Project Report On

Summary Generation using NLP Techniques

By

Ms. Sweta Gupta

Mr. Yash Jobalia

Mr. Isheet Shetty

Under Guidance of

Prof. Anagha Patil

Department of Information Technology

Vidyavardhini's College of Engineering & Technology

University of Mumbai

2021-2022

Vidyavardhini's College of Engineering & Technology
Department of Information Technology

Certificate

This is to certify that the following students

Ms. Sweta Gupta
Mr. Yash Jobalia
Mr. Isheet Shetty

have submitted project report entitled

Summary Generation using NLP Techniques

*as a part of their project-work in partial fulfillment of Semester VIII for the award of degree of **Bachelor of Engineering** in **Information Technology** during academic year 2021-2022.*

Internal Guide : ____________________ (   )

External Guide : ____________________ (   )

Internal Examiner : ____________________ (   )

External Examiner : ____________________ (   )

Dr. Ashish Vanmali
HOD - IT,
VCET, Vasai

Dr. Harish Vankudre
Principal,
VCET, Vasai

Declaration

We declare that this written submission represents our ideas in our own words and where others' ideas or words have been included, we have adequately cited and referenced the original sources. We also declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in our submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Ms. Sweta Gupta (**)**

Mr. Yash Jobalia (**)**

Mr. Isheet Shetty (**)**

Date : ____________________

Acknowledgment

The authors would like to thank our guide Prof. Anagha Patil and all the staff members of Information Technology for their constant support and the encouragement they gave us. We are very thankful for all the faith they had in us and last but not the least, to all our friends and family for their support. Finally, it was the dedication of our team members and enthusiasm helped us to move forward.

Ms. Sweta Gupta
Mr. Yash Jobalia
Mr. Isheet Shetty

Abstract

Nowadays, the world is plummeting by the escalation of the amount of data, with such a bombardment of data wandering aimlessly in a high-tech real-time virtual universe, it appears necessary to summarize extra long texts and serve on a target compendium that can cogently deliver the intended messages. Therefore, synopsis generation is the need of the hour. It is the generation of incisive summaries unescorted by human assistance while conserving the genuine sense of the overlong document. Summarization creation is critical in the goal of saving time. Amongst the various NLP summarization systems and techniques which exists in the world, we have focused on the two broad methods of summarization which are extractive text summarization as well as abstractive text summarization. Each of these methods serves a monolithic purpose depending on the needs of the user and hence our project tries to cater to a larger portion of the population saving loads of time.

Contents

List of Figures

List of Tables

List of Abbreviations

STT	Speech To Text
API	Aplication Programming Interface

Chapter 1

Introduction

1.1 Overview

The human attention span is less than 8 seconds so, if you need to capture someone's attention or highlight an important topic, you need to have a strong headline or summary. A prime example of this phenomenon is that whenever you open a newspaper you glance through the headlines and read through only the ones which have captivating headlines or an interest which aligns with yours. Hence, in this ever growing and vastly expanding world with abundance of data, you need to make sure that the important topics gain your priority attention and you get precise information as well as knowledge from the vast abyss of data. Information is knowledge and knowledge is power. With the huge increase in the amount of data available the fine distinction between the knowledge information and data is thinning. The motive is to use the huge amount of data provided, abstract the required information and try to highlight emphasize on the necessary knowledge that it contains.

1.2 Motivation

The amount of data these days is exponentially increasing through the internet and various other sources. To avoid browsing through these overutilized and long-drawn-out documents and converting them to succinct summary, we stand in need of a tool that can help withdraw summary by clipping of the data in these documents and giving the foremost sentences with pivotal meaning from the prolix document or from a cluster of documents. It is difficult for the human mind to reminisce all this data, so a synopsis generator plays a vital role to save the human effort and time. Our work aims to build a synopsis generator that provides the user the liberty to make a selection from n number of summarization methods in accordance with their needs. Furthermore humans are highly intellectual species who can communicate through sophisticated measures and in order to make a futuristic tool which can utilise these highly efficient communicative methods the tool must have additional features like converting human dictated or vocalised speech to a written readable and legible texts and hence giving them a boon of summarising verbose and elongated speeches and conversations to minimalistic, concise and to the point summaries containing the heart of the entire passage all the while keeping them limited upto a certain precise point.

1.3 Problem Statement

In the modern day, time is not only considered as valuable but is also sometimes regarded as the most important aspect because once spent it cannot be taken back. Moreover no one has the luxury of spending their precious time on reading reports which are more than a few pages long just of the regular day to day meeting or a general conversation. Hence in order to save your valuable time the meeting summarizer can summarise the entire meeting into a few paragraphs and also highlight the entire gist of the content in a few lines. The project's purpose is to learn about natural language processing concepts and construct a text summary machine learning tool that only incorporates the most important information from the material. The process takes place as follows. We offer various methods by which the user can provide input. The different types of inputs can be in various formats such as plain text, document files like text file (.txt), pdf file (.pdf) and word file (.docx), audio files (.wav). When using audio file as an input, we use highly sophisticated STT (speech To Text) algorithms which recognise speech and convert them to text as our first input. These algorithms have the capacity to convert huge audio files with large number of recorded minutes into smaller chunks of audio files slashed at regular intervals of time in order to reduce discrepancy and for better time complexity and management with better results. On the other hand, if we use document files as an input, then the text is extracted from the document file. The converted text will be then encoded and decoded by algorithms by either extractive summarisation or abstractive summarisation based on the users requirements in order to convert to convert the lengthy conversation or meetings into small summaries of the entire topic containing of the most important topics and covering the gist of the text.

1.4 Organization and Contributions of the Report

The last section of the first chapter should have organization of the report. It will briefly tell the reader what lies ahead in the report.

