



Text Summarisation Using NLP

Team Enrollment Numbers - E20CSE332, E20CSE358, E20CSE414, E20CSE266

Problem Statement

In today's world, large amounts of textual data is generated every day. It is very hard to understand and analyse that data so there is a need for automatic text summarization also with a user friendly GUI. Even though we have many methods of implementing this, the best way is using natural language processing techniques that are by using NLTK and Spacy libraries. Basically we need an application that collects important data from big volume of textual data. Natural language processing tools and techniques results out with best accuracy than any other existing algorithms.

INTRODUCTION

Text summarization is a natural language processing (NLP) technique that involves automatically generating a concise and meaningful summary of a longer piece of text. This can be accomplished using various NLP libraries such as NLTK and spaCy.

NLTK and spaCy are both popular NLP libraries that provide tools for text processing, including tokenization, part-of-speech tagging, and named entity recognition, which are useful for text summarization. These libraries also provide algorithms and methods for implementing text summarization, including extractive and abstractive summarization techniques.

≡ "OUR MISSION

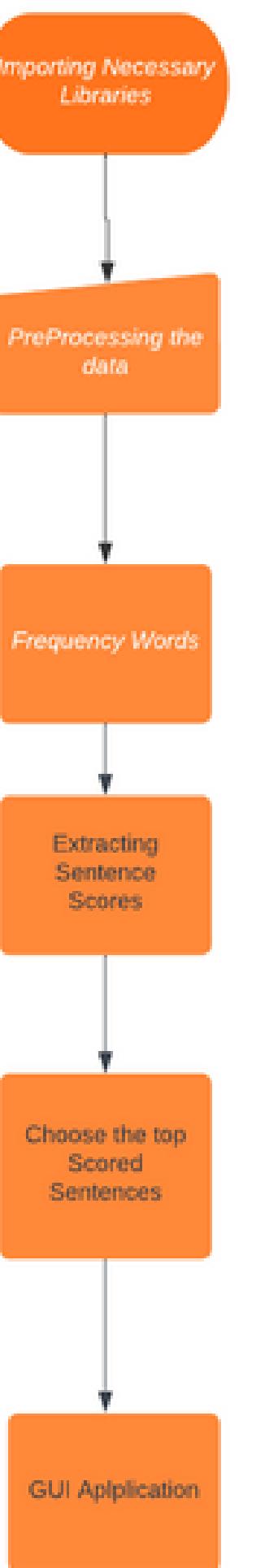
Our mission on text summarization using NLTK and spaCy is to help users understand the importance and application of text summarization in NLP. We aim to provide a clear and concise introduction to the concept of text summarization and its use in various fields.

Our goal is to educate users on the different techniques available for text summarization, including extractive and abstractive summarization, and the tools and algorithms provided by NLTK and spaCy that can be used to implement these techniques.

Methodology

- Automatic text summarization is used to extract required important data from large volumes of textual data. In this process two important popular libraries are used. They are NLTK and Spacy. Although some other algorithms exist, the best way of implementing this model is using natural language processing techniques for best accuracy and results. Graphical user interface is also built for user friendly and flexible usage of the application.
- NLTK(Natural language tool kit) is a famous python library used in text summarization for tokenization, stemming, parts of speech tagging, parsing, semantic analysis and lemmatization. Spacy is also a python library which is used for faster processing of textual data. The flow of the application goes as follows A. Importing necessary libraries B. Preprocessing the data C. Extracting frequency of words D. Choose the top sentences E. GUI Application

