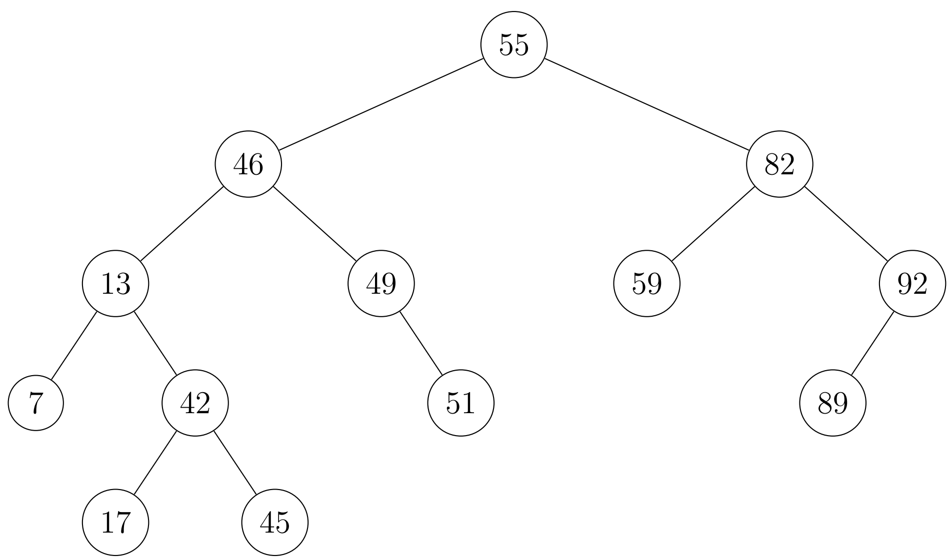
University of Waterloo
CS240 Fall 2017
Assignment 3

Due Date: Wednesday, November 1, at 5pm

Please read `http://www.student.cs.uwaterloo.ca/~cs240/f17/guidelines.pdf` for guidelines on submission. This assignment contains written questions and a programming question. Submit your written solutions electronically as a PDF with file name a03wp.pdf using MarkUs. We will also accept individual question files named a03q1w.pdf, a03q2w.pdf, a03q3w.pdf, a03q4w.pdf and a03q5w if you wish to submit questions as you complete them.

Problem 1 AVL Tree Operations [2+4+4=10 marks]

This problem will concern operations on the following AVL tree T:



a) Show that T is an AVL tree by writing in the balance at each node.

b) Show the process of inserting a KVP with key 29 into T. Specifically, indicate every single rotation performed, in order (show any double rotations as two separate single rotations). Draw the tree, with balance factors, after each rotation.

c) Show the process of deleting key 49 from the *original* tree T above. Specifically, indicate every single rotation performed, in order (show any double rotations as two separate single rotations). Draw the tree, with balance factors, after each rotation.

Problem 2 AVL-2 trees [3+4+3+3=13 marks]

We consider a modified version of AVL trees, where the height difference between the left and right subtrees of any node is in $\{-2, -1, 0, 1, 2\}$ instead of $\{-1, 0, 1\}$. These are called AVL-2 trees. We let m_h be the minimum number of nodes of an AVL-2 tree with height h.

a) For $h = 1, \ldots, 6$, determine m_h and give an example of an AVL-2 tree with m_h nodes.

b) Find a recurrence relation for m_h and give its initial conditions.

c) Using your recurrence, prove that $m_h \geq 2^{h/3}$ by induction on h.

d) Prove that the height of an AVL-2 tree with n nodes is $\Theta(\log(n))$.

Problem 3 Dynamic Ordering [5+5=10 marks]

Consider the list of keys:

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

and assume we perform the following searches:

$$10, 7, 2, 5, 3^\star, 3, 1, 6, 4, 9^\star, 1, 8, 1, 9^\star$$

a) Using the move-to-front heuristic, give the list ordering after the starred ($\star$) searches are performed. Additionally, record the number of comparisons between keys after each search, as well as the total number of comparisons.

