COSC 320 – 001

Analysis of Algorithms

2022/2023 Winter Term 2

Project Topic Number: 1

Project Third Milestone (Group 32)

Keyword Replacement in Corpus

Group Lead: Anitej Isaac Sharma

Group Members:

Anitej Isaac Sharma

Abdirahman Hajj Salad

Yuki Isomura

Abstract

This is the Third Milestone for our Keyword Replacement (in Corpus) Algorithm. In this milestone, we succeeded in coming up with the Dataset details, Implementation of our algorithm, Results, Unexpected Cases/Difficulties, and Task Separation and Responsibilities.

Dataset

The dataset consists of a large collection of abbreviations along with their corresponding keywords/phrases that should replace those abbreviations. The abbreviations may consist of various formats including letters, numbers, or punctuation marks. The abbreviations can be highly variable in terms of style and context, and the algorithm will attempt to be flexible enough to handle this variability while identifying and replacing the abbreviations correctly.
Some details about the dataset we need to consider:

- Abbreviation , Full form: The first column contains the abbreviations while the second column contains the corresponding keyword/phrases
- Size: The dataset is very large, containing millions of rows hence the need for efficient algorithm for processing
- Format: The abbreviation can be in various formats including single words (e.g. ASAP, LOL) or multiple words (e.g. FYI) or alphanumeric characters (e.g. 24/7, 2moro)
- Variability: The abbreviations can vary in terms of their capitalization, punctuation, and context. For example, ASAP can be written as A.S.A.P, Asap, ASAP, or as soon as possible.
- Accuracy: The dataset is assumed to be accurate, i.e., each abbreviation is correctly paired with its corresponding keyword/phrase.

Implementation

- Explain how you implemented the algorithm and tested it. All the subtle details should be included.
- This is just an explanation and you do not need to copy paste your implementation here

Result

Subtitle:

- Include the plots and the interpretation of the plots as input grows.
- Compare it to the big O function of the running time. For example, if your algorithm runs in O(n^2), show the graph for the n^2 function in the same plot as well.

- Explain if this is what you expected, and how the implementation of your algorithm might have affected the constant values. How the choice of data structure might have affected this result?

Unexpected Cases/Difficulties

Task Separation and Responsibilities

Anitej Isaac Sharma	**Yuki Isomura** Implementation of Algorithm	**Abdirahman Hajj Salad** Dataset Description