Data Structures and Algorithms

# Definitions

Definition of Data Structure

It organizes and stores data, they all differ in the way that they store the data.

Definition of Algorithm

It described the steps you must perform to accomplish a specific task.

Definition of Implementation

The code you write to perform the task, an algorithm can have many implementations.

# Big-O Notation

Definition

The Big-O Notation gives a way of comparing the time and space complexity of different algorithms in an objective manner. It measures the worst-case time or space complexity relative to the input.

* **O (1) – Constant**
* **O (log2n) – Logarithmic**
* **O (n) – Linear**
* **O (nlog2n) – Linearithmic**
* **O (n2) – Quadratic**
* **O (2n) – Exponential**
* **O (n!) – Factorial**

By using Big-O Notation, developers can predict how algorithms will perform as the input size grows and select the most efficient algorithm based on scalability and resource use.

# Stable and Unstable Sort

Definition

A stable sort is when the algorithm preserves the relative ordering of duplicate items, an unstable sort does not preserve the relative order.

# Arrays

Definition

They are not a dynamic data structure, which means once you create an array you cannot change its size, it is a continuous block in memory. Every element occupies the same amount of space in memory.

If we know the index of an element, the time to retrieve the element will be the same.

Using the *formula* **x + i \* y** to reach the memory address of the **ith** element, the size of each element being **y** and starting at the memory address **x**.

For arrays:

**O (1) –** Retrieve with index

**O (n) –** Retrieve without index

**O (n) –** Add an element to a full array

**O (1) –** Add an element to the end of an array (has space)

**O (1) –** Insert or update an element at a specific index

**O (1) –** Delete an element by setting it to null

**O (n) –** Delete an element by shifting elements

The general rule:

If it doesn’t have a loop, it’s *Constant Time*

If it requires a single loop, it’s *Linear Time*

If it requires a loop inside of a loop, it’s *Quadratic Time*

# Bubble Sort

It’s an in-place and stable algorithm and it has a time complexity of **O (n2)**.

# Selection Sort

It’s an in-place and unstable algorithm and it has a time complexity of **O(n2)**, but does not require as much swapping as the *Bubble Sort*.

# Insertion Sort

It’s an in-place and stable algorithm and it has a time complexity of **O (n2)**

# Shell Sort

It’s an in-place and unstable algorithm and it has a time complexity of **O (n2)** but it can perform much better because of the gap, it doesn’t require as much shifting as insertion sort so it usually performs better and it doesn’t require as much swapping as bubble sort so it usually performs better, can utilize the Knuth Sequence to calculate the gap **(3k** **– 1) / 2** where k is the length of the array and the result should be a close value to the length of the array, without being greater than the array.

# Bucket Sort

It’s a not in place algorithm and its stability depends on the sort algorithm used to sort buckets (usually uses insertion sort). Uses hashing and makes assumptions about the data, achieving **O (n)**, but only if there’s only one item per bucket. It distributes the items into buckets based on their hashed values, sorts them inside each bucket and merges them.

# Heap Sort

We know the root has the largest value, swap root with last element in the array, heapify the tree but exclude the last node, after heapify, second largest element is at the root, then repeat.

# Recursion

Factorial Algorithm

A method is a recursive method when it calls itself

**Iterative** implementation usually runs faster and uses less memory than the **Recursive** one

Seach up*: Tail recursion*

# Merge Sort

It’s a not in-place and stable alrogithm and it has a time complexity of **O (nlog2n)**.It’s a Divide and Conquer algorithm, divided into two phases: Splitting (logical) and Merging. An unsorted array will be divided into two smaller arrays (extra goes right) until all the arrays have one element each – these are sorted. Every left/right pair of sibling arrays will be merged into a sorted array, then again until you have a single sorted array. The merging phase is not in-place, it uses temporary arrays.

# Quick Sort

It’s an in-place and unstable algorithm and it has a time complexity of **O (nlog2n)**.Same as Merge Sort it’s also a Divide and Conquer algorithm, it uses a pivot element to partition the array into two parts, smaller elements will go to the left and right elements will go right, pivot ends up being in sorted position, process is now repeated for left and right array and eventually every element has been the pivot.

