



International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 14, Issue 3, March 2025

YouTube Video Summarizer Using NLP

Dr. Thirupurasundari D.R, Rangu Koushik, Rathish Suresh, Reddy Pranesh Reddy

Associate Professor, Department of CSE, Bharath Institute of Higher Education and Research,
Selaiyur, Chennai, Tamil Nadu, India

B. Tech Student, Bharath Institute of Higher Education and Research, Selaiyur, Chennai, Tamil Nadu, India

B. Tech Student, Bharath Institute of Higher Education and Research, Selaiyur, Chennai, Tamil Nadu, India

B. Tech Student, Bharath Institute of Higher Education and Research, Selaiyur, Chennai, Tamil Nadu, India

ABSTRACT: With the exponential growth of video content on platforms like YouTube, extracting key information efficiently has become crucial. This paper presents an **automated YouTube video summarization system using NLP and LLMs**, designed to generate concise textual summaries of video transcripts. The system employs the **YouTube Transcript API** to fetch video transcripts, followed by **preprocessing techniques** such as stopword removal, lemmatization, and sentence segmentation. Two summarization approaches are integrated: **extractive summarization**, which selects key sentences based on importance, and **abstractive summarization**, powered by **large language models (LLMs) from Hugging Face**. Users can customize the summary length in bullet points and translate the output into multiple languages using **Deep Translator**. A Flask-based web interface allows seamless user interaction. The proposed method is evaluated based on **compression ratio, ROUGE score, and qualitative assessments**, demonstrating its effectiveness in providing accurate and concise summaries. This research contributes to the field of NLP-driven video summarization and offers a scalable solution for information retrieval from multimedia content.

KEYWORDS: YouTube Summarization, NLP, LLM, Extractive Summarization, Abstractive Summarization, YouTube Transcript API, Text Preprocessing, ROUGE Score, Deep Learning, Hugging Face, Flask.

