



IJIRCCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 11, Issue 5, May 2023

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.379

 9940 572 462

 6381 907 438

 ijircce@gmail.com

 www.ijircce.com

Youtube Transcript Summarizer

Ms S Ranjani ¹, Preethi Prakash ², Sherlin Retna R ³, Suja Marshalin P ⁴, Vaishnavi B ⁵

¹ Assistant Professor, Dept of Computer Science & Engineering, Saranathan College of Engineering, Tiruchirappalli,
Tamil Nadu, India

²⁻⁵ Student, Dept of Computer Science & Engineering, Saranathan College of Engineering, Tiruchirappalli,
Tamil Nadu, India

ABSTRACT: Our project delivers the summary for the YouTube transcript by highlighting the key elements. An important issue of extracting information from videos, where data is collected from each frame. The viewer watches the entire video to get the full content but it takes too much time. So, our project aims to shorten the time takes to watch the entire video by only reading the summary of transcripts. The summarizer analyses the video's transcription details, identifies important topics, and generates a precise and clear summary that captures the important keywords of the video. In recent years the accuracy of YouTube video transcripts increased with the help of Natural Language Processing (NLP), and we make it a valuable tool for viewers and content creators.

KEYWORDS: Transcript, Transformers, Pipeline, YouTube Transcript API, Text Summarizer, Natural Language Processing

I. INTRODUCTION

Welcome to the world of YouTube Summarizer and Transcripts. With over 2 billion monthly active users, YouTube is one of the most popular video-sharing platforms on the internet. Millions of videos are uploaded to the platform every day, covering a wide range of topics from entertainment to education, news, sports, and much more. However, with so much content available on the platform, it can be challenging to keep up with everything that is being shared. That is where YouTube summarizers and transcripts come in.

A YouTube summarizer is a tool that extracts the key points and highlights from a video, making it easier for viewers to get a quick overview of the content without having to watch the entire video.

On the other hand, a YouTube transcript is a written record of everything that is said in a video, including dialogue, narration, and even background noise.

Both YouTube summarizers and transcripts can be incredibly helpful for a variety of reasons. They can help people save time by allowing them to quickly skim through a video or read a transcript to get the main points without having to watch the entire thing. They can also be useful for people with hearing impairments who may have difficulty following the audio in a video.

In short, YouTube summarizers and transcripts are valuable tools for anyone who wants to make the most of the vast amount of content available on the platform. So, whether you are a casual viewer or a content creator looking to reach a wider audience, these tools can help you get the most out of YouTube.

II. LITERATURE SURVEY

There have been several studies and research conducted on YouTube summarizers and transcripts, highlighting their importance and usefulness. Here are a few examples:

1. "Towards Automatic Generation of YouTube Video Transcripts" by J.R. Cho and Y.-H. Kim (2018)

The paper introduces a method for automatically producing written records of YouTube videos using a combination of techniques in speech recognition and natural language processing. The system described in the paper comprises three primary elements: audio pre-processing, speech recognition, and post-processing. The audio pre-processing component involves modifying and standardizing the audio signal to enhance the accuracy of speech recognition. The speech recognition component employs a deep neural network model to convert the audio into written text. Lastly, the post-processing element utilizes techniques in natural language processing to rectify errors and enhance the clarity of the transcript.

The authors assessed the effectiveness of their system by applying it to a dataset of YouTube videos that already had manually created transcripts. The findings indicated that the system was capable of producing highly accurate

transcripts, achieving a word error rate (WER) of 10.3% and a character error rate (CER) of 4.3%. The proposed system has potential applications in various domains, including education, entertainment, and accessibility for individuals with hearing impairments. By automatically generating video transcripts, it becomes easier for users to search for specific information and comprehend the video's content, thus improving the search capability and accessibility of online video content.

2."Video Summarization of YouTube for Effective Educational Content Retrieval" by S. K. Dwivedy and S. K. Gupta (2019)

The research paper introduces an innovative system designed for condensing YouTube videos into concise summaries, specifically targeting educational content retrieval. The system proposed in the paper utilizes a blend of computer vision and natural language processing methodologies to extract crucial frames and subtitles from YouTube videos. These extracted components are then employed to generate a summary of the video, encompassing the most significant information conveyed within it.

