

COSC 320 – 001

Analysis of Algorithms

Milestone 3

2022/2023 Winter Term 2

Project Topic Number: 1

Title of project:

Keyword replacement in a corpus

Group Number 12:

Anna Ciji Panakkal

Erfan Kazemi

Sahil Chawla

Abstract:

In this milestone, we reviewed our pseudocode and algorithm analysis last milestone 1 to work on the implementation of the same. We met via discord to split the responsibilities. Once completed, we combined our work and looked at the results to check if it returned what we expected.

Dataset:

We used the dataset provided by the professor which includes 2.8 million app reviews to implement and test the algorithm that we created.

Implementation:

The algorithm consists of four classes that work together to achieve the desired functionality. The Node class creates nodes by taking an abbreviation and its expansion as inputs. The AbbTree class creates a tree structure, generates a root node, and creates 28 basic nodes for each of the 28 letters. It then inserts all new nodes to the tree based on their first letter.

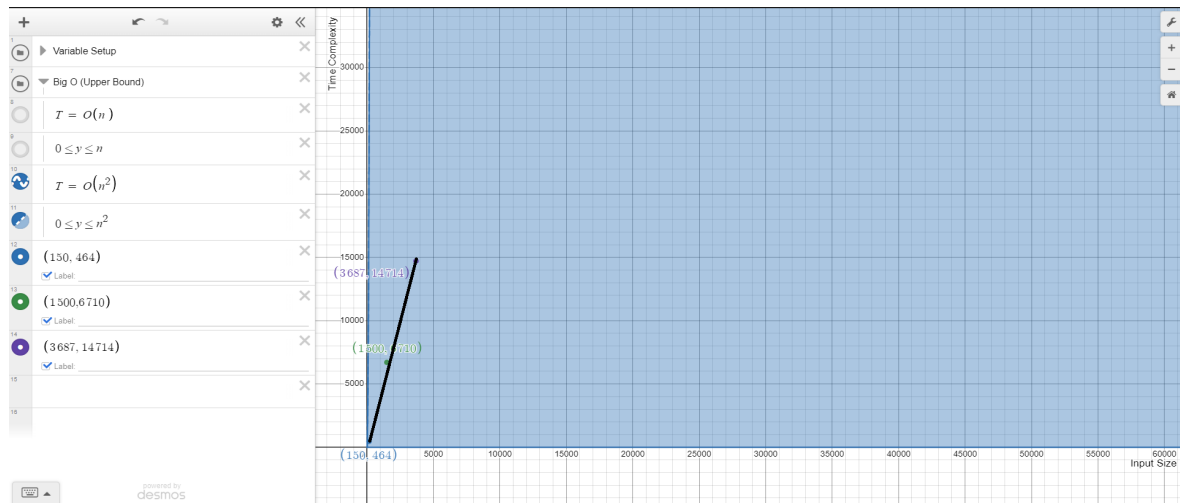
The krc class takes the tree and a word from the input text as inputs and begins searching the tree to find the corresponding abbreviation for the given word. Lastly, there is a main menu that serves as a testing platform, allowing users to input files and texts to be processed by the algorithm.

By breaking down the implementation into these four classes, the algorithm can be easily maintained and updated in the future. Additionally, the modular structure allows each component to be tested individually, leading to more efficient troubleshooting and debugging.

Result:

Expected time complexity : $O(n^2)$

Actual time complexity : $O(n^2)$



It is not exactly what we expected, due to coding restrictions but we got a satisfactory result.

For data structures, we used trees and nodes in our code, this helped us go through and replace the abbreviations faster than it would in a normal word by word search as it stores the words in a tree and we could search branches parallel and find the word and replace it

Unexpected Cases/Difficulties:

1. Our initial plan was to code in using C++ but due had to switch to java due to restrictions faced while coding
2. Debugging the logical errors in the code was another big difficulty that we faced. This was solved by peer reviewing code written to figure out what was the mistakes

Task Separation and Responsibilities:

Erfan, Anna - Implementation, Results

Sahil - Unexpected Cases/Difficulties