



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJIRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 12, Issue 2, February 2023

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.118

 9940 572 462

 6381 907 438

 ijirset@gmail.com

 www.ijirset.com

Survey on YouTube Transcript Summarizer

**Prof. V. R. Parkhede, Renuka Nemane, Ankita Wankhade, Nikita Patil, Tejas Tathod,
Mayur Pande**

Department of Computer Science and Engineering, Mauli Group of Institutions College of Engineering and
Technology, Shegaon, India

ABSTRACT -This project proposes a video summarizing system based on natural language processing (NLP) and Machine Learning to summarize the YouTube video transcripts without losing the key elements. Now-a-days the quantity of videos available on web platforms is steadily expanding and the content is made available globally, primarily for educational purposes. Additionally, educational content is available on YouTube, Facebook, Google, and Instagram. A significant issue of extracting information from videos is that unlike an image, where data can be collected from a single frame, a viewer must watch the entire video to grasp the context. This study aims to shorten the length of the transcript text of the given video. The suggested method involves retrieving transcripts from the video link provided by the user and then summarizing the text by using Hugging Face Transformers and Pipelining. The built model accepts video links and the required summary duration as input from the user and generates a summarized transcript as output. According to the results, the final translated text was obtained in less time when compared with other proposed techniques. Furthermore, the video's central concept is accurately present in the final text without any deviations.

KEYWORDS: Machine learning, Natural Language Processing, Abstractive text summarization

I. INTRODUCTION

Video transcript summarizer has got a lot of scope in today's world. It highlights the important topics from the video. Every minute, 300 hours of YouTube videos are uploaded. Almost one-third of the YouTube viewers in India access videos on their mobiles and spend over 48 hours a month on the website, a Google study said. It is frustrating and time consuming to search for the videos that contain the information we are actually looking for.

If you wish to find a video to get any important information about a topic, it is a very difficult task to achieve as most of the videos are filled with insignificant buffer material. In most of the cases, the overall intent is to obtain some form of quality information from the video. This project brings forward a video summarization system based on Natural Language Processing and Machine Learning to generalize YouTube video transcripts for abstractive text summarization without losing the main elements and content. This project focuses on reducing the length of the script for the videos. As there are many videos available online in which the speaker talks for a long time on a given topic, but it is hard to find the content the speaker wants to convey to the audience unless we watch the entire video. Python has various packages which are very helpful. Currently accessing the YouTube content has been made easier through the API in the python library like the transcripts of the videos etc. By taking this advantage we directly access the transcripts of the video and summarize to view them to the user. This can be implemented by using Hugging face transformer which is one of the text summarization techniques.

II. LITERATURE SURVEY

In 2021, "Natural Language Processing (NLP) based Text Summarization - A Survey" was published by Ishitva Awasthi, Kuntal Gupta, Prajbot Singh Bhojal, Anand, Piyush kumar. The techniques used are Extractive and abstract methods for summarizing texts. The advantages are: Based on linguistic and statistical characteristics, the implications of sentences are calculated. The disadvantage is each type of summarization technique is useful in different situations. One cannot say which technique is more promising [1].



“Review of automatic text summarization techniques & methods” is developed by AdhikaPramita, SupriadiRustad, Abdul Shukur, Affandy. It was published in 2020. They have used Text summarization, Systematic review techniques. The drawback is that the Fuzzy based approach is weak in semantic problems. Many loopholes have to be improved in extractive methods [2].

“Study on Abstractive Text Summarization Techniques” is developed by Parth Rajesh Dedhia, Hardik Pradeep, Meghana Naik. It was published in 2020. They have used Seq2Seq, Encoder-Decoder, and Pointer Mechanism. The drawback is that the current model does not work if multiple documents are passed to the model[3].

“Abstractive Summarization of video sequences” is developed by AniqDilawari, Muhammad Usman Ghani Khan. They have used multi-line video description, RCNN deep neural network model. The drawback is It focuses only on the conciseness of the summary. Memory efficiency and time constraints are not under consideration[4].

III. SYSTEM DESIGN

In this project we get transcripts/subtitles for a given YouTube video Id using a Python API, perform text summarization on obtained transcripts using HuggingFace transformers, build a Flask backend REST API to expose the summarization service to the client and develop a chrome extension which will utilize the backend API to display summarized text to the user.

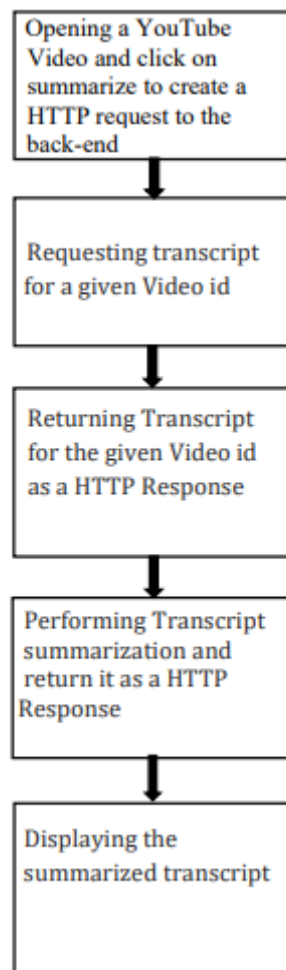


Figure 1 System architecture of YouTube transcript summarizer



V. CONCLUSION

This project has proposed a YouTube Transcript summarizer. The system takes the input YouTube video from the Chrome extension of the Google Chrome browser when the user clicks on the summarize button on the chrome extension web page, and access the transcripts of that video with the help of python API. The accessed transcripts are then summarized with the transformers package. Then the summarized text is shown to the user in the chrome extension web page. This project helps the users a lot by saving their valuable time and resources. This helps us to get the gist of the video without watching the whole video. It also helps the user to identify the unusual and unhealthy content so that it may not disturb their viewing experience.

This project also ensures great user interface experience in finding out the summarized text as chrome extensions have been used. This helps in getting the summarized text without copying the URL and pasting at terminals or by any third-party applications.

REFERENCES

- [1] I. Awasthi, K. Gupta, P. S. Bhogal, S. S. Anand and P. K. Soni, "Natural Language Processing (NLP) based Text Summarization - A Survey," *2021 6th International Conference on Inventive Computation Technologies (ICICT)*, 2021, pp. 1310-1317, doi: 10.1109/ICICT50816.2021.9358703.
- [2] AdhikaPramitaWidyassari, SupriadiRustad, GuruhFajarShidik, Edi Noersasongko, Abdul Syukur, Affandy, De Rosal Ignatius Moses Setiadi, Review of automatic text summarization techniques & methods, *Journal of King Saud University - Computer and Information Sciences*, 2020, ISSN 1319-1578, <https://doi.org/10.1016/j.jksuci.2020.05.006>.
- [3] P. R. Dedhia, H. P. Pachgade, A. P. Malani, N. Raul and M. Naik, "Study on Abstractive Text Summarization Techniques," *2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE)*, 2020, pp. 1-8, doi: 10.1109/ic-ETITE47903.2020.087.
- [4] A. Dilawari and M. U. G. Khan, "ASoVS: Abstractive Summarization of Video Sequences," in *IEEE Access*, vol. 7, pp. 29253-29263, 2019, doi: 10.1109/ACCESS.2019.2902507.
- [5] <https://huggingface.co/transformers/installation.html>
- [6] <https://atmamani.github.io/blog/building-restful-apis-with-flask-in-python/>



INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 8.118



ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 **9940 572 462**  **6381 907 438**  **ijirset@gmail.com**



www.ijirset.com

Scan to save the contact details