Problem Set 7 (Revision 1)

Data Structures and Algorithms, Fall 2022

Due: October 27, in class.

Problem 1

(a) Recall that in the last problem set, we have created a red-black tree by successively inserting the keys 99, 80, 77, 11, 31, 2 into an initially empty red-black tree. Now, show the red-black trees that result from the successive deletion of the keys in the order 2, 31, 11, 77, 80, 99. You need to show the red-black tree after *each* deletion.

(b) Suppose that a node x is inserted into a red-black tree and then is immediately deleted. Is the resulting red-black tree the same as the initial red-black tree? You need to prove your answer.

Problem 2 [OJ Problem]

(Solve this problem on the OJ platform, do not hand in written solutions!)

In this problem, you are required to design a data structure to maintain a multiset S. When S contains n elements, your data structure should complete each of the following operations in $O(\log n)$ time.

- Given element x, add x to the multiset S;
- Given element x, output the number of occurrences of x in S, then remove *one single copy* of x from S if $x \in S$;
- Given L and R, output $\sum_{L \leq x \leq R} x \cdot o_x$, where o_x is the number of occurrences of x in S.

Problem 3

You are given a function RAND7, which generates a uniform random integer in the range 1 to 7 in $O(1)$ time. Use it to produce a function RAND10, which generates a uniform random integer in the range 1 to 10. Remember to analyze the running time of your RAND10 function.

Problem 4

You are given n unbiased coins, and perform the following process to generate all heads. Toss all n coins independently at random onto a table. Each round consists of picking up all the tails-up coins and tossing them onto the table again. You repeat until all coins are heads.

(a) What is the expected number of rounds performed by the process? Give an asymptotic upper bound as tight as possible. You need to prove your answer.

(b) What is the expected number of coin tosses performed by the process? Give an asymptotic upper bound as tight as possible. You need to prove your answer.

Problem 5

(a) Consider a version of the division method in which $h(k) = k \bmod m$, where $m = 2^p - 1$, k is a character string interpreted in radix 2^p, and $p > 1$ is an integer. (For example, if we use the 7-bit ASCII encoding, then $p = 7$ and string "AB" has key value $65 \times 128 + 66$.) Show that if we can derive string x from string y by permuting its characters, then x and y hash to the same value.

(b) Consider inserting the keys 10, 22, 31, 4, 15, 28, 17, 88, 59 into a hash table of length $m = 13$ using open addressing with the auxiliary hash function $h'(k) = k$. Illustrate the result of inserting these keys using linear probing, using quadratic probing with $c_1 = 1$ and $c_2 = 3$, and using double hashing with $h_1(k) = k$ and $h_2(k) = 1 + (k \bmod (m-1))$. (It suffices to show the eventual hash table.)

Problem 6

Define a family $\mathcal{H}$ of hash functions from a finite set U to a finite set B to be ϵ-universal if for all pairs of distinct elements k and l in U,

$$\Pr[h(k) = h(l)] \leq \epsilon$$

where the probability is over the choice of the hash function h drawn uniformly at random from the family $\mathcal{H}$. Show that an ϵ-universal family of hash functions must have

$$\epsilon \geq \frac{1}{|B|} - \frac{1}{|U|}$$