Chapter 2

Review of Literature

This segment elucidates about the techniques that have been used for text. It is one of the branches in natural processing language. Text summarization is bifurcated into categories as Extractive Summarization and Abstractive Summarization. Extractive text summarization is to handpick must-have and imperative sentence from the text. This necessitous and important paragraph can be picked out by using linguistic and statistical features of paragraphs. Abstractive summarization learns the main concept of the long-drawnout document and meaning of the same document or text. It discovers the new concept from the document by using linguistic method by interpreting the text. In preliminary researches, text summarization which is a part of natural processing language was carried out on scientific documents focused on the proposed features like sentence ranking. J.N. Madhuri and Ganesh Kumar R [3] have done analysis on extractive summarization by extracting upmost weighted frequency sentences. In this paper we see that after finishing the pre-processing step they calculate the frequency of each keyword like how habitually that keyword has arisen, from that greatest frequency of the keyword is taken. Then weighted frequency of the word is calculated by dividing frequency of the keywords by maximum frequency of the keywords. In this step, they calculate the sum of weighted frequencies. Finally, the summarizer extracted the high weighted frequency sentences. Similarly, Aakanksha Sharaff, Amit Siddharth Khaire and Dimple Sharma [4] analyze rule-based logic for extractive text summarization in which first the dataset is pre-processed which is used to calculate frequency and position. This is done by using hash map and formulae after which the weight of sentences is calculated which is done by using frequency and position of words formula. The last step is analysis where the calculated values are used to find the mean for that sentence and are fed to the triangular membership function which gives values between 0 and 1 to each sentence and the fuzzy rule is applied. These values are arranged in descending order and they are picked according to the percentage of the original text needed. Finally, the ROUGE score is given to each and every summary to be compared efficiently. In the next paper, Siya Sadashiv Naik, Manisha Naik Gaonkar [5] research about extractive text summarization by feature-based sentence extraction using rule based. The main focus of this paper is to summarize a single document and create its extractive summary. After pre-processing step, each sentence of document is represented as attribute vector of features. Seven features are calculated for each sentence and each feature is given a value from 0 to 1 after normalization. Features considered are Sentence Position, Title Feature, Numerical Value, Keyword Weight, Proper Noun, Sentence-To-Sentence Similarity and Sentence Length. Based on their ratings, all of the sentences are arranged in ascending order. Finally, the extractive summary of the document will be generated and displayed. In the following paper, Kaiz Merchant and Yash Pande [6] use latent semantic analysis approach for creating short summaries on basis of similar words. They use 2 approaches depending on the type of case if it is a criminal case, they used single document untrained approach and for civil case they used multi-document trained approach. They first preprocessed the data which is cleaning it, lemmatization and removing stop words then they pass it through the model and depending on the type of case (civil case or a criminal case) it is decided which process is used. Then using the appropriate model, they generate a summary and finally add sentence selection, in which the final line is always added because it is the judgement passed and hence the final summary output is generated. Followed by Parth Rajesh Dedhia, Hardik Pradeep Pachgade, Aditya Pradip Malani, Nataasha Raul, Meghana Naik [7] the goal of this paper is to highlight and examine current models for abstractive text summarization, as well as to identify topics for future research. This study explains the foundations of RNN models used to construct attention models, as well as a quick overview of feature selection, attention models, pointer mechanisms, and how they work together to produce abstractive text summaries. When numerous documents are supplied to the model, the existing model fails. A system could be invented in the future to preserve the context of the previous document before moving on to the next. Arunlfo and Ledeneva[8] using tf-idf, they suggested a method for term selection and weighting. They created a non-redundant summary using an unsupervised learning system. Mofifiz Mojib Haider and Md. Arman Hossin[9] in it for a single text, this study presents a sentence-based clustering approach (K-Means). They utilised Gensim word2vec for feature extraction, which is designed to extract semantic concepts from documents in the most efficient way possible. Das, D. and Martins, A. F. [10] they demonstrated extractive and abstractive text summarizing techniques for single and multi-document summarization. Various strategies, such as the Naïve Bayes approach, Rich features and Decision trees, Hidden

Markov methods, and Long Linear models, were used to manifest the performance depending on the data set. Partha Mukherjee, Soumen Santra [11] they created a basic application that turns inputted text into synthetic speech and reads it out to the user, which can then be saved as an mp3 file. Daksha Singhal [12] they used a transformer model to reduce n-gram blocking to reduce repetition and successfully presented supervised abstractive summarization on the Switchboard Dataset. The dialogues were summed up Model will be used in future projects. with a pointer generator and training on a transformer for a state-of-the-art hyperparameter tweaking summarizer. Putting the model through its paces and assessing it on several platforms. Google Dialogue Dataset is an example of a dialogue. Narendra Andhale [13] they gave an overview of various text summarization techniques which includes abstractive as well as extractive methods. A few of the methods which they inspect are text summarization with fuzzy logic, text summarization with neural network, which are renowned extractive summarization techniques, as well as rule based method, template based method and tree based method, which are abstractive summarization techniques. Meena S M, Ram Kumar M P [14] they used text frequency ranking sentence prediction which is a combination of both, abstractive as well as extractive text summarization. They used rouge score, a standardized text summarization scoring system and gained good precision Fmeasure and recall with higher precision in abstractive text summarization. M Indu, Kavitha K V [15] they reviewed various text summarization evaluation methods in order to evaluate the informativeness of the automatic summaries by comparing them to human made models. The two evaluation methods could be either intrinsic or extrinsic, each containing of various methods like utility method, BLEU scores, content similarity, the question game, Shannon game and last but not the least, keyword association, which were competitive candidates to rob scoring method. Mhasa Afsharizadeh, Hossein Ebrahimpour-Komleh [16] the pre-prcoessing used by their query oriented text summarization uising sentence extraction technique extracts sentences which contain useful information and display them in the summary. They prepared data using various preprocessing methods such as tokenization, stop word removal, stemming, etc. Then they used feature extraction to extract various features such as normalized sentence length, numerical data, proper noun, topic frequency, headline feature and sentence scoring. They also used rogue scores as a summarization evaluation technique.

Chapter 3

Report on The Present Investigation

This segment elucidates about the techniques that have been used for text summarization.It is one of the branches in natural processing language. Text summarization is bifurcated into categories as follows :

- Extractive Summarization
- Abstractive Summarization.

Extractive text summarization is to handpick must-have and imperative sentence from the text. This necessitous and important paragraph can be picked out by using linguistic and statistical features of paragraphs. Abstractive summarization learns the main concept of the long-drawn-out document and meaning of the same document or text. It discovers the new concept from the document by using linguistic method by interpreting the text. In preliminary researches, text summarization which is a part of natural processing language was carried out on scientific documents focused on the proposed features like sentence ranking.
J.N.Madhuri and Ganesh Kumar.R have done analysis on extractive summarization by extracting upmost weighted frequency sentences. In this paper we see that after finishing the pre-processing step they calculate the frequency of each keyword like how habitually that keyword has arised, from that greatest frequency of the keyword is taken. Then weighted frequency of the word is calculated by dividing frequency of the keywords by maximum frequency of the keywords. In this step, they calculate the sum of weighted frequencies. Finally, the summarizer extracted the high weighted

frequency sentences. Similarly, Aakanksha Sharaff, Amit Siddharth Khaire and Dimple Sharma analysed rule based logic for extractive text summarization in which first the dataset is pre-processed which is used to calculate frequency and position. This is done by using hash map and formulae after which the weight of sentences is calculated which is done by using frequency and position of words formula. The last step is analysis where the calculated values are used to find the mean for that sentence and are fed to the triangular membership function which gives values between 0 and 1 to each sentence and the fuzzy rule is applied. These values are arranged in descending order and they are picked according to the percentage of the original text needed. Finally, the ROUGE score is given to each and every summary to be compared efficiently.
In the next paper, Siya Sadashiv Naik, Manisha Naik Gaonkar [3] research about extractive text summarization by feature based sentence extraction using rule based. The main focus of this paper is to summarize a single document and create its extractive summary. After pre-processing step, each sentence of document is represented as attribute vector of features. Seven features are calculated for each sentence and each feature is given a value from 0 to 1 after normalization. Features considered are Sentence Position, Title Feature, Numerical Value, Keyword Weight, Proper Noun, Sentence-To- Sentence Similarity and Sentence Length. Based on their ratings, all of the sentences are arranged in ascending order. Finally, the extractive summary of the document will be generated and displayed. In the following paper, Kaiz Merchant and Yash Pande [4] use latent semantic analysis approach for creating short summaries on basis of similar words. They use 2 approaches depending on the type of case if it is a criminal case, they used single document untrained approach and for civil case they used multi-document trained approach. They first pre processed the data which is cleaning it, lemmatisation and removing stop words then they pass it through the model and depending on the type of case (civil case or a criminal case) it is decided which process is used. Then using the appropriate model, they generate a summary and finally add sentence selection, in which the final line is always added because it is the judgement passed and hence the final summary output is generated.

