CompSci Journal
Double Blind Peer Reviewed, Open Access International Research Journal.
ISSN: 3067-3089 | Website: https://csjour.com | Volume-1, Issue-1 | February - 2025

Original Article

SummaTube: AI-Based YouTube Summarization

Manoj Shivaji Koli[1], Rutuja Raut[2], Nikita Pawar[3], Sakshi Pawar[4], Gauri Pawar[5]
[1,2,3,4,5]SKN Sinhgad College of Engineering, Korti Pandharpur

Manuscript ID:
CSJ-2025-010106

ISSN: 3067-3089

Volume 1

Issue 1

Pp. 37-43

February 2025

Submitted: 30 Dec. 2024

Revised: 28 Jan. 2025

Accepted: 20 Feb. 2025

Published: 28 Feb. 2025

Correspondence Address:
Manoj Shivaji Koli, 5SKN Sinhgad College of Engineering, Korti Pandharpur

Quick Response Code:

Web. https://csjour.com/

DOI:
10.5281/zenodo.15302811

DOI Link:
https://doi.org/10.5281/zenodo.15302812

Abstract

This paper introduces "YouTube Video Transcript," a web application designed to improve access to YouTube content by providing automatic transcription and multilingual support. Developed using the Flask framework, the web application generates transcripts from video URLs, with translation options for Marathi, Hindi, and English. Complex terms are highlighted for users, linking them to relevant Google searches for additional context. The web application also allows downloading transcripts as PDF files, making it easier to engage with video content across language barriers and simplifying complex information. This tool addresses both educational and accessibility challenges. The paper details the system's architecture, technical implementation, and future development opportunities.

Keywords: *YouTube Transcription, Flask, Multilingual Translation, Keyword Highlighting, YouTubeTranscriptApi, Google Search Integration.*

Introduction:

The BRIEFTUBE MADE SUMMARIES EASY is designed to streamline the process of summarizing YouTube video transcripts using NLP techniques. It interacts with a backend REST API that handles the heavy lifting of natural language processing. The extension sends a request to this API, which processes the YouTube transcript to extract key insights and information. The summarized version of the transcript is then sent back to the extension for display to the user. This integration of AI and NLP technologies aims to enhance user experience by condensing lengthy video
content into concise summaries, thereby saving time and improving information retrieval efficiency.
This research paper presents the development of the YouTube Video Transcripter, a web application created using the Flask framework.

a) What is Artificial Intelligences?

Artificial Intelligence (AI) refers to the development of computer systems of performing tasks that require human intelligence. AI aids, in processing amounts of data identifying patterns and making decisions based on the collected information. This can be achieved through techniques like Machine Learning, Natural Language Processing, Computer Vision and Robotics. AI encompasses a range of abilities including learning, reasoning, perception, problem solving, data analysis and language comprehension.

b) Liabraries and modules:

flask: A popular web framework for Python that allows you to create web applications. It provides tools for handling requests, rendering templates, and managing sessions.
render_template: A function from the flask library that renders HTML templates and returns the rendered content as a response.
request: A function from the flask library that provides information about the incoming HTTP request, such as the URL, HTTP method, and request data.
make_response: A function from the flask library that creates an HTTP response object.
reportlab: A Python library for generating PDF documents. It provides tools for creating pages, adding content, and formatting the layout.
io.BytesIO: A Python class that represents an in-memory file-like object, which is useful for creating PDF documents in memory before downloading them.
youtube_transcript_api: A Python library that provides an interface for interacting with the YouTube Data API and retrieving video transcripts.

How to cite this article:
Koli, M. S., Raut, R., Pawar, N., Pawar, S., & Pawar, G. (2025). SummaTube: AI-Based YouTube Summarization. Compsci Journal, 1(1), 37–43. https://doi.org/10.5281/zenodo.15302812

nltk: The Natural Language Toolkit, a Python library for natural language processing tasks such as tokenization, part-of-speech tagging, and stemming.
collections.Counter: A Python class for counting the frequency of elements in a sequence

c) What is NLP?