10	7	2	5	3	3	1	6	4	9	1	8	1	9	Total

b) Repeat part (a), using the transpose heuristic instead of the move-to-front heuristic.

Problem 4 Skip Lists [6+2+3+3+3=17 marks]

a) Show the skip list that is created by inserting the keys

23, 12, 56, 231, 3, 16, 8, 10, 15, 7, 0, 11

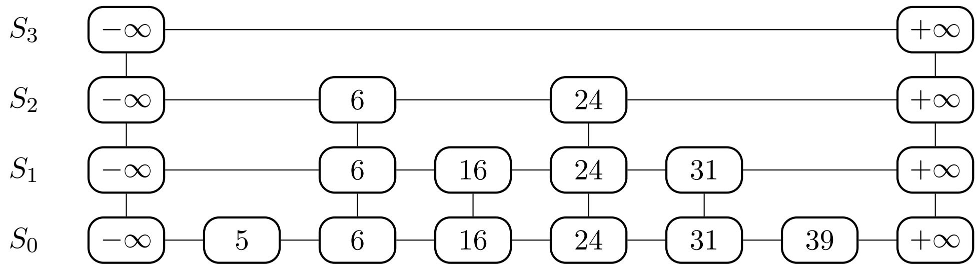
into an empty skip list. You must use the following coin flip sequence:

$$THTHHHTHTHHTHTHTHHTHHHHTHTHTHHTT$$

Note that each coin flip in the sequence will only be used once, in order and there may be some unused coin flips.

After inserting these keys, determine the exact number of comparisons required to search for each key (that is, indicate what the search cost would be for 23, and then the search cost for 12, and so on).

The following diagram is an example of a skiplist that may be useful in order to review the required format:



b) Assume that the probability of adding a level to a tower is p ($0 < p < 1$), as opposed to $\frac{1}{2}$.

- **i)** Explain why the probability that a node in the skip list has level at least i is p^i.
- **ii)** Show that the expected number of nodes (i.e., the sum of all heights of all towers) is $O(n)$, where the constant $c = 1/(1-p)$. Therefore, the space requirements for this skip list are linear in the number of keys being stored.
- **iii)** Similar to what was done in lecture, let $C(k)$ be the expected total path length which rises k levels when working back to the top-most, left-most node. Find the recurrence relation for $C(k)$ in terms of k and p.
- **iv)** Assuming $C(0) = 0$, show that your recurrence relation from part (iii) has closed form $C(k) = k/p$.

Problem 5 Skip lists with two levels [5+4=9 marks]

Suppose you have a skip list with only two levels: the lower one has n entries $a_0, \ldots, a_{n-1}$, and the top one has k entries. For simplicity, we assume k divides n, so that $n = km$, for some integer m. We assume that the k top entries are evenly spread out, so they correspond to $a_0, a_m, a_{2m}, \ldots, a_{(k-1)m}$.

a) What is the worst time for a query? Give a Θ expression involving k and n.

b) Given n, how should you choose k to minimize this worst case, and what does the worst case become? Give Θ expressions.

Problem 6 Tries [4+3+4+3+3+5=22 marks]

a) Insert the following binary keys into an initially empty (uncompressed) binary trie:

1001, 001, 1111, 10110, 10, 11, 10100, 1, 000, 101, 00

Be sure to indicate which nodes are flagged and which are unflagged.

b) From your answer to part (a), delete the following keys, and show the trie after each deletion:

$$10100, 11, 1001$$

Note that LaTeX will save you some writing for this question.

c) Repeat part (a), except use a compressed trie.

d) Repeat part (b), starting from your answer to part (c).

e) Find the exact height of the trie obtained when the keys representing the binary numbers from 0 to $2^k - 1$ without leading 0s are inserted into a trie in increasing order. That is, insert 0, 1, 10, 11, etc.

f) Repeat part (e) using compressed tries. Also, for this part, prove your result, including any structural properties of the compressed tries, using mathematical induction.

Note that in order to obtain full marks for proofs involving mathematical induction, solutions must clearly state:

- what you are doing induction on;
- the base case(s);
- the inductive hypothesis;
- the inductive step.