

# 

# 

# CZ2001 Algorithm

Example Class 3 Report

|  |  |
| --- | --- |
| Members - SE1 (GROUP 2) | |
| U1620986A | Lee Yi Zhuo |
| U1721978B | Ngo Jun Hao Jason |
| U1722308D | Thomas Stephen Felix |
| U1722841L | Tan Yong Heng Kenrick |
| N1804770J | Yung-Hsueh Lee |

Algorithm Implementation

Insertion sort:

Insertion sort is a sorting algorithm in which the elements are transferred one at a time to the right position.

Code used for Insertion Sort:

Merge sort (using auxiliary array in merge):

Merge Sort is a divide and conquer algorithm. It works by recursively breaking down a problem into two or more sub-problems, until the problems becomes simple enough to solve. The solutions to the sub-problems are then combined to give a solution to the original problem. So mergesort(int n, int m) first divides the array into equal halves and then combines them in merge(int n, int m) a sorted manner.

Code used for Merge Sort:

Generating Input Data

* Array of increasing sizes from 1000 to 50\_000
  + Random order
  + Ascending order
    - Generated using a simple for loop
  + Descending order
    - Generated using a simple for loop

Measuring Time Complexity

* For each array type and size, we did 60 insertion sorts and 60 merge sorts
  + Counted number of key comparisons using an int variable as a counter
    - Incremented this counter above key comparison statements
  + Recorded CPU time
    - Set current time after entering sorting function as start time
    - Set current time before returning from sorting function as end time
    - CPU time = end time - start time
* For each array type and size, the interquartile mean of 60 recorded CPU times is used as the CPU time
  + Interquartile mean obtained by
    - First sorting the 60 CPU times using a combination of quick sort and insertion sort
    - Followed by taking the middle 30 CPU times (the bottom 15 and top 15 CPU times are ignored)
    - And finding their average (sum them up and divide by 30)
* For each sort (insertion and merge), we placed their statistical results of sorting through different array types (random, ascending and descending order) in separate excel files
  + Each file lists number of key comparisons and CPU time in increasing array sizes

Analysis Of Results

Scatter plots (empirical results):

* Placeholder text

Theoretical Analysis of the Time Complexity of Insertion Sort:

1. **Best Case [ *n*-1 = O(*n*) ]**

The best case for insertion sort occurs when the array is already sorted in the desired order. In this case, each key in the array, except for the first key, only needs to be compared to the key right before it. So for an array of length *n*, only *n*-1 number of key comparisons need to be made, resulting in a time complexity of O(*n*).

1. **Worst Case [ *n*\*(*n*-1) /2 = O(*n*2) ]**

The worst case for insertion sort occurs when the array is already sorted, but in the reverse order. In this case, each key in the array, except for the first key, needs to be compared to all the keys that come before it.

For a key at index *i*, there are *i*-1 keys before it, so it has to go through *i*-1 number of key comparisons. If there are *n* number of keys in the array (i.e. array of size n), there will be (*n*-1) + (*n*-2) + … + 2 + 1 = (*n*\*(*n*-1) /2) number of key comparisons when sorting through this array, resulting in a time complexity of O(*n*2).

1. **Average Case [ (*n*-1)\*(*n*+2) /4 = O(*n*2) ]**

The average case for insertion sort occurs when the keys in the array are in a completely random order. For each key in a randomly-ordered array, there is an equal probability of a key at a particular index being compared to each key that comes before it, because there is an equal probability of that particular key being smaller or larger than the key that is in front (ie ½) . Mathematically, this means that for each key at index *i* in the array, there is a 1/*i* chance of it being compared to each key at index 0 through *i*-1, with 1 comparison at each index.

In other words, for a key at index i, there is a 1/*i* chance of it going through 1 key comparison, 1/*i* chance of it going through 2 key comparisons, …, 1/*i* chance of it going through i key comparisons. So, on average, a key at index *i* goes through (1/*i*)(1+2+...+*i*) = (1/*i*)( *i*\*(*i*+1)/2 ) = (*i*+1)/2 number of key comparisons.

For an array of length *n*, each key in the array, except for the first key, goes through the above process. Therefore, there are

[( (*n*-1) + 1 ) /2] + [( (*n*-2) + 1 ) /2] + … + [( 2+1) /2] + [(1+1) /2]

= (½)\* [( (*n*-1) + 1 ) + ( (*n*-2) + 1 ) + … + ( 2+1) + (1+1)]

= (½)\* [ *n* + (*n*-1) + … + 3 + 2 ]

= (½)\* [ (*n*-1)\*(*n*+2) /2 ]

= (*n*-1)\*(*n*+2) /4

number of key comparisons in an array of length *n*, resulting in a time complexity of O(*n*^2).

Previous answer:

Assuming the probability of going through each iteration is 1/i, the average number of comparison in the i th iteration is the sum of probability of going through each iteration(1/i) times the number of iteration (1, 2, 3, …, i). Since there will be n-1 number of iteration, it will be the total sum of each iteration from 1 to n-1, for the average number comparison in the i th iteration. This will give the time complexity of O(n^2).

Theoretical Analysis of the Time Complexity of Merge Sort

Merge sort functions by splitting the original array into 2 subarrays of equal length, and recursively splits the subarrays till each subarray contains only 1 key. When a subarray has a length of 1, it is considered trivially sorted so there is no need for any key comparison. This is the base case (W(1) = 0) for the recurrence equation which solves the time complexity of merge sort..