The meaning of NLP is Natural Language Processing (NLP) which is a fascinating and rapidly evolving field that intersects computer science, artificial intelligence, and linguistics. NLP focuses on the interaction between computers and human language, enabling machines to understand, interpret, and generate human language in a way that is both meaningful and useful. With the increasing volume of text data generated every day, from social media posts to research articles, NLP has become an essential tool for extracting valuable insights.

d) Benefits of AI :-

Automate repetitive tasks:-

Many of our days are filled with repetitive – and often mundane – tasks that must be completed in order to keep things running smoothly.

Quickly analyze big data sets:-

Data is essential to the daily operations of countless organizations worldwide. Yet, while many businesses and individuals know the value of big data, few are able to effectively analyze the data at their disposal to produce the kinds of insights they need to make the most impactful decisions

Improve decision making:-

Leveraging data analytics, AI applications can equip organizations with near real time, actionable insights that can help them make informed decisions

Quickly generate new material:-

Writing, coding, and designing visuals are time-consuming processes. Today, many writers, programmers, designers, and artists spend a large amount of their time completing tedious tasks

Reduce operational costs:-

Running a business can be expensive. In addition to requiring a skilled workforce, employers must also invest in programs, systems, and infrastructure (among other things) to ensure that they can operate effectively.

Enhanced Healthcare:-

AI-based solutions could prove invaluable in the field of healthcare, in so many ways. AI could be used, for example, to assist researchers in developing cures and treatments for illnesses that have plagued mankind for many years.

Literature Review:

The advent of online video platforms has revolutionized the way information is consumed, making multimedia content more accessible to a global audience. However, the lack of proper subtitles and translations in many videos poses significant barriers to comprehension, particularly for non-native speakers and individuals with hearing impairments. This literature review examines existing tools and research related to video transcription, translation, and keyword extraction, highlighting the gaps that the YouTube Video Transcript aims to address.

Video Transcription Technologies

Automated video transcription has gained traction in recent years, with various tools developed to convert spoken language into written text. According to Dufour et al. (2021), the effectiveness of automatic speech recognition (ASR) systems has improved dramatically due to advances in machine learning and natural language processing. Tools like Google Cloud Speech-to-Text and IBM Watson Speech to Text are widely used for generating transcripts, but they often require significant customization and integration efforts (Chung et al., 2020).

The YouTubeTranscriptApi is a Python library that facilitates the extraction of subtitles from YouTube videos, providing an efficient method for acquiring transcripts when subtitles are available. However, many videos lack comprehensive subtitle support, leading to a need for more robust solutions that can generate accurate transcripts across diverse content types.

Translation and Multilingual Support

Language barriers remain a significant hurdle in global content consumption. Research by Hu et al. (2019) emphasizes the importance of automated translation tools in enhancing content accessibility. While established translation services like Google Translate and DeepL provide multilingual support, they often lack contextual understanding, leading to inaccuracies in translation (Zhou et al., 2021).

The integration of contextual translation techniques is crucial, as highlighted by Tran and Nguyen (2020), who discuss how semantic analysis can enhance translation accuracy by considering the broader context rather than relying solely on wordfor-word translations. The YouTube Video Transcripter addresses this gap by incorporating multiple translation options (English, Marathi, and Hindi) to cater to diverse audiences.

Keyword Extraction and Highlighting

Keyword extraction has been a focal point in natural language processing, enabling the identification of significant terms within a text. Several studies have explored various algorithms for effective keyword extraction, including term frequency-inverse document frequency (TF-IDF) and Latent Semantic Analysis (LSA) (Manning et al., 2008; Salton & Buckley, 1988). Highlighting complex or technical terms can significantly enhance user understanding, especially in educational contexts where students may encounter unfamiliar vocabulary. The research conducted by Zhang and Yang (2022) demonstrates that interactive keyword highlighting can improve information retention and comprehension. By allowing users to click on highlighted terms for additional information, as implemented in the YouTube Video Transcript, the application fosters an engaging learning experience.

