**Course: Computational Thinking and Algorithms**

**Student Name: Niamh O’Leary**

**Student ID: G00376339**

**Project: Benchmarking Sorting Algorithms**

1. **Introduction**

“Numerous computations and tasks become simple by properly sorting information in advance. The search for efficient sorting algorithms dominated the early days of computing. Indeed, much of the early research in algorithms focused on sorting collections of data that were too large for the computers of the day to store in memory. Because today’s computers are so much more powerful than the ones of 50 years ago, the size of the data sets being processed is now on the order of terabytes of information. Although you may not be called on to sort such huge data sets, you will likely need to sort large numbers of items. In this chapter, we cover the most important sorting algorithms and present results from our benchmarks to help you select the best sorting algorithm to use in each situation.”[[1]](#footnote-1)

In computer science, a sorting algorithm is an algorithm that puts elements of a list in a particular order. The most frequently used orders are numerical and lexicographical orders. Sorting is very useful for canonicalising data and for producing human-readable output.

Essentially sorting arranges data in a sequence which makes searching easier, for example in ascending or descending order. As humans realised the importance of searching quickly, the need for efficient sorting became more sought after.[[2]](#footnote-2) Therefore sorting was the most fundamental algorithmic problem that was faced in the early days of computing. The development of sorting algorithm helped specify the way to arrange data in a particular order. Most common orders are in numerical or lexicographical order. The importance of sorting lies in the fact that data searching can be optimised to a very high level, if data is stored in a sorted manner. Sorting is also used to represent data in more readable formats. Following are some of the examples of sorting in real-life scenarios −

* Telephone Directory − The telephone directory stores the telephone numbers of people sorted by their names , so that the names can be searched easily.
* Dictionary − The dictionary stores words in an alphabetical order so that searching of any word becomes easy.[[3]](#footnote-3)

Since the explosion of modern technologies, computer algorithms have expanded and can be found in nearly every aspect of life, hence he need to find the most efficient method of sorting.

* 1. **Conditions for sorting**
* A collection of items is deemed to be “sorted” if each item in the collection is less than or equal to its successor
* To sort a collection A, the elements of A must be reorganised such that if A[i] < A[j], then i < j
* If there are duplicate elements, these elements must be contiguous in the resulting ordered collection – i.e. if A[i] = A[j] in a sorted collection, then there can be no k such that i < k < j and A[i] ≠ A[k].
* The sorted collection A must be a permutation of the elements that originally formed A (i.e. the contents of the collection must be the same before and after sorting)

* 1. **Complexity**

The complexity of an algorithm is a function describing the efficiency of the algorithm in terms of the amount of data the algorithm must process. There are two main complexity measures of the efficiency of an algorithm:

**Space complexity** is a function describing the amount of memory (space) an algorithm takes in terms of the amount of input to the algorithm. When we say “this algorithm takes constant extra space,” because the amount of extra memory needed doesn’t vary with the number of items processed.[[4]](#endnote-1)

* Space Complexity is a function describing the total space used by the algorithm to compete its execution in terms of input size and auxiliary space.

***Space Complexity = Auxiliary Space + Input Size***

* Auxiliary space refers to the extra or temporary space that is required by the algorithm during its execution.
* An algorithm requires space during execution for the following for 3 reasons:
  + Instruction Space refers to the amount of memory used to save the compiled version of instructions. The space is fixed but varies depending on the number of lines of code in the program.
  + Environmental Stack- At times an algorithm(function) may be called inside another algorithm(function). If this occurs, the current variables are pushed onto the system stack, where they wait for further execution and then the call to the inside algorithm (function) is made. Essentially it is the space needed to store the environment information to resume the suspended function.
  + Data Space - is the amount of space needed to store all the variables and constants values.

Note while calculating Space Complexity of an algorithm normally Data Space is used while the others (Instruction Space and Environmental Stack) are ignored.[[5]](#footnote-4)

**Time complexity** is a function describing the amount of time an algorithm takes in terms of the amount of input to the algorithm. In layman’s terms, We can say time complexity is sum of number of times each statement gets executed.

* 1. **Asymptotic Notations**

Asymptotic Notations are languages that allow us to analyse an algorithm’s running time by identifying its behaviour as the input size for the algorithm increases. This is also known as an algorithm’s growth rate.[[6]](#footnote-5)

The 3 asymptotic notations that will be discussed in terms of time complexity:

**Big Oh (O) – Upper Bound**

Big O represents the maximum amount of resources in terms of space and time that an algorithm would require to run. It provides an upper bound on the number of operations an algorithm uses to solve a problem with input of a particular size. Describes the complexity of an algorithm in the worst case. It tells us that a function grows slower or at least as slow as another.

**Big Omega (Ω) – Lower Bound**

Omega represents the minimum amount of resources in terms of space and time that an algorithm would require to run. It describes the complexity of an algorithm in the best case. It tells us that a function grows at least as fast as another.

**Big Theta (Θ)- Average Bound**

Theta represents the functions that lie in both O and Ω expression. It describes the complexity of an algorithm in the average case. It is usually much more difficult to determine the average case time complexity of an algorithm. It represents the average number of operations an algorithm uses to solve a problem over all inputs of a particular size.[[7]](#footnote-6)

There is a strong correlation between run time of a sorting algorithm and the number of inversions in the input array, where the inversion number relates to one measure of how far it is from being sorted.

The different sorting algorithms are a perfect showcase of how algorithm design can have such a strong effect on program complexity, speed, and efficiency.[[8]](#footnote-7)

When identifying the complexity of an algorithm, it is important to identify the most expensive computation within an algorithm to determine its classification. The overall performance of the algorithm must be classified as quadratic. Figure 1., shows the typically, algorithms complexity of a number of polynomial and exponential algorithms. The most popular metric for calculating time complexity is Big O notation. Time complexity is measured in terms of the number of operations an algorithm uses.



It is important to consider the characteristics of the data type in the input as well as the size of the input. An algorithm which takes an array as an input and returns the first item in the array will run in constant time (denoted by O(1)), regardless of the size of the input, it will always take the same amount of time to run. An algorithm which takes an input array and loops through each item to find the sum of the items in the array will have a run time that varies in direct proportion to the size of the input data. This is denoted by O(n). Finally, an algorithm that uses a nested loop to determine if a dataset contains duplicates can have complexity that varies in proportion to the square on the size of the input. This is denoted by O(n2).

**1.4 Desirable properties for a sorting algorithm:**

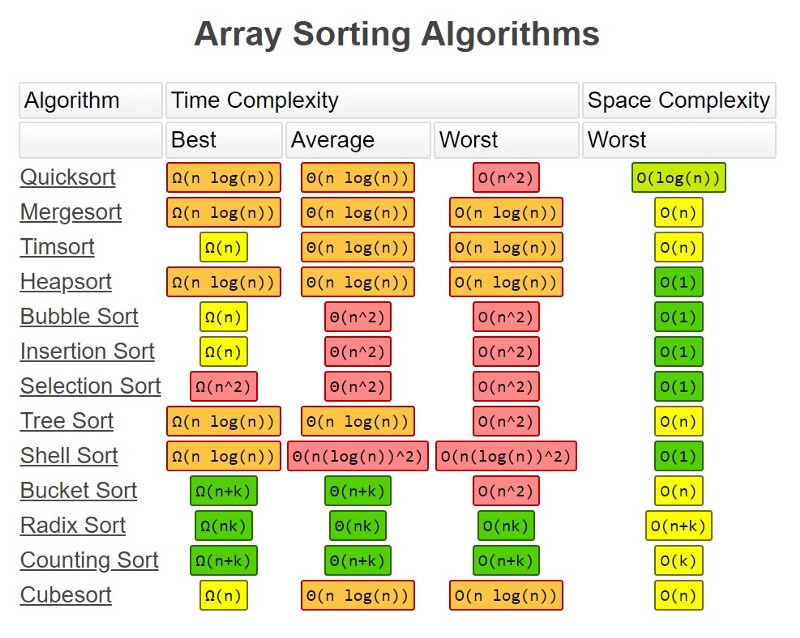
1. Stable sorting algorithms retains the relative order of elements with equal keys. Using an unstable sorting algorithm changes the relative order of elements with equal keys.[[9]](#footnote-8)



1. In-place sorting (Internal sorting): uses only a fixed additional amount of working space, independent of the input size. Algorithms within this category: Selection Sort, Insertion Sort, Bubble Sort, Quick Sort. Not-in place (External Sorting): algorithms that require additional working memory, the amount of which is often related to the input size. Algorithms within this category: Merge Sort. This is a desirable property if you are concerned about the availability of space.
2. Run time efficiency (Best, Average and Worst case)
3. Suitability – properties of sorting algorithm are well matched to the class of input instances which are expected (determine the advantages and disadvantages of each algorithm when choosing them)

**1.5 Performance**

While complexity can be considered the theoretical measure of algorithm efficiency, performance can be seen as the practical measure. Performance of an algorithm is measured by implementing the algorithm. The amount of time, disk space and memory consumed when a program is run. Performance can be affected by the computer specification, the compiler used to run the code and the efficiency of the code itself.



The figure above shows a summary if the performance of different algorithms.

In general cases, we mainly used to measure and compare the worst-case theoretical running time complexities of algorithms for the performance analysis. The fastest possible running time for any algorithm is O(1), commonly referred to as Constant Running Time. In this case, the algorithm always takes the same amount of time to execute, regardless of the input size. This is the ideal runtime for an algorithm, but it’s rarely achievable. In actual cases, the performance (Runtime) of an algorithm depends on n, that is the size of the input or the number of operations is required for each input item.[[10]](#footnote-9)

The algorithms can be classified as follows from the best-to-worst performance

* A logarithmic algorithm – O(logn)- Runtime grows logarithmically in proportion to n.
* A linear algorithm – O(n) - Runtime grows directly in proportion to n.
* A superlinear algorithm – O(nlogn) - Runtime grows in proportion to n.
* A polynomial algorithm – O(nc) - Runtime grows quicker than previous all based on n.
* An exponential algorithm – O(cn) - Runtime grows even faster than polynomial algorithm based on n.
* A factorial algorithm – O(n!) - Runtime grows the fastest and becomes quickly unusable for even small values of n.

