MULTIMEDIA SUMMARIZATION USING SPACY

M. Priyanka[1], K. Sindhuja[2], V. Madhuvani[3], K. Prasoona Sowpthika[4], K. Kranthi Kumar[5]

[1,2,3,4]*Students,* [5]*Associate Professor*
[1]*Department of Information Technology*
[1]*Vasireddy Venkatadri Institute of Technology, Guntur, India.*

ABSTRACT

The main objective of this paper is to propose a summarizing system based on natural language processing to summarize the data that can be in either video, audio or textual format. Video summarization is to generate a short summary of the content of a video by selecting and presenting the most informative or interesting materials for potential users. With the arrival of high-speed Internet and low-cost storage, the amount of data has been generated at a rocket pace, most of it is in the form of visual or video data. This makes them an important information dispensing tool. Searching and analyzing such a large number of videos is an extremely tedious task. Unlike images, where information may be gathered from a single frame, videos need the user to watch the full content, regardless of duration, which presents a challenge to information extraction. Similarly, we save time by extracting crucial information from audio recordings and text. The system focuses on the development of a module using Natural Language Processing with python to summarize the YouTube videos, audio files and textual data.

KEYWORDS - *Multimedia, Natural Language Processing, Spacy, Summarization, Tokenization*

I. INTRODUCTION

In this multimedia era, large number of videos, audios and text are generated and shared across internet 24/7. YouTube is a popular online video sharing platform which includes wide range of purposes from education to entertainment along with other genres like documentaries, speeches and more, ensuring that people of all age groups can easily access its content.[1]

But one of the potential drawbacks observed in such platform is recommending abundant videos for a query which deviates from overall user experience. In many cases users visit misleading or irrelevant information. As a consequence, users unknowingly end up investing their time in the videos which they provide minimal to no valuable information.[2]

We make use of Natural Language Processing (NLP) which trains the computer to get the ability to understand text and spoken words in the same way as human. NLP acts as a game changer for the users who lack in time to learn new languages or become proficient in them. The different applications are language translation, Tokenization, Named Entity Recognition (NER), Optical Character recognition (OCR), Part of Speech Tagging that plays a major role in real world applications.[3]

Summarization can be categorized as Abstractive Summarization and Extractive Summarization. In Abstractive Summarization, new sentences are formed by combining sentences in original text where as in Extractive Summarization, we extract important sentences based on some criteria to generate the relevant summary.[4] The most popular python libraries like NLTK (Natural Language Toolkit) and SpaCy are used for summarization. It is observed that quality of summary is better while using SpaCy rather than NLTK (Natural Language Toolkit).[5]

In this research paper, we propose a model to provide summarization for multimedia inputs like video, audio or text using Natural Language Processing. We make use of SpaCy in python to do tasks such as tokenization using English model like en-core- web-sm. Summarization of videos with and without transcripts is possible by converting video to audio, audio to text and later to summarized text. Other multimedia inputs like audio and text can also be summarized using our model. Audio content that maybe lengthy such as speeches, interviews or podcasts can be summarized or the user can directly provide text to be summarized.

II. OBJECTIVES
- ➤ Develop a user-friendly system with scalable processing capabilities for generating concise and informative summaries.
- ➤ Create summarization algorithms that can effectively work with various media types, including video, audio, and text

EPRA International Journal of Research and Development (IJRD)

which is capable to work with any length.
- ➢ Ensure that the model is versatile and can handle different content genres, such as education, motivational speeches, sports, and entertainment.

