Computational Thinking with Algorithms: Project 2021

## Eoin Lees – G

## Higher Diploma in data

# Introduction

Concept of sorting and sorting algorithms

Sorting is the action of “arranging a collection of items according to some pre-defined ordering rules.” (Mannion, 2021)

In computing it is an important concept as a large porting of computing work is spent on sorting. It has therefore since the beginning of modern computing been an important filed of study for optimization. By sorting information in advance, many tasks can be simplified greatly and take up less computational power. It is often an important part of other algorithms too, so its optimization to a specific task is important ??waffle

Sorting algorithms in computing began in 1945 with the development of merge sort by John von Neumann. Multiple other sorting methods were also developed in the early days of computing e.g Radix sort (1954), Counting sort (1954), Quick sort (1962) etc.

### Comparator Functions

Takes two elements of any type and compares one to the other.

Which input, a or b should appear first according to predefined definition of comparator function.

E.g < = or >

This project is focused on integer sorting. So greater than or less than will be used.

It is possible to sort strings using lexicographical methods. i.e how letters appear in alphabet, or items appear in dictionary.

### Comparison-based-sorts

“A comparison sort is a type of sorting algorithm which uses comparison operations only to determine which of two elements should appear first in a sorted list.” (Mannion, 2021)

A sorting algorithm is called ***comparison-based*** if the only way to gain information about the total order is by comparing a pair of elements at a time via the order ≤. (Mannion, 2021)

In comparison-based sorting no algorithm that sorts by this method can do better than n log n performance in the average or worst cases. This is reflected in table 1.

There are many sorting algorithms that are in this category (e.g. Bubble Sort, Insertion Sort, Selection Sort, Merge Sort, Quicksort, Heapsort).

They are the most widely applicable sorts for a diverse type of input data.

There are no assumptions made about the data. They compare all elements against each other.

O(n log n) is the ideal worst case scenario for a comparison-based sort.

### Non-comparison-based-sorts

“Under some special conditions relating to the values to be sorted, it is possible to design other kinds of non-comparison sorting algorithms that have better worst-case times (e.g. Bucket Sort, Counting Sort, Radix Sort)” (Mannion, 2021)

These methods can achieve better “worst case” times as shows in the table below however they are not as stable.

O(n) time is possible if assumptions are made about the data. This linear result comes from not having to compare every element against each other.

Examples include: Counting sort, Bucket sort and radix sort.

O(n) is the minimum sorting time possible since each element must be examined at least once.

### Stability

Stability in a sorting algorithm as it is important to preserve the order of already sorted data. This is a desirable property in sorting algorithms.

### In-place sorting

Sorting algorithms have different memory requirements, which depend on how the specific algorithm works.

A sorting algorithm is called in place if it uses only a fixed additional amount of working space, independent of the input size.

Other sorting algorithms may require additional working memory, the amount of which is often related to the size of the input n

In place sorting is a desirable property if the availability of memory is a concern (Mannion, 2021)

Rewrite all this in my words

### complexity (time and space)

Write about it here… find notes

### Performance

Table shows run time effeciancy

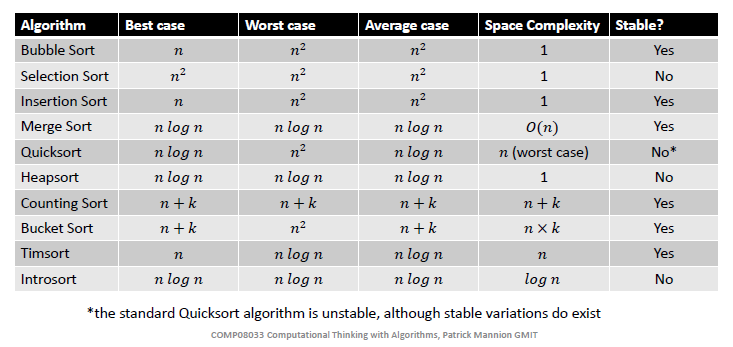


Table : Overview of sorting algorithms

Show graph from notes

# Sorting Algorithms

## Bubble Sort

Bubble sort takes its name from having larger values in a list “bubble up” to the end of the list as the sorting takes place. It was first analysed formally as early as 1956. It is a comparison-based sorting method.

It has a time complexity of n (linear) in best case and n2 (quadratic) in the worst and average cases. (table 1)

It is an in-place sorting algorithm. It uses only the size of the input data set plus a constant amount of additional working space to function. The additional working space is just a few variables required to make the swaps of the elements in different places. A temporary variable is used here.

It is a slow sorting method. It is impractical for most problems. It only really works well on small data sets, or on data that is already nearly sorted.

How it works:

* Compare each element to the element on it right (exclude last element)
  + If out of order, swap. Then repeat to next element.
  + Largest element ends up at end of list in its final position.
* Repeat process (exclude last two elements)
  + If out of order, swap. Then repeat to next element.
  + Second largest element ends up in its final position.
* Continue until there are no more remaining unsorted elements on the left.

