BLG 335E – Analysis of Algorithms I

Fall2020 Homework 1

Deliver Date: 26.11.2020

Due Date: 10.12.2020

Important Notes:

- Please write your own codes, copying code parts from books, websites or any other sources including your friends is considered as plagiarism and results in minus points.
- Do not upload your codes any public platform (e.g. Github) until the deadline of homework passes.
- Use C++ language and do not forget to compile your codes on Linux using g++ command before sending them.
- You may use STL but do not use built-in functions for sorting.
- Do not forget to comment your code.
- Submit your source codes and report files on Ninova before the deadline, late submissions and submissions via e-mail will not be accepted.

Part 1. Implementation (40 points)

You are given a dataset where the global sales of different products are reported. There are following attributes in dataset:

Country: Country that sells the product

Item Type: Type of the product

Order ID: Unique ID for each order

Units Sold: Number of units that is sold

Total Profit: Profit obtained from order

You are supposed to sort the orders in alphabetical order in terms of the name of **country**. For the orders that have the same country name you should sort them in **descending** order of **total profits**. You must use regular **Quicksort** as the sorting algorithm.

Your code must run with the following command: **./a.out N**

N: number of the sales to be sorted (You can just take the first N entries from the file)

After the execution, a message including the elapsed time of execution should be printed out and you should write the sales in sorted order into sorted.txt file with the same format.

Part 2. Report (60 points)

a) **(15 points)** Write down the asymptotic upper bound for the Quicksort for best case, worst case and average case. Prove them solving the recurrence equations.

b) **(10 points)** In implementation, we wanted to sort the sales by alphabetical order of country names and then by their total profits. Let's assume that we are having this kind of method:

1) Sort the sales.txt data by the total profits and write it into sorted_by_profits.txt
2) Sort the sorted_by_profits.txt data according to country names using QuickSort

Does this solution give us the desired output for all cases?

1. Explain why or why not and give a simulation on a small fraction from the dataset (you may modify the results if necessary).

2. Give 3 examples for the sorting algorithms that give the desired output.

c) **(15 points)** Run the algorithm for different **N** values {10, 100, 1000, 10K, 100K, 500K, 1M} on **sales.txt** data and calculate the average time of running. (Run the algorithm 10 times for each N value and take the average execution time for each N value). Report the average execution times in a table and prepare an Excel plot which shows the N – runtime relation. Comment on the results considering the asymptotic bound that you have found in **(a). (3-4 sentences)**

Note: You can use the clock() function under ctime library for calculating time of execution for the search functions. Refer to http://www.cplusplus.com/reference/clibrary/ctime/clock for more details.

d) **(20 points)** Run the algorithm for different **N** values {10, 100, 1000, 10K, 100K, 500K, 1M} on **sorted.txt** data and calculate the average time of running as you have done in **(c).** Report the results in a table and plot them similarly.

1. Compare the results with the results you have obtained at **(c)** and explain the difference in detail referring the equations you have given in **(a). (4-5 sentences)**

2. Which other input cases would give us the similar results? **(1 sentence)**

3. Propose a solution to this case. **(1 sentence)**