CS2040C Semester 2 2021/2022

Data Structures and Algorithms

Tutorial+Lab 05
Midterm Quiz/First Half Review; Hash Function

For Week 07

Document is last modified on: February 24, 2022

1 Introduction and Objective

Welcome back from recess week =). I hope that you are (a bit) fresher now.

In the early part of the tutorial component of this session, we will discuss the solutions and a few common mistakes that were found during grading.

Then, we will discuss a bit of `https://visualgo.net/en/hashtable` in this tutorial that were only covered on Tuesday of Week 07 (more Hash Table on Week 08). Thus, we will do longer a lab component today (warm-up PE...).

2 Tutorial 05 Questions

Midterm Quiz Solutions + Review

Q1). See separate solution folder: `CS2040C-2021-22-S2-midterm-hard.zip` at LumiNUS Files.
TA will do one quick presentation of the solutions for all 3 questions of Midterm Quiz,
then open a 5-10m AMA (Ask Me Anything) session about that Quiz to give closer to all.

Hash Function Basics

Q2). Which of the following is the best (string) hash function?

1. `int index = (rand() * (key[0]-'A')) % N;`
2. `int index = (key[0]-'A') % N;`
3. `int index = hash_function(key) % N;`

where

- `rand()` is a function that returns a pseudo-random integral number in the range between 0 and `RAND_MAX` (This value is library-dependent, but is guaranteed to be at least 32767 on any standard library implementation).
- `key` is a C++ std::string
- `N` is the hash table size, usually a prime number
- `hash_function(v)` is as shown in `https://visualgo.net/en/hashtable?slide=4-7`

Q3). A good hash function is essential for good Hash Table performance. A good hash function is easy/efficient to compute and will evenly distribute the possible keys. Comment on the flaw (if any) of the following (integer) hash functions. Assume that for this question, the load factor α = number of keys N / Hash Table size $M = 0.3$ (i.e., low enough) for all cases below:

1. $M = 100$. The keys are positive even integers. The hash function is h(key) = key % 100.
2. $M = 100$. The keys are non-negative integers in the range of [0, 10 000]. The hash function is h(key) = floor(sqrt(key)) % 100.
3. $M = 101$. The keys are integers in the range of [0, 1 000]. The hash function is h(key) = floor(key * random) % 101, where $0.0 \leq$ random ≤ 1.0.

Hands-on 5

TA will run the (slightly longer) second half of this session with a few to do list:

- PS3 Debrief,
- Finally, live solve TWO chosen Kattis problem involving material from the **first half** of CS2040C (please treat this as a warm-up exercise for the upcoming Practical Exam (PE) that will be harder than this)

Problem Set 4

We will end the tutorial with **short algorithmic** discussion of PS4.
As we still have Week 08 before PS4 is due, then TAs are not supposed to reveal the algorithmic ideas of the near 100+100 solutions publicly (yet).