The algorithm assumes every element on the left is unsorted, and the right is sorted. It then has two separate arrays of data. Sorted and un-sorted.

(Diagram of explination) \* Save until tomorrow

Different input instances – graph made already – single out each

## Merge Sort

Efficient comparison sort

Proposed by John von Neumann in 1945. Is a recursive algorithm father than using for loops or while loops.

Its worst case running time is 0(n log n). Its best and average cases are also n log n.

It is a good all-around performance algorithm. It is a stable algorithm.

There are versions of merge sort that are particurarly good for sorting data with slow access times, e.g. data that cannot be held in RAM.

Discuss space and time complexity

Merge sort must take the least time that is linear in the total size of the two lists in the worst case. Every element must be looked at so the correct order can be determined.

How it works:

* When called
  + If size of input is 0 or 1, return
  + Otherwise separate input array into two arrays approx.. equal in size
* Apply merge sort to left and right sides.
  + Repeat until every element is 0 or 1.
* Call merge
  + Positions swap relevant to one another in sorted position.
  + Combine sub arrays
* Call merge
  + Positions swap relevant to one another in sorted position.
  + Combine sub arrays
* Sorted array achieved.

(Diagram of explination) \* Save until tomorrow

Different input instances – graph made already – single out each

## Counting Sort

A non-comparison-based sort. It was proposed by Harold H. Seaward in 1954, who also proposed radix sort.

It allows a collection of items to be sorted in close to linear time.

Certain assumptions are made in order to achieve this. These assumptions are also limitations on the type of input data the sorting algorithm can handle.

The assumption with counting sort is: assuming an input of n, where each item is a non-negative integer key, with a range of k. If using zero-indexing, the keys are in the range [0,…,k-1].

The beast, worst and average case time complexity is n + k, the space complexity is also n + k.

The running time advantage comes at the cost o the algorithm not being as widely applicable as comparison sorts.

It is a stable algorithm if used correctly.

How it works:

* The key range k, of the input array, is determined
  + Ini
* Step 2
  + a

(Diagram of explination) \* Save until tomorrow

Different input instances – graph made already – single out each

<https://www.youtube.com/watch?v=OKd534EWcdk&ab_channel=CSDojo>

## Insertion Sort

A simple comparison-based sorting algorithm. It is similar to the method used usually by card players sorting cards. It is intuitive to understand. It is easy to implement, is a stable sorting algorithm, it works in-place (efficient with memory) and is very good for small and partially sorted lists.

In best case scenarios it runs in n + d time, where d is the number of inversions in the input instance.

It is very inefficient for large lists that are randomly ordered.

It is an iterative approach. It splits the list into sorted and unsorted sublists. ( head and tail)

Complexity:

The total number of data comparisons is the number of inversions d plus n-1. On average

How it works:

* Starting on the left of the array, the key is set as the element at index 1.
  + Any elements on the left greater than the key get moved right by one position and the key gets inserted.
* Set the key as the element at index 2.
  + Any elements on the left greater than the key get moved right by one position and the key gets inserted.
* Set the key as the element at index 3.
  + Any elements on the left greater than the key get moved right by one position and the key gets inserted.
* Continue until the key is set at n-1 index.
  + Any elements on the left greater than the key get moved right by one position and the key gets inserted.
* The array is sorted.

(Diagram of explination) \* Save until tomorrow

Different input instances – graph made already – single out each

## Quick Sort

One of the most important algorithms developed. It was developed in 1959 by C.A.R Hoare. It is a recursive/divide and conquer algorithm. It is on average one of the fastest known sorting algorithms in practice.

It is not a stable algorithm, however stable modifications do exist.

The performance is 𝑛2 in the worst case, however this is rare. The average and best cases are n log n.

If the pivot chosen is poor then the n2 performance may happen. E.g if the first or last element is chosen in an array where the data is already nearly sorted.

The median element is usually the best choice, as it makes it so that a fairly equal distribution will occur in the left and right arrays.

Its memory usage is O(n) – linear memory usage in standard usage. Some variations exist with O(n log n) memory performance.

How it works:

* Pick an element from the array – the “pivot”
* Partitioning – takes every array element that has a value less than pivot and positions them to the left of the pivot. It also takes every array element greater than the pivot value and positions them to the right. The pivot is now in its final position with two subarrays on either side.
* Recursion – the above steps are then applied to each sub array. It is repeated until a base case is reached: a subarray of length 1 or 0.

(Diagram of explination) \* Save until tomorrow

Different input instances – graph made already – single out each

# Implementation and Benchmarking

Describe process followed when implementing the application

Present results:

Table of results

Results graphed against each other

Discuss how measured results differed

Were they similar to what you expected - Given time complexity of each chosen algorithm

Table and graph labelled appropriately

# References

Mannion, P., 2021. Sorting Algorithms, Part 1. In: *COMP08033 Computational Thinking with Algorithms,.* s.l.:GMIT.

Wikipedia, 2021. *Wikipedia.* [Online]   
Available at: https://en.wikipedia.org/wiki/Sorting  
[Accessed 12 May 2021].