The effectiveness of the system was evaluated by the authors using a dataset of YouTube videos about educational content. The findings demonstrated that the system successfully produced accurate video summaries, encompassing the most vital information delivered in the original videos. This proposed system holds promising applications within the realm of education, as it can greatly enhance the retrieval process of educational content on YouTube. By providing users with concise and informative video summaries, the system facilitates efficient access to desired educational materials. Additionally, the system serves to save time for users seeking specific information within videos, as they can quickly review the video summary to determine its relevance to their needs

3."Transcribing YouTube Videos: A Case Study of Using Crowdsourcing for Speech Recognition Training Data" by H. Lu and J. L. Gwizdka (2018)

The research paper presents an empirical investigation on the utilization of crowdsourcing to transcribe YouTube videos, specifically to acquire training data for speech recognition. The authors employed Amazon Mechanical Turk to enlist workers who transcribed a designated set of YouTube videos, subsequently utilized for training a speech recognition system. The authors assessed the transcription quality and analysed the impact of different transcription guidelines on the quality of the acquired training data.

The findings indicated that the quality of the transcriptions was influenced by various factors, including the transcription guidelines, audio quality, and the workers' level of experience. Additionally, the authors discovered that employing a combination of multiple workers and implementing a majority voting system enhanced the accuracy of the transcriptions. This approach holds potential applications in several domains such as speech recognition, natural language processing, and machine learning. Crowdsourcing presents a cost-effective and efficient means of generating large volumes of training data for these applications. In summary, this research paper contributes to the advancement of crowdsourcing-based methodologies for generating training data for speech recognition systems. Furthermore, it offers valuable insights into the factors that can impact the quality of transcriptions, thus enhancing our understanding of this domain.

4."A Review of Video Summarization Techniques" by S. S. Khattar and S. S. Agrawal (2021)

The research paper presents an extensive survey of various video summarization techniques proposed in the existing literature. The authors categorize these techniques into three main groups: keyframe-based, representative frame-based, and key-segment-based approaches. Keyframe-based techniques involve selecting a subset of frames that effectively represent the video content, whereas representative frame-based techniques focus on selecting a representative frame for each shot within the video. On the other hand, key-segment-based techniques aim to identify and select segments that contain the most vital information in the video.

Furthermore, the authors delve into the evaluation metrics commonly used to assess video summarization techniques, such as F-measure, precision, and recall. They also shed light on the challenges associated with evaluating video summarization, including the subjective nature of the task and the absence of standardized datasets and evaluation metrics. The paper extensively reviews a range of video summarization techniques proposed in the literature, encompassing unsupervised, supervised, and hybrid approaches. Additionally, the authors discuss the diverse applications of video summarization, including video browsing, indexing, and retrieval.

In conclusion, this paper offers a comprehensive examination of video summarization techniques, their applications, and the challenges involved in their evaluation. It serves as a valuable resource for researchers and practitioners interested in exploring video summarization in-depth.



III. METHODOLOGY

A YouTube summarizer can be created using natural language processing (NLP). Here is a general methodology for creating such a system:

Input Data: Accept the YouTube video link and language code of that particular video.

Transcription: This step is to retrieve the transcription details from the YouTube link given by the user, a transcript for each duration is obtained and sent to the transcript for the entire duration of the video is generated and displayed.

Transcripts can be collected through the Python API YouTube-transcript-API.

Summarization: Finally, the system can use summarization algorithms to generate a summary of the video. This can be done using techniques like extractive summarization, where the system selects the most important sentences from the video transcript, where the system generates new sentences that capture the essence of the transcription details. Transformers is a natural language processing library from Hugging Face. It provides APIs and tools to easily download and train state-of-the-art pre-trained models. Pretrained models can reduce computation costs, saving the time and resources required to train a model from scratch. These models support Natural Language Processing - summarization and text generation.