III. METHODOLOGY

A. Proposed System Architecture

For summarizing video, user need to give the URL of the video they want to summarize. Using pytube library, we can download YouTube videos using python language. Using pytube we can check whether the URL is valid or not. Along with that it provides additional features like authentication to avoid inappropriate videos from summarization. The downloaded video which is stored in.mp4 format is now converted into .wav file (Waveform Audio File Format) using ffmpeg. It can handle multimedia files by supporting various video and audio formats. Using Speech Recognition models like google, we can convert .wav file to text. Likewise, we can also provide audio as input which will be converted to text or can directly provide text for summarization. Now to obtain summary from original text we will make use spacy library of python along with some procedure as showed in figure2. While working with videos the steps involved are video to audio, audio to text and text to summary.
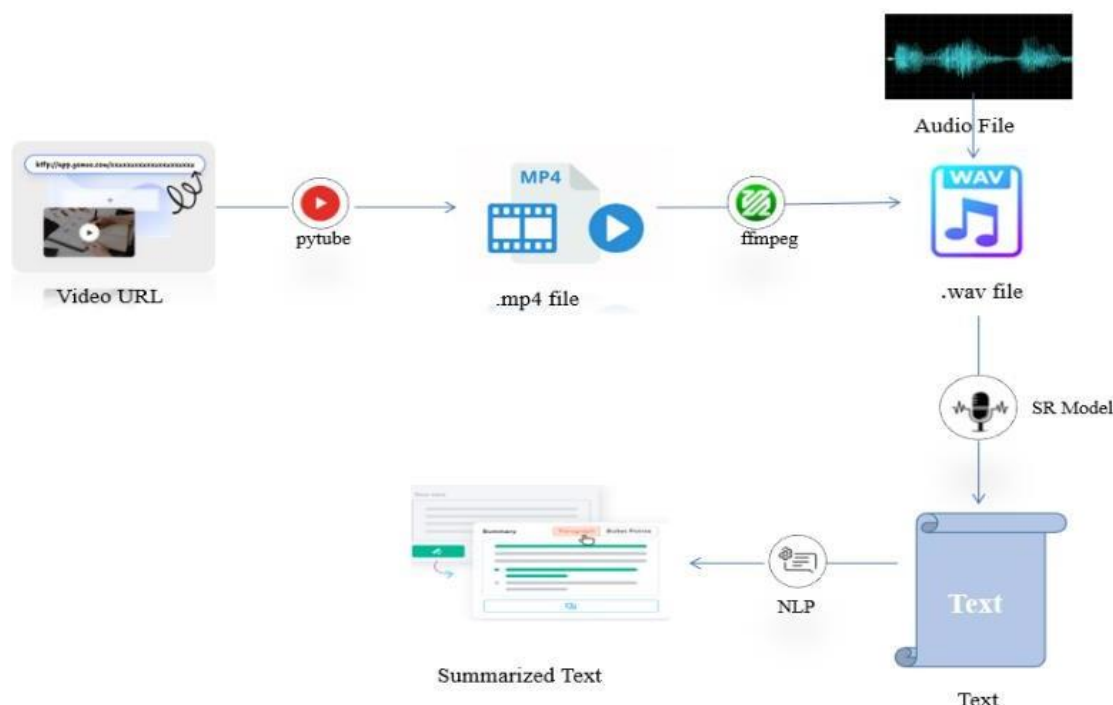

Fig.1: Proposed System Architecture

B. Algorithm for Summarization

Text Cleaning: - In Text Cleaning, we avoid stop words, punctuations and perform altering text case. Stop words like "the", "and", "in" are removed because they do not carry any importance during summarization. We convert the entire text into single format either in lowercase or uppercase so that "Hello"," HELLO" or "hello" are treated as one to avoid unnecessary conflicts during summarization.

Word Tokenization- Word tokenization means breaking the entire text into words or tokens. We use the en-core-web-sm model of SpaCy to perform tokenization. This model considers spaces and punctuation as boundaries to recognize words in sentences.

Word Frequency Table- A frequency Table consists of words present in the text along with its frequency. Frequency means number of times the word has appeared in the text. It helps us to find out important words that might be significant for the summary.

Normalization- In order to bring the word frequencies within a particular range, we perform normalization. The frequency of each word is divided by the maximum frequency observed. After normalization, the frequency falls within the range of 0 to 1.
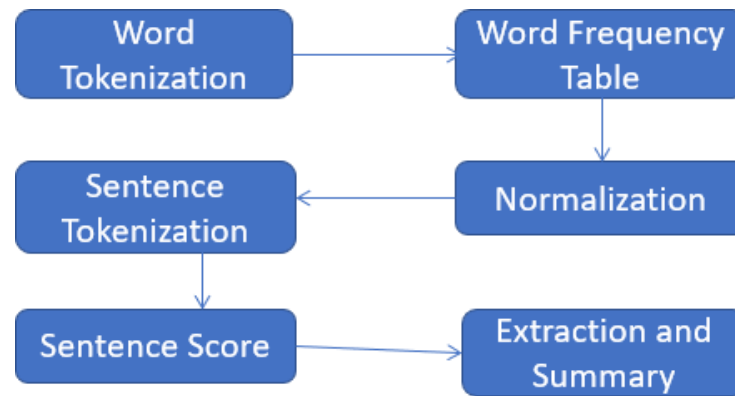
Fig.2: Algorithm For Summarization

Sentence Tokenization– The entire text is divided into sentences using SpaCy which provides effective sentence tokenization. By mainly considering punctuations as sentence boundary, tokenization is performed using SpaCy.