3.1 Overview

In this section, we represent our methodology for making an effective text summarizer for english to english text document. The project uses a top-down approach in which each segment is divided into smaller components and each component performs a particular role. The objective of the project is to understand the concept of natural processing language and creating a tool for text summarization containing only main points described in the document.

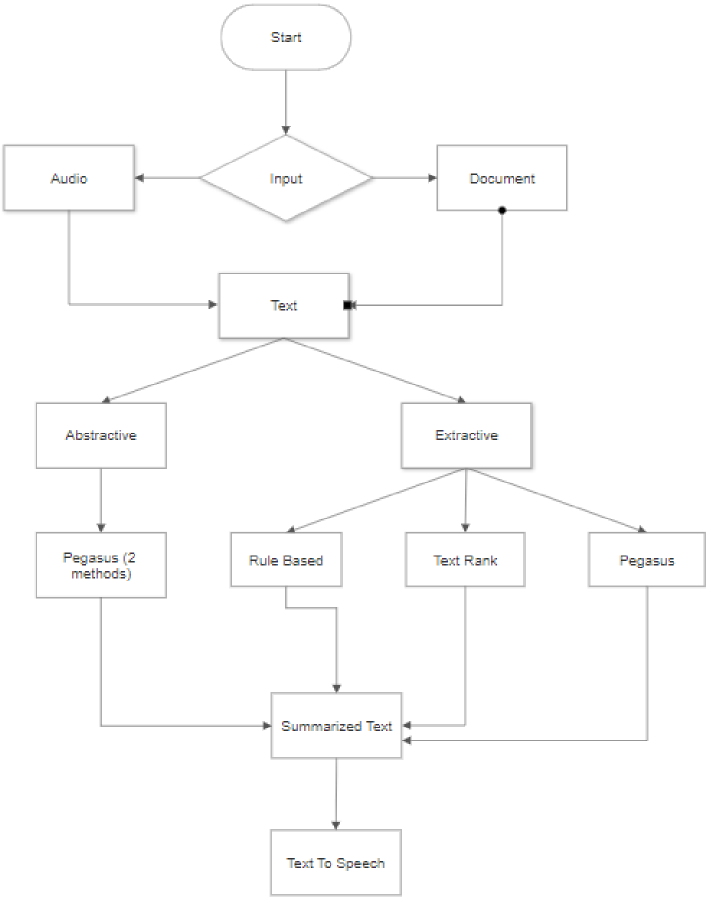


Figure 3.1: Flowchart of the project

For example, the entire project is divided into three phases:

- Speech to text phase
- Summarization phase
- Text to speech phase

The first phase aims at converting the audio files into texts. The output generated from the first phase is further used as an input for the succeeding phase which consists of summarizing these verbose texts into short excerpts along with exterminating discrepancies which may have been introduced in the previous phase due to disturbances in the audio file. Finally, the third phase is an optional feature provided to the users especially with special needs which helps them convert the short summary into an audio output. The users can use the model in various ways depending on their specific requirements and the numerous options provided in each and every phase. Furthermore, each phase also has various options provided to the users which have been discussed in detail below.

3.1.1 Speech To Text Phase

The first phase is the speech to text phase. In this phase, the given audio input file is converted into text. This process is done through passing the audio file through an algorithm which uses google speech recognition in order to identify the audio speech and convert it into text.

3.1.2 Document To Text Phase

The alternative first phase can be the extraction of raw text from the input document which can be in either formats (pdf, word document or text file). This extraction is done through an algorithm which uses the PyPDF2 library to extract from PDF files and docx library to extract text from document files.

3.1.3 Text Summarization Phase

The second phase is the core of the project. It deals with the summarization of the text which were the output from the previous phase. This segment gives the user the ability to select several choices depending on his/her needs of summarization. The first set of options from which the user can select the type of summarization is either abstractive or extractive. In abstractive method, the user gets futher options to either summarize the gist of the text in limited lines or generate an abstractive headline. These methods try to understand the context of the input and generate the desirable output in complete new words either using pegasus xsum or pegasus reddit tifu model. On the counter part, the extractive method gives the user another set of choices consisting of different algorithms based on extractive summarization in which key pivotal lines are extracted from the input text and used as it is for the output. There are several extractive methods listed below which are explained in great detail in the working chapter. The extractive methods used are as follows: Rule based, gensim, text rank and pegasus.

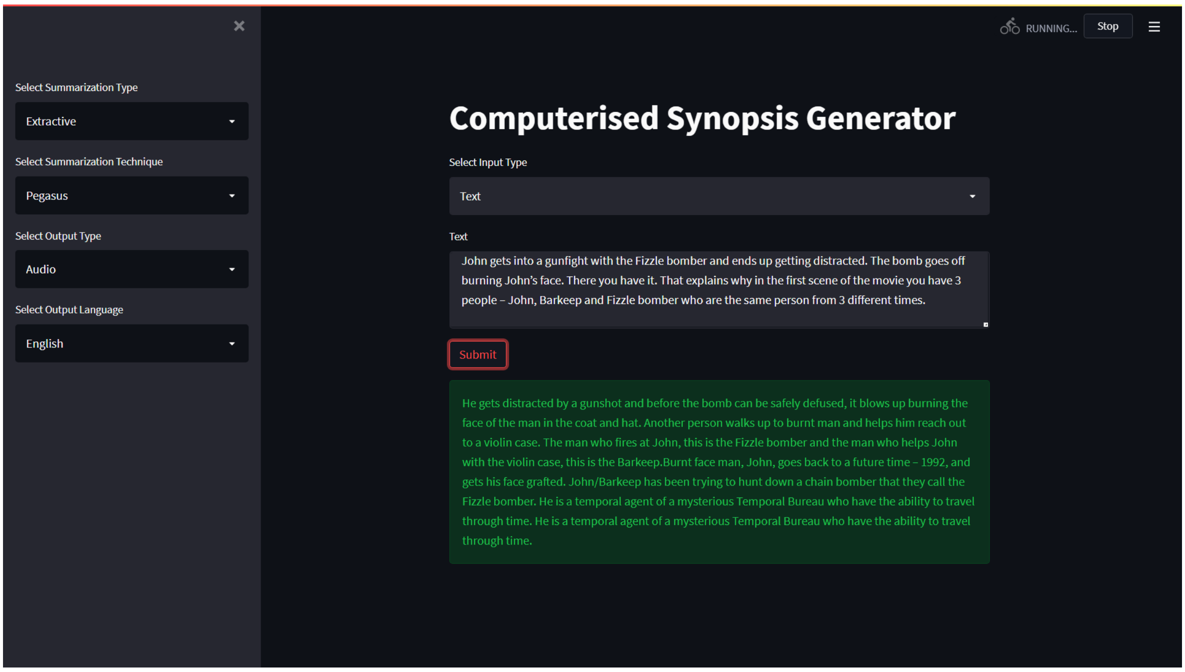
3.1.4 Text To Speech Phase

The final phase consists of the optional functionality which the user can use depending on his/her needs. This phase has the main purpose of converting the summarised text outputs into audio form. This is done by using various libraries like gTTs and pyttsx3 of python language. These additional features help the user in getting the output in any format he/she desires. This can be useful for users with special needs.