Therefore the ideal sorting algorithm would have the following properties:

* Stable: Equal keys aren’t reordered.
* Operates in place, requiring O(1) extra space.
* Worst-case O(n·lg(n)) key comparisons.
* Worst-case O(n) swaps.
* Adaptive: Speeds up to O(n) when data is nearly sorted or when there are few unique keys.

1. **Sorting Algorithms**

As mentioned earlier sorting can be defined as arranging a collection of items according to some pre-defined ordering rules.[[11]](#footnote-10) Sorting algorithms should have the following properties;

* Stability
* Good run time efficiency
* In place sorting
* Suitability

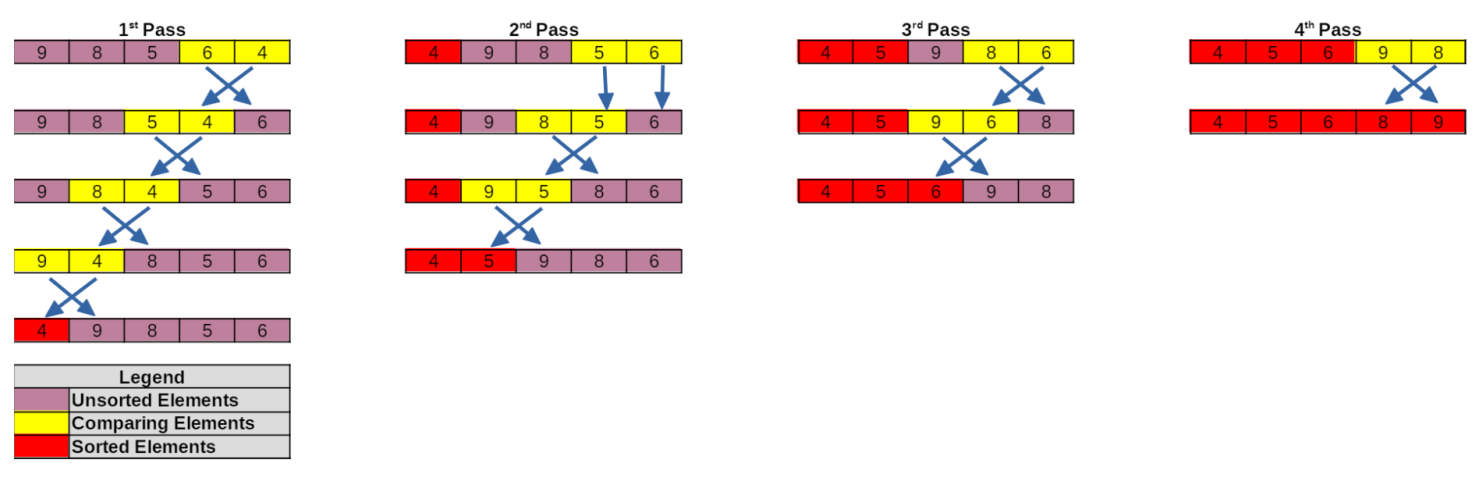
It is also important to analyse each

* 1. **Bubble Sort (a simple comparison based algorithm)[[12]](#footnote-11)**

Bubble sort was first analysed in 1956. It is a comparison based algorithm, in that it uses comparison operations only to determine which of the two elements should appear first in a sorted list. It keeps repeatedly swaps the adjacent element if they are in the wrong order.

If there are n items in the list, then there are n−1n−1 pairs of items that need to be compared on the first pass. It is important to note that once the largest value in the list is part of a pair, it will continually be moved along until the pass is complete. At the start of the second pass, the largest value is now in place. There are n−1n−1 items left to sort, meaning that there will be n−2n−2 pairs. Since each pass places the next largest value in place, the total number of passes necessary will be n−1n−1. After completing the n−1n−1 passes, the smallest item must be in the correct position with no further processing required.

Sorting takes place by stepping through all the elements one by one and comparing with the adjacent element and swapping them if required. This algorithm gets its name from the way the larger values in a list “bubble up” to the end as the sorting takes place.



Implementation of the above diagram:

First Pass:

* Start at the last two elements in the array (6 and 4), compare them and swap them.
* Move to the next two elements (4 and 5), compare them and swap them.
* Move to the next two elements (8 and 4), compare them and swap them.
* Move on to the first two elements (4 and 9), compare them and swap them.

Second pass:

* Compare the last two elements in the array (5 and 6) there is no need to swap them this time. We then
* move to the next two elements in the array (8 and 5), compare them and swap them. Compare the next 2 elements (9 and 5) and swap them.
* At the end of this pass the first 2 elements are in their correct positions.

Third pass:

* Compare the last 2 elements (8 and 6) and swap them.
* Cove onto the next 2 elements (9 and 6) compare them and swap them (the first 3 elements in the array are sorted)
* final run through the array, the last two elements (9 and 8) are compared and swapped.

The array is now sorted in ascending order

The Python code to implement this algorithm is shown below. There are two loops, the outer which runs from the second last item to the first in increments of one, and an inner loop, which from the first item in the list to the item in the list being currently run by the outer loop.

def bubbleSort(alist):

    for passnum in range(len(alist)-1,0,-1):

        for i in range(passnum):

            if alist[i]>alist[i+1]:

                temp = alist[i]

                alist[i] = alist[i+1]

                alist[i+1] = temp

alist = [54,26,93,17,77,31,44,55,20]

bubbleSort(alist)

print(alist)

Output:

[17, 20, 26, 31, 44, 54, 55, 77, 93]

The Python code uses a nested loop to run the comparisons and sorts. It is possible to perform simultaneous assignment. The statement a,b=b,a will result in two assignment statements being done at the same time. Using simultaneous assignment, the exchange operation can be done in one statement.[[13]](#footnote-12)

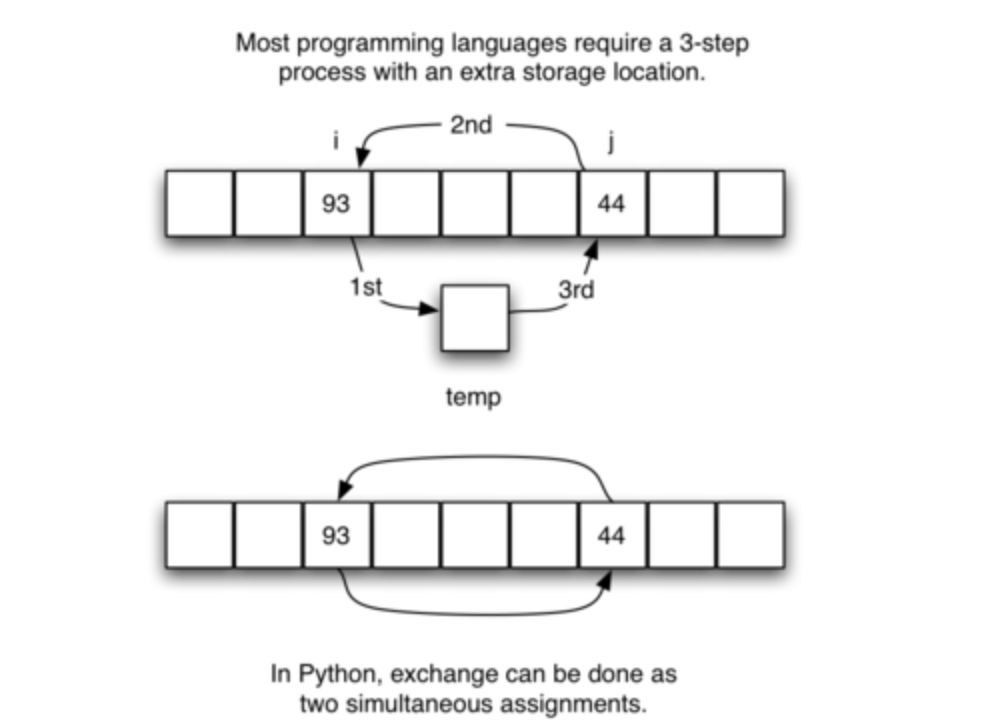


Figure 4. Exchanging Two Values in Python

* + 1. **Complexity of Bubble Sort**

Run time of Bubble Sort is very much dependent on the given input. If the given numbers are sorted, this algorithm runs in *O(n)* time. If the given numbers are in reverse order, the algorithm runs in *O(n2)* time

* **Worst and Average Case Time Complexity:**O(n\*n). Worst case occurs when array is reverse sorted.
* **Best Case Time Complexity:** O(n). Best case occurs when array is already sorted.
* **Auxiliary Space:** O(1)
* **Boundary Cases:** Bubble sort takes minimum time (Order of n) when elements are already sorted.
* **Sorting In Place:**Yes
* **Stable:** Yes

