Summarization using NLP

by Dr. Ashish Vanmali

Submission date: 03-May-2022 08:26AM (UTC-0600)
Submission ID: 1827203664
File name: Summarization_using_NLP_Techniques_6.pdf (786.08K)
Word count: 5732
Character count: 31220

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

The world is plummeting by the increasing 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 the extra large 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 the abstractive text summarization. Each of these methods serves a monolithic purpose depending on the necessity 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 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 over utilized 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 user's 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

This section gives a short glimpse of the contents which are enclosed in each of the chapters lying ahead. The report comprises of several sections which contain detailed information of topics which start from the literature survey in the chapter entailing this one followed by a report on the present investigation. The results are shown in the penultimate chapter with a detailed discussion about them and finally the conclusions which are made and the future work which can be done on this project is stated in the fifth and the ultimate chapter of this thorough report.

Chapter 2: This chapter contains an array of literature papers which have been referred to for information related to the project topics and have been cited at various places in the entire report. The N number of paper with different approaches have been examined at length and given a small gist in this section.

Chapter 3:This section of the report gives a detailed analysis on the investigations and the developments in the current project.The in's and out's of the project with different kinds of compendium techniques, their input methods, their output methods, the different languages that are supported by the project and then their custom additional features like the number of lines for the output are all explained in the different working phases of the project.

Chapter 4:This segment of the report gives a few glimpses of project in full-fledged working condition. There are various screenshots of the project implementing various summarization techniques such as abstractive summarization using pegasus heading and summary algorithms,extractive summarization techniques like rule-based gensim etc. There are also screen captures of the different extractive summarization methods , abstractive summarization methods, types of output supported and the output languages supported.

Chapter 5:This final ultimate chapter of the report not only gives the final conclusion but also discusses the future work which can be done on it. It gives a detailed analysis

of the learnings and the findings on the topics and the various implementations which can be done in future in order to enhance, optimise and deliver a more accurate summarized version of the current project.

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 valuable lines can be picked out by using linguistic and statistical features of paragraphs. Abstractive summarization learns the important concept of the long-drawn out document and the meaning of the same document.The newer concepts in the document are discovered by using various linguistic methodologies by interpreting the text in it . 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 et al. [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 et.al [4] analyze rule-based logic for extractive text summarization in which first the dataset is pre-processed that is used to find out the

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 [5] research about extractive text summarization by feature-based sentence extraction using rule based. The paper primarily focuses on summarizing a single document and creating its extractive summary. After performing the pre-processing step, each and every sentence of document is entirely represented as an attribute vector of its features.Several of the features from the original seven features are calculated for each and every sentence.Each of these features is given a number from 0 to 1 after normalization.Few of the Features which are taken into consideration are as follows, 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 et al. [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 pre-processed 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 method 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 et al. [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 et al. [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 [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 [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 et al. [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 [16] the pre-prcoessing used by their query oriented text summarization using sentence extraction technique extracts sentences which contain useful information and display them in the summary. They prepared data using various pre-processing methods such as tokenization, stop word removal, stemming, etc. Then they used feature extraction to extract various features such as numerical data, sentence scoring, proper noun, topic frequency, normalized sentence length, and headline feature. 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. The discovery of newer concepts from the given document is done by interpreting the texts by using some of the linguistic methods. 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 have done analysis on extractive summarization by extracting upmost weighted frequency sentences. In this paper we see that after finishing the preprocessing step the frequency of each and every keyword is calculated and tried to determined how habitually that keyword has been raised, from that calculation the 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 summation of weighted frequencies. Finally, the summarizer extracted the high frequency sentences. Similarly, Aakanksha Sharaff et al. analysed rule based logic for extractive text summarization in which first the dataset is pre-processed that is used to calculate the 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 [3] research about extractive text summarization by feature based sentence extraction using rule based. The paper primarily focuses on summarizing a single document and creating its summary. After performing the pre-processing step, each and every sentence of the document is entirely represented as an attribute vector of its features. Several of the features from the original seven features are calculated for each and every sentence. Each of these features is given a value from 0 to 1 after normalization. Few of the Features which are taken into consideration are as follows, 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 [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 are representing our methodology for making an effective text summarizer for a 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.
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 Language Translation Phase