Chapter 4

Results and Discussions

As a result of implementing various extractive and abstractive methods of text summarization, each having its own perks and swindles we can conclude in this chapter the efficiencies and the conclusions of enacting various techniques. Following are the implementation results, each compared to the output to the same input.



Select Summarization Type: Extractive

Select Summarization Technique: Pegasus

Select Output Type: Audio

Select Output Language: English

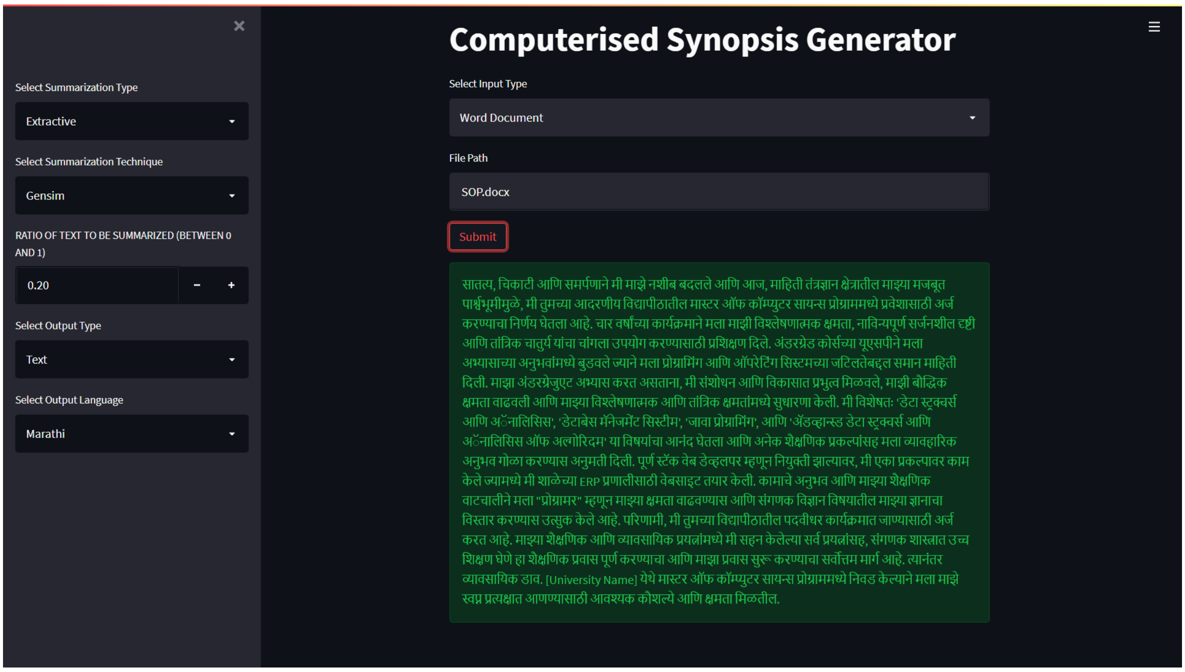
Computerised Synopsis Generator

Select Input Type: Text

Text: John gets into a gunfight with the Fizzle bomber and ends up getting distracted. The bomb goes off burning John's face. There you have it. That explains why in the first scene of the movie you have 3 people – John, Barkeep and Fizzle bomber who are the same person from 3 different times.

He gets distracted by a gunshot and before the bomb can be safely defused, it blows up burning the face of the man in the coat and hat. Another person walks up to burnt man and helps him reach out to a violin case. The man who fires at John, this is the Fizzle bomber and the man who helps John with the violin case, this is the Barkeep.Burnt face man, John, goes back to a future time – 1992, and gets his face grafted. John/Barkeep has been trying to hunt down a chain bomber that they call the Fizzle bomber. He is a temporal agent of a mysterious Temporal Bureau who have the ability to travel through time. He is a temporal agent of a mysterious Temporal Bureau who have the ability to travel through time.