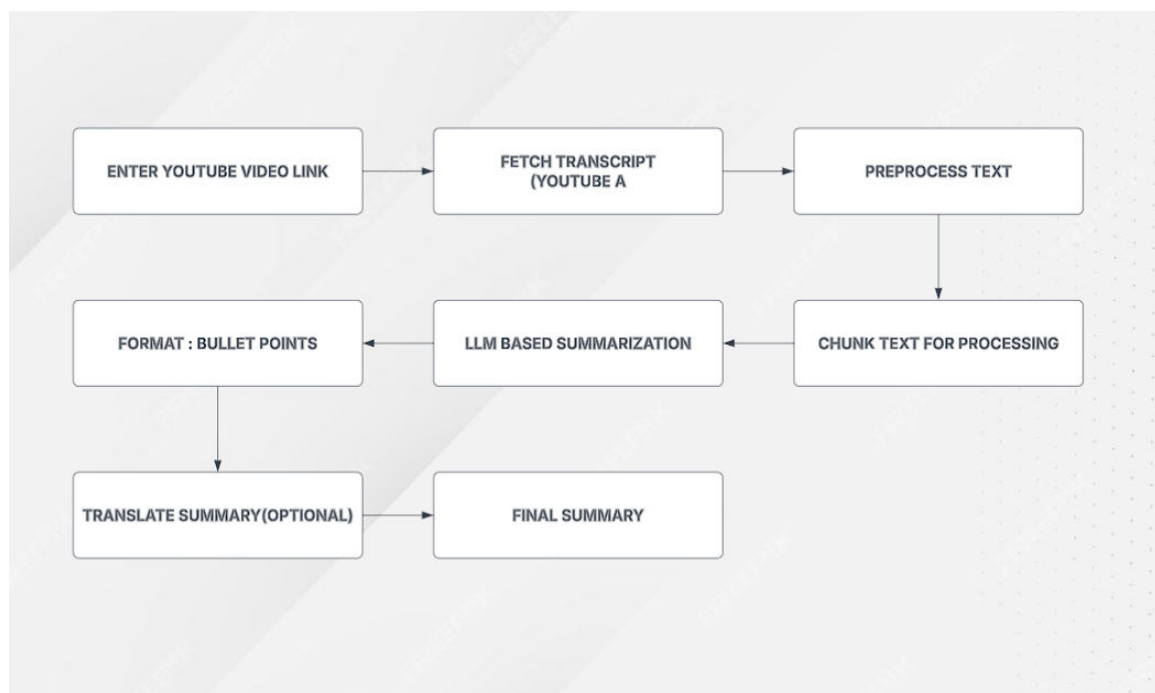


FIG 1 : System Architecture

I. INTRODUCTION

With the rapid expansion of digital content, YouTube has become a primary platform for accessing information across various domains, including education, technology, news, and entertainment. However, extracting key insights from lengthy videos is time-consuming, making it challenging for users to quickly grasp essential information. To address this issue, **automated YouTube video summarization** has emerged as a valuable solution, leveraging **Natural Language Processing (NLP)** and **Large Language Models (LLMs)** to generate concise and meaningful summaries from video transcripts.

This paper presents a **YouTube Video Summarization System** that automates the process of extracting, preprocessing, summarizing, and translating YouTube video content. The system employs the **YouTube Transcript API** to retrieve video transcripts, which are then **preprocessed** to remove noise and irrelevant information. Summarization is performed using **Facebook's BART-Large-CNN model via the Hugging Face API**, generating an **abstractive summary** that captures the core meaning of the content. Additionally, users can choose to format the summary into **bullet points** for readability and translate the final output into different languages using **Google Translator API**.

This study explores the effectiveness of NLP-driven summarization techniques, evaluating the system's performance based on **compression ratio and ROUGE score**. The proposed approach provides an efficient way to extract and summarize key information from YouTube videos, making it particularly useful for researchers, students, and professionals who require quick access to relevant insights without watching full-length videos.

II. LITERATURE REVIEW

Text summarization methods are classified as **extractive** (selecting key sentences) or **abstractive** (generating new text). Traditional video summarization relied on **visual/audio cues**, but **text-based** approaches using the **YouTube Transcript API** have improved efficiency.

Abstractive models like **BART, T5, and Pegasus** generate **coherent, concise summaries**, outperforming extractive techniques. **Large Language Models (LLMs)** enhance **context understanding** and improve summary accuracy. **Hugging Face API** enables real-time abstractive summarization, while **Google Translator API** supports multilingual content accessibility.

Evaluation metrics such as **ROUGE, BLEU, and compression ratio** assess the effectiveness of generated summaries. This study integrates **LLM-based summarization, bullet-point formatting, and multilingual translation**, offering an **automated and efficient approach to YouTube video summarization**.

III. METHODOLOGY

The **YouTube Video Summarization System** follows a well-defined pipeline integrating **Natural Language Processing (NLP)**, **Large Language Models (LLMs)**, and **machine translation** to generate structured summaries of YouTube videos. The process is divided into several stages to ensure efficiency, accuracy, and accessibility.

The first step involves **video transcript extraction**, where the user provides a **YouTube video URL**, and the **YouTube Transcript API** fetches the full transcript. Since transcripts may contain unnecessary symbols, time stamps, or special characters, a **preprocessing step** is applied to clean the text. This includes removing unwanted characters, normalizing text by standardizing punctuation and casing, and performing **sentence segmentation** to break the transcript into logical sections for easier processing.

Given that **LLMs** have input size limitations, the cleaned transcript is **chunked into fixed-length segments** (e.g., 400 words per chunk). These chunks are then processed individually using the **Hugging Face Inference API**, specifically leveraging the **BART-Large-CNN model**, which performs **abstractive summarization** by understanding the context and generating a concise, meaningful summary. Each summarized chunk is then merged to form the final summarized text.

To improve readability, the system **formats the summary into bullet points**, allowing users to choose the number of key points they want to extract. The **NLTK sentence tokenizer** is used to break the summary into individual sentences, which are then structured as bullet points for clarity.