This phase is the one of the core phases which gives the language functionality to the project. The english generated output in the previous stages is translated using googletrans module which supports various languages such as hindi, marathi, gujurati, punjabi, telugu, tamil and kannada. The user can select any one of the following languages and the summary text will be converted to that language and given as an output. Some of the languages also have the voice output option to provide more

feasible and supportive features for users with specific needs.

3.1.5 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 different types of 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.

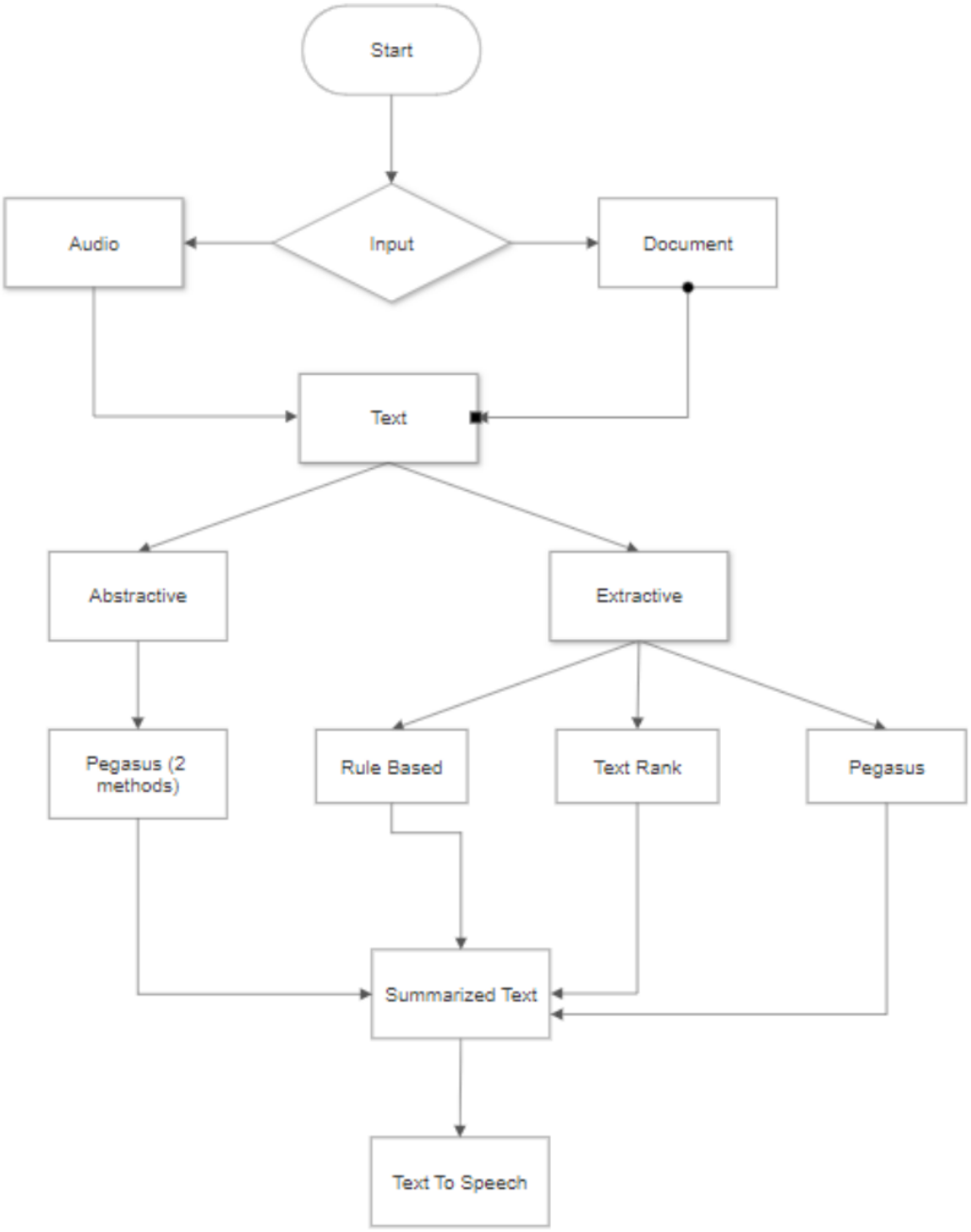

Figure 3.1: Flowchart of the project

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.