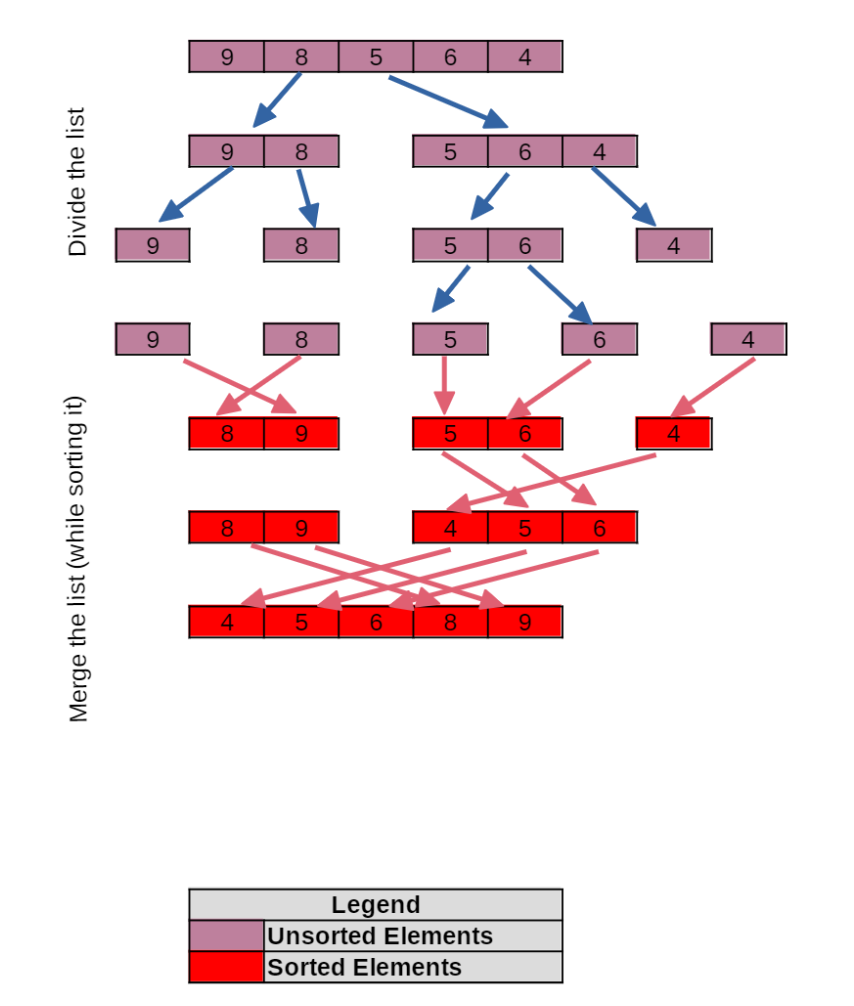
**Bubble Sort is easy to understand and therefore it is often use as a teaching tool.** It is simple to write, easy to understand and it only takes a few lines of code. The data is sorted in place so there is little memory overhead and, once sorted, the data is in memory, ready for processing.[[14]](#footnote-13) Bubble sort, if used when a data set is very small, can be an efficient algorithm to use.

The main disadvantage is the amount of time the algorithm takes to sort. The average time increases almost exponentially as the number of elements increases. In ten times the number of items, takes almost one hundred times as long to sort. Therefore, it will not deal well with large sets of data.

* 1. **Merge Sort (an efficient comparison based algorithm) [[15]](#footnote-14)**

Merge sort was first proposed by John von Neumann in 1945 and is often described as a divide and conquer algorithm. In that it divides input array in two halves, calls itself for the two halves and then merges the two sorted halves. **The merge () function** is used for merging two halves. The merge(arr, l, m, r) is key process that assumes that arr[l..m] and arr[m+1..r] are sorted and merges the two sorted sub-arrays into one. See following C implementation for details.[[16]](#footnote-15)

The following diagram from shows the complete merge sort process for an example array.



* 8 and 9 are merged in the correct order (a swap is required here)
* 5 and 6 are merged in the correct order (no swap required)
* 4 does nothing
* In the next step the 8 and 9 do nothing, but 4 is merged with 5 and 6
* In the last step the entire list is merged back together in the correct ordering

The Python code to implement this algorithm is shown below. It begins by asking the base case question. If the length of the list is less than or equal to one, then we already have a sorted list and no more processing is necessary. If, on the other hand, the length is greater than one, then we use the Python slice operation to extract the left and right halves. It is important to note that the list may not have an even number of items. That does not matter, as the lengths will differ by at most one. Once the MergeSort function is invoked on the left half and the right half it is assumed they are sorted. The rest of the function is responsible for merging the two smaller sorted lists into a larger sorted list. Notice that the merge operation places the items back into the original list (alist) one at a time by repeatedly taking the smallest item from the sorted lists. Note that the statement lefthalf[i] <= righthalf[j] ensures that the algorithm is stable. A stable algorithm maintains the order of duplicate items in a list and is preferred in most cases.[[17]](#footnote-16)

**MergeSort(arr[], l, r)**

If r > l

**1.** Find the middle point to divide the array into two halves:

middle m = (l+r)/2

**2.** Call mergeSort for first half:

Call mergeSort(arr, l, m)

**3.** Call mergeSort for second half:

Call mergeSort(arr, m+1, r)

**4.** Merge the two halves sorted in step 2 and 3:

Call merge(arr, l, m, r)

The mergeSort function has been augmented with a print statement to show the contents of the list being sorted at the start of each invocation. There is also a print statement to show the merging process. The transcript shows the result of executing the function on our example list. Note that the list with 44, 55, and 20 will not divide evenly. The first split gives [44] and the second gives [55,20]. It is easy to see how the splitting process eventually yields a list that can be immediately merged with other sorted lists.

   def mergeSort(alist):

    print("Splitting ",alist)

    if len(alist)>1:

        mid = len(alist)//2

        lefthalf = alist[:mid]

        righthalf = alist[mid:]

        mergeSort(lefthalf)

        mergeSort(righthalf)

        i=0

        j=0

        k=0

        while i < len(lefthalf) and j < len(righthalf):

            if lefthalf[i] <= righthalf[j]:

                alist[k]=lefthalf[i]

                i=i+1

            else:

                alist[k]=righthalf[j]

                j=j+1

            k=k+1

        while i < len(lefthalf):

            alist[k]=lefthalf[i]

            i=i+1

            k=k+1

        while j < len(righthalf):

            alist[k]=righthalf[j]

            j=j+1

            k=k+1

    print("Merging ",alist)

alist = [54,26,93,17,77,31,44,55,20]

mergeSort(alist)

print(alist)

Output:

Splitting [54, 26, 93, 17, 77, 31, 44, 55, 20]

Splitting [54, 26, 93, 17]

Splitting [54, 26]

Splitting [54]

Merging [54]

Splitting [26]

Merging [26]

Merging [26, 54]

Splitting [93, 17]

Splitting [93]

Merging [93]

Splitting [17]

Merging [17]

Merging [17, 93]

Merging [17, 26, 54, 93]

Splitting [77, 31, 44, 55, 20]

Splitting [77, 31]

Splitting [77]

Merging [77]

Splitting [31]

Merging [31]

Merging [31, 77]

Splitting [44, 55, 20]

Splitting [44]

Merging [44]

Splitting [55, 20]

Splitting [55]

Merging [55]

Splitting [20]

Merging [20]

Merging [20, 55]

Merging [20, 44, 55]

Merging [20, 31, 44, 55, 77]

Merging [17, 20, 26, 31, 44, 54, 55, 77, 93]

[17, 20, 26, 31, 44, 54, 55, 77, 93]

The mergeSort function has been augmented with a print statement to show the contents of the list being sorted at the start of each invocation. There is also a print statement to show the merging process. The transcript shows the result of executing the function on our example list. Note that the list with 44, 55, and 20 will not divide evenly. The first split gives [44] and the second gives [55,20]. It is easy to see how the splitting process eventually yields a list that can be immediately merged with other sorted lists.[[18]](#footnote-17)

* + 1. **Complexity of Merge Sort**

As the worst, best and average case time complexity is O(n\*log n), it is a very efficient algorithm. It starts with small sub files and finishes with larger aubfiles, therefore it does not neet to stack. Merge sort is highly parallelisable and can be used to impairment a stable sort.

**Time Complexity:**  Time complexity of Merge Sort is O(n\*log n) in all 3 cases (worst, average and best) as merge sort always divides the array into two halves and take linear time to merge two halves.

**Auxiliary Space:** O(n)

**Algorithmic Paradigm:**Divide and Conquer

**Sorting In Place:** No in a typical implementation

**Stable:** Yes

Merge Sort is marginally slower that Quick Sort in practice, and it is not as efficient as Block Sort. Therefore

2.3 **Counting Sort (a non-comparison sort)**

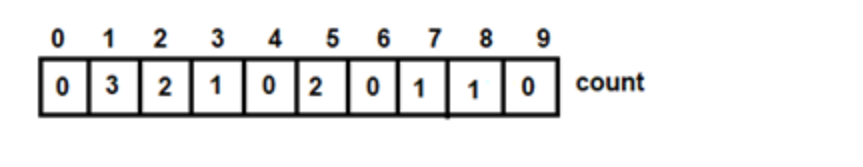
Counting sort was first proposed by Harold H.Seward in 1954. Counting sort allows us to do something which seems impossible, to sort a collection of items, in close to linear time. In essence Counting sort is a sorting algorithm that sorts the elements of an array by counting the number of occurrences of each unique element in the array. The count is stored in an auxiliary array and the sorting is done by mapping the count as an index of the auxiliary array.[[19]](#footnote-18)

There are a few assumptions when implementing Counting Sort. **it assumes that the input elements are**n**integers in the range [0,**k**].** When k = O(n), then the counting sort will run in O(n) time.

