CSCI 3104 Fall 2022
Instructors: Prof. Grochow and Chandra Kanth Nagesh

Problem Set 10

Due Date .... November 14, 2022
Name .... **Aidan Reese**
Student ID **108418975**
Collaborators .... **Miles Sanders**

Contents

Instructions

- The solutions **must be typed**, using proper mathematical notation. We cannot accept hand-written solutions. Useful links and references on LaTeXcan be found here on Canvas.
- You should submit your work through the **class Canvas page** only. Please submit one PDF file, compiled using this LaTeX template.
- You may not need a full page for your solutions; pagebreaks are there to help Gradescope automatically find where each problem is. Even if you do not attempt every problem, please submit this document with no fewer pages than the blank template (or Gradescope has issues with it).
- You are welcome and encouraged to collaborate with your classmates, as well as consult outside resources. You must **cite your sources in this document. Copying from any source is an Honor Code violation. Furthermore, all submissions must be in your own words and reflect your understanding of the material.** If there is any confusion about this policy, it is your responsibility to clarify before the due date.
- Posting to **any** service including, but not limited to Chegg, Reddit, StackExchange, etc., for help on an assignment is a violation of the Honor Code.
- You **must** virtually sign the Honor Code (see Section Honor Code). Failure to do so will result in your assignment not being graded.

Honor Code (Make Sure to Virtually Sign the Honor Pledge)

Problem HC. On my honor, my submission reflects the following:

- My submission is in my own words and reflects my understanding of the material.
- Any collaborations and external sources have been clearly cited in this document.
- I have not posted to external services including, but not limited to Chegg, Reddit, StackExchange, etc.
- I have neither copied nor provided others solutions they can copy.

In the specified region below, clearly indicate that you have upheld the Honor Code. Then type your name.

Honor Code. Aidan Reese □

26 Standard 26 – Hash Tables

26.1 Problem 1

Consider the following hash function. Let U be the universe of strings composed of the characters from the alphabet $\Sigma = [\texttt{A}, \ldots, \texttt{Z}]$, and let the function $f(x_i)$ return the index of a letter $x_i \in \Sigma$, e.g., $f(\texttt{A}) = 1$ and $f(\texttt{Z}) = 26$. Finally, for an m-character string $x \in \Sigma^m$, define $h(x) = ([\sum_{i=1}^{m} f(x_i)] \mod \ell)$, where ℓ is the number of buckets in the hash table. That is, our hash function sums up the index values of the characters of a string x and maps that value onto one of the ℓ buckets.

Suppose this is going to be used to hash words from a large body of English text.

List at least 4 reasons why $h(x)$ is a bad hash function relative to the ideal behavior of uniform hashing.

Answer.
- The first problem with the hashing function is that it assumes that all 26 letters to be the same, when in reality the enlgish language will be using the letter e far more than any other language, this is ineffeceint as so many more of the hashes will be placed at e.
- Using the mod function to do the hashing will lead to a large amount of collisions compared to uniform hashing.
- H(x) is inconsitanct compared to uniform hashing, as H(x) is heavily dependant on the sting that it is fed.
- Any word that contains more than two of the same letter garentees a collision. □

26.2 Problem 2

Consider a chaining hash table A with b buckets that holds data from a fixed, finite universe U. Recall the definition of worst-case analysis, and consider starting with A empty and inserting n elements into A under the assumption that $|U| \leq bn$.

(a) What is the worst case for the number of elements that collide in a single bucket? Give an exact answer and justify it. **Do not assume the uniform hashing assumption for this question.**

Answer. Worst case scenario would be if every element is put into the same bucket. Therefore the worst case scenario for the hashing table would be all the elements so n elements in a single bucket. □

(b) Calculate the worst-case total cost of these n insertions into A, and give your answer as $\Theta(f(n))$ for a suitable function f. Justify your answer.

Answer. Each insertion into the hash table is O(1), and you must do this (n) times so the cost of this will be $\Theta(n)$ □

(c) **For this part only, assume the uniform hashing assumption, and that the elements added were chosen uniformly at random from U.** After the n insertions, suppose that m `find` operations are performed. What is the total cost of these m find operations? Give your answer as $\Theta(f(n))$ for a suitable function f, and justify your answer.

Answer. We know that a find operation is based off O(1 + α) and is done m times so f = $\Theta(m(1+\alpha))$ □

26.3 Problem 3

Hash tables and balanced binary trees can be both be used to implement a dictionary data structure, which supports insertion, deletion, and lookup operations. In balanced binary trees containing n elements, the runtime of all operations is $\Theta(\log n)$.

For each of the following three scenarios, compare the average-case performance of a dictionary implemented with a hash table (which resolves collisions with chaining using doubly-linked lists) to a dictionary implemented with a balanced binary tree.

(a) A hash table with hash function $h_1(x) = 1$ for all keys x.

Answer. We can treat this scenario of hash table to be the same complexity of a linked list. This is because all of the elements will be in one bucket. For delete and search the average search time will be O(n) so longer than a BST. However for insert the cost will be O(1), so faster than BST. □

(b) A hash table with a hash function h_2 that satisfies the Simple Uniform Hashing Assumption, and where the number m of buckets is $\Theta(n)$.

Answer. For a hash table with uniform hashing, the cost to insert will be O(1), so faster than a BST. For both delete and search the cost to delete and search will be O(1 + α). Alpha is calulated with $\frac{n}{m}$ where in this case m = $\Theta(n)$ so alpha = 1. Final cost being O(1) for all operations. □

(c) A hash table with a hash function h_3 that satisfies the Simple Uniform Hashing Assumption, and where the number m of buckets is $\Theta(n^{3/4})$.

Answer. Alpha is defined as $\frac{n}{m}$, where m $\simeq O(n^{3/4})$, so the cost for deleting and searching would be $O(\frac{n}{n^{3/4}}) \Rightarrow \Theta(n^{1/4})$, lastly the cost of insert operation continues to be $\Theta(1)$. □