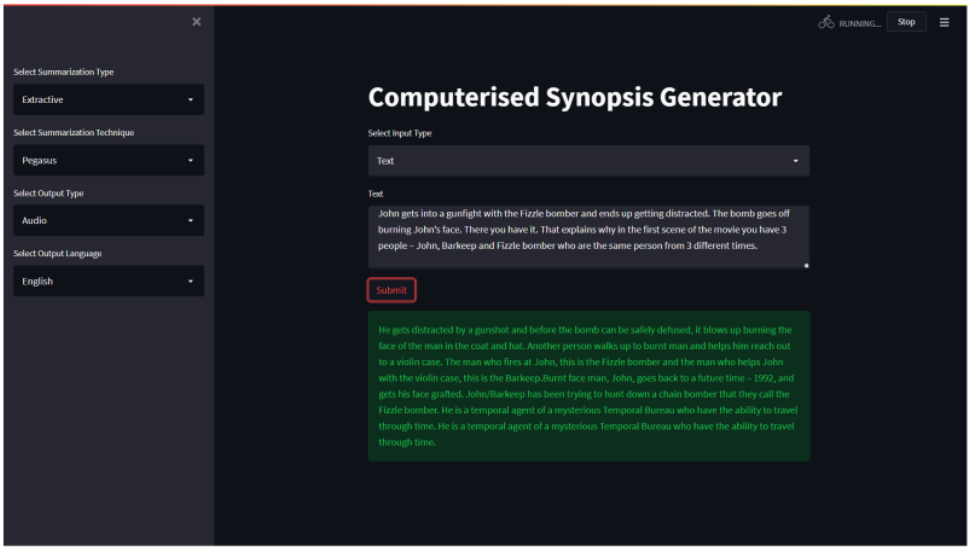

Figure 4.1: Pegasus Extractive Method With Audio Output in English

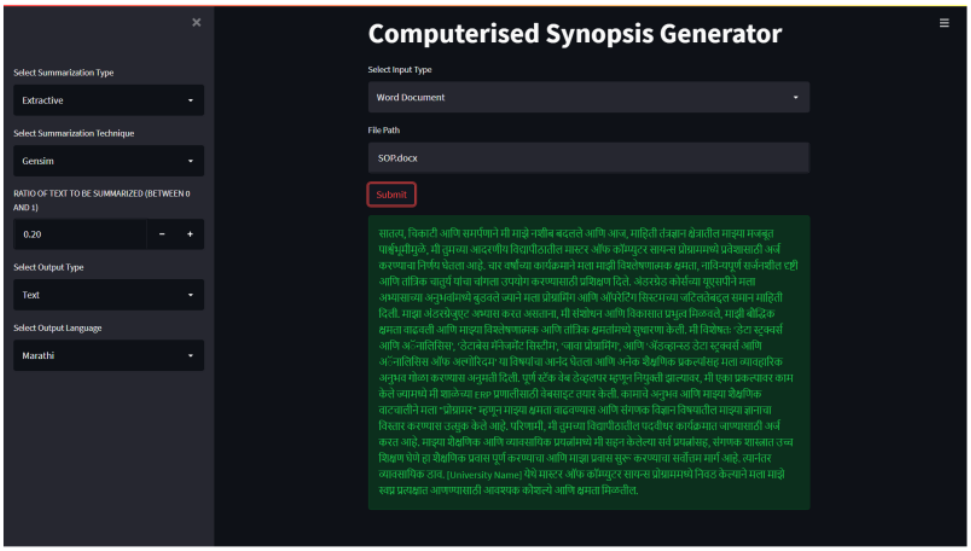

Figure 4.2: Gensim Extractive Method With 0.2 Ratio Text Output in Marathi

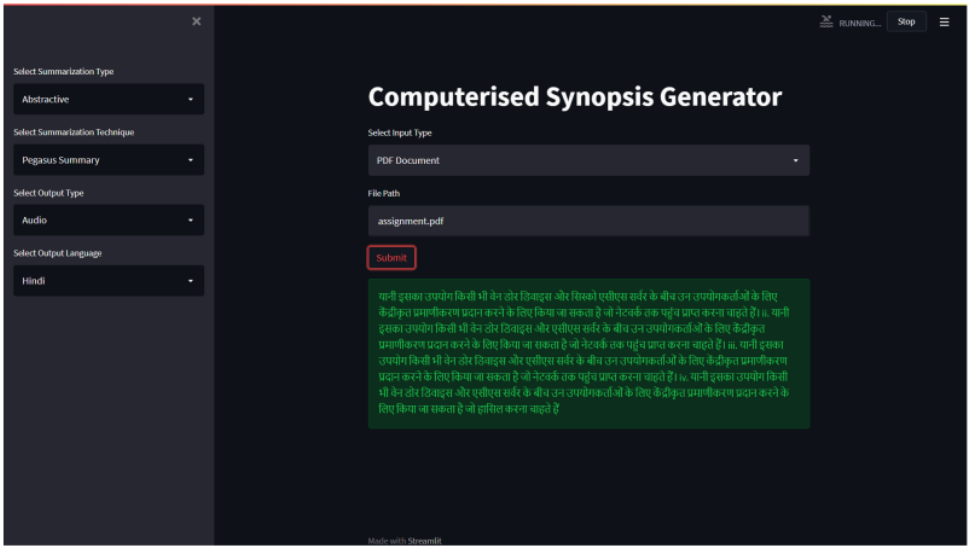

Figure 4.3: Pegasus Summary Abstractive Method with Audio Output in Hindi

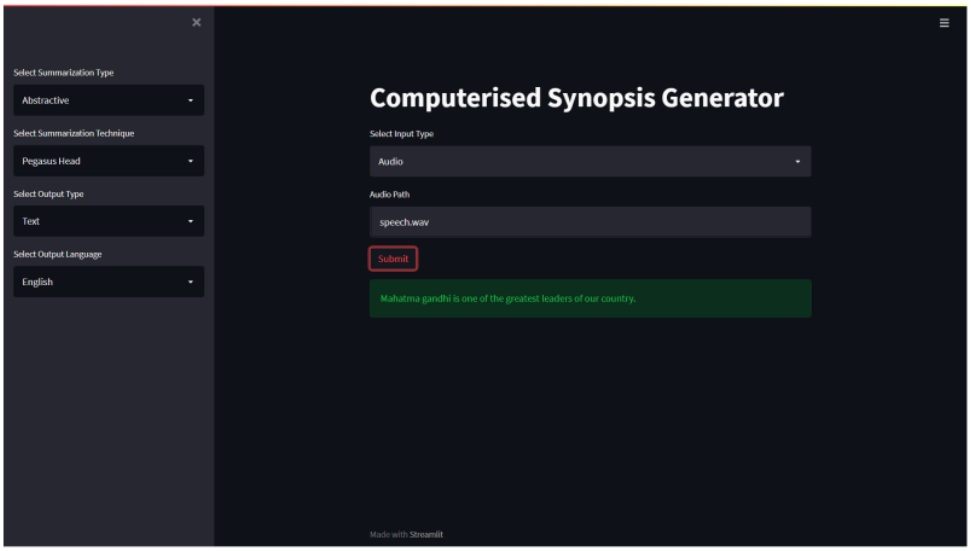

Figure 4.4: Pegasus Heading Abstractive Method with Text Output in English

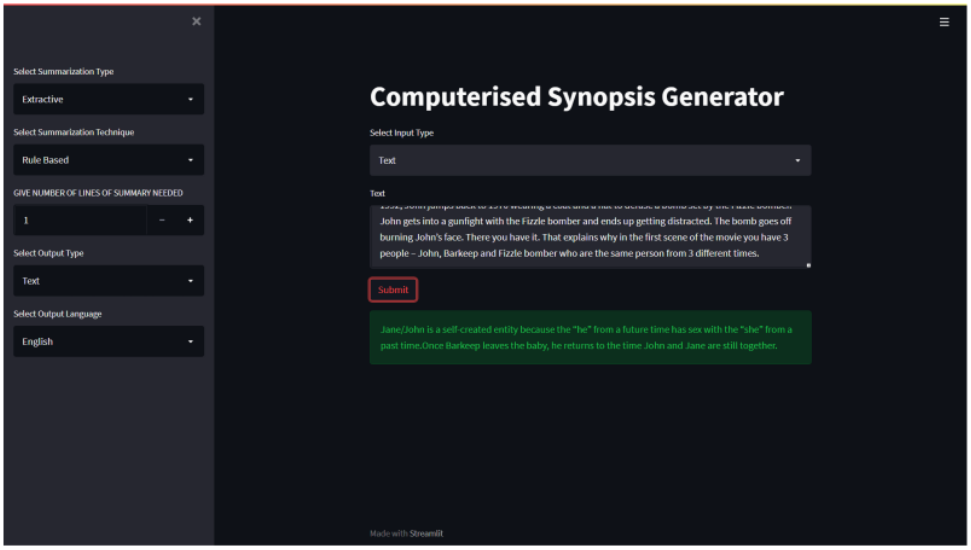

Figure 4.5: RuleBased Extractive Method with Text Output of 1 Line

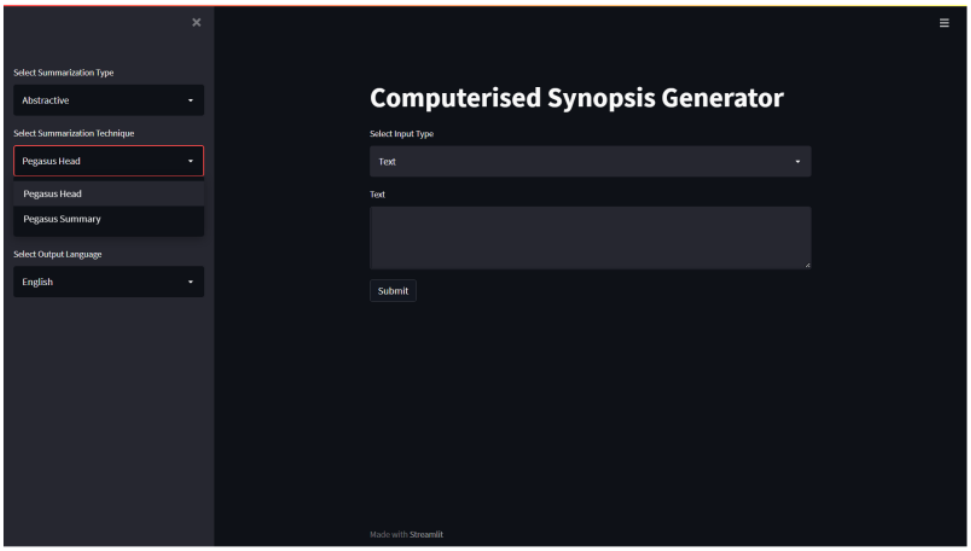

Figure 4.6: Abstractive Summarization Types

Figure 4.7: Extractive Summarization Types

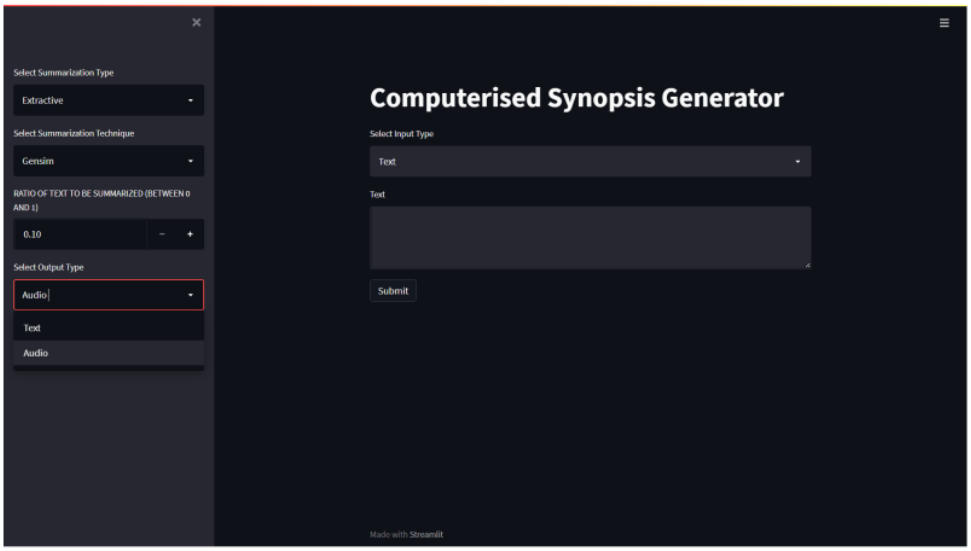

Figure 4.8: Types of Output Supported

Figure 4.9: Output Languages Supported

Chapter 5