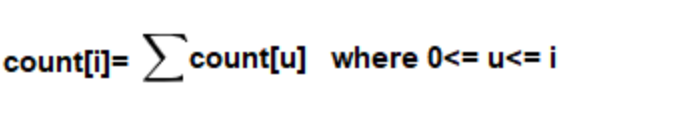
Consider the following input array A to be sorted. All the elements are in range 0 to 9

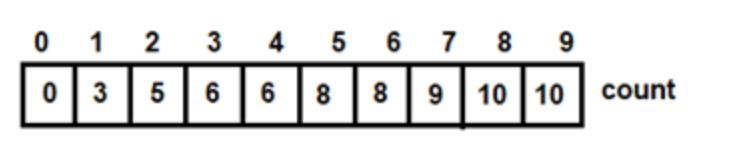


Step 1: Initialise an auxiliary array, say count and store the frequency of every distinct element. Size of count is 10 (k+1, such that range of elements in A is 0 to k)



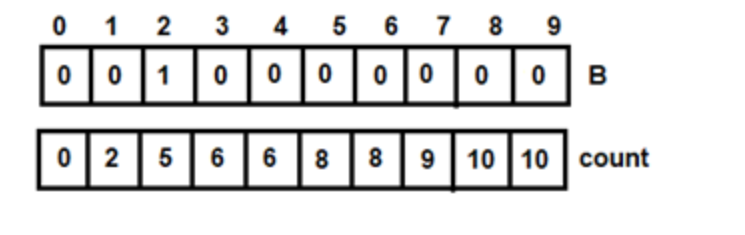
Step 2: Using the formula, updated count array is -



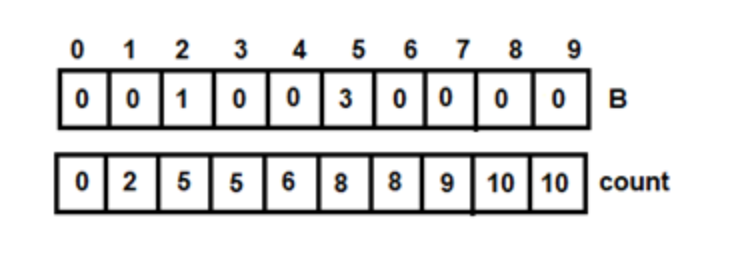


Step 3: Add elements of array A to resultant array B using the following steps:

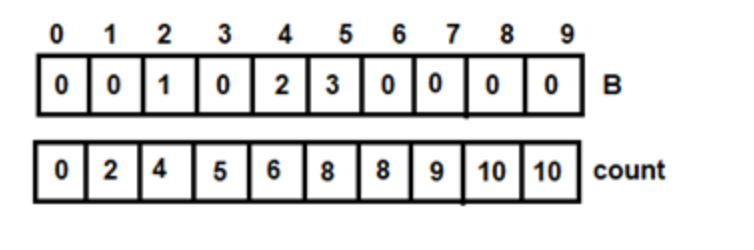
For, i=0, t=1, count[1]=3, v=2. After adding 1 to B[2], count[1]=2 and i=1



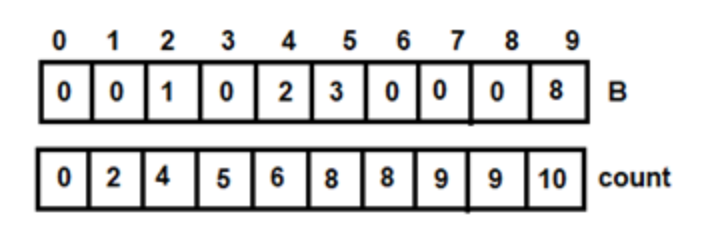
For i=1, t=3, count[3]=6, v=5. After adding 3 to B[5], count[3]=5 and i=2



For i=2, t=2, count[2]= 5, v=4. After adding 2 to B[4], count[2]=4 and i=3

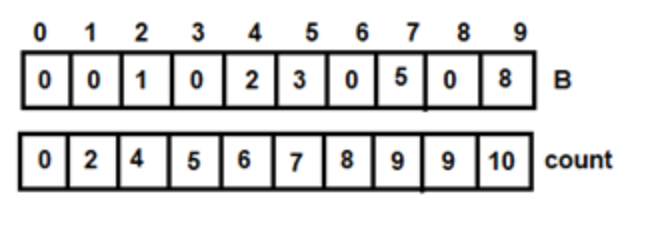


For i=3, t=8, count[8]= 10, v=9. After adding 8 to B[9], count[8]=9 and i=4

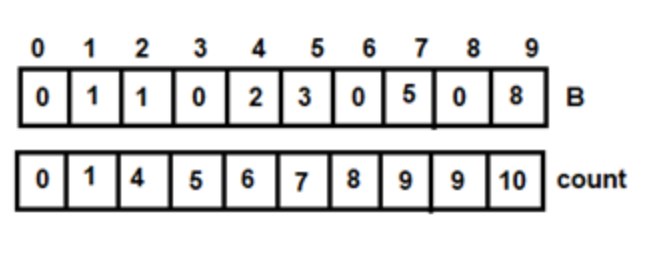


On similar lines, we have the following:

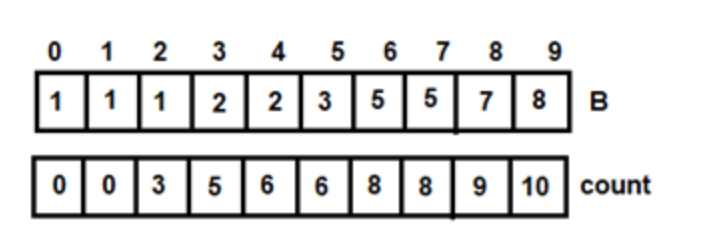
For i=4



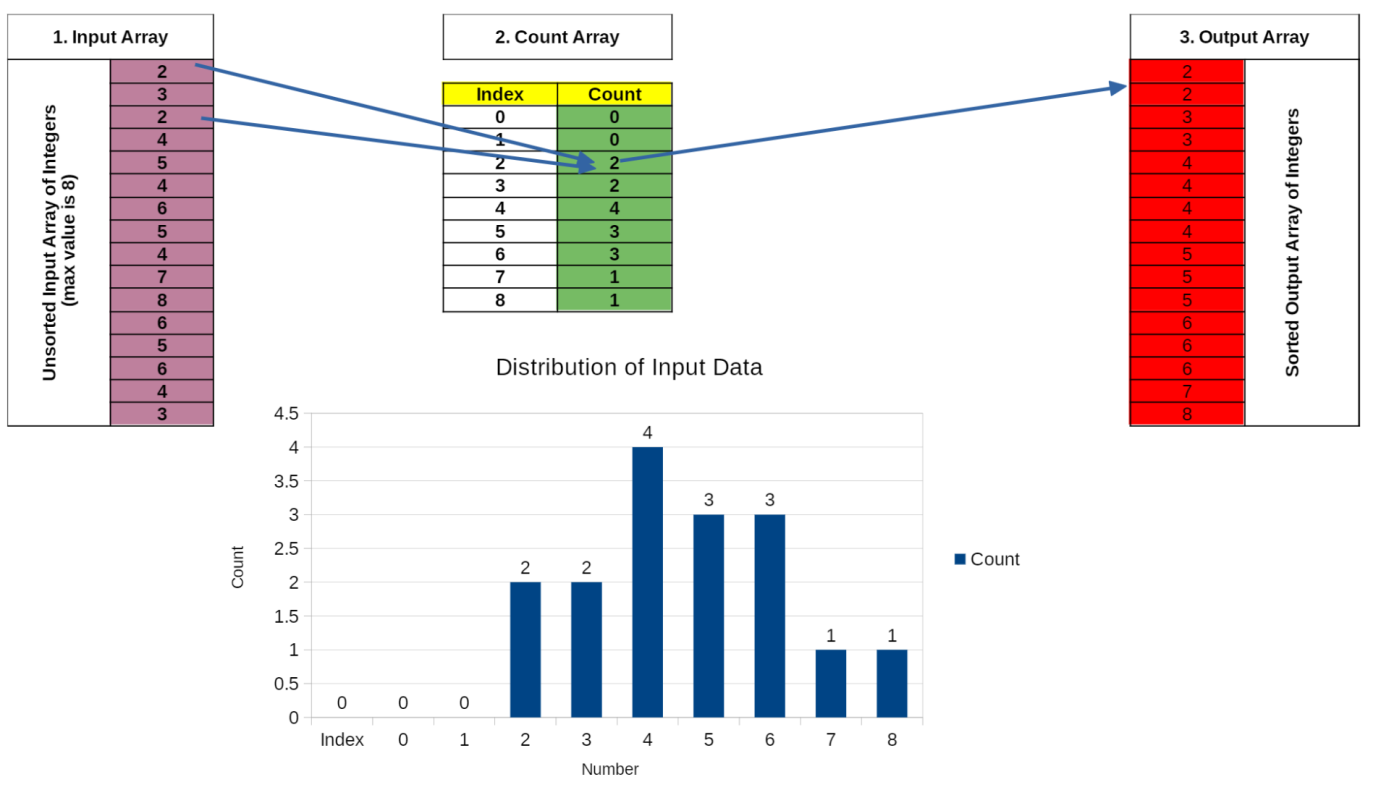
For i=5



All the way until i=9



Thus, array B has the sorted list of elements.[[20]](#footnote-19)



* The above demonstrates how counting sort can be used on an array of 16 integers with a maximum value of 8. As there the maximum value in the input array is 8, an array with 9 elements is needed. The values at each index in the count array are initially set to zero.
* The next step is to iterate through the input array and check the value of each element. Then increment the element at that index in the count array.
* The final step is to construct an output array using the information we have in the count array

The Python code to implement this algorithm is shown below.

def counting\_sort(array1, max\_val):

    m = max\_val + 1

    count = [0] \* m

    for a in array1:

    # count occurences

        count[a] += 1

    i = 0

    for a in range(m):

        for c in range(count[a]):

            array1[i] = a

            i += 1

    return array1

print(counting\_sort( [1, 2, 7, 3, 2, 1, 4, 2, 3, 2, 1], 7 ))

Output:

[1, 1, 1, 2, 2, 2, 2, 3, 3, 4, 7]

**2.3.2 Complexity of Counting Sort**

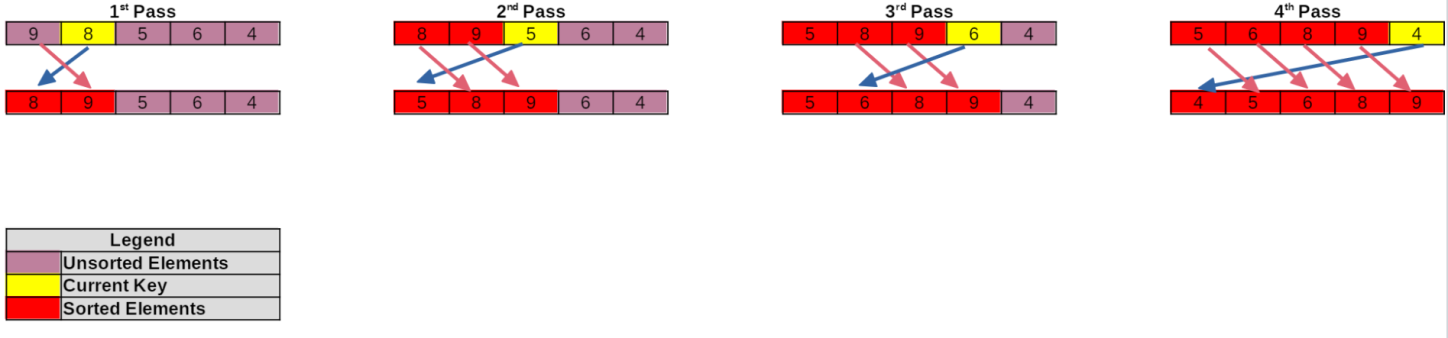
Counting sort has a O(k+n)*O*(*k*+*n*) running time. The first loop goes through A*A*, which has n*n* elements. This step has a O(n)*O*(*n*) running time. The second loop iterates over k*k*, so this step has a running time of O(k)*O*(*k*). The third loop iterates through A*A*, so again, this has a running time of O(n)*O*(*n*). Therefore, the counting sort algorithm has a running time of O(k+n)*O*(*k*+*n*). Counting sort is efficient if the range of input data, k*k*, is not significantly greater than the number of objects to be sorted, n*n*. Counting sort is a stable sort with a [space complexity](https://brilliant.org/wiki/space-complexity/) of O(k + n)*O*(*k*+*n*).[[21]](#footnote-20)

**Time Complexity:** O(n+k) where n is the number of elements in input array and k is the range of input.  
**Auxiliary Space:** O(n+k)

Counting Sort is both a fast and stable algorithm. However, is it not the most ideal for sorting large datasets or strings values.