Google's Pegasus- Building upon earlier breakthroughs in the natural language processing (NLP) field, Google's PEGASUS further improved the state-of-the-art (SOTA) results for abstractive summarization, in particular with low resources. Architecture on a high level, PEGASUS uses an encoder-decoder model for sequence-to-sequence learning. In such a model, the encoder will first take into consideration the context of the whole input text and encode the input text into something called context vector, which is basically a numerical representation of the input text.

mT5_multilingual_XLSum- This repository contains the mT5 checkpoint fine-tuned on the 45 languages of the XLSum dataset. For fine tuning details and scripts.

XLSum- This repository contains the code, data, and models of the paper titled "XL-Sum: Large-Scale Multilingual Abstractive Summarization for 44 Languages"

distilbart-cnn-12-6- This checkpoint should be loaded into BartForConditionalGeneration from_pretrained.

IndicBART- IndicBART is a multilingual, sequence-to-sequence pre-trained model focusing on Indic languages and English. It currently supports 11 Indian languages and is based on the mBART architecture.

User interface: The user has to input the video URL and the language. The collected transcript and summary generated by the system can be presented to the user through the web browser.

IV. EXPERIMENTAL ANALYSIS

Input:

Give the YouTube video URL below <https://www.YouTube.com/watch?v=A4OmtyaBHF6>.

Output:

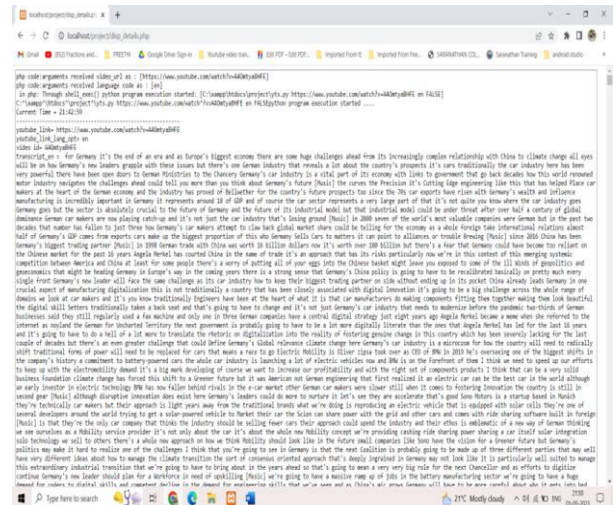
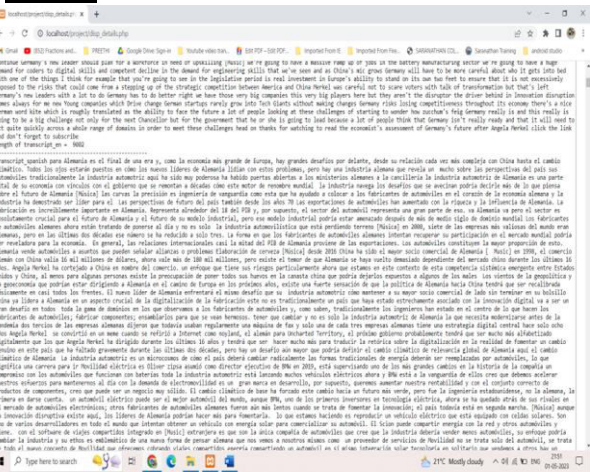