Figure 4.1: Pegasus Extractive Method With Audio Output in English



Computerised Synopsis Generator

Select Summarization Type
Extractive

Select Summarization Technique
Gensim

RATIO OF TEXT TO BE SUMMARIZED (BETWEEN 0 AND 1)
0.20

Select Output Type
Text

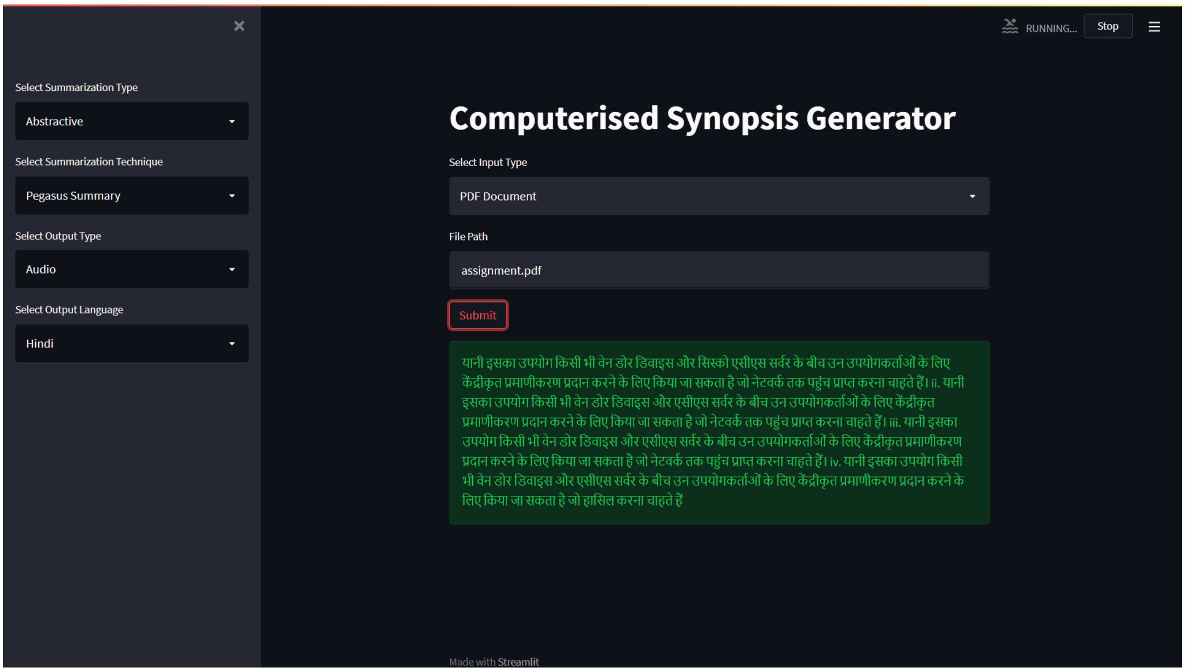
Select Output Language
Marathi

Select Input Type
Word Document

File Path
SOP.docx

सातत्य, चिकाटी आणि समर्पणाने मी माझे नशीब बदलले आणि आज, माहिती तंत्रज्ञान क्षेत्रातील माझ्या मजबूत पार्श्वभूमीमुळे, मी तुमच्या आदरणीय विद्यापीठातील मास्टर ऑफ कॉम्प्युटर सायन्स प्रोग्राममध्ये प्रवेशासाठी अर्ज करण्याचा निर्णय घेतला आहे. चार वर्षांच्या कार्यक्रमाने मला माझी विश्लेषणात्मक क्षमता, नाविन्यपूर्ण सर्जनशील दृष्टी आणि तांत्रिक चातुर्य यांचा चांगला उपयोग करण्यासाठी प्रशिक्षण दिले. अंडरग्रेड कोर्सच्या यूएसपीने मला अभ्यासाच्या अनुभवांमध्ये बुडवले ज्याने मला प्रोग्रामिंग आणि ऑपरेटिंग सिस्टमच्या जटिलतेबद्दल समान माहिती दिली. माझा अंडरग्रेजुएट अभ्यास करत असताना, मी संशोधन आणि विकासात प्रभुत्व मिळवले, माझी बौद्धिक क्षमता वाढवली आणि माझ्या विश्लेषणात्मक आणि तांत्रिक क्षमतांमध्ये सुधारणा केली. मी विशेषतः 'डेटा स्ट्रक्चर्स आणि अॅनालिसिस', 'डेटाबेस मॅनेजमेंट सिस्टीम', 'जावा प्रोग्रामिंग', आणि 'ॲडव्हान्स्ड डेटा स्ट्रक्चर्स आणि अॅनालिसिस ऑफ अल्गोरिदम' या विषयांचा आनंद घेतला आणि अनेक शैक्षणिक प्रकल्पांसह मला व्यावहारिक अनुभव गोळा करण्यास अनुमती दिली. पूर्ण स्टॅक वेब डेव्हलपर म्हणून नियुक्ती झाल्यावर, मी एका प्रकल्पावर काम केले ज्यामध्ये मी शाळेच्या ERP प्रणालीसाठी वेबसाइट तयार केली. कामाचे अनुभव आणि माझ्या शैक्षणिक वाटचालीने मला "प्रोग्रामर" म्हणून माझ्या क्षमता वाढवण्यास आणि संगणक विज्ञान विषयातील माझ्या ज्ञानाचा विस्तार करण्यास उत्सुक केले आहे. परिणामी, मी तुमच्या विद्यापीठातील पदवीधर कार्यक्रमात जाण्यासाठी अर्ज करत आहे. माझ्या शैक्षणिक आणि व्यावसायिक प्रयत्नांमध्ये मी सहन केलेल्या सर्व प्रयत्नांसह, संगणक शास्त्रात उच्च शिक्षण घेणे हा शैक्षणिक प्रवास पूर्ण करण्याचा आणि माझा प्रवास सुरू करण्याचा सर्वोत्तम मार्ग आहे. त्यानंतर व्यावसायिक डाव. [University Name] येथे मास्टर ऑफ कॉम्प्युटर सायन्स प्रोग्राममध्ये निवड केल्याने मला माझे स्वप्न प्रत्यक्षात आणण्यासाठी आवश्यक कौशल्ये आणि क्षमता मिळतील.

Figure 4.2: Gensim Extractive Method With 0.2 Ratio Text Output in Marathi



Computerised Synopsis Generator

Select Summarization Type
Abstractive

Select Summarization Technique
Pegasus Summary

Select Output Type
Audio

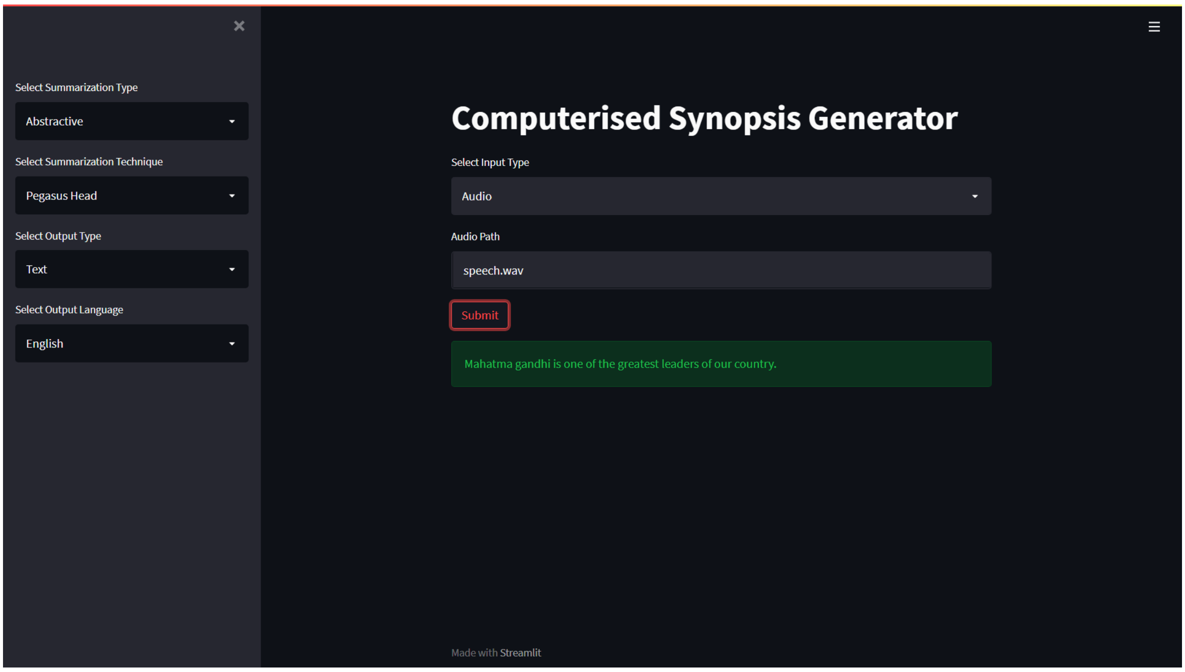
Select Output Language
Hindi

Select Input Type
PDF Document

File Path
assignment.pdf

यानी इसका उपयोग किसी भी वेन डोर डिवाइस और सिस्को एसीएस सर्वर के बीच उन उपयोगकर्ताओं के लिए केंद्रीकृत प्रमाणीकरण प्रदान करने के लिए किया जा सकता है जो नेटवर्क तक पहुंच प्राप्त करना चाहते हैं। ii. यानी इसका उपयोग किसी भी वेन डोर डिवाइस और एसीएस सर्वर के बीच उन उपयोगकर्ताओं के लिए केंद्रीकृत प्रमाणीकरण प्रदान करने के लिए किया जा सकता है जो नेटवर्क तक पहुंच प्राप्त करना चाहते हैं। iii. यानी इसका उपयोग किसी भी वेन डोर डिवाइस और एसीएस सर्वर के बीच उन उपयोगकर्ताओं के लिए केंद्रीकृत प्रमाणीकरण प्रदान करने के लिए किया जा सकता है जो नेटवर्क तक पहुंच प्राप्त करना चाहते हैं। iv. यानी इसका उपयोग किसी भी वेन डोर डिवाइस और एसीएस सर्वर के बीच उन उपयोगकर्ताओं के लिए केंद्रीकृत प्रमाणीकरण प्रदान करने के लिए किया जा सकता है जो हासिल करना चाहते हैं

