INTERVIEW PR3P

•••

~kryddaður~

Newsflash

- Apple Confirms it is a Google Cloud Customer (link)
- Six Degrees of Wikipedia (link)

Some terms

- In computer science, we are concerned with the efficiency in which we use resources, which are oftentimes scarce.
  - Time complexity: resource = time
  - Space complexity: resource = memory, disk space (includes call stack)
- O(X) means the operation you are conducting takes up, *at most*, c * X resources, where c is some constant number (whereas X can be asymptotically yuge)

- O(1): Constant, not dependent on size of input data (n)
- O(n): Linear, every time n doubles, resource usage doubles
- $O(\log_k n)$: Logarithmic, generally involves some kind of partitioning of the data
- $O(n^2)$: Quadratic, every time n doubles, resource usage quadruples
- $O(n^k)$ for some constant k: Polynomial
- $O(k^n)$ for some constant k: Exponential, every time n increases by 1, resource usage **is multiplied by k**, *avoid at all costs*

Collections

Arrays

- Most rudimentary collection data structure
- Association between key-value pairs, where keys must be integers

- O(1) indexed element access
- O(log n) search **if sorted**, O(n) otherwise
- O(n) insertion/deletion

- O(n) space

Pros:

- Good cache locality
- Easy to make, minimal additional information other than your data

Cons:

- Any unused indices between 0 - [max index] is wasted space
- Layout does not give additional information about the contents of the data

Linked Lists

- No indexed element access, instead they can be iterated through by following pointers between them.
- Can augment pointing structure
  - e.g. doubly-linked, circular, etc.

- O(n) search
- O(1) insertion/deletion

- O(n) space

Pro:

- Simple insertion and deletion

Con:

- Pointers is small space overhead
- Can’t perform binary search

Hash Sets & Tables

- Map any arbitrary *key* to some associated *value*
- Values are transformed into a key, which is used to index and find given value
- Generally ok to take their functionality for granted during coding problem, but be prepared to explain its inner workings
    - Probing, hash function, load factor, etc.

- O(1) insertion/deletion
- O(1) search *if searching by key*

- O(n) space

Pros:

- Most languages provide an easy-to-use built-in hash set/table
- Easy to use

Cons:

- Generally tricky to *implement* right
    - Delete is hard
- Many assumptions need to be met for time complexity to be true
- Higher overhead per-element (not good for small collections)

Practice Problems

1. Write a function that takes an input string and returns the minimum substring which contains every letter of the alphabet set at least once.
2. Given two arrays representing two individuals' rankings of restaurants, find the mutually highest ranked restaurant, or return None if no such restaurant is found.
3. You are given an int array of size n with unique values ranging from 0 to (n - 1). Suppose one element is replaced with a -1. Find the value of the original element. Explore ways to do this in-place, with an extra array, and an extra hash table.

Recursive things

Trees

- Every “node” has information, but also points to other nodes in an acyclical manner
- Binary Search Trees are most common, in which every “parent” node is greater than its left “child” and less than its right “child”
- BSTs can be “augmented” to ensure balanced-ness
  - Red-black
  - AVL
  - Treap

Pros:

- Lots of flexibility by being able to augment nodes
- Keep searching as fast as with arrays, but with less maintenance work

Cons:

- Rotations can be a pain

- O(log n) insert/search/delete **if balanced**, O(n) otherwise
- O(n) space

Recursion

- Call a function within the function itself.
- Helpful on problems that require solving a smaller problem first.

Pro:

- Work on complex, multi-part problems without taking up memory space
- Often a really nice way to model problems

Con:

- Can cause a function call “stack overflow"
- Slow if not tail recursion

Stacks & Queues

- Can only access the first and/or last elements

- O(1) insertion/deletion
- Generally can't iterate/search

Pro:

- Can simulate recursion without fear of stack overflow

Con:

- Searching is expensive and needs external structure to avoid deleting data

Heaps

- Tree structure, but with different relationship rules
  - Parents are greater (or smaller) than *both* children
- Has a sorting algorithm based off of it

- O(1) find min
- O(log n) insertion/deletion/search
- O(n) build

Pro:

- Obtain minimum super fast
- Can be represented through special indexing on an array

Con:

- Occasionally too complicated for the problem at hand
- Better get upHeap() right!

