International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 04 | April - 2023 Impact Factor: 8.176 ISSN: 2582-3930

YOUTUBE VIDEO SPEECH-TO-TEXT SUMMARIZER

Tummanapelli Bhuwaneshwar[1], Thota Jeevika[2] , Shaik Rehana[3], Puppala Archana[4], Mr.B.Kiran Kumar[5]

[1,2,3,4]*B.Tech. Student, Department of Computer Science and Engineering,*

bhuwaneshwarthummanapelli@gmail.com, jeevikathota25@gmail.com,RehanaShaik.reha5@gmail.com, puppalaarchana17@gmail.com

[5]*Assistant Professor, Department of Computer Science and Engineering,*

Nalla Malla Reddy Engineering College, Hyderabad, India

***Abstract*—A YouTube speech-to-text summarizer is a tool that can automatically transcribe the audio content of a YouTube video and create a concise summary of the key points covered in the video. This tool can be useful for quickly reviewing the content of a video without watching it in its entirety, or for individuals who may have difficulty hearing or understanding the audio content of a video. By using advanced speech recognition technology, this tool can accurately transcribe the audio content of a video and then analyze that text to identify the most important concepts and themes covered in the video. Overall, a YouTube speech-to-text summarizer can help users save time and improve their understanding of the content of YouTube videos.**

***Keywords*— Transcripts, Text Summarization, Hugging face tool, REST API, Browser Extension, Automatic speech recognition.**

1. INTRODUCTION

Introduction of the YouTube speech-to-text summarizer project. In today's world, videos have become an increasingly popular way of sharing information and knowledge on various topics. However, it can be time-consuming to watch a full video to understand its content, and for some individuals, it may be difficult to comprehend the audio content of a video. This is where a YouTube speech-to-text summarizer can be incredibly helpful. The goal of this project is to create a tool that can automatically transcribe the audio content of a YouTube video and generate a concise summary of the video's key points.

By leveraging advanced speech recognition technology, the tool will be able to accurately transcribe the audio content and analyze the text to identify the most important concepts and themes covered in the video. This will help users to quickly review the content of a video without having to watch it in its entirety and improve their understanding of the video's content.

This project has the potential to benefit a wide range of users, from students and researchers who need to quickly review a large number of videos on a particular topic, to individuals who may have difficulty hearing or understanding the audio content of a video. By providing a more efficient and accessible way of accessing video content, this tool can help to democratize access to information and knowledge.

2. PROPOSED SYSTEM

The proposed YouTube Speech-To-Text Summarizer system will use speech recognition technology to transcribe the audio content of the video and generate a text transcript. The proposed system is to run the browser extension and the time duration to which the video has to summarize as shown in the below Figures 1, 2, 3. After clicking on the button the summarization process is started. The user can easily know that summarization is started by changing the color of the button to green. Finally, the system will use summarization algorithms to generate a concise summary of the video's content. After generating the summary of the video, The text

is displayed on the pop-up screen as shown in Figure 4.
These are the steps to summarize the video.

Open any browser and load the customized extension. Select the YouTube video to perform the summarization.

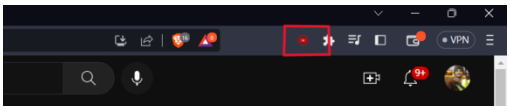

Fig 2.1: The User is to click the icon to run the Extension.

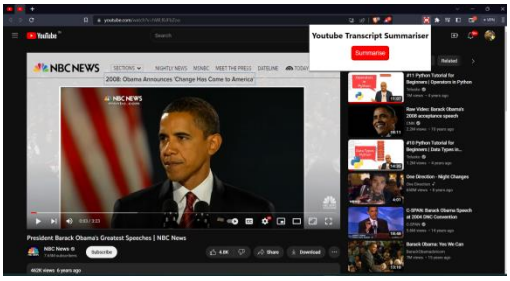

Fig 2.2: Click on the shown pop-up to start the summarization process.

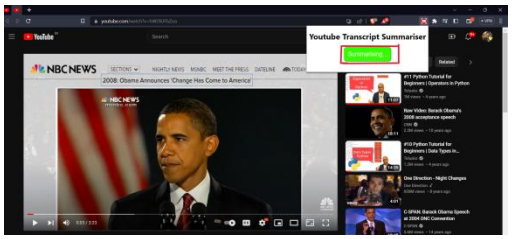

Fig 2.3: When the extension pop button changes to color green the summarization is started.

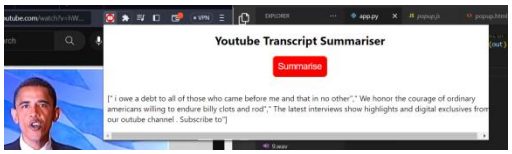

Fig 2.4: The summarization is completed and the final result is displayed on the screen.

3. LITERATURE REVIEW

SURVEY ON ABSTRACTIVE TRANSCRIPT SUMMARIZATION OF YOUTUBE VIDEOS.

