CSC 501 - Fall
Quiz #1

Zhe(Kevin) Chen
V00819544

Question 1
You would like to store key-value pairs of the form (student id → student name) in a way that permits quickly retrieving student names, given their ids. Which of the four core data models (array, list, hash, tree) would you use? Give a one sentence explanation for each. (Hint: you could use all four, but two offer distinct advantages over the others.)

Hash is one of the ideal data model for storing (student id, student name) key-value pairs, The student id can be hashed and point to the student name. The time access seems to be O(1).

When space becomes the problem with array. I would use Tree model to store key-value pair with student id(key) to be the comparator. First I would build a binary search tree using their student id, then the time complexity of one student information retrieval will be O(log n).

Linked list will result a time complexity of O(n) after sorting the list which is not ideal in this case. And Array will eventually take very large space to store pairs,the size of array is up to the largest student id, but with O(1) time complexity.

Question 2
You have one collection S of n=1000 unique items, represented by a linked list, and one collection T of m=100 unique items, represented by a balanced, binary tree with sorted leaves.

(a) What is the complexity of finding all the pairs of elements from S and T that match? Be as precise as possible, and explain your response with one sentence.

Since collection S is a linked list that holds 1000 unsorted items, and collection T is a sorted binary tree that holds 100 items. To find all pairs from S and T, the time complexity will be 1000 x log100. Because for every item in S, it will need approximately O(log 100) to find the pair.

(b) What is a simple way to improve the response time of this search?

Sort linked list S from the smallest to largest, then keep track of the position of the last item in T that is larger than the previous item in S.

Question 3
You have a collection of n genomic strings (i.e., sequences of the letters A, T, G, and C) and a

hash table with 16 buckets. Which hash function would you use to map these strings from the set $\{ATGC\}_{2k}$ (for some arbitrary, fixed, constant k) to the range [0,15]?

I think this is a multiple choice question.