Sentence Scoring– To determine important sentences from the list of sentences obtained through sentence tokenization, we make use of sentence scoring. Sentence scoring is done based on the various factors like word frequency of the words present in the sentence, position of the sentence in the text.

Extraction and Summary- The sentences with high sentence score are selected for summarization. As a result, summary is formed using sentences that provide concise representation of the main points of the text. Informative summary of a longer text.

IV. RESULTS
In our project we have designed user-friendly website which consists of 4 webpages. The initial webpage is homepage which is followed by video input page, audio input page and text input page.

A. Home Page
In Home Page, we have provided 3 options which are video, audio and text. The user will choose the option which they want to summarize.
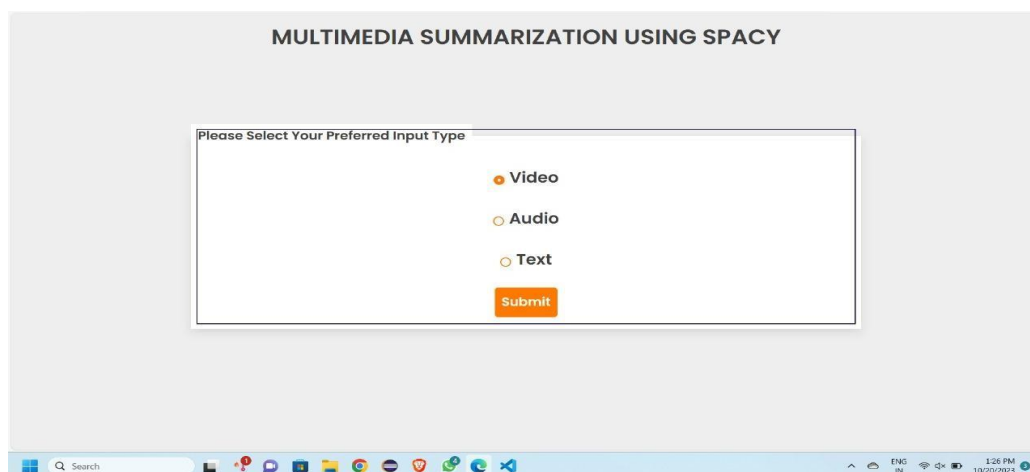
Fig.3: Home Page

B. Video Input Page

The user needs to give the video URL that they want to summarize as input. After clicking submit, the summarized output will be displayed.
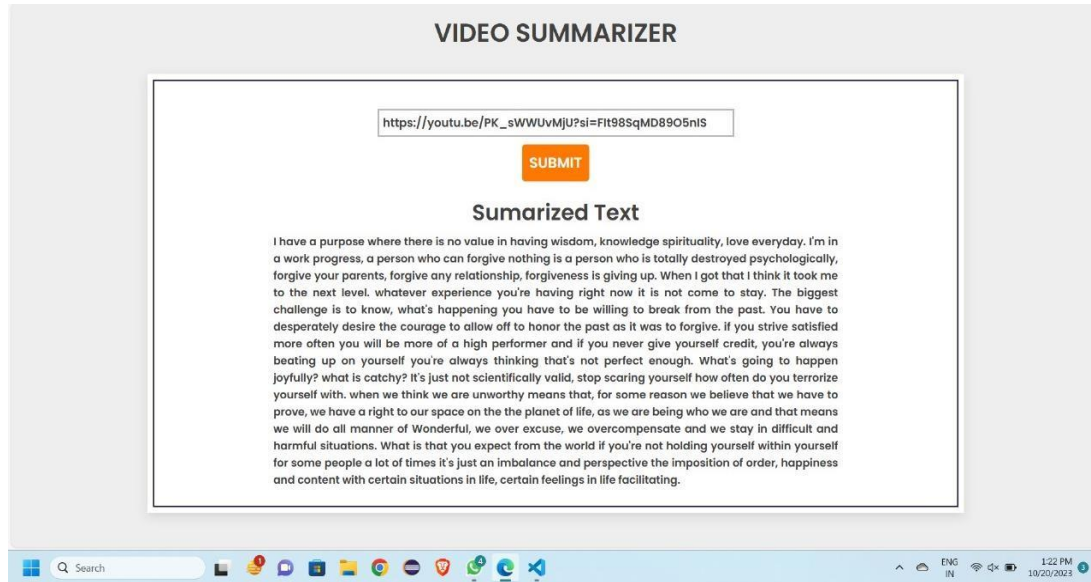