* 1. **Insertion Sort (a simple comparison-based sort)[[22]](#footnote-21)**

Insertion sort is another example of a simple comparison based sorting algorithm. It always maintains a sorted subsist in the lower positions of the list. Each new item is then “inserted” back into the previous subsist such that the sorted subsist is one item larger. The image below shows the insertion sorting process. The shaded items represent the ordered subsists as the algorithm makes each pass. A real life example of Insertion Sort would be how card players sort a hand of cards. For example, a player has five cards, and they were already sorted from smallest to largest. The player gets a sixth card and inserts this card into the correct position. This is the idea behind **insertion sort**. Loop over positions in the array, starting with index 1. Each new position is like the new card handed to you by the dealer, and you need to insert it into the correct place in the sorted sub array to the left of that position.[[23]](#footnote-22)



First Pass:

* The element at index 1 is set as the key (8). It is compared to the item at index 0 (9). As it is smaller than 9, the 9 is pushed right by one place and the 8 is inserted at index 0.

Second pass:

* The item at index 2 (5) is the key. It is smaller than both 8 and 9 so they are both shifted right by one position and the 5 is inserted at index 0.

Third pass:

* The element at index 4 is the key. Note that in this case, the 6 is greater than 5, so 5 stays where it is. It is less than 8, so the 8 and everything to the right of it in the sorted portion of the list is shifted right by one position. The 6 is inserted at position once held by the 8.

Fourth Pass:

* The key (4) is smaller than all other elements in the array, so everything is shifted right by one position and the 4 is inserted at index 0.

The list is now sorted

The Python code to implement this algorithm is shown below. The implementation shows that there are again n−1n−1 passes to sort *n* items. The iteration starts at position 1 and moves through position n−1n−1, as these are the items that need to be inserted back into the sorted subsists. Line 8 performs the shift operation that moves a value up one position in the list, making room behind it for the insertion. Remember that this is not a complete exchange as was performed in the previous algorithms. The maximum number of comparisons for an insertion sort is the sum of the first n−1n−1 integers. Again, this is O(n2)O(n2). However, in the best case, only one comparison needs to be done on each pass. This would be the case for an already sorted list. One note about shifting versus exchanging is also important. In general, a shift operation requires approximately a third of the processing work of an exchange since only one assignment is performed. In benchmark studies, insertion sort will show very good performance.[[24]](#footnote-23)

      def insertionSort(alist):

   for index in range(1,len(alist)):

     currentvalue = alist[index]

     position = index

     while position>0 and alist[position-1]>currentvalue:

         alist[position]=alist[position-1]

         position = position-1

     alist[position]=currentvalue

alist = [54,26,93,17,77,31,44,55,20]

insertionSort(alist)

print(alist)

Output:

[17, 20, 26, 31, 44, 54, 55, 77, 93

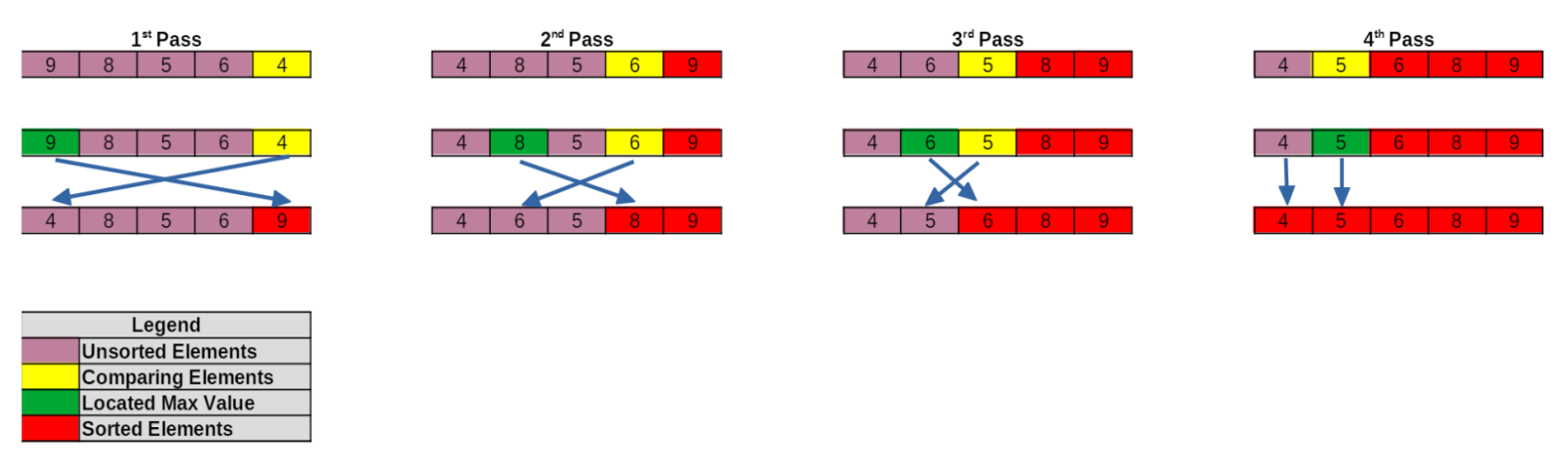
* + 1. **Complexity of Insertion Sort**

The time complexity of insertions sorted is related to the number of inversions in the input array. The total number of comparisons made by insertion sort is the number of inversions plus n-1 (the number of times the for loop must run).

* **Time Complexity:** O(n\*2)
* **Auxiliary Space:**O(1)
* **Boundary Cases**: Insertion sort takes maximum time to sort if elements are sorted in reverse order. And it takes minimum time (Order of n) when elements are already sorted.
* **Algorithmic Paradigm:** Incremental Approach
* **Sorting In Place:** Yes
* **Stable:** Yes
* **Online:** Yes
* **Uses:** Insertion sort is used when number of elements is small. It can also be useful when input array is almost sorted, only few elements are misplaced in complete big array.
  1. **Selection Sort (a simple comparison-based sort)**

Selection Sortimproves on the bubble sort by making only one exchange for every pass through the list. In order to do this, a selection sort looks for the largest value as it makes a pass and, after completing the pass, places it in the proper location. As with a bubble sort, after the first pass, the largest item is in the correct place. After the second pass, the next largest is in place. This process continues and requires n−1n−1 passes to sort *n* items, since the final item must be in place after the (n−1)(n−1) set pass.

The figure below shows the entire sorting process.



First pass:

* Start with the last element in the array, we then check the entire array for the largest value (9). We can say that 9 is now in the sorted portion of the array, while the rest of the array can be called the unsorted sub-array. We swap these values so the maximum value is in the last position in the array.

Second Pass:

* Consider the element in the second last position in the array (6), find the maximum value in the unsorted sub-array (8) and swap these elements.

Third pass:

* Consider the element in the third last position in the array (5). find the maximum number in the unsorted sub-array (6) and swap it with this element.

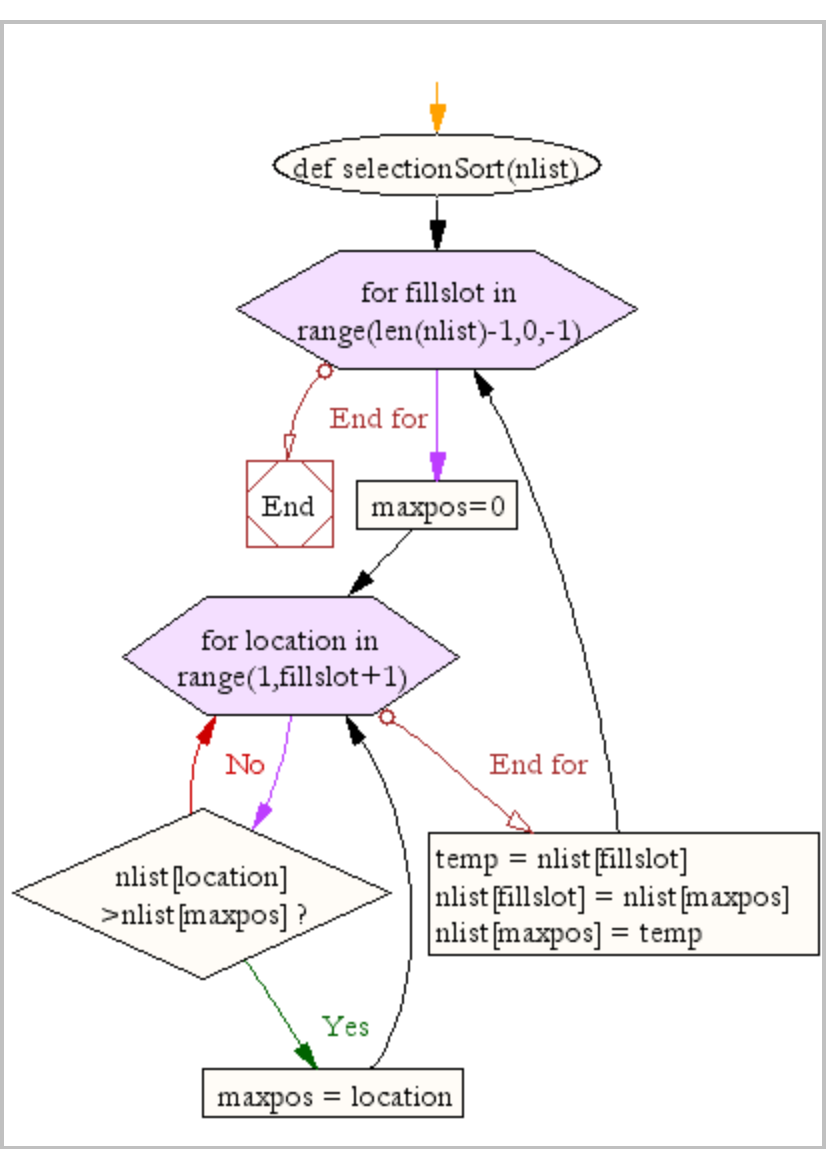
Fourth Pass:

