

## CS 2100: Data Structures, Spring 2019

### Final Worksheet

Note: Do not submit! This is a final worksheet, so the questions are just to help you practice.

1. Draw the treap that results when the following pairs are inserted, where we form a BST over the letters and a min heap over the numbers:  
 $(R, 5), (S, 9), (E, 7), (H, 1), (W, 13), (D, 8), (J, 2), (K, 4), (P, 11)$ .
  
2.
  - (a) Draw the frequency array and Huffman tree for the following string: “dogs can not spot hot pots or cats”.
  - (b) Using the Huffman tree from the previous part, write the binary encoding of the message “spot soda too”.
  - (c) How many bits did your encoding of the message in the previous problem take? How many bits would a standard ASCII encoding have taken (assuming 8 bits per character)?
  
3.
  - (a) Draw the 11-entry hash table that results from using the hash function  $h(i) = (3i + 5) \bmod 11$  to hash the keys 12, 44, 13, 53, 23, 71, 8, 39, 20, 16, and 5, assuming collisions are handled by *chaining* (or inserting into a list when there are collisions).
  - (b) What is the result of the previous part, assuming collisions are handled by linear probing?
  - (c) Show the result from part (a) assuming collisions are handled by quadratic probing (up to the point where the method fails).
  - (d) Show the result from part (a) assuming collisions are handled by double hashing using a secondary hash function  $h'(k) = 7 - (k \bmod 7)$ .
  
4. Consider the adjacency matrix and adjacency list representations of a graph. Which choice would be better in each of the following situations? Justify your answer.
  - (a) The graph has 10,000 vertices and 20,000 edges, and it is important to use as little space as possible.
  - (b) The graph has 10,000 vertices and 20,000,000 edges, and it is important to use as little space as possible.
  - (c) You need to answer queries to find if two vertices are adjacent as fast as possible, no matter how much space you use.

5. A phone company, RT&T, has designed a network of  $n$  switching stations connected by  $m$  high-speed communication links. Each phone connects directly to one of the switching stations in his or her area. The engineers of RT&T developed a prototype video chat system on the phone, but in order for it to have acceptable image quality, the number of links to transmit video cannot exceed 4. Design an efficient algorithm that computes, for each station, the set of stations it can link to for a video chat. (Note: be sure to tell me the running time for your algorithm and justify it!)
  
6. Set implementation questions:
  - (a) Given two sets  $A$  and  $B$  which are stored in sorted vectors, how would you implement the union operation, to compute a new set vector that contains all elements from  $A$  and  $B$  (with no duplicates)? What's the running time of your function?
  - (b) If you have a set that is over values from 0 to some maximum value  $N - 1$ , one popular way to represent a set is as a bit vector: a boolean vector of size  $N$ , where entry  $i$  is equal to true if and only if  $i$  is present in your set. How would you implement the union function in this representation? What would the running time be, and how does it compare to the previous implementation in the last question?