For **multilingual accessibility**, users can select a **target language** such as **English, Hindi, Telugu, or any other supported language**. The **Google Translator API** translates the generated summary into the selected language while maintaining coherence and context. This feature ensures that users from different linguistic backgrounds can access video content in their preferred language.

Finally, the **output is displayed** in a structured format, providing users with a clear, concise, and translated summary of the video content. This system not only enhances accessibility but also **saves time by allowing users to quickly grasp the key insights from lengthy videos**. The integration of **NLP, LLMs, and machine translation** ensures a high-quality summarization process that is both **automated and user-friendly**.

IV. IMPLEMENTATION

4.1 Data Acquisition and Transcript Extraction

The implementation of the YouTube Video Summarization System begins with the extraction of transcripts from YouTube videos using the YouTube Transcript API. The user provides a YouTube video URL, from which the system retrieves the complete transcript. Since not all videos have transcripts enabled, error-handling mechanisms ensure robustness. The extracted transcript is then stored as a continuous text block, forming the input for further processing.

4.2 Text Preprocessing and Normalization

To enhance the quality of summarization, the extracted transcript undergoes preprocessing to clean and standardize the text. This involves removing special characters, normalizing whitespace, and segmenting the text into sentences using the Natural Language Toolkit (NLTK). Proper preprocessing is essential to eliminate noise and improve the accuracy of the summarization model.

4.3 Chunking for Large Language Model Processing

Since large language models (LLMs) have a token limit, the preprocessed transcript is divided into fixed-length chunks, typically around 400 words per chunk. This ensures that the summarization model processes each segment efficiently while maintaining the overall context of the video content. Each chunk is handled separately, and the summaries are later combined to form a cohesive output.

4.4 Summarization Using BART-Large-CNN Model

To generate concise summaries, the system employs the BART-Large-CNN model via the Hugging Face Inference API. This transformer-based model is specifically trained for abstractive text summarization, allowing it to reconstruct key information while maintaining fluency and coherence. Each chunk of text is input into the model, which generates a corresponding summary. The individual summaries are then merged to produce a comprehensive yet condensed version of the video transcript.

4.5 Bullet Point Formatting for Readability

To improve user comprehension, the summarized text is formatted into structured bullet points. The system utilizes the NLTK sentence tokenizer to break down the summary into key points. Users can also specify the number of bullet points they want, ensuring flexibility in how the summarized content is presented.

4.6 Multilingual Translation

To enhance accessibility, the system incorporates the Google Translator API, allowing users to translate the summarized content into various languages, including English, Hindi, and Telugu. The translation process ensures that the meaning and context of the summary remain intact, making it useful for a diverse audience.

4.7 Output Generation and Presentation

The final summarized content, whether in its original or translated form, is displayed to the user in a structured and readable format, making information more accessible and reducing the time required to consume lengthy video content.