Conclusion and Future Work

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 increasing 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 project 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.

Appendix

.1 Python

Python is a high-level programming language for general-purpose programming. It supports numerous programming types like object-oriented, functional programming and procedural programming, and has a dynamic type system and automated memory management. It contains a sizable and well-rounded standard library. Python interpreters are available for many operating systems like Windows, Linux and Mac OS, which allows Python code to execute on a number of platforms. The basic version of Python is CPython, as well as virtually all of its variant implementations, is open source software with a community-based development strategy.

.2 Natural Language Processing

Natural Language Processing (NLP) is one of the branches of various sciences like information technology, computer science as well as data science, artificial intelligence and machine learning that includes aspects of human language and artificial intelligence.It focuses more on the branch of artificial intelligence and data science. 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 software that is used for conversion, editing, and streaming solution of audios and videos. It is a command-line video program that runs on Windows, Mac OS, and Linux. It can not only convert between a wide range of video but also audio formats.

References

[1] U. Hahn and I. Mani, "of Automatic Researchers are investigating summarization tools and methods that," in IEEE Computer 33.11, no. November, pp. 29–36, IEEE, 2000.

[2] K. Sparck Jones, "Automatic summarising: The state of the art," Information Processing Management, vol. 43, pp. 1449–1481, Nov 2007.

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

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

[5] 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,

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

[7] Parth Rajesh Dedhia, Hardik Pradeep Pachgade, Aditya Pradip Malani, Nataasha Raul, Meghana Naik, "Study on Abstractive Text Summarization Techniques", International Conference on Emerging) Trends in Information Technology and Engineering, 2020.

[8] R. A. Garcıa-Hernandez and Y. Ledeneva, "Word sequence models for single text summarization," in Proceedings of the 2nd International Conferences on Advances in Computer-Human Interactions, ACHI 2009, pp. 44–48, IEEE, 2009.