Figure 4.3: Pegasus Summary Abstractive Method with Audio Output in Hindi



Computerised Synopsis Generator

Select Summarization Type: Abstractive

Select Summarization Technique: Pegasus Head

Select Output Type: Text

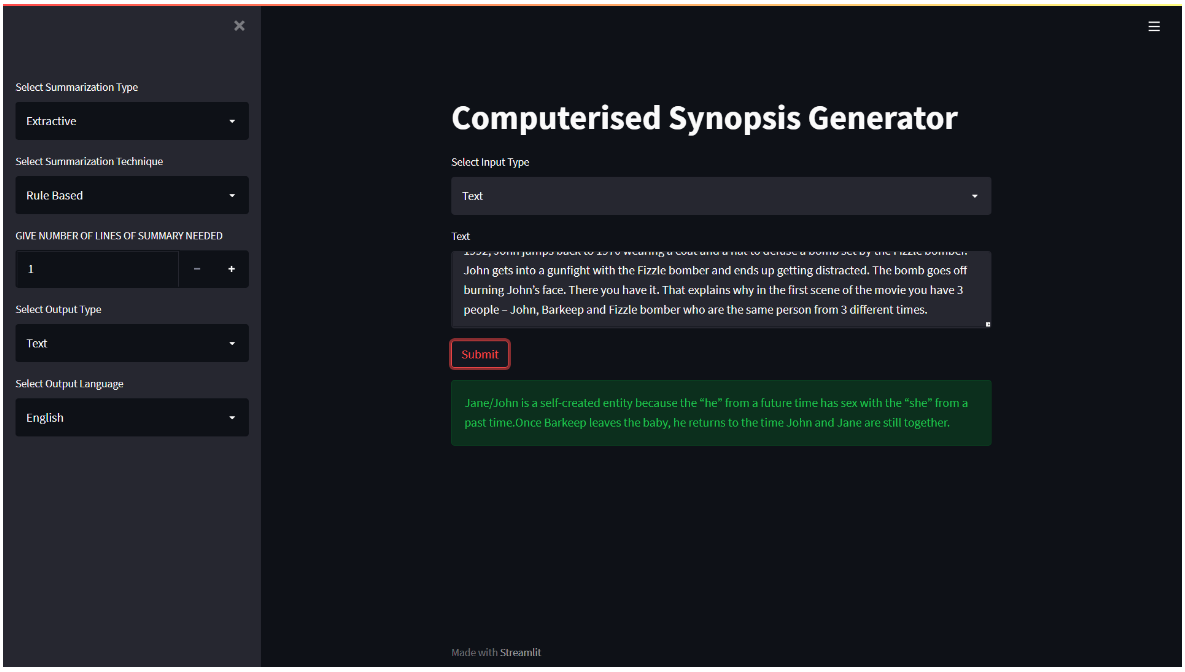
Select Output Language: English

Select Input Type: Audio

Audio Path: speech.wav

Mahatma gandhi is one of the greatest leaders of our country.

Figure 4.4: Pegasus Heading Abstractive Method with Text Output in English



Computerised Synopsis Generator

Select Summarization Type: Extractive

Select Summarization Technique: Rule Based

GIVE NUMBER OF LINES OF SUMMARY NEEDED: 1

Select Output Type: Text

Select Output Language: English

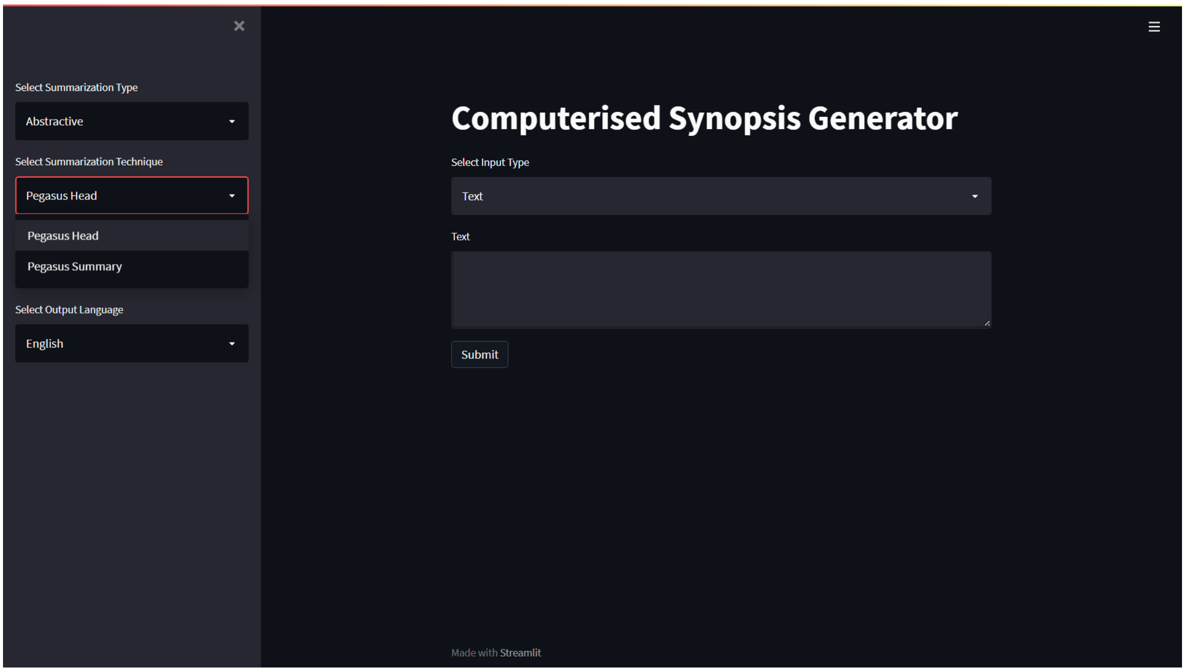
Select Input Type: Text

Text:

1992, John jumps back to 1970 wearing a coat and a hat to defuse a bomb set by the Fizzle bomber. John gets into a gunfight with the Fizzle bomber and ends up getting distracted. The bomb goes off burning John’s face. There you have it. That explains why in the first scene of the movie you have 3 people – John, Barkeep and Fizzle bomber who are the same person from 3 different times.