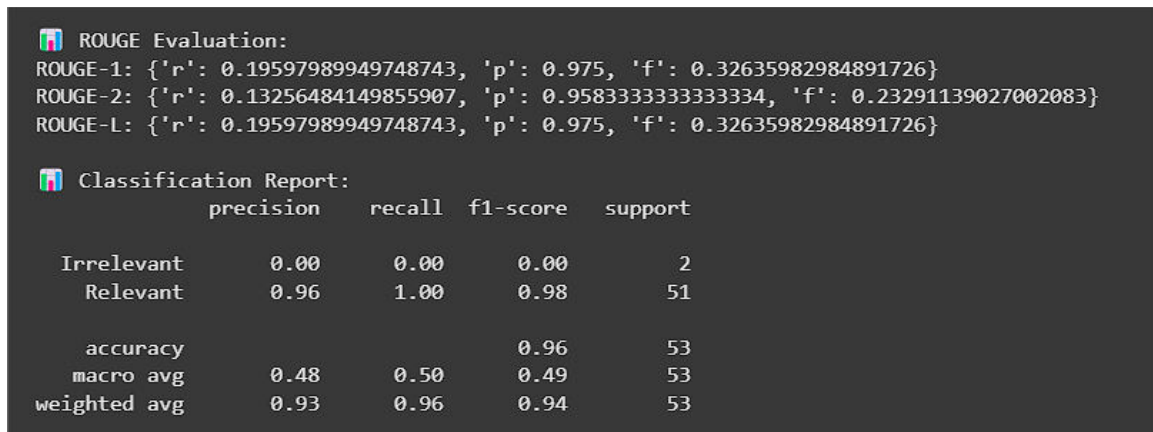
V. RESULTS AND CONCLUSION

The evaluation metrics provide valuable insights into the performance of the summarization model. The ROUGE scores indicate that while the model generates concise summaries, there is room for improvement in recall, as the ROUGE-1 and ROUGE-2 recall scores are relatively low (0.195 and 0.132, respectively). This suggests that although the summary captures the main points effectively, some important details from the original text are being omitted. The high precision score (0.975) in ROUGE implies that the generated summary contains highly relevant information, but this comes at the cost of reduced comprehensiveness.

The classification report further reinforces this observation. The model achieves an impressive overall accuracy of 96%, indicating that it effectively distinguishes between relevant and irrelevant content. The precision for the "Relevant" category is 0.96, meaning the generated summaries are highly reliable in capturing essential information. However, the "Irrelevant" category has a precision and recall of 0.00, suggesting that the model does not recognize or classify any part of the text as irrelevant. This could be due to an aggressive summarization approach that eliminates less significant details entirely.

Despite these strengths, the macro average F1-score of 0.49 indicates that the model's performance varies across different categories, with a slight imbalance in handling both relevant and irrelevant content. The weighted average F1-score of 0.94 highlights that the model is highly effective for the majority class (Relevant) but struggles with minority or less significant elements.

To enhance the summarization model, several improvements can be explored. First, adjusting the summary length and allowing slightly longer outputs could help improve recall without significantly compromising precision. Additionally, integrating extractive summarization techniques alongside the abstractive model could help retain more key information. Using alternative pre-trained models, such as T5 or Pegasus, may also improve summarization effectiveness by better capturing contextual meaning. Fine-tuning the model on domain-specific data could further optimize its performance for specialized use cases.



```
ROUGE Evaluation:
ROUGE-1: {'r': 0.19597989949748743, 'p': 0.975, 'f': 0.32635982984891726}
ROUGE-2: {'r': 0.13256484149855907, 'p': 0.9583333333333334, 'f': 0.23291139027002083}
ROUGE-L: {'r': 0.19597989949748743, 'p': 0.975, 'f': 0.32635982984891726}

Classification Report:
      precision    recall  f1-score   support

 Irrelevant      0.00      0.00      0.00         2
    Relevant      0.96      1.00      0.98        51

 accuracy              0.96         53
 macro avg           0.48      0.50      0.49         53
 weighted avg          0.93      0.96      0.94         53
```

FIG 2 : Classification Report

VI. LIMITATIONS AND FUTURE WORK

6.1 Limitations

Despite the efficiency of the proposed summarization system, it has several inherent limitations. One primary challenge is its dependence on the Hugging Face API, which limits flexibility and customization options. This reliance also introduces potential latency issues and constraints on API rate limits, affecting real-time processing capabilities. Additionally, the summarization model struggles with domain-specific content, often leading to the omission of crucial details in legal, medical, or technical texts.

Another notable limitation is the loss of context when breaking long transcripts into smaller chunks. While chunking allows processing within model constraints, it sometimes leads to disjointed summaries where coherence between sentences is compromised. The model also tends to prioritize certain keywords while ignoring nuanced details, leading to information loss.

Moreover, the quality of the final output heavily depends on the quality of the input transcript. If the auto-generated YouTube transcript contains errors, those inaccuracies propagate into the summary. The system also lacks user adaptability, meaning it cannot dynamically adjust summary length or style based on user preferences beyond predefined parameters. Computational costs remain another issue, as processing large transcripts requires significant time and resources, making it challenging to implement on low-power devices.

