CSc545 - fall 2018 - Homework #1.
Due: Sep 22 2018 11:59pm

September 11, 2018

**Instructions.**

1. Solution may **not** be submitted by students in pairs.

2. You may submit a pdf of the homework, either printed or hand-written and scanned, as long as it is **easily** readable.

3. If your solution is illegible not clearly written, it might not be graded.

4. Unless otherwise stated, you should prove the correctness of your answer. A correct answer without justification may be worth less.

5. If you have discussed any problems with other students, mention their names clearly on the homework. These discussions are not forbidden and are actually **encouraged**. However, you must write your whole solution yourself.

6. Unless otherwise specified, all questions have the same weight.

7. You may refer to data structures or their properties studied in class without having to repeat details, and may reference theorems we have studied without proof. If your answer requires only modifications to one of the algorithms, it is enough to mention the required modifications, and the effect (if any) on the running time and on other operations that the algorithm performs.

8. In general, a complete solution should contain the following parts:

   (a) A high level description of the data structures (if needed). *E.g.* We use a binary balanced search tree. Each node contains, a key and pointers to its children. We augment the tree so each node also contains a field...

   (b) A clear description of the main ideas of the algorithm. You may include pseudocode in your solution, but this may not be necessary if your description is clear.

   (c) Proof of correctness (e.g. show that your algorithm always terminates with the desired output).

   (d) A claim about the running time, and a proof showing this claim.
1. Consider an infinite sequence of functions \( f_i \). It is known that for every \( i \), it is true that \( f_i(n) = O(n) \). We define a new function \( g(n) = \sum_{i=1}^{n} f_i(n) \). Is it true that \( g(n) = O(n^2) \) ?

2. Assume that we insert the keys \( 1, 2, \ldots, n \) into a binary search tree \( T \). The algorithm for inserting a key \( k_i \) is simple. Initially the tree is empty. Then for \( i = 1, 2, \ldots, n \), to insert key \( k_i \) we perform a search for \( k_i \) starting at the root of \( T \), and once reaching a NULL pointer, we insert \( k_i \) as a new leaf. For simplicity, assume \( k_i = i \).

We do not balance the tree after each insertion.

As usual \( h(T) \), the height of \( T \) denote the number of edges from the root of \( T \) to the leaf that is the most remote from the root.

(a) warm-up

i. show that the height of \( T \) is \( \Theta(n) \).

ii. Show that the time to perform all \( n \) insertions is \( \Theta(n^2) \)

iii. next we insert the same keys into \( T \) but we could pick different permutations. Show that there is an order for which the time to perform all insertions is \( O(n \log n) \)

iv. During the process of constructing the tree, we define random variable \( Y_{ij} \) to be 1 if \( i \) and \( j \) were compared to each other. (For every \( 1 \leq i, j \leq n \).) Otherwise \( Y_{ij} = 0 \). Show that the work to construct the tree is proportional to

\[
\sum_{1 \leq i,j \leq n} Y_{ij}.
\]

For example if \( j \) is the first key that was inserted, then it is in the root of \( T \) and it is compared to every other key inserted afterward.

(b) Next assume that the keys are inserted in a random order, where each permutation is equally likely.

i. What is the expected time

\[
E \left( \sum_{1 \leq i,j \leq n} Y_{ij} \right) ?
\]

ii. Prove that the expected distance of a key \( j \) from the root is \( O(\log n) \)

iii. Prove that for large enough \( \text{constant } c \), and large enough value of \( n \), the probability that \( h(T) \leq c \log_2 n \) is very close to 1. You could pick what is \( c \) and how to formalize 'close to 1'

3. Show the preference list of a set of \( n \) men and \( n \) women, for which more than one single stable pairing exists.

4. Suggest an \( O(n^2) \)-time algorithm that determines if, given the preference lists of all the men and women, there exists more than one stable matching.
5. Suggest a modification of the SkipList structure, such that in addition to the operations \texttt{insert}(x), \texttt{find}(x) and \texttt{delete}(x), you could also answer the operation \texttt{avg}(x_1, x_2) which reports the average of all keys stored in the skip list whose value is \( \geq x_1 \) but \( \leq x_2 \). The \textit{expected} time for each operation should be \( O(\log n) \).

Hint: Start by storing at each node \( v \) of the SL another field, called size\([v]\), containing the number of keys in the SL between \( v \) and the next node at the same level as \( v \). Note that maintaining the values of these fields might imply extra work while performing other operations.

6. (a) Let \( d \) be a fixed positive integer. Next consider a perfect SkipList constructed as follows: In order to create the \( i \)th level \( L_i \) of the SkipList, we scan the keys of level \( L_{i-1} \), and promote to \( L_i \) every \( d \)'th key. So for example, the perfect SkipList discussed in class uses the value \( d = 2 \). The case \( d = 3 \) implies that every third key is promoted, and so on.

i. What exactly is the worst case running time of \texttt{find}(x), as a function of both \( d \) and \( n \) ?

ii. Which value(s) of \( d \) will do you think would lead to good performances, and which are poor choices? Why?

(b) In this question, consider a SkipList \( L \) created by inserting a set \( S \) of \( n \) keys into an (initially empty) SkipList. As seen in class, if a key \( x \) appears in level \( i \), the probability that it also appears in level \( i + 1 \) is \( \frac{1}{2} \).

Assume that we re-create a SkipList by inserting the same keys, in the same order, but this time this probability \( p \) is 0.01. (if a key \( x \) appears in level \( i \), the probability that it also appears in level \( i + 1 \) is 0.01. ) Will the expected time to perform \texttt{find}(x) operation increase or decrease, compared to the expected time for the same operation in the original SkipList \( L \) ?

7. What is the expected value of \( \texttt{cnt} \) after the following function is executed?

\[
\begin{align*}
\text{cnt} &= 0; M = -\infty; \\
\text{for } i = 1 \text{ to } n \{ \\
\quad x_i = \text{rand}(); \text{Comment: returns a random float number of between 0 and 1.} \\
\quad \text{if } (x_i > M) \text{ then } \{ \\
\quad \quad M = x_i; \\
\quad \quad \text{cnt}++; \\
\quad \} \\
\}\end{align*}
\]

Hint: Define the random variable \( Y_i \) which is 1 if

\[ x_i > \max\{x_1, x_2 \ldots x_{i-1}\} \]

and \( Y_i = 0 \) otherwise. What is \( E(Y_1 + Y_2 + \cdots + Y_n) \).

8. Create a SkipList for a set of \( n \) keys. The keys are known in advance, and are sorted. The height of the Skiplist is \( \leq 3 \). We define \( T(n) \), the “performance” of the Skiplist as the maximum search time.
That is, we perform 'search' for each key \( k_i \), always start at the upper top node, and we define \( T(n) \) as the maximum search time obtain for each one of the keys.

9. Let \( L \) be a perfect SkipList constructed on a set of \( n \) keys, and let \( X = \{k_1 \ldots k_m\} \) be another set of keys. Assume \( m < n \). Suggest an \( O(n \log(\frac{n}{m})) \) time algorithm for computing which key(s) of \( X \) appear in \( L \).

You could not assume that the keys of \( X \) are equally spaced in \( L \). Yet understanding this special case could shed some intuition.