Fig.4: Video Input Page

C. Audio Input File

For summarizing audio file, user need to choose .mp3 or .wav file as input. After clicking submit, the summarized output will be displayed.
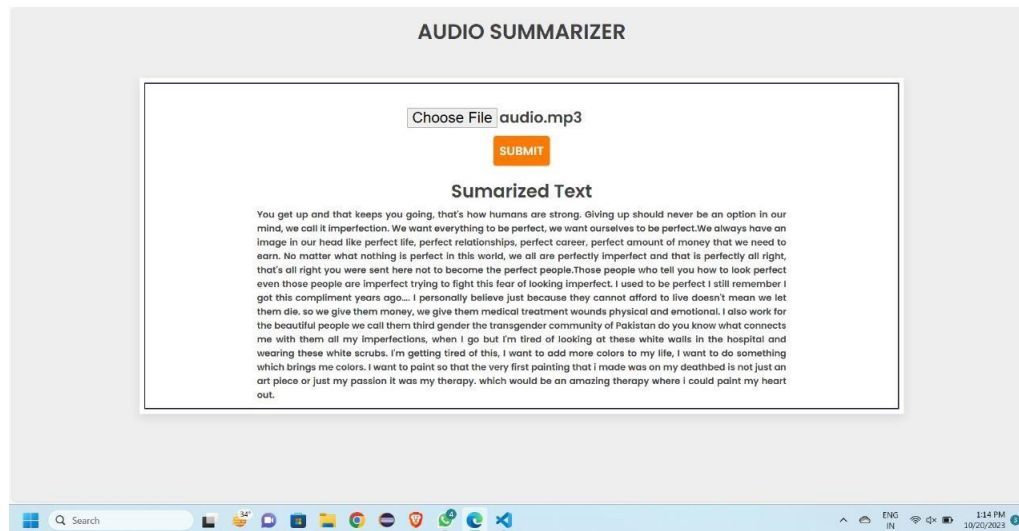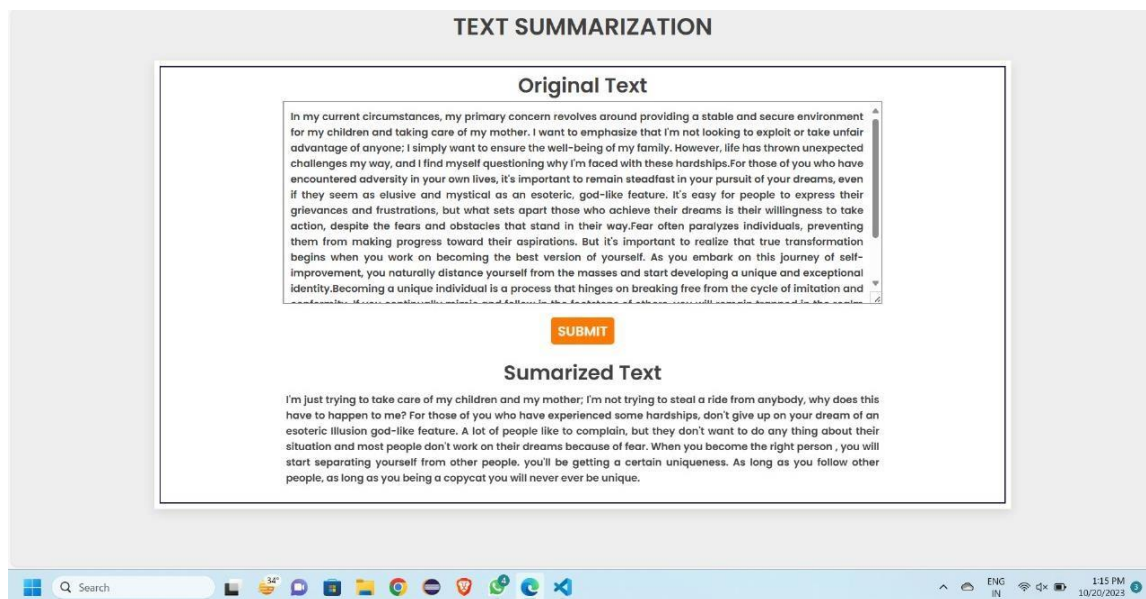

