CSCI 6220/4030: Homework 4

Assigned Thursday October 26 2017. Due at beginning of class Thursday Nov 9 2017.

Remember to typeset your submission, and label it with your name. Please start early so you have ample time to see me during office hours. Provide mathematically convincing arguments for the following problems. Ask me if you are unclear whether your arguments are acceptable.

1. Consider a random treap T with n vertices. As in the lectures, order the vertices in decreasing order of their search keys, so $x_1 > \cdots > x_n$, and the priorities of the vertices are i.i.d Uniform$[0, 1]$ random variables.

    (a) What is the expected number of leaves in T?

    (b) What is the expected number of nodes in T with two children?

    (c) What is the expected number of nodes in T with exactly one child?

    (d) Prove that the expected number of proper descendants of any node in T is exactly equal to the expected depth of that node.

    (e) What is the probability that x_j is a common ancestor of x_i and x_k?

    (f) What is the expected length of the unique path from x_i to x_k in T?

2. Suppose we are given two skip lists, one storing a set A of m keys, the other storing a set B of n keys. Describe and analyze an algorithm to merge these into a single skip list storing the set $A \cup B$ in $O(n + m)$ expected time. We do not assume that every key in A is smaller than every key in B; the two sets may be arbitrarily intermixed.

3. Consider the following variant of multiplicative hashing, which uses slightly longer salt parameters. For any integers $a, b \in [2^{w+\ell}]$ where a is odd, let

$$h_{a,b}(x) = \left\lfloor \frac{(ax+b) \bmod 2^{w+\ell}}{2^w} \right\rfloor,$$

and let $\mathcal{MB}^+ = \left\{ h_{a,b} \mid a, b \in [2^{w+\ell}] \text{ and } a \text{ is odd} \right\}$. Prove that the family of hash functions $\mathcal{MB}^+$ is strongly near-universal (aka near-uniform):

$$\mathbb{P}\left[\{h(x) = i\} \cap \{h(y) = j\} \right] \leq \frac{2}{m^2}$$

for all items $x \neq y$ and all (possibly equal) hash values i and j.

4. Suppose that we are using an open-addressed hash table of size m to store n items, where $n \leq m/2$. Assume an ideal random hash function. For any i, let X_i denote the number of probes required for the ith insertion into the table, and let $X = \max_i X_i$ denote the length of the longest probe sequence.

    (a) Prove that $\mathbb{P}[X_i > k] \leq \frac{1}{2^k}$ for all i and k.

    (b) Prove that $\mathbb{P}[X_i > 2\ln n] \leq \frac{1}{n^2}$ for all i.

    (c) Prove that $\mathbb{P}[X > 2\ln n] \leq \frac{1}{n}$.

    (d) Prove that $\mathbb{E}[X] = O(\ln n)$.

5. A radix tree over an alphabet of size m is a tree data structure where each node has up to m children, each corresponding to one of the letters in some alphabet. A word is represented by a node at the end of a path whose edges are labeled with the letters in the word in order. The only nodes created in the radix tree are those corresponding to stored keys or ancestors of stored keys. Radix trees are particularly useful for string matching and string completion problems. See Figure 1 for an example.

    Suppose you have a radix tree into which you have already inserted n strings of length k from an alphabet of size m, generated uniformly at random with replacement. What is the expected number of new nodes you need to create to insert a new string of length k?
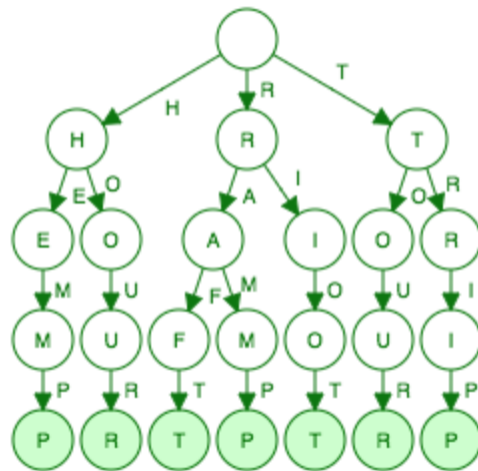
Figure 1: A radix tree containing the words "HEMP", "HOUR", "RAFT", "RAMP", "RIOT", "TOUR", and "TRIP".