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

Problem Set 10

Due Date .... November 14, 2022
Name .... **Your Name**
Student ID .... **Your Student ID**
Collaborators .... **List Your Collaborators Here**

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. □

e.g. List API

	LinkedList
- new list	$O(1)$
- add item	$O(1)$
- remove item @ position i	$O(n)$ ← n= len list
- get item at position i	$O(n)$

Assuming uniform hashing, average-case runtime $O(1+\alpha)$

	Balanced Bin. Tree	Hash Table avg.	Hash Table worst
Add	$\Theta(\log n)$	$O(1)$	$O(1)$
Find	$\Theta(\log n)$	$O(1+\alpha)$	$O(n)$
Remove	$\underbrace{\Theta(\log n)}_{\text{wrst-case}}$	$O(1+\alpha)$	$O(n)$

Key: Design hash fn h so that the uniform hashing assumption is as close to true as possible. (for your input data).

When to use BBT vs HashTable

- HashTable Add always better $(\overset{HT}{O(1)} \text{ vs } \overset{BBT}{\Theta(\log n)})$
- If doing lots of finds/remove ops

HT better when $1+\alpha \ll \log n$

$$n/\ell < o(\log n)$$

$$\Omega\left(\frac{n}{\log n}\right) < \ell$$

[Source: Prof. Joshua's lecture note]   Click here for the "HASHING" material.

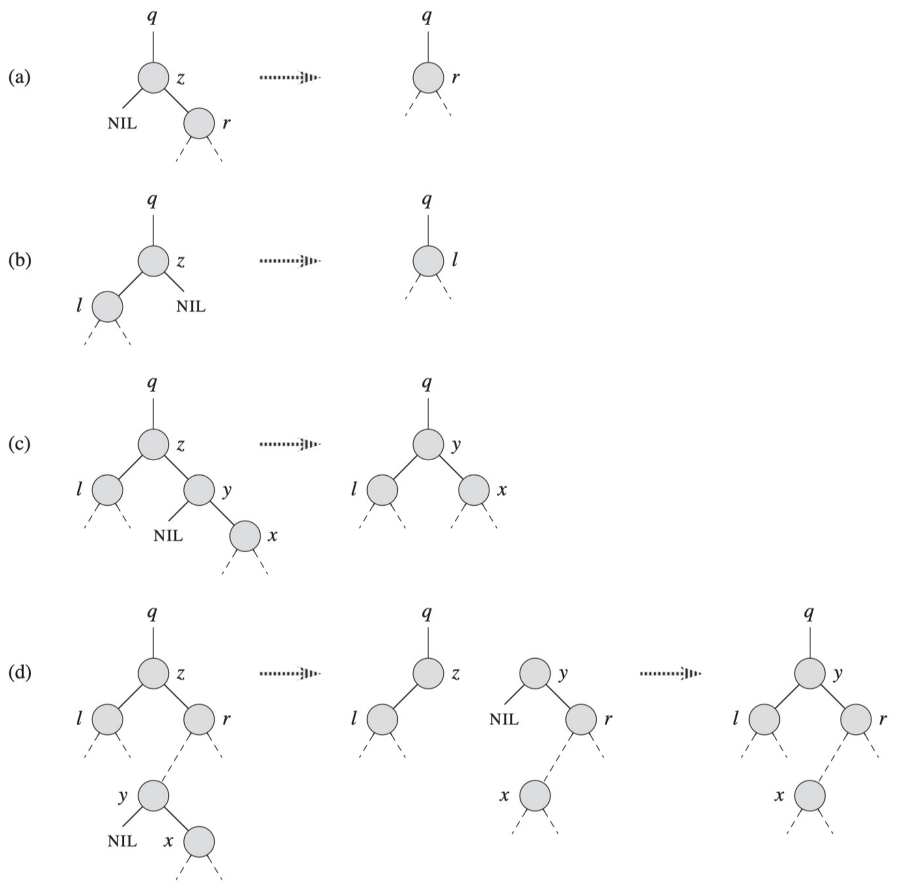

Figure 12.4 Deleting a node z from a binary search tree. Node z may be the root, a left child of node q, or a right child of q. **(a)** Node z has no left child. We replace z by its right child r, which may or may not be NIL. **(b)** Node z has a left child l but no right child. We replace z by l. **(c)** Node z has two children; its left child is node l, its right child is its successor y, and y's right child is node x. We replace z by y, updating y's left child to become l, but leaving x as y's right child. **(d)** Node z has two children (left child l and right child r), and its successor $y \neq r$ lies within the subtree rooted at r. We replace y by its own right child x, and we set y to be r's parent. Then, we set y to be q's child and the parent of l.

[Source: Introduction to Algorithm]

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. □

Click here for the "HASHING" material.

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. □

(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. □

(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. □

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. □

(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. □

(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. □