

DSEI1020 Fall 2018, Intro to Data Science, Michael Grossberg

HW2:P4 Reviewing Data Structures on Safari

In this assignment you will be reviewing the kind of data Structures we did in class. You will be asked to implement something in a later problem so make sure you are comfortable. For now this assignment will require you to make 2 good multiple choice questions for all the sections I am assigning you. For this assignment you will need the student membership of the acm.org (\$20/yr) and using that get to Oreilly Safari through the acm learning center.

Why learn this?

Reason 1: Choosing the right data structures are important both for data engineering and data science.

In CS/Engineering/Life we always have constraints. Time is limited, space is limited, computational cost can be an issue (particularly when you pay for it in the cloud). Choosing the right data structure can sometimes help or even sometime move something from impossible to easy.

Reason 2: Industry contacts are telling us that they are finding that many people interviewing for data scientists positions don't know data structures. Data structures questions are easy to ask and so they are assuming as technical people you can answer them:

- <https://www.geeksforgeeks.org/commonly-asked-data-structure-interview-questions-set-1>
- <https://medium.freecodecamp.org/the-top-data-structures-you-should-know-for-your-next-coding-interview-36af0831f5e3>

Assignment

Please go to Oreilly Safari though the learning center of the acm. This requires an acm.org student membership (\$20/yr) but not the very expensive Safari membership. This is a great deal. Please watch the sections

- Introduction
- Fundamentals
- Ubiquitous Lists
- Pointer Structures
- Recursive Structures (through Traversal)

in the course "Designing Data Structures in Python" by George T. Heineman

For each subsection (all 17) up to and including Traversal create a multiple choice question about that subsection.

Guide on the questions

Each of the question 17 should cover an important idea or topic in that section. Each of the multiple choice questions should be a plausible answer. One of the 4 should be a closely related but not-quite-right answer. Indicate the correct answer. It should go without saying that your questions should be original and **not** something I can find by searching web (eg. google) or from looking at a classmate.

Example:

Section: Detect Cycles in a Link List

What is the problem with a cycle in a link list?

- a: it is difficult to insert a new node
- b: it is difficult to delete a new node
- C: iterating through the list goes on forever <- correct
- d: data is over-written <- sort of correct