# Counting Sort

It’s a not in-place algorithm and can be a stable algorithm with a few extra steps, the time complexity is **O (n)** since assumptions are being made, a linear time complexity can be achieved. An algorithm that makes assumptions about the data, doesn’t use comparison, only works with non-negative discrete values and counts the occurrence of each value which need to be within a specific range.

# Radix Sort

It’s an in-place and stable algorithm, it has a time complexity of **O (n)**, but often runs slower because of the overheac. It’s an algorithm that makes assumptions about the data, and it has to have the same width and radix, starts at the rightmost position and must use a stable sort algorithm at each stage.

# Lists

Definition

It’s an ordered collection of elements that allows duplicates and permits elements to be inserted or accessed by their position (index) in the Iist.

ArrayList

Dynamic array implementation of the List Interface. Provides fast retrieval but slower insertion and deletion in comparison to LinkedList

ArrayList<String> arrayList = new ArrayList<>();

LinkedList

Each item in the list is called a node, the first item in the list is the head of the list.

If you have a singly (one link between nodes) linked list you should insert and delete items to the front of the list to maintain **O (1)**, otherwise you have to traverse the array

If you have a doubly linked list, the list has a head and tail and for them the time complexity is **O (1)** but for inbetween nodes time is **O (n)**. Allows for fast insertion and deletion at the cost of slower random access.

LinkedList<Integer> linkedList = new LinkedList<>();

Vector

Similar to ArrayList but synchronized, making it thread-safe, often considered legacy class

Vector<Double> vector = new Vector<>();

# Abstract Data Type

Definition

It doesn’t dictate how the data is organized but it does dictate the operations you can perform, abstract data types are usually an interface.

# Stacks

They are an abstract data type, LIFO – last in, first out so there is no random access. The item at the top of the stack is the only one we’re allowed to access. Operations are *push/pop/peek.* The ideal data structure to back a stack is linked list. The time complexity will be **O (1)** when using a linked list and **O (n)** for an array (should be used when either memory is tight or if you know the maximum number of items that will ever be on the stack).

Stack<Integer> stack = new Stack<>();

# Queues

They are an abstract data type, FIFO – first in, first out, item is added at the end of the queue, remove or peek the item at the front of the queue. Time complexity depends on what is the back of the interface.

A **Deque (Deck)** is a double ended queue, supports insertion and removal at both ends.

Queue<String> queue = new LinkedList<>();

# Hashtables

Also an abstract data type, provides access to data using keys and consists of pais key/value pair, it is optimized for retrieval (when you know the key). The **load factor** tells how full a hash table is, it’s used to decide when to resize the array backing the hash table, can’t be too low (empty space) and can’t be too full (increase likelihood of collisions).

Map<String, Employee> hashMap = new HashMap<String, Employee>();

# Trees

It’s a hierarchical data structure in which every item is a node, the node at the top of the tree is the root, every non-root node has one and only one parent, a leaf node has no children and a singleton tree has only one node – the root.

Binary Tree

Every node has 0, 1 or 2 children, being refered to as left or right child.

Binary Search Tree (BST)

Can perform insertions, deletions, and retrievals in **O (log2­­­n)** time. The left child always has a smaller value than its parent, the right child always a higher value, because of that Binary Search is possible.

TreeSet<Integer> treeSet = new TreeSet<>();

# Heaps

A complete binary tree that satisfies the heap property (which depends whether it’s a Max or Min Heap, with a *Max Heap*: every parent is greater than or equal to its children, with a *Min Heap*: every parent is less than or equal to its children. The maximum or minimum value will always be at the root of the tree.

Heapify is the process of converting a binary tree into a heap. Usually implemented as arrays.

Heaps as Arrays

To store a heap as an array, the root is put at array[0], traverse each level from left to right

So for the node at array[i]:

left child = **2i + 1**

right child = **2i + 2**

parent = **floor((i – 1) / 2)**

To insert into Heap, always add new items to the end of the array, then heapify, compare the new item against its parent, if the item is greater than its parent, we swap it with its parent, then repeat.

Priority Queues

Not like stacks or queues, the item with highest priority gets removed first, so it’s not a FIFO. It’s basically a max-heap

PriorityQueue<Integer> pq = new PriorityQueue<Integer>();