Fig1: Transcript of the video in English

Fig2: Transcript of the video in Spanish

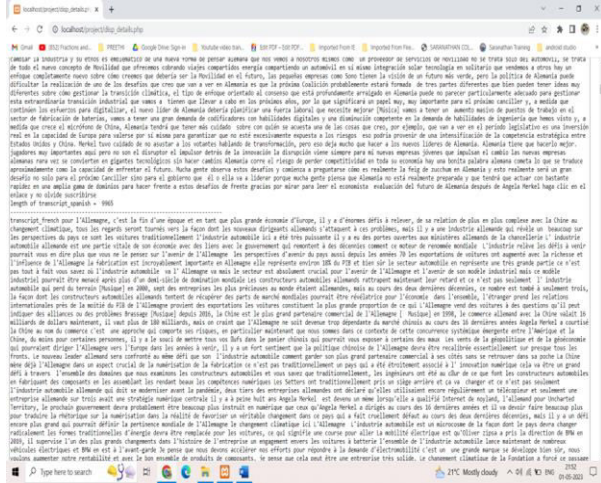


Fig3: Transcript of the video in French

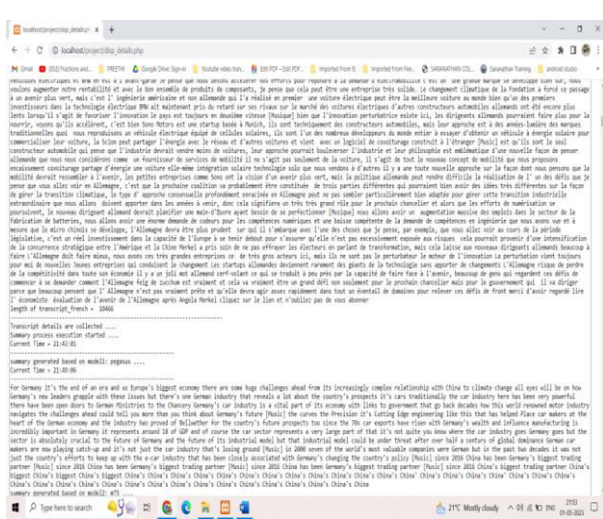


Fig4: Summary of the video generated based on pegasus ,mt5 model

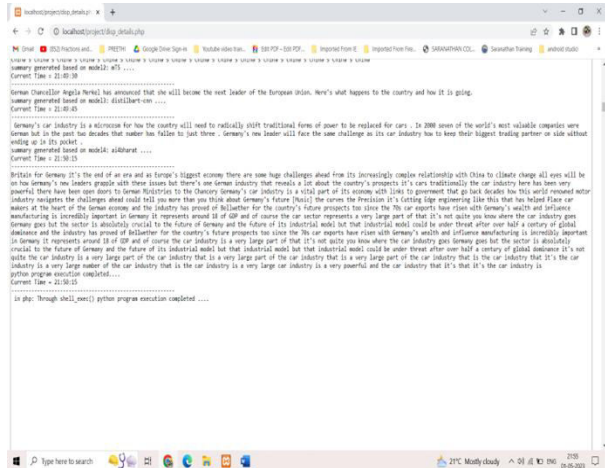


Fig5: Summary of the video generated based on distilbart-cnn,ai4Bharat

V. CONCLUSION

This project embarked on a comprehensive exploration and implementation of various video summarization and transcription techniques, intending to shed light on their effectiveness and potential applications. Through extensive analysis and evaluation, valuable insights were gained into the capabilities of these techniques in handling video content and extracting crucial information. By delving into their strengths and limitations, this project has significantly contributed to the understanding of video summarization and its wide-ranging applications.

The project entailed a meticulous examination of different video summarization methods, including keyframe-based, representative frame-based, and key-segment-based approaches. Each technique was carefully studied and evaluated to assess its ability to distill the most salient elements and capture the essence of the video. By classifying and categorizing these methods, the project offers a comprehensive overview of the video summarization landscape, providing researchers and practitioners with a valuable resource for further exploration and experimentation. Overall, this project has significantly contributed to the field of video summarization and transcription by exploring and implementing different techniques and methodologies. The knowledge gained through this endeavour serves as a valuable resource for researchers and practitioners seeking to delve deeper into these areas and unlock their full potential.