The suggested Latent Dirichlet Allocation (LDA) summarizing model is divided into three stages.
The first phase prepares the subtitle file for modeling by deleting stop words and doing other pre-processing tasks. The subtitles are used to train the LDA model. The second step is to generate the list of keywords that will be employed to extract relevant sentences. The summary is prepared based on the list of keywords generated in the third phase.

The quality of LDA-based generated summaries beats that of TF-IDF and LSA summaries. Stream Hover which is a platform for explaining and summarizing transcripts of live-streamed videos was presented. They looked at a neural extractive summarization model that learns vector representations of an audio file and extracts significant observations from subtitles to construct summaries using a vector quantized auto encoder.

YOUTUBE TRANSCRIPT SUMMARIZER

The YouTube videos are the data for our proposed videotape summarization algorithm. Using the link, YouTube paraphrases API will prize mottoes from that particular videotape.
Downloading videos from YouTube is delicate. To do so first, we've to copy the link of the videotape we want to download also bury the link in the YouTube videotape downloader website.
This system of downloading is time-consuming. Pytube is a feather-light, reliance-free Python library that is used to download YouTube videos fluently.
This can be achieved with just one or two lines of the law.
Pytube library creates the object of the YouTube module by passing a YouTube link of the videotape as the parameter. Also, it gets the applicable extension and resolution of the

videotape. The name of the train can be kept grounded on stoner convenience. After that, download the train using the download function of the pytube library. This download function takes only one parameter the position where downloaded lines need to be saved.

In Python, URLs are handled using the urllib system, which calls a particular URL and handles results after visiting the URL. We're using urllib to get the title of a videotape using the YouTube link.

LSA Algorithm: Latent Semantic Analysis (LSA) is an unsupervised approach fashion in Natural Language Processing. It's an Algebraic-Statistical system that excerpts the features of the rulings that cannot be directly mentioned. These features are essential to data but aren't original features of the dataset.

4. METHODOLOGY

The methodology for the YouTube speech-to-text summarizer project involves several steps:

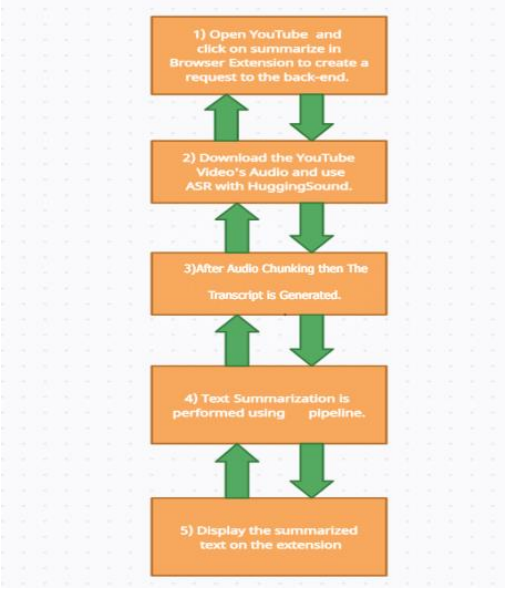

Fig 3.1: Workflow of YouTube Speech-To-Text Summarizer

Run Extension: The first step is to open any browser and load the customized extension into the browser. Now select the video on YouTube and start summarizing by clicking on the Extension.

Download File: The second step is to download the selected video's audio. Some space is required to download audio files. This file is in mp4 format.

Pytube is a lightweight, dependency-free Python library that is used to download YouTube videos easily. This can be achieved with just one or two lines of code.

Speech recognition: The next step is to transcribe the audio content of each video using speech recognition technology. This technology should be advanced enough to accurately transcribe the audio content even with variations in accents, dialects, and speech patterns. Automatic Speech Recognition [ASR] technology is used with the Hugging Sound tool.

Audio Chunking: The download audio file is chunked into finite parts according to the video length. For every single chunked part of file is separately Transcript is generated. And combine all the audio's files text.

Text analysis: Once the audio content has been transcribed, the tool will analyze the text to identify the most important concepts and themes covered in the video. This could involve techniques such as natural language processing and machine learning to automatically identify keywords and topics.

Summary generation: Using the results of the text analysis, the tool will generate a concise summary of the video's key points. The summary should be brief and easy to understand, capturing the most important information covered in the video.

Evaluation: The final step is to evaluate the accuracy of the tool's transcriptions and summaries. This could involve comparing the tool's output to human-generated summaries and measuring the precision and recall of the tool's text analysis algorithms.

Overall, the methodology for the YouTube speech-to-text summarizer project involves

using advanced speech recognition and text analysis technologies to automatically transcribe and summarize the audio content of YouTube videos, to improve accessibility and efficiency of accessing video content.

RESULTS

The results for the proposed system are shown below:

Video Title	Total time of video (minutes)	Summary Time requested by the user (minutes)	Processing time (Minutes)	Memory usage (MB)
The Batman| Cast Interview |DC	5	2	2	50
#3Python Tutorial for Beginners |Getting Started with Python	15	5	6	112
Top 5 New Netflix Original Series Released In 2023| Best Netflix Web Series	7	3	3	90

TABLE 3.1.1: Table of results.

Table 1 shows the result of the proposed model used to obtain the video summarization using the speech of the video. The algorithm gives less than 5 seconds of error video as output for the input given.