* Consider the item in the second position (or index 1 – value is 5 here). In this case it is the maximum value in the sub array so there is no swap required

Selection sort is conceptually the simplest sorting algorithm. It first finds the smallest element in the array and then swaps it with the element in first position, it will then find the second smallest element and swap it with the element in the second position, and it will keep on repeating this process until the entire array is sorted.

The flow diagram shows the how Selection Sort algorithm works:

1. Consider the first element to be sorted and the rest to be unsorted.
2. Assume the first element is the smallest.
3. Check if the first element is smaller than each of the other elements:
   1. Yes – do nothing
   2. No – choose the other smaller element as the min and repeat step 3
4. After completion of one iteration through the list, swap the smallest element with the first element of the list.
5. Consider the second element in the list to the the smallest and repeat the process until all elements are covered.



The python code to implement the algorithm is shown below.

def selectionSort(alist):

   for fillslot in range(len(alist)-1,0,-1):

       positionOfMax=0

       for location in range(1,fillslot+1):

           if alist[location]>alist[positionOfMax]:

               positionOfMax = location

       temp = alist[fillslot]

       alist[fillslot] = alist[positionOfMax]

       alist[positionOfMax] = temp

alist = [54,26,93,17,77,31,44,55,20]

selectionSort(alist)

print(alist)

Output:

[17, 20, 26, 31, 44, 54, 55, 77, 93]

* + 1. **Complexity of Selection Sort[[25]](#footnote-24)**

Number of comparisons: (n-1)+(n-2)+(n-3)+…….+1=n(n-1)/2 nearly equals n2.

**Complexity** = O(n 2) Also we can analyse the complexity by simply observing the number of loops. There are 2 loops so the complexity is n\*n=n2

**Time Complexities:**

**Worst Case Complexity:** O(n 2)  
If we want to sort in ascending order and the array is in descending order then, the worst case occurs.

**Best Case Complexity:**  O(n 2)  
It occurs when the array is already sorted

**Average Case Complexity:**  O(n 2)  
It occurs when the elements of the array are in jumbled order (neither ascending nor descending).[[26]](#footnote-25)

**Auxiliary Space:** O(1)  
The good thing about selection sort is it never makes more than O(n) swaps and can be useful when memory write is a costly operation.

1. **Implementation and Benchmarking**

**3.1 Python Code**

The python application used for carrying out this benchmarking project is included with the submission in a python file called *CTA project python.py.*

**3.1.1 Import the required libraries for the project**

import time

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

* time: used for timing each of the sorting algorithms
* numpy: used for generating lists of random integers
* pandas: used to create a dataframe to store the output from the trials
* matplotlib.pyplot: used for creating a graphical representation of the benchmark tests

**3.1.2 Python code for five Sorting Algorithms to be Used**

* Bubble sort
* Selection sort
* Insertion sort
* Merge sort
* Counting sort

**3.1.3 Timer functions to benchmark the algorithms**

**array\_create():** Takes an integer as argument and returns a list of randomly generated integers (between 1 and 1000) of the specified length. It uses the numpy package to generate the list.

**timer():** Takes an input array and a sorting algorithm as arguments. Runs the sorting algorithm on the input array and returns the time it takes in seconds.

**average\_time():** Takes a number of runs, test size and sorting algorithm as inputs. It runs for the specified number of times, each time it creates a new random array of the specified length and times how long it takes to sort using the specified algorithm. It returns the average run time.

**3.1.4 Formatting output**

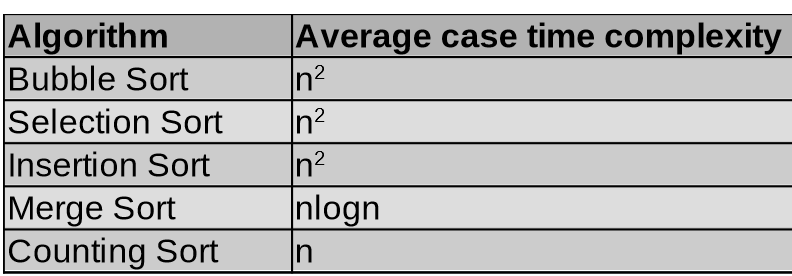
**algo\_trial():** This is the function that actually carries out the trials. It takes a sorting algorithm and test size as arguments. It creates a random array of the specified size, and uses the average\_time function to find the average run time for 10 runs and returns the average time in milliseconds formatted to 3 place of decimal.

**col\_create():** This is a function used to create columns for a pandas dataframe. Each column will contain the data for a particular test size. So it takes a list of sorting algorithms and a test size as arguments. It then loops through the list of algorithms and calls the algo\_trial function for each for the specified test size. It returns a column of data for the dataframe.

**df\_create():** takes a data dictionary as an input and converts it to a pandas dataframe. This is done because it is simple to format the output of a dataframe

**3.2 Results**

In this final section of the project, the results of the benchmarking exercise are presented for discussion. Before doing this however, we should first consider what the expected outcome of this should be. The application for this project compared run times for each algorithm that were averaged over 10 runs each. Therefore when discussing expected performance we will concern ourselves with the expected average case performance of the sorting algorithms. The table below shows a list of the five sorting algorithms together with their average case time complexity.



Benchmarking also known as posteriori analysis which evaluates efficiency empirically (using experiments). The relative performance is analysed by comparing the actual measurements that has been collected during experiment for the algorithms implemented. It is a measure of the real world performance of your algorithm but is affected by various hardware and software factors which include:

* System architecture
* CPU design
* Choice of Operating System
* Background processes
* Energy saving
* Performance enhancing technologies

Hence it is recommended that multiple runs are performed using same experimental setup to establish an average expected performance. Benchmarking can be used to verify the priori analysis (theory) of algorithms.[[27]](#footnote-26)

Each algorithm has a unique performance profile, with some as bad as average-case *n^2* performance. This means that it's infeasible to measure algorithm performance for *N*=80,000,000 for all algorithms, due to time constraints.

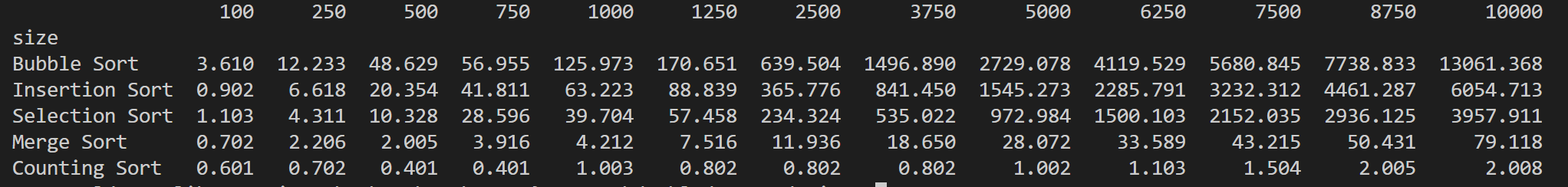
At a minimum, algorithm performance will be measured at a few significant intervals i for each order of magnitude m up through 10,000.