Fig.5: Audio Input Page

D. Text Input Page

If the user is directly working with text instead of video or audio then he can directly enter the text which has to be summarized. As a result, we get the summarized output.

Fig.6: Text Input Page

V. SUGGESTIONS
➢ Parallel Processing: Improve your system for parallel processing to handle a large volume of data efficiently, to create a impact on real world applications where we can summarize multiple videos, audios or text simultaneously.
➢ Output Format: To display output using images along with textual summarized content to provide visual content for better understanding of the end-users.

VI. CONCLUSION
In conclusion, the 'Multimedia Summarization using SpaCy' project represents effective and efficient summarization by harnessing the power of Natural Language Processing. The model accepts a diverse range of inputs, including video URLs, audio files, and text, showcasing its versatility and adaptability for real-world applications. It empowers users to quickly grasp key concepts and insights, effectively utilizing their time in a user-friendly manner to meet the demands of our multimedia-rich digital world which has been generated in both visual, audio and textual format.

VII. REFERENCES
1. *Ilampiray P, Naveen Raju D, Thilagavathy A, Mohamed Tharik M, Madhan Kishore S, Nithin A.S, Infant Raj I. (2023)," VIDEO TRANSCRIPT SUMMARIZER", E3S Web of Conferences 399, 04015.*
2. *Sulochana Devi(B), Rahul Nadar, Tejas Nichat, and Alfredprem Lucas," Abstractive Summarizer for YouTube Videos"*
3. *, Department of Information Technology, Xavier Institute of Engineering, Mumbai, India.*
4. *Diksha Khurana, Aditya Koli, Kiran Khatter, and Sukhdev Singh, (2022)," Natural Language Processing: State of The Art, Current Trends and Challenges", under exclusive license to Springer Science Business Media, LLC, part of Springer Nature 2022, corrected publication 2022.*
5. *Prerna Mishra, Kartik Garg, Naveen Rathi UG, (2023)," Video-to-Text Summarization using Natural Language Processing", International Journal of Advanced Research in Science, Communication and Technology (IJARSCT) Volume 3, Issue 2.*
6. *Abhilasha Sharma, Raghav Aggarwal, Raghav Alawadhi, "A Comparative Study of Text Summarization using Gensim, NLTK, Spacy, and Sumy Libraries" Journal of Xi'an Shiyou University, Natural Science Edition.*
7. *Ishitva Awasthi, Kuntal Gupta, Prabjot Singh Bhogal, Sahejpreet Singh Anand, Prof. Piyush Kumar Soni, (2021)," Natural Language Processing (NLP) based Text Summarization - A Survey", Proceedings of the Sixth International Conference on Inventive Computation Technologies IEEE Xplore.*
8. *SWARANJALI JUGRAN, ASHISH KUMAR, BHUPENDRA SINGH TYAGI, Mr. VIVEK ANAND, (2021)," Extractive*
9. *Automatic Text Summarization using Spacy in Python & NLP", International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE).*
10. *Achyutha Prasad N, Piyush Kumar Pareek, (2022)," Implementation of NLP based automatic text summarization using spacy", publication at: https://www.researchgate.net/publication/362581063, article in International Journal of Health Sciences.*
11. *AAKASH SRIVASTAVA, KAMAL CHAUHAN, HIMANSHU DAHARWAL, NIKHIL MUKATI, PRANOTI SHRIKANT*

KAVIMANDAN, (2022)," Text Summarizer Using NLP (Natural Language Processing)", IRE Journals Volume 6 Issue 1.

12. Pradeepika Verma, Anshul Verma, (2022)," A Review on Text Summarization Techniques", publication at: https://www.researchgate.net/publication/339146200, Article in Journal of Scientific Research.