Video Link	Total time of video (minutes)	Summary Time requested by the user (minutes)	Summary time formed by our algorithm (minutes)
https://www.youtube.com/watch?v=h-JVjs9AAmQ	5	2	2min 33sec
https://www.youtube.com/watch?v=DWgzHbglNIo&list=PLsyeobzWxl7poL9JTVyndKe62ieoN-MZ3&index=4	15	5	4min 47sec
https://www.youtube.com/watch?v=OHH4cvB6Wcs	7	3	2min 51sec

TABLE 3.1.2: Table of processing time and Memory usage.

Table 2 shows the results with memory usage and processing time. Here the memory usage and processing time results depend on the total time of the input video and the summary time requested by the user.

5. DISCUSSION

Increased accessibility: The tool can make video content more accessible to individuals who may have difficulty hearing or understanding the audio content of a video. Improved efficiency: The tool can save users time by allowing them to quickly review the content of a video without watching it in its entirety.

Democratization of knowledge: The tool can help to democratize access to information and knowledge by providing a more efficient and accessible way of accessing video content.

However, there are also several challenges and limitations to consider, including:

Accuracy: The accuracy of the tool's transcriptions and summaries will depend on the quality of the speech recognition and text analysis technologies used. Variations in accents, dialects, and speech patterns could also affect accuracy.

Context: The tool may have difficulty understanding the context of certain words or phrases, which could lead to inaccuracies or misunderstandings in the generated summaries.

Legal and ethical considerations: Depending on the content of the videos, there may be legal and ethical considerations to take into account, such as copyright infringement or privacy concerns.

User experience: The tool's effectiveness will depend on its usability and user experience, which will need to be carefully designed and tested to ensure user satisfaction.

Overall, while the YouTube speech-to-text summarizer project has the potential to be a valuable tool for improving accessibility and efficiency in accessing video content, careful consideration of accuracy, context, legal and ethical considerations, and user experience will be important in its development and implementation.

6. CONCLUSION

The YouTube speech-to-text summarizer using a browser extension is a promising tool for improving accessibility and efficiency in accessing video content. By leveraging advanced speech recognition and text analysis technologies, the extension can automatically transcribe and summarize the audio content of YouTube videos in real time.

One of the key advantages of this approach is the ability to use the tool within the context of the YouTube platform, without having to leave the website or use a separate application. This can help to streamline the user experience and make the tool more accessible to a wider range of users.

However, there are still challenges and limitations to consider, including accuracy, context, legal and ethical considerations, and user experience. These factors will need to be carefully considered and addressed in the development and implementation of the Chrome extension to ensure its effectiveness and user satisfaction.

Overall, the YouTube speech-to-text summarizer using a Chrome extension is an exciting development in the field of speech recognition and text analysis technologies. By continuing to explore and develop innovative approaches to improving accessibility and efficiency in accessing video content, we can help to democratize access to information and knowledge for a wide range of users.

REFERENCES

[1] Alrumiah, S. S., Al-Shargabi, A. A. (2022). Educational Videos Subtitles' Summarization Using Latent Dirichlet Allocation and Length Enhancement. CMC-Computers, Materials & Continua 70(3), 6205–6221.

[2] Sangwoo Cho, Franck Dernoncourt, Tim Ganter, Trung Bui, Nedim Lipka, Walter Chang, Hailin Jin, Jonathan Brandt, Hassan Foroosh, Fei Liu, "StreamHover: Livestream Transcript Summarization and Annotation", arXiv: 2109.05160v1 [cs.CL] 11 Sep 2021

[3] S. Chopra, M. Auli, and A. M. Rush, ''Abstractive sentence summarization with attentive recurrent neural networks,'' in Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics Hum. Lang. Technol., June 2016, pp. 93–98.

[4] Ghadage, Yogita H. and Sushama Shelke. "Speech to text conversion for multilingual languages." 2016 International Conference on Communication and Signal Processing (ICCSP) (2016): 0236-0240.

[5] Pravin Khandare, Sanket Gaikwad, Aditya Kukade, Rohit Panicker, Swaraj Thamke, " Audio Data Summarization system using Natural Language Processing, " International Research

[6] Journal of Engineering and Technology (IRJET) Volume 06, Issue 09, [September - 2019], e-ISSN: 2395-0056; p-ISSN: 2395-0072.

[7] Hugo Trinidad and Elisha Votruba, " Abstractive Text Summarization Methods "

[8] C. M. Taskiran, Z. Pizlo, A. Amir, D. Ponceleon, and E. J. Delp, "Automated video program summarization using speech transcripts," IEEE Transactions on Multimedia, vol. 8, no. 4, pp. 775–790, Aug. 2006, doi: 10.1109/TMM.2006.876282.

[9] You J., Liu G., Sun L., and Li H. A multiple visual models-based perceptive analysis framework for multilevel video summarization. IEEE Trans. Circuits Syst. Video Tech., 17(3), 2007.

[10] Ferman A.M. and Tekalp A.M. Two-stage hierarchical video summary extraction to match low-level user browsing preferences. IEEE Trans. Multimedia, 5(2):244–256, 2003