CPSC 319

ASSIGNMENT #2

JOHN EZEKIEL JULIANO

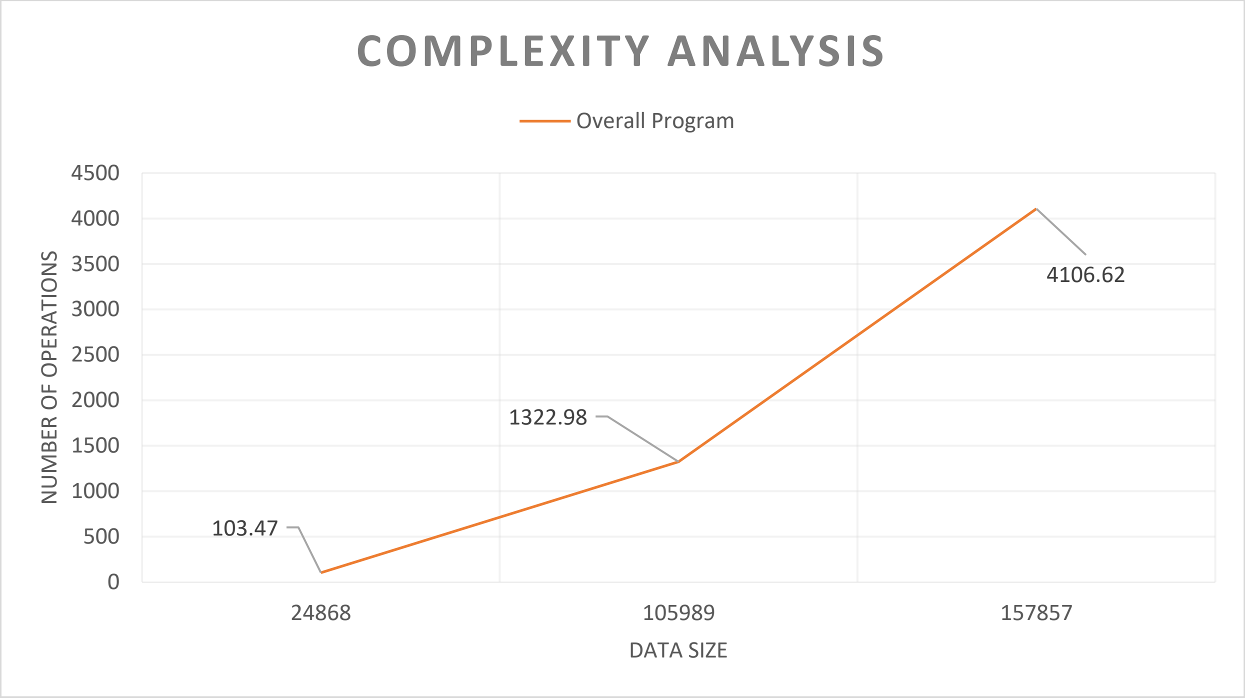
30000523 T05

RASHID MAMUNUR [TA]

FEBRUARY 16, 2018

I. Data

Data Size	Overall Program
24868	103.47 seconds
105989	1322.98 seconds
157857	4106.62 seconds



II. Complexity Analysis

1. The worst-case complexity analysis of all the individual methods involved in the program can be found in the pages in the end of the file. These theoretical results will be used to compare with the experimental runs. The method used in determining whether two words are anagrams of the other is isAnagram(). Within in is another method charSort(). The total big-O value for this is $O(k^2n)$. It is assumed that 'k' is the maximum length of any word from the input file.

(Refer to the next few pages for the theoretical analysis of the program)

2. Using the same complexity analysis results however in this case the variables are:

$$n \rightarrow N \quad \text{and} \quad k \rightarrow L$$

The total theoretical big-O running time of the program is:

$$O(k^2n^2)$$

This is mainly generated by the file reader method which also contains the method to determine whether two words are anagrams or not. Even though there is a method that generated a $O(n^4)$ – this is insertionSort() inside a for-loop. It is a known fact that reading from files take longer than most operations. Therefore, I assumed that most of the program's running time will be invested in reading the input file and not on sorting the items and arrived at a conclusion that the total big-O of the program is $O(k^2n^2)$.

Observing the graph of the complexity analysis of the whole program, it can be approximated that the relationship between the data size and the number of operations is quartic (x^4). This yielded similar results as the theoretical ones; $O(k^2n^2)$ and is also true for $O(n^4)$.