REFERENCES

- [1] Dipanjan Das, F.T Andr´e Martins, “A survey on automatic text summarization”Das, Dipanjan & Martins, Andr´e. (2007). A survey on automatic text summarization.
- [2] H.Lawrence Reeve¹, Hyoil Han¹, V.Saya Nagori², C. Jonathan Yang², A.Tamara Schwimmer², D.Ari Brooks², “Concept Frequency Distribution in Biomedical Text Summarization”, Vaishali P, Kalpana B and Mahender C. (2022). A Survey on Automated Text Summarization System for Indian Languages. Intelligent Data Communication Technologies and Internet of Things. 10.1007/978-981-16-7610-9_67.(921-941). https://link.springer.com/10.1007/978-981-16-7610-9_67
- [3] R. Dragomir Radev*, Eduard Hovy†, Kathleen McKeown‡, “Introduction to the Special Issue on Summarization”, Krishnaveni P and Balasundaram S. (2023). Generating fuzzy graph-based multi-document summary of text-based learning materials. *Expert Systems with Applications: An International Journal*. 214: C. Online publication date: 15-Mar-2023. <https://doi.org/10.1016/j.eswa.2022.119165>
- [4] Mehdi Allahyari, Seyedamin Pouriyeh, Mehdi Assefi, Saeid Safaei, D.Elizabeth Trippe, B. Juan Gutierrez, Krys Kochut, “Text Summarization Techniques: A Brief Survey”, [arXiv:1707.02268](https://arxiv.org/abs/1707.02268)[cs.CL] (or arXiv:1707.02268v3[cs.CL] for this version) <https://doi.org/10.48550/arXiv.1707.02268>
- [5] Mohamed Abdel Fattah, and Fuji Ren, “Automatic Text Summarization “, International Journal of Information, Control and Computer Sciences ISSN: 2517-9942 Vol:2, No:1, 2008
- [6] 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), 5-7 June 2020, Dhaka, Bangladesh
- [7] Ani Nenkova, Kathleen McKeown, “A SURVEY OF TEXT SUMMARIZATION TECHNIQUES”, Nenkova, A., McKeown, K. (2012). A Survey of Text Summarization Techniques. In: Aggarwal, C., Zhai, C. (eds) Mining Text Data. Springer, Boston, MA. https://doi.org/10.1007/978-1-4614-3223-4_3
- [8] Horacio Saggion, Thierry Poibeau, “Automatic Text Summarization: Past, Present and Future”, Horacio Saggion, Thierry Poibeau. Automatic Text Summarization: Past, Present and Future. T. Poibeau; H. Saggion. J. Piskorski, R. Yangarber. Multi-source, Multilingual Information Extraction and Summarization, Springer, pp.3-13, 2012, Theory and Applications of Natural Language Processing, 978-3-642-28569-1. final-00782442f
- [9] H. P. EDMUNDSON, “New Methods in Automatic Extracting”, Gunes, Erkan., Dragomir, R., Radev. (2004). LexRank: graph-based lexical centrality as salience in text summarization. Journal of Artificial Intelligence Research, 22(1):457-479. doi: 10.1613/JAIR.1523
- [10] Julian Kupiec, Jan Pedersen and Francine Chen, “A Trainable Document Summarizer”, @inproceedings {Kupiec1995ATD, title= {A trainable document summarizer},author={Julian Kupiec and Jan O. Pedersen and Francine R. Chen},book title={Annual International ACM SIGIR Conference on Research and Development in Information Retrieval},year={1995}}
- [11] A. Bookstein, S. T. Klein, T. Raita, “Detecting Content-Bearing Words by Serial Clustering — Extended Abstract*”, @inproceedings {Bookstein1995DetectingCW, title= {Detecting content-bearing words by serial clustering—extended abstract}, author={Abraham Bookstein and Shmuel Tomi Klein and Timo Raita}, book title={Annual International ACM SIGIR Conference on Research and Development in Information Retrieval}, year={1995}}
- [12] RONALD BRANDOW, J. KARL MITZE 2 and E. LISA RAU 3*, “@article {Brandow1995AutomaticCO, title={Automatic Condensation of Electronic Publications by Sentence Selection},author={Ronald Brandow and K. Mitze and Lisa F. Rau},journal={Inf. Process. Manag.},year={1995},volume={31},pages= {675-685}}”



Impact Factor: 8.379



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

 9940 572 462  6381 907 438  ijircce@gmail.com



www.ijircce.com

Scan to save the contact details