The comparisons of keys in the subarrays are done in the merging of a pair of subarrays that were split from an original array. This forms the recurrence equation: W(n) = 2\*W(n/2) + (number of key comparisons in the merging of subarrays of length n/2) for number of key comparisons for a merge sort on an array of length *n*.

For an array of length *n* where *n* = 2k (k ≥ 0), there will be k number of splits. This is because each split results in a subarray of length of *n*/2 = 2k-1 and every subarray is split recursively till each of them has a length of 1 = 20. k number of splits would result in k number of pairs of subarrays, which in turn results in *k* number of merges. Therefore, the number of merges in a merge sort on an array of length *n* is k = log2 *n*.

The number of key comparisons in a merge of 2 subarrays varies, depending on the case.

1. **Best Case [ (*n*/2)\*log2 *n* = O( *n*\*(lg *n*) ) ]**

The best case for merge sort occurs when every key in one subarray are smaller than or equal to every key in the other subarray. Each key of the first subarray will be compared against the first (smallest) key of the second subarray. After comparing and finding that the last (biggest) key of the first subarray is smaller than the first (smallest) key of the second subarray, the rest of the keys in the second subarray are assumed to be bigger than all the keys in the first subarray, and are merged without a need for comparison.

As such, the number of key comparisons in this case is the length of the first subarray. If the original array is of length *n* and the recursive splitting is such that 2 subarrays resulting from the split are of approximately equal length, then the length of the first subarray is *n*/2. This means *n*/2 number of key comparisons for each merge, so there are *n*/2 number of key comparisons.

The best case can also occur when each key at index i in the first subarray is equal to each key at index i in the second subarray, in which case both keys are merged simultaneously. When the keys from both subarrays are all merged pair-wise, simultaneously, the number of key comparisons is equal to the number of keys in either subarray (i.e. length of subarray) which is also *n*/2.

Therefore, the number of key comparisons is given by the recurrence equation:

W(*n*) = 2\*W(*n*/2) + *n*/2, or

W(2k) = 2\* W(2k-1) + 2k-1

= 2\*( 2\*W(2k-2) + 2k-2 ) + 2k-1

= 22 \* W(2k-2) + 2k-1 + 2k-1

…

= 2k \* W( 2k-k ) + k \* ( 2k-1 )

= 2k \* W(1) + k \* ( 2k-1 )

= 0 + (log2 *n*) \* *n* / 2

After simplification, this gives us a time complexity of O( *n* \* lg *n* ).

1. **Worst Case [ *n*\*log2 *n* - *n* + 1 = O( *n*\*(lg *n*) ) ]**

The worst case occurs when the largest and second largest keys among both subarrays are in separate subarrays or, if there are more than one key with the largest value, two of the largest keys are in separate subarrays

This arrangement is such that all keys up to the largest keys in each subarray of length *n*/2 have to be compared to another key. The largest keys in both subarrays would be compared against each other before being merged simultaneously. As such, the number of comparisons in one merge is *n*/2 + *n*/2 - 1 = *n*-1. Note that the number of key comparisons is one less than the total length of the subarrays as the last elements of both subarrays are merged together.

Therefore, the number of key comparisons is given by the recurrence equation:

W(*n*) = 2\* W(*n*/2) + (*n*-1), or

W(2k) = 2\* W(2k-1) + (2k - 1)

= 2\* [ 2\* W(2k-2) + (2k-1 - 1) ] + (2k - 1)

= 22 \* W(2k-2) + (2k - 2) + (2k - 1)

= 22 \* [ 2\* W(2k-3) + (2k-2 - 1) ] + (2k - 2) + (2k - 1)

= 23 \* W(2k-3) + (2k - 22) + (2k - 21) + (2k - 20)

…

= 2k \* W(2k-k) + k \* 2k - ( 2k-1 + … + 22 + 21 + 20 )

= 2k \* W(1) + (log2 *n*) \* *n* / 2 - ( 2k -1 )

= 0 + (log2*n*) \* *n* / 2 - *n* + 1

After simplification, this gives us a time complexity of O( *n* \* lg *n* )

1. **Average Case [ c1\**n*\*log2 *n* + c2\**n* + c3 = O(n\*(lg n) ]**

The average case for merge sort occurs when keys of the input array are randomly-ordered. In the average case, the probability of a key in the first subarray having to be compared to a key in the second subarray in each merge has to be taken into account when calculating the average total number of key comparisons in merge sort.

Based on page 43 of *(Vitter and Flajolet, 1990)*, the probability of a key in the first subarray having to be compared a key in the second subarray is given by the conditional probability:

P( (key of first subarray > key of second subarray | (last keys of either first and second subarray have not been reached) )

As such, the average number of key comparisons in one merge is the sum of each possible number of key comparisons multiplied by the probability for said number of key comparisons.

Hence, the average number of key comparisons in a merge sort is given by:

*n*\*log2 *n* + α\**n* + O(1), where α = 1.2645 approximately

, according to page 43 of *(Vitter and Flajolet, 1990)*.

Comparison of empirical results with theoretical analysis of time complexity

* Placeholder text

Conclusion

* Placeholder text

References

Vitter, J. and Flajolet, P. (1990). *Average-Case Analysis of Algorithms and Data Structures*. [online] Algo.inria.fr. Available at: <http://algo.inria.fr/flajolet/Publications/ViFl90.pdf> [Accessed 6 Oct. 2018].