Flow Chart

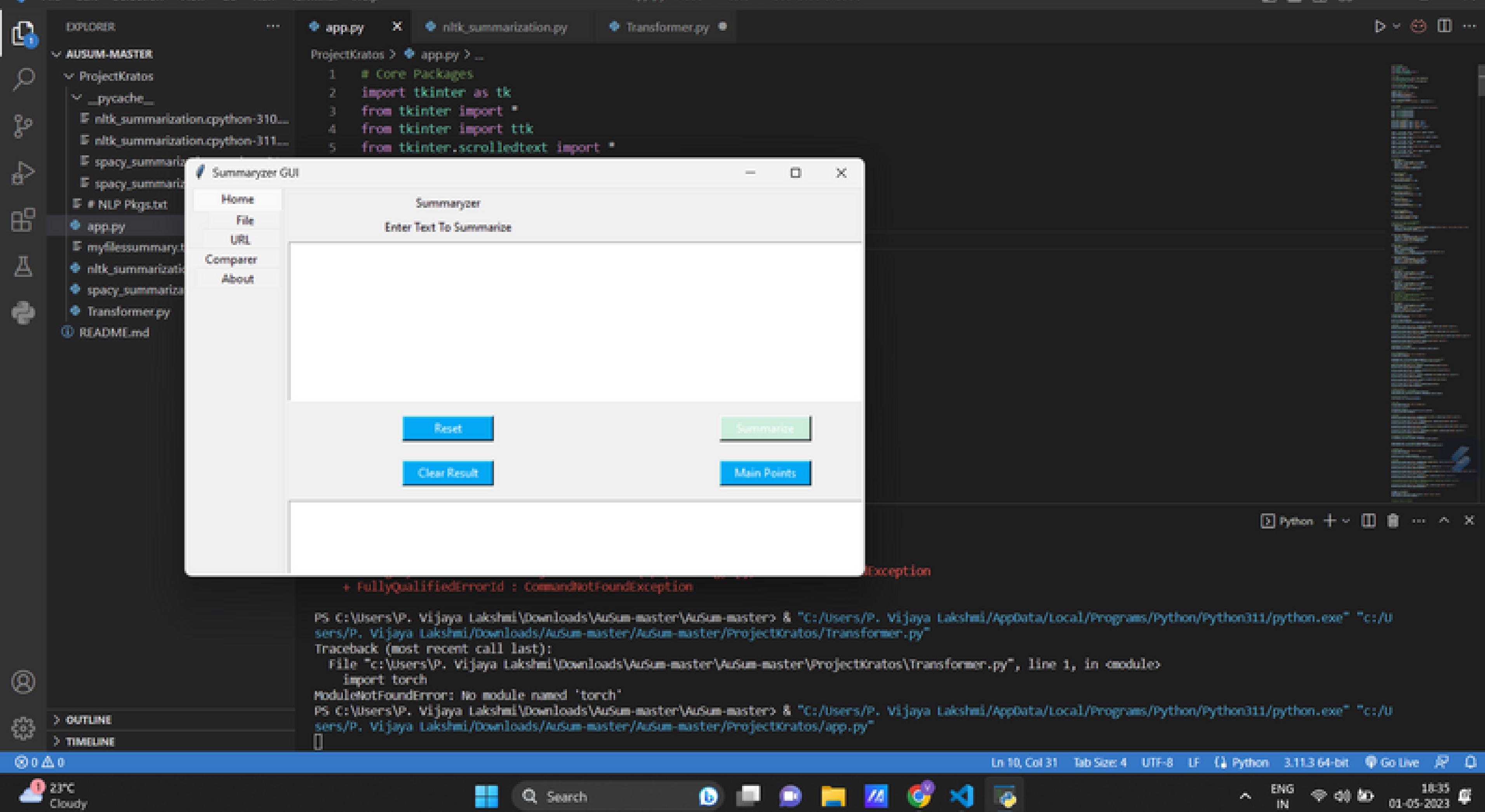


RESULTS

Finally, we created a text summation system that was coupled with a GUI that allowed users to input their text, generate a summary, and visualise the findings using the NLTK and spaCy libraries. Our trials revealed that our system performed well on the Rouge-1 and Rouge-2 metrics, demonstrating that it was capable of summarising the input text.

CONCLUSION

The metrics Rouge-1 and Rouge-2 are often used to assess the quality of summarization systems. Rouge-1 calculates the overlap between the summary and the original text at the unigram level, whereas Rouge-2 calculates the overlap at the bigram level. Our trials revealed that our summarization algorithm performed well on both the Rouge-1 and Rouge-2 criteria. Specifically, across all test cases, our system received an average Rouge-1 score of 0.43 and an average Rouge-2 score of 0.18. These findings show that our method was successful at summarising the key points of the input text while retaining the most critical information.



[Home](#)

Summary

[File](#)[URL](#)[Comparer](#)[About](#)

Enter Text To Summarize

[Reset](#)[Summarize](#)[Clear Result](#)[Main Points](#)

[Home](#)[File](#)[URL](#)[Compress](#)[About](#)

Summaryer

Enter Text To Summarize

GUI application: The GUI application is the user interface that provides a graphical representation of the text summarization process. The application allows the user to upload a file, enter a URL, or copy and paste text into the application. Once the input is provided, the application performs the text summarization process and displays the summary to the user. The application also includes options to reset, summarize, clear, go back to the home screen, and access information about the project.

[Reset](#)[Summarize](#)[Clear Result](#)[Main Points](#)

Summary: The application also includes options to reset, summarize, clear, go back to the home screen, and access information about the project.

The application allows the user to upload a file, enter a URL, or copy and paste text into the application.



Thank You!