Practice Problems

1. Find the nth Fibonacci number, once using recursion, again with a stack or queue.
2. Given pointers to two nodes of a BST (nodes have parent pointers but you are not given a root), find the common ancestor (if it exists).
3. Given a two dimensional array of pixels, and a pixel that the user clicked on, implement a “bucket-filling” operation.
4. Print a binary tree in “zig-zag” order, in which the tree is printed level-by-level, where odd levels are read left to right and even levels are right to left.

Graphs

- Great for modeling relationships between entities
- Have nodes, edges
- Edges can be undirected, directed
- Edges can be weighted
- Can have cycles
- Many representations
  - Edge list
  - Adjacency matrix
  - Adjacency lists
- BFS, DFS

Practice Problems

1. Given a directed graph, design an algorithm to find out whether there is a route between two nodes.
2. Given a function that returns a list of URLs that can be visited from a given URL, write a function that returns all distinct pages on a website.
3. You are given a function that returns all friends of a given user id. We define a user as “recommendable” to another user if they are second degree connections. Write a function that returns an array of user id’s recommendable to a given user id, sorted by the number of mutual friends the two have.

Newsflash

- Spotify files F-1 (link)
- Hangouts Chat, Google's Slack competitor, comes out of beta (link)
- Discord shutters alt-right and Nazi servers, but they're not completely gone (link)

Data layout & search

Sorting

- Lower bound of O(n log n) *for comparison-based sorts*
- Some freaky shit happens with radix sort (not comparison-based) and it ends up being O(dn)
- Good to know when to use bubble, merge, quick, insertion sorts & their respective time complexities
    - Also be prepared to implement them with a certain constraint in place

- What is the space complexity of a traditional merge sort?

Searching

- How can you best lay out your data such that it is searchable in some key way?
- Often a trade-off between initial build time and search time

- What properties do you need to query quickly and often? What data structure do you use? How can you use the relations in the data to store less?

Practice problems

Given a log like:

T1 requested A @ 1

T2 requested B @ 3

T1 requested C @ 4

T1 freed A @ 6

Write a function that can determine who owned a given server S at time H

How fast can you do this?

How do you index your data?

- What is the worst/best case time complexity of quick sort? Under what conditions will either occur?
- You are given an array of N elements and a function f that can in O(1) time place an element into one of constant K categories. How fast can you stratify the list into the categories? In place? (DNF)

Logic

- Doesn’t involve programming but similar thought process
- Usually best solved by trial-and-error

Design

- Usually involves designing some kind of class or struct that needs to handle certain operations with constraints
  - Brush up on your knowledge on classes
- Specification may be vague so *be sure* to ask many clarification questions
- Think about your operation implementation before finalizing what attributes your class will need.

- *What am I storing? What do I need to do?*

Practice Problems

- Design a data structure for Least Recently Used Cache. It should support the following operations:
  - get(key) - Get the value of the key if the key exists in the cache, otherwise return None
  - set(key,value) - Set or insert the value if the key is not already present. When the cache reached its capacity, it should invalidate the least recently used item before inserting a new item.
- Design a URL shortening service
- Design a type-ahead search tool (autocomplete, dropdown, etc)

- Given a village of n people where you can identify each person by some number i < n, you are looking for a person eligible to be president. To be eligible to be president, that person must fulfill two criteria:
  - That person must NOT know anybody else in the village
  - Every person in the village MUST know that person.
- You can ask whether person A "knows" person B or not (but doesn't present any information for B knowing A). What is the minimum times total you have to ask people before finding if the village has a person eligible to be president?