n towers) in the skip list, give the expected number of nodes at level i, for $i \geq 0$.

c) Show that the expected number of nodes in a skip list with n keys is $n/(1-p)$.

d) Show that the expected height of a skip list is at most

$$\log_{1/p} n - 1 + 1/(1-p).$$

You can assume that n is a power of $1/p$ (for $p = 1/2$, this would mean that n is a power of 2, for instance).

Hint: use indicator variables V_i, with $V_i = 0$ if level i is empty and 1 otherwise, and prove that you have both $E(V_i) \leq 1$ and $E(V_i) \leq np^i$, where $E(\ )$ denotes the expected value.

Question 5 Interpolation search [5 marks]

Prove that the worst case of interpolation search in an array of size n is $\Theta(n)$.

Question 6 Tries [5 marks]

Draw the uncompressed trie obtained by inserting the following strings into an initially empty trie:

$$1001\$, 001\$, 1111\$, 10110\$, 10\$, 11\$, 10100\$, 1\$, 000\$, 101\$$$