Jane/John is a self-created entity because the “he” from a future time has sex with the “she” from a past time.Once Barkeep leaves the baby, he returns to the time John and Jane are still together.

Figure 4.5: RuleBased Extractive Method with Text Output of 1 Line



Select Summarization Type
Abstractive

Select Summarization Technique
Pegasus Head

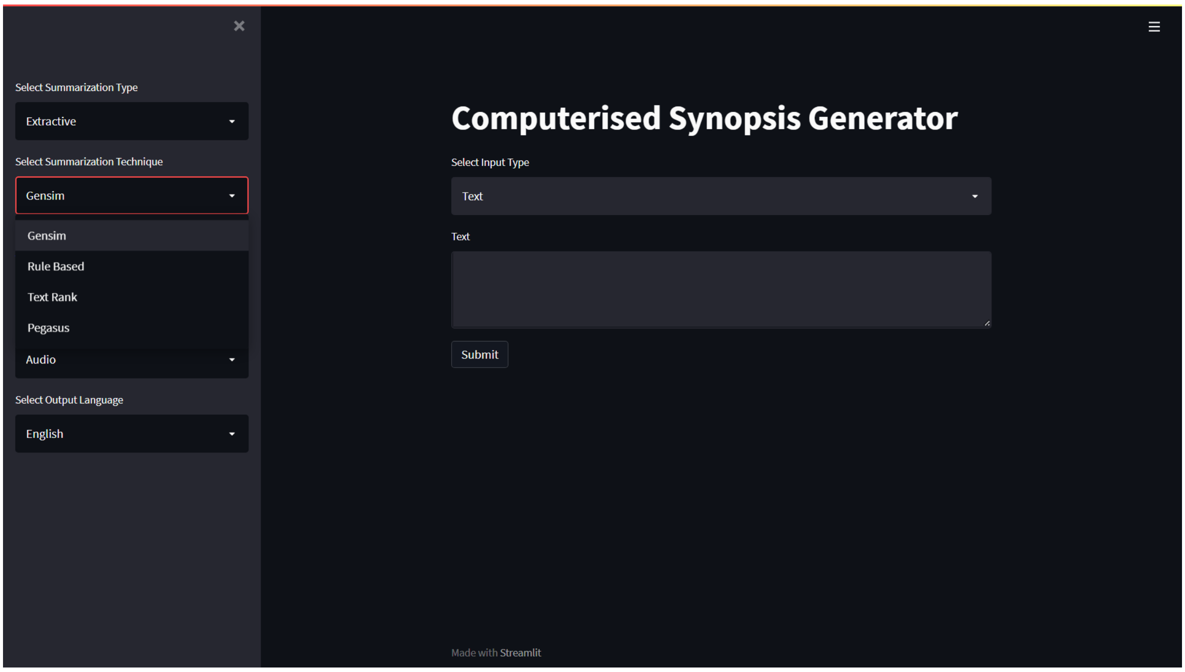
Select Output Language
English

Computerised Synopsis Generator

Select Input Type
Text

Text

Figure 4.6: Abstractive Summarization Types



Select Summarization Type
Extractive

Select Summarization Technique
Gensim

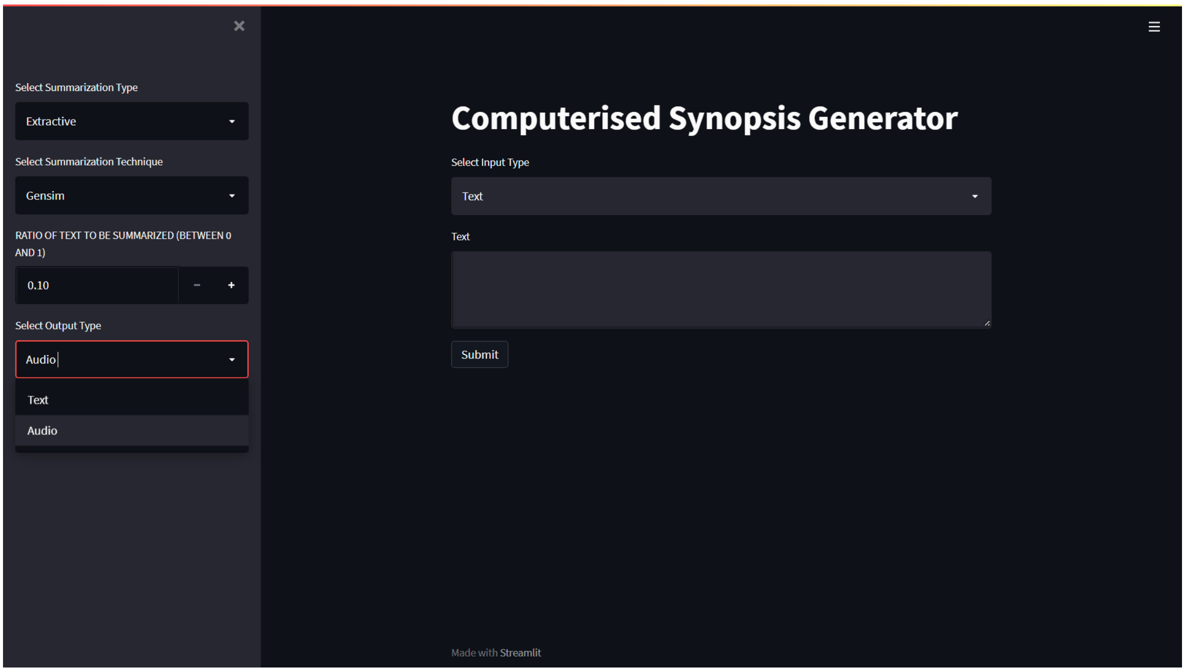
Select Output Language
English

Computerised Synopsis Generator

Select Input Type
Text

Text

Figure 4.7: Extractive Summarization Types



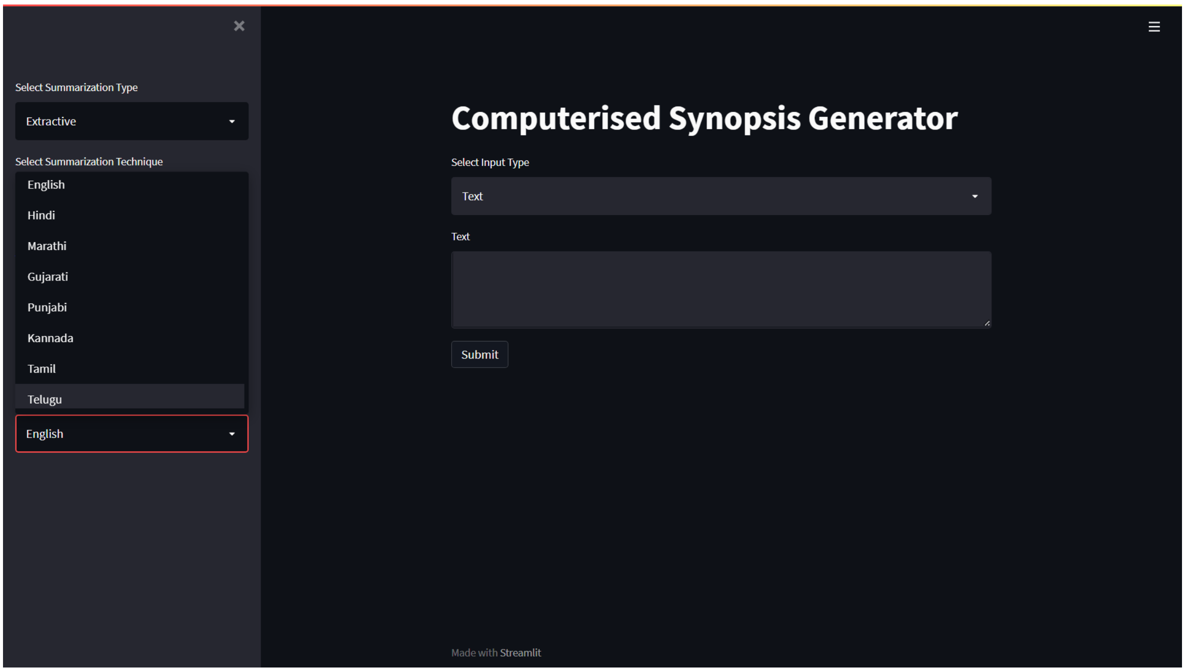
Select Summarization Type
Extractive

Select Summarization Technique
Gensim

RATIO OF TEXT TO BE SUMMARIZED (BETWEEN 0 AND 1)
0.10

Select Output Type
Audio

Figure 4.8: Types of Output Supported



Select Summarization Type
Extractive

Select Summarization Technique
English

Figure 4.9: Output Languages Supported

Chapter 5

Conclusion and Future Work

This is the final chapter of the thesis. The first portion of the Chapter will be a concise summary of the work that has been completed. The results of the logical analysis provided in the Results and Discussions Chapter must be presented and stated in full, with each argument articulated separately. The scope of future work should be explicitly indicated in the chapter's final section.

In recent days, several studies on the creation of summaries from numerous documents have been done. Text summarising generates a summary comprising essential phrases and all pertinent important information from the source material automatically. According to the summary data, extractive and abstractive techniques are among the most common. Text summary has a number of key advantages, including the following: They make reading easier. It saves you time. It makes it easier to remember information. It improves work rate efficiency.The basic objectives of text summarization are as follows:-Optimal topic coverage Maximum readability. There are a few evaluation standards to assure these two aspects. One of these is salience, or the retention of the most significant component. To capture the most significant information from the original document, a summarizer must be programmed. The final summary must be exactly the right length. It should be neither too long nor too short. The structure must be user-friendly. The sentences must be logical and understandable.It should not include any unusual pronouns. The summary as a whole must be balanced. This implies that it must include all of the main components of the paper and, of course, be grammatically perfect throughout. Finally, the phrases should not be repetitive. If a summarizer meets all of these characteristics, it will be able to generate reader-friendly summaries that can benefit us in a variety of ways. Text summary has become an important component of the daily lives of academics, students, and anyone who work with large amounts of text. With the advancement of technology and AIs, it is likely that one day, automatic text summarising may be as excellent and clear as manual text summarization.

5.1 Future Work

For future works, the current prokect works only and only for English language but in future we can expand this in such a way that it can support not only various languages but also inter language summarization.
Apart from the single document summarization, we can enhance this project to summarize multiple documents at the same time.
The time complexity of converting speech to text can be modified to achieve a faster and more accurate level.