[9] Mofiz Mojib Haider, Md. Arman Hossin," Automatic Text Summarization Using Gensim Word2Vec and K-Means Clustering Algorithm", IEEE, 2020.

[10] D. Das and A. Martins, "A survey on automatic text summarization. literature survey for language and statistics," II Course at CMU, 2007.

[11] Partha Mukherjee, Soumen Santra, Subhajit Bhowmick,"Development of GUI for Text-to-Speech Recognition using Natural Language Processing", IEEE, 2018.

[12] Daksha Singhal, Kavya Khatter, Tejaswini A , Jayashree R," Abstractive Summarization of Meeting Conversations", IEEE, 2020.

[13] Narendra Andhale, L.A. Bewoor, "An Overview of Text Summarization Techniques", IEEE, 2016.

[14] Meena S M, Ramkumar M P, "Text Summarization Using Text Frequency Ranking Sentence Prediction", ICCCSP, 2000.

[15] M Indu, Kavitha K V, "Review on text summarization evaluation methods", IEEE, 2020.

[16] Mahsa Afsharizadeh, Hossein Ebrahimpour-Komleh, Ayoub Bagheri "Query-oriented Text Summarization using Sentence Extraction Technique", ICWR, 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`

Summarization using NLP

ORIGINALITY REPORT

16%	8%	9%	7%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	J.N. Madhuri, R. Ganesh Kumar. "Extractive Text Summarization Using Sentence Ranking", 2019 International Conference on Data Science and Communication (IconDSC), 2019 Publication	3%
2	www.coursehero.com Internet Source	2%
3	Siya Sadashiv Naik, Manisha Naik Gaonkar. "Extractive text summarization by feature-based sentence extraction using rule-based concept", 2017 2nd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), 2017 Publication	2%
4	www.ijert.org Internet Source	1%
5	Submitted to Gitam University Student Paper	1%

6	Submitted to Institute of Technology, Nirma University Student Paper	1 %
7	Parth Rajesh Dedhia, Hardik Pradeep Pachgade, Aditya Pradip Malani, Nataasha Raul, Meghana Naik. "Study on Abstractive Text Summarization Techniques", 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE), 2020 Publication	1 %
8	dspace.bracu.ac.bd:8080 Internet Source	1 %
9	Sonam Gandotra, Bhavna Arora. "Chapter 65 Feature Selection and Extraction for Dogri Text Summarization", Springer Science and Business Media LLC, 2021 Publication	1 %
10	Daksha Singhal, Kavya Khatter, Tejaswini A, Jayashree R. "Abstractive Summarization of Meeting Conversations", 2020 IEEE International Conference for Innovation in Technology (INOCON), 2020 Publication	<1 %
11	Submitted to Rochester Institute of Technology Student Paper	<1 %

12	Submitted to Taylor's Education Group Student Paper	<1 %
13	Submitted to Coventry University Student Paper	<1 %
14	Prathamesh P. Churi, Vaishali Ghate, Kranti Ghag. "Jumbling-Salting: An improvised approach for password encryption", 2015 International Conference on Science and Technology (TICST), 2015 Publication	<1 %
15	Submitted to Sardar Patel College of Engineering Student Paper	<1 %
16	www.onlyinfotech.com Internet Source	<1 %
17	Submitted to University of Sheffield Student Paper	<1 %
18	www.redbooks.ibm.com Internet Source	<1 %
19	utpedia.utp.edu.my Internet Source	<1 %
20	www.uop.edu.jo Internet Source	<1 %
21	documents.mx Internet Source	<1 %

22	Mofiz Mojib Haider, Md. Arman Hossin, Hasibur Rashid Mahi, Hossain Arif. "Automatic Text Summarization Using Gensim Word2Vec and K-Means Clustering Algorithm", 2020 IEEE Region 10 Symposium (TENSYMP), 2020 Publication	<1 %

Exclude quotes	On	Exclude matches	Off
Exclude bibliography	On		