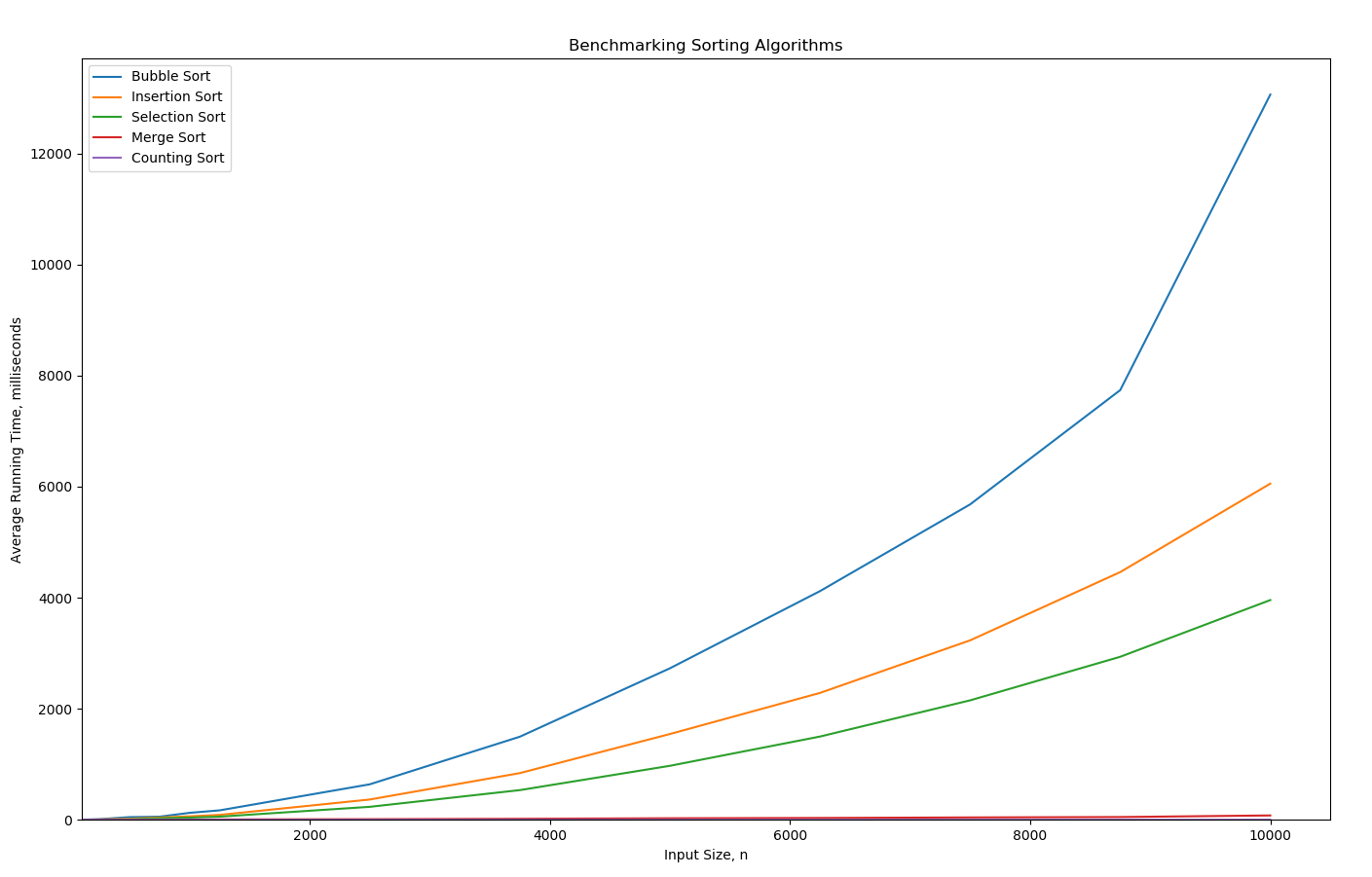
That is, we will measure performance at every input size N=i×10^m, where

* i is an element of {1, 2, 4, 8}
* m is an element of {1, 2, 3, 4}

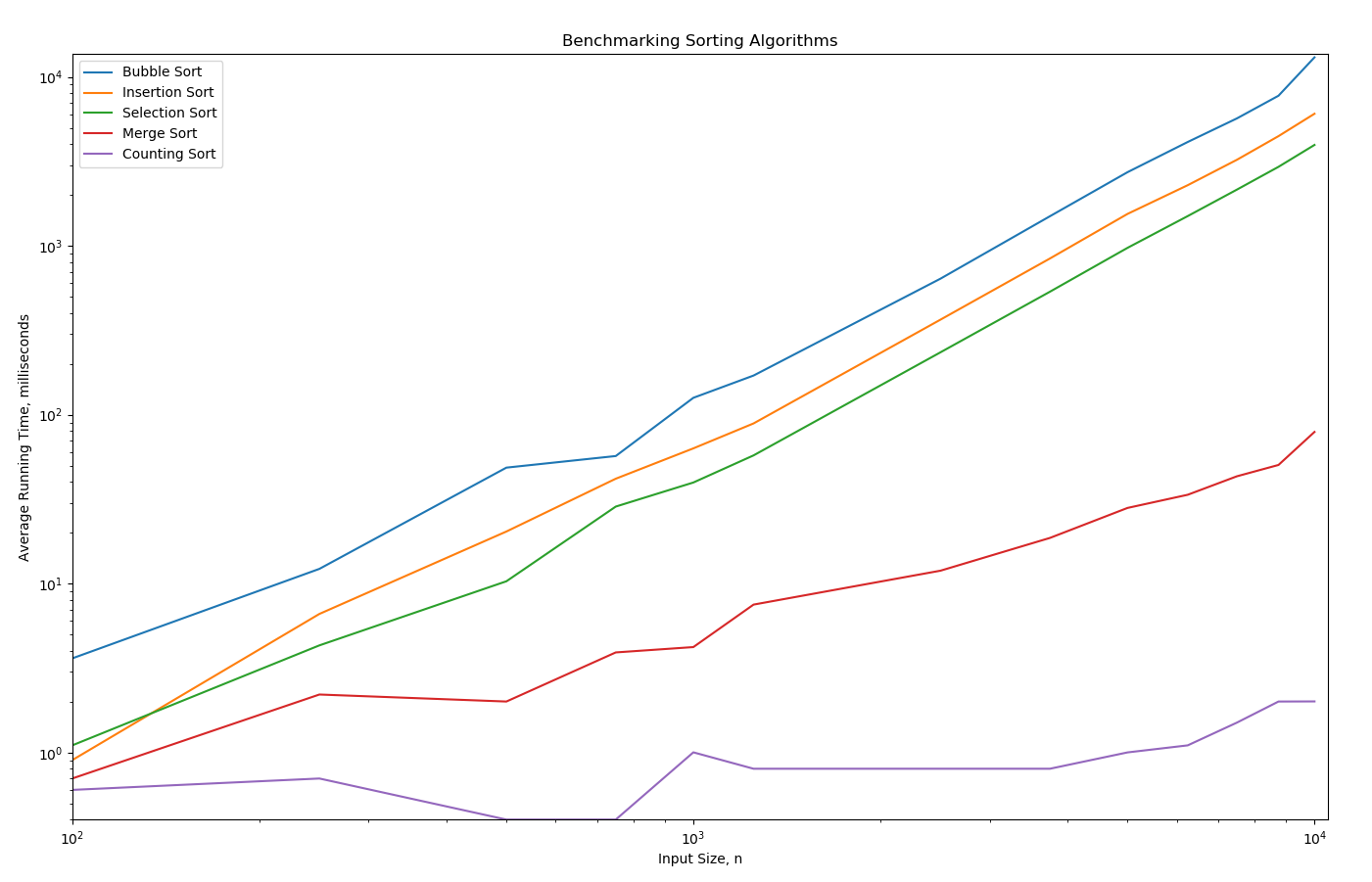
If algorithm performance allows, data will be collected for larger values of m. For example, Merge Sort can probably be sorted through m=7.[[28]](#footnote-27)

The table and graphs below show the results of the benchmarking exercise, as generated using the project.py Python application.





Graph – linear scale



Graph – log scale

As expected, bubble sort is worst performing sorting algorithm over all. It takes, on average, over 13 seconds to sort a random list of 10,000 numbers. Insertion sort fared slightly better at just over 6 seconds. In this trial, selection sort performed the best of the three simple comparison based algorithms taking just under 4 seconds, on average to sort the largest lists. While there are differences between the actual times taken for the three algorithms – they are all of the same order of magnitude – approximately 10 seconds. Merge sort was, on average, 50 times faster than selection sort – in other words a completely different order of magnitude. Likewise counting sort was nearly 40 times faster that merge sort, taking just 2 milliseconds to sort lists containing 10,000 random numbers. However, it should be noted that the implementation of counting sort here can only be used on lists of integers whose values are between 1 and 1000. If an array containing a number larger that this is passed to it, it will throw an error. The trade-off for the better performance is that the algorithm can only be used in certain circumstances

Both merge and insertion sort employ a divide and conquer approach with merge sort performing the better of the two. This matches the expectation that as the input size increases merge sort is the superior algorithm. Interestingly from the above table, in the worst case (Big O) scenario Quick Sort is *n2*and Merge Sort is *n log n*. Merge Sort is an excellent choice if predictable runtime is important as its best, worst and average time complexity is *O (n log n)*.

Finally counting sort performed the best of the five algorithms selected for this project with a potential runtime of *n+k*in the best, worst and average case, (where n refers to the input size and each item has a non-negative integer key, with a range of k). However it is important to be aware that this is not as widely as applicable as comparison sorts.

The results of the exercise are broadly in line with what we would have expected to achieve. Bubble sort would have been expected to perform the worst, and it did. Insertion might be expected to perform better that selection sort in some instances, However, if the input data is very unordered and has lots of inversions in it (as randomly generated data would be expected to), insertion sort can be very inefficient which is what the results here are showing. Merge sort would be expected to perform much better the three of these and it has. Counting sort would be expected to perform the best and it has.

1. Conclusion

1. Heineman et al., (2015), Algorithms in a Nutshell [↑](#footnote-ref-1)
2. <https://www.tutorialspoint.com/data_structures_algorithms/sorting_algorithms.htm> [↑](#footnote-ref-2)
3. <https://medium.com/@info.gildacademy/time-and-space-complexity-of-data-structure-and-sorting-algorithms-588a57edf495> [↑](#footnote-ref-3)
4. Heineman et al., (2015), Algorithms in a Nutshell

   <https://www.tutorialspoint.com/data_structures_algorithms/sorting_algorithms.htm3>.

   <https://medium.com/@info.gildacademy/time-and-space-complexity-of-data-structure-and-sorting-algorithms-588a57edf495>

   4. <https://www.studytonight.com/data-structures/space-complexity-of-algorithms>.

   <https://codeburst.io/algorithms-i-searching-and-sorting-algorithms-56497dbaef20>.

   6. [https://player.slideplayer.com/32/9926197/#](https://player.slideplayer.com/32/9926197/).

   <https://medium.com/@george.seif94/a-tour-of-the-top-5-sorting-algorithms-with-python-code-43ea9aa02889>.

   <https://www.robasworld.com/sorting-algorithms/>.