The input type for speech to text currently only accepts wav files. This drawback needs to be amended such that it can support various file types.
A wide array of input types such as pdf, docx, jpg, etc can also be used as a supplement to the current input types.

Appendix

Detailed information, lengthy derivations, raw experimental observations etc. are to be presented in the separate appendices, which shall be numbered in Roman Capitals (e.g. “Appendix IV”). Since reference can be drawn to published/unpublished literature in the appendices these should precede the “ Lit. . . .erature Cited” section.

Sample file is shown below...

.1 Python

Python is a widely used high-level programming language for general-purpose programming that was initially introduced in 1991 by Guido van Rossum. Python is an interpreted language with a design philosophy that prioritises code readability (notice the use of whitespace indentation to delimit code blocks rather than curly brackets or keywords) and a syntax that allows programmers to express concepts in fewer lines of code than languages like C++ or Java. Python supports numerous programming paradigms, including object-oriented, imperative, functional programming, and procedural approaches, and has a dynamic type system and automated memory management. It contains a sizable and well-rounded standard library. Python interpreters for a number of operating systems are available, allowing Python code to execute on a wide range of platforms. The standard version of Python, CPython, as well as virtually all of its variant implementations, is open source software with a community-based development strategy. The Python Softwar eFoundation, a non-profit organisation, is in charge of CPython.

.2 Natural Language Processing

Natural Language Processing (NLP) is a branch of computer science that includes aspects of human language and artificial intelligence. Machines utilise this technology to comprehend, analyse, manipulate, and interpret human languages. It aids developers in organising their information in order to execute tasks like translation, automated summarization, Named Entity Recognition (NER), audio recognition, relationship extraction, and topic segmentation.

.3 FFmpeg

FFmpeg is a full audio and video recording, conversion, editing, and streaming solution. It's a command-line video programme that runs on Windows, Mac OS X, and Linux. It can convert between a wide range of video and audio formats.

References

[1] J.N.Madhuri , Ganesh Kumar.R “Extractive Text Summarization Using Sentence Ranking”, Institute of Electrical and Electronics Engineers (IEEE), 2019,

[2] Aakanksha Sharaff, Amit Siddharth Khaire, Dimple Sharma,” Analysing fuzzy based approach for extractive text summarization”, International Conference on Intelligent Computing and Control Systems (ICICCS 2019),

[3] Siya Sadashiv Naik, Manisha Naik Gaonkar,” Extractive Text Summarization by Feature based sentence extraction using rule based”, IEEE International Conference On Recent Trends in Electronics Information Communication Technology (RTEICT),2017,

[4] Kaiz Merchant, Yash Pande ,” NLP Based Latent Semantic Analysis for Legal Text Summarization”,IEEE,2018.

Publications and Awards

Publications

The following list of publications, presented at scientific conferences and published in reputed journals, contains work that is part of this report:

1. Gupta, S. D.; Jobalia, Y. H.; Shetty, I. H.; Patil, A. J., “Summary Generation using NLP Techniques,” *P*roceedings of the 16th INDIACom; INDIACom-2022; IEEE Conference ID: 54597 2022 9th International Conference on “Computing for Sustainable Global Development”, 23rd - 25th March, 2022 Bharati Vidyapeeth’s Institute of Computer Applications and Management (BVICAM), New Delhi (INDIA)
   DOI: 10.1007/s12046-017-0673-1
   URL: `http://bvicam.in/INDIACom/news/INDIACom%202022%20Proceedings/Main/papers/259.pdf`