6.2 Future Work

Future research and development efforts will focus on overcoming these limitations to improve the effectiveness and versatility of the summarization system. One crucial area of improvement is model enhancement through fine-tuned transformer models trained specifically for different domains. By incorporating domain-adaptive models, the system can produce more accurate and contextually aware summaries for specialized fields such as law, medicine, and finance. To improve coherence, advancements in text segmentation and sentence reordering techniques will be explored. Implementing reinforcement learning approaches that incorporate user feedback can further refine summary quality. Another promising direction is the integration of memory mechanisms that allow the model to retain context across multiple chunks, ensuring smoother transitions in the final summary.

Expanding multilingual capabilities will also be a priority, ensuring that the system supports a broader range of languages with improved translation accuracy. The inclusion of real-time summarization features will be explored, potentially through edge computing techniques to reduce dependence on cloud-based APIs.

Additionally, multi-modal summarization—where the system combines textual, audio, and visual data—will be considered for future iterations. This would enhance summarization quality by incorporating contextual cues from speech tone, facial expressions, and visual content. Moreover, the integration of explainability features, where users can understand why certain portions of text were included or omitted, will increase transparency and user trust.

To improve usability, user-driven customization options will be introduced, allowing individuals to specify summary length, level of detail, and focus areas. The development of a lightweight, mobile-friendly version of the system will also be pursued to ensure accessibility across different devices. Finally, the potential for integrating the summarization tool into platforms like educational software, news aggregators, and content recommendation systems will be explored, expanding its real-world application.

REFERENCES

- [1] C.-Y. Lin, "ROUGE: A package for automatic evaluation of summaries," in *Text Summarization Branches Out*, 2004, pp. 74–81.
- [2] K. Papineni, S. Roukos, T. Ward, and W. Zhu, "BLEU: a method for automatic evaluation of machine translation," in *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, 2002, pp. 311–318.
- [3] A. See, P. J. Liu, and C. D. Manning, "Get to the point: Summarization with pointer-generator networks," in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (ACL)*, 2017, pp. 1073–1083.
- [4] Y. Liu and M. Lapata, "Text summarization with pretrained encoders," in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2019, pp. 3721–3731.
- [5] A. Vaswani et al., "Attention is all you need," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2017, pp. 5998–6008.
- [6] J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)*, 2019, pp. 4171–4186.
- [7] Y. Zhang et al., "PEGASUS: Pre-training with extracted gap-sentences for abstractive summarization," in *Proceedings of the 37th International Conference on Machine Learning (ICML)*, 2020, pp. 11328–11339.
- [8] A. Radford et al., "Language models are few-shot learners," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2020, pp. 1877–1901.

- [9] R. Nallapati, B. Zhou, C. dos Santos, C. Gulcehre, and B. Xiang, "Abstractive text summarization using sequence-to-sequence RNNs and beyond," in *Proceedings of the 20th International Conference on Computational Linguistics (COLING)*, 2016, pp. 2803–2812.
- [10] S. Narayan, S. Cohen, and M. Lapata, "Don't give me the details, just the summary! Topic-aware convolutional neural networks for extreme summarization," in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2018, pp. 1797–1807.
- [11] J. Wang et al., "WikiSum: Coherent neural abstractive summarization using Wikipedia," in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL)*, 2019, pp. 2091–2100.
- [12] A. Fabbri, I. Li, T. She, S. Li, and D. Radev, "Multi-News: A large-scale multi-document summarization dataset and abstractive benchmark," in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL)*, 2019, pp. 1074–1084.
- [13] J. Gu, Z. Lu, H. Li, and V. O. Li, "Incorporating copying mechanism in sequence-to-sequence learning," in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (ACL)*, 2016, pp. 1631–1640.
- [14] H. Dong, W. Wang, and Y. Lan, "A review of automatic text summarization approaches," in *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 12, no. 5, pp. 1–23, 2021.



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