   <https://www.geeksforgeeks.org/analysis-algorithms-big-o-analysis/>

   Lecture notes – Sorting Algorithms

   <https://runestone.academy/runestone/books/published/pythonds/SortSearch/TheBubbleSort.html>

   <https://runestone.academy/runestone/books/published/pythonds/SortSearch/TheBubbleSort.html>

   **References:**

   **Websites:**

   * Dictionary of Algorithms and Data Structures: <https://xlinux.nist.gov/dads/>

   <https://www.tutorialspoint.com/data_structures_algorithms/sorting_algorithms.htm>

   <https://medium.com/@info.gildacademy/time-and-space-complexity-of-data-structure-and-sorting-algorithms-588a57edf495>

   <https://www.geeksforgeeks.org/sorting-algorithms/>

   <https://www.tutorialspoint.com/data_structures_algorithms/sorting_algorithms.htm>

   <https://medium.com/@info.gildacademy/time-and-space-complexity-of-data-structure-and-sorting-algorithms-588a57edf495>

   <https://www.geeksforgeeks.org/sorting-algorithms/>

   <http://code.jnickg.net/algorithm-benchmarking/>

   <https://www.slideshare.net/JamesStevens13/stevensbenchmarking-sorting-algorithms>

   **Videos:**

   * What on Earth is Recursion? Computerphile, 2014. <https://www.youtube.com/watch?v=Mv9NEXX1VHc>
   * Programming Loops vs Recursion. Computerphile, 2017. <https://www.youtube.com/watch?v=HXNhEYqFo0o>
   * XoaX.net Video tutorials on Algorithms, covering Complexity, Sorting and Searching. <http://xoax.net/comp_sci/crs/algorithms/index.php>
   * Visualization of 24 sorting algorithms in 2 minutes. Viktor Bohush, 2016. <https://www.youtube.com/watch?v=BeoCbJPuvSE>

   **Books:**

   * Harel D. and Feldman Y. (2012). Algorithmics - [The Spirit of Computing](http://www.hdip-data-analytics.com/_media/resources/pdf/algorithmics-the_spirit_of_computing_3rd_by_david_harel.pdf) (3rd Edition). Springer.
   * Pollice G., Selkow S. and Heineman G. (2016). [Algorithms in a Nutshell](http://www.hdip-data-analytics.com/_media/resources/pdf/algorithms_nutshell_.pdf), 2nd Edition. O' Reilly.
   * Goodrich M.T. and Tamassia R. (2014). [Data Structures and Algorithms in](http://www.hdip-data-analytics.com/_media/resources/pdf/data_structures_and_algorithms_in_python_pdfdrive.com_.pdf) Java (6th edition). John Wiley & Sons Inc.
   * Cormen T.H. (2013). [Algorithms Unlocked](http://www.hdip-data-analytics.com/_media/resources/pdf/algorithms_unlocked.pdf). MIT Press.
   * Cormen T.H., Leiserson C.E., Rivest R.L. and Stein C. (2009). [Introduction to Algorithms](http://www.hdip-data-analytics.com/_media/resources/pdf/introduction_to_algorithms_-_3rd_edition.pdf) (3rd Edition). MIT Press.
   * MacCormick J. (2013). Nine Algorithms That Changed the Future: The Ingenious Ideas That Drive Today's Computers. Princeton University Press.

   **Documentaries:**

   * The Secret Rules of Modern Living: Algorithms. BBC Four, first broadcast on September 24 2015.
   * The Wall Street Code. vpro backlight, 2013. <https://www.youtube.com/watch?v=kFQJNeQDDHA>

   **Papers:**

   * Furia C. A., Meyer B. and Velder S. (2014). Loop invariants: analysis, classification, and examples. ACM Computing Surveys. vol. 46, no. 3. <http://se.ethz.ch/~meyer/publications/methodology/invariants.pdf>

   **Github Repositories**

   <https://github.com/shkyler/gmit-cta-project>

   <https://github.com/RitRa>

   <https://github.com/G00364778>

   1. A brief history of algorithms (and why it's so important in automation, machine learning, and everyday life). [ONLINE] Available at: <https://e27.co/brief-history-algorithms-important-automation-machine-learning-everyday-life-20161207/>. [Accessed 18 April 2019].
   2. P.Mannion (2019) Week 9: Sorting Algorithms Part 2, Galway-Mayo Institute of Technology
   3. Karuna Sehgal. 2019. An Introduction to Bubble Sort – Karuna Sehgal – Medium. [ONLINE] Available at: <https://medium.com/karuna-sehgal/an-introduction-to-bubble-sort-d85273acfcd8>. [Accessed 2 May 2019].
   4. Wikipedians & Creutzburg, Reiner & Knackmuß, Jenny. (2013). Lecture Notes - Algorithms and Data Structures - Part 4: Searching and Sorting. 10.13140/2.1.5158.0486. [ONLINE] Available at:<https://www.researchgate.net/publication/259398088_Lecture_Notes_-_Algorithms_and_Data_Structures_-_Part_4_Searching_and_Sorting/citation/download>. [Accessed 2 May 2019].
   5. What are the advantages and disadvantages to an insertion sort? - Quora. 2019. What are the advantages and disadvantages to an insertion sort? - Quora. [ONLINE] Available at: <https://www.quora.com/What-are-the-advantages-and-disadvantages-to-an-insertion-sort>. [Accessed 6 May 2019].
   6. P.Mannion (2019) Week 10: Sorting Algorithms Part 3, Galway-Mayo Institute of Technology
   7. Sorting Algorithms in Software Development | RoBa's World. 2019. Sorting Algorithms in Software Development | RoBa's World. [ONLINE] Available at: <https://www.robasworld.com/sorting-algorithms/>. [Accessed 11 May 2019].
   8. GeeksforGeeks. 2019. QuickSort - GeeksforGeeks. [ONLINE] Available at: <https://www.geeksforgeeks.org/quick-sort/>. [Accessed 11 May 2019].
   9. Meet Zaveri. 2019. An intro to Algorithms: Searching and Sorting algorithms. [ONLINE] Available at: <https://codeburst.io/algorithms-i-searching-and-sorting-algorithms-56497dbaef20>. [Accessed 11 May 2019].
   10. George Seif. 2019. A tour of the top 5 sorting algorithms with Python code. [ONLINE] Available at: <https://medium.com/@george.seif94/a-tour-of-the-top-5-sorting-algorithms-with-python-code-43ea9aa02889>. [Accessed 11 May 2019].
   11. P.Mannion (2019) Week 8: Sorting Algorithms Part 1, Galway-Mayo Institute of Technology [11]
   12. Sorting Algorithms in Software Development | RoBa's World. 2019. Sorting Algorithms in Software Development | RoBa's World. [ONLINE] Available at: <https://www.robasworld.com/sorting-algorithms/>. [Accessed 11 May 2019].
   13. Time Complexity of Algorithms | Studytonight. 2019. Time Complexity of Algorithms | Studytonight. [ONLINE] Available at: <https://www.studytonight.com/data-structures/time-complexity-of-algorithms>. [Accessed 11 May 2019].
   14. Algorithms Chapter 3 [ONLINE] Available at: [https://player.slideplayer.com/32/9926197/#](https://player.slideplayer.com/32/9926197/). [Accessed 11 May 2019].
   15. Space Complexity of Algorithms | Studytonight. 2019. Space Complexity of Algorithms | Studytonight. [ONLINE] Available at: <https://www.studytonight.com/data-structures/space-complexity-of-algorithms>. [Accessed 04 May 2019].
   16. P.Mannion (2019) Week 12: Benchmarking Algorithms in Python, Galway-Mayo Institute of Technology

   [↑](#endnote-ref-1)
5. <https://www.studytonight.com/data-structures/space-complexity-of-algorithms>. [↑](#footnote-ref-4)
6. <https://codeburst.io/algorithms-i-searching-and-sorting-algorithms-56497dbaef20>. [↑](#footnote-ref-5)
7. [https://player.slideplayer.com/32/9926197/#](https://player.slideplayer.com/32/9926197/).  [↑](#footnote-ref-6)
8. <https://medium.com/@george.seif94/a-tour-of-the-top-5-sorting-algorithms-with-python-code-43ea9aa02889>. [↑](#footnote-ref-7)
9. <https://www.robasworld.com/sorting-algorithms/>. [↑](#footnote-ref-8)
10. <https://www.geeksforgeeks.org/analysis-algorithms-big-o-analysis/> [↑](#footnote-ref-9)
11. Lecture notes – Sorting Algorithms Part II [↑](#footnote-ref-10)
12. <https://runestone.academy/runestone/books/published/pythonds/SortSearch/TheBubbleSort.html> [↑](#footnote-ref-11)
13. <https://runestone.academy/runestone/books/published/pythonds/SortSearch/TheBubbleSort.html> [↑](#footnote-ref-12)
14. <https://www.techwalla.com/articles/advantages-disadvantages-of-bubble-sort> [↑](#footnote-ref-13)
15. <https://runestone.academy/runestone/books/published/pythonds/SortSearch/TheMergeSort.html> [↑](#footnote-ref-14)
16. <https://www.geeksforgeeks.org/merge-sort/> [↑](#footnote-ref-15)
17. <https://runestone.academy/runestone/books/published/pythonds/SortSearch/TheMergeSort.html> [↑](#footnote-ref-16)
18. <https://runestone.academy/runestone/books/published/pythonds/SortSearch/TheMergeSort.html> [↑](#footnote-ref-17)
19. <https://www.programiz.com/dsa/counting-sort> [↑](#footnote-ref-18)
20. <https://www.studytonight.com/data-structures/counting-sort> [↑](#footnote-ref-19)
21. <https://brilliant.org/wiki/counting-sort/> [↑](#footnote-ref-20)
22. <https://runestone.academy/runestone/books/published/pythonds/SortSearch/TheInsertionSort.html> [↑](#footnote-ref-21)
23. <https://www.khanacademy.org/computing/computer-science/algorithms/insertion-sort/a/insertion-sort> [↑](#footnote-ref-22)
24. <https://runestone.academy/runestone/books/published/pythonds/SortSearch/TheInsertionSort.html#lst-insertion> [↑](#footnote-ref-23)
25. <https://runestone.academy/runestone/books/published/pythonds/SortSearch/TheSelectionSort.html> [↑](#footnote-ref-24)
26. <https://www.programiz.com/dsa/selection-sort> [↑](#footnote-ref-25)
27. Dominic Carr (2020) Week 12: Benchmarking Algorithms in Python, Galway-Mayo Institute of Technology [↑](#footnote-ref-26)
28. <http://code.jnickg.net/algorithm-benchmarking/> [↑](#footnote-ref-27)