Existing Tools and Applications

Several existing applications and tools offer similar functionalities, yet often lack comprehensive features that address the needs of diverse users. For instance, platforms like Kapwing and VEED.IO provide basic transcription and editing features but do not offer robust translation or contextual keyword highlighting capabilities. The YouTube Video Transcript differentiates itself by integrating all these features into a single application, thereby addressing the limitations found in existing tools. By providing automated transcription, multi-language translation, and keyword insights in one cohesive platform, the application aims to enhance the accessibility and usability of video content.

Proposed System:

The YouTube Video Transcripter is designed to provide an accessible and userfriendly platform for extracting and translating video content. The proposed system aims to enhance the understanding of YouTube videos through automated transcription, multilingual support, keyword extraction, and contextual information. This section outlines the system architecture, key features, and workflow of the application.

This project aims to develop a robust and efficient tool for extracting, analyzing, and summarizing YouTube video transcripts. By leveraging natural language processing (NLP) techniques

Here's a suggested Proposed System section for your YouTube Video, the project will provide valuable insights into the video content, including key terms and a concise summary.

Transcripter research paper:

1. **System Architecture:-**

The architecture of the YouTube Video Transcripter is based on a client-server model, leveraging the Flask framework to handle user requests and responses. The main components of the system include:

•**Frontend:** A user interface that allows users to input a YouTube video URL, select translation options, and view the generated transcript. The frontend is built using HTML, CSS, and JavaScript for dynamic user interactions.

•**Backend**: The Flask server processes incoming requests, extracts video transcripts, performs translations, and manages keyword extraction. It interacts with external libraries and APIs, such as: o YouTubeTranscriptApi: For fetching subtitles from YouTube videos.

•**YouTube Data API**: This API will be used to access YouTube video data, including captions (if available), metadata (title, description, etc.), and video content for further processing.

2. Key Features:-

The proposed system includes the following features:

•**Video URL Input**:

Users can input a YouTube video URL, which the system processes to extract the corresponding transcript.

• **Automated Transcript Generation:**

The application fetches subtitles from the provided video URL using the
YouTube Transcript Api, allowing for the automatic generation of a complete transcript.

•**Multi-language Translation**:

Users can convert the transcript into English, Marathi, or Hindi. The translation feature aims to cater to a diverse audience, enhancing accessibility for non-native speakers.

•**Keyword Extraction and Highlighting:**

Complex or technical terms within the transcript are analyzed and highlighted in blue.
When clicked, these keywords provide additional contextual information via Google search results, facilitating a deeper understanding of the subject matter.

•**PDF Generation:** Users can download the generated transcript in PDF format, providing a convenient way to access and share content offline.

Methodology:

The development of the YouTube Video Transcripter involved a structured approach that encompassed system design, implementation, testing, and evaluation. This section outlines the methodology followed during the project, detailing the tools and

techniques used at each stage of the development process.

1. **User Input**

The project begins with the user entering a YouTube video URL into a designated input field on the application's interface. This serves as the starting point for the entire workflow.

2. Transcript Extraction

Once the user submits the video URL, the backend processes this input by utilizing the YouTubeTranscriptApi. This library fetches the subtitles from the specified video, generating a complete transcript if subtitles are available. The extraction process involves:

Validating the URL to ensure it points to a valid YouTube video.

Calling the API to retrieve the transcript associated with the video, handling any exceptions if transcripts are not available.

3. Translation Options

After the transcript is successfully generated, the user is presented with translation options for the transcript. The application supports translations into three languages: English, Marathi, and Hindi. The translation process involves:

Utilizing the mtranslate library to convert the original transcript into the selected language.

Displaying the translated text alongside the original transcript, allowing users to compare and understand both versions.

4. Keyword Extraction and Highlighting

To enhance the user's comprehension of the content, the application analyzes the generated transcript for complex or technical terms. This step includes:

•Implementing Natural Language Processing (NLP) techniques using the nltk library to identify significant keywords.

•Highlighting these keywords in blue within the displayed transcript. Users can click on these highlighted terms to access additional contextual information through Google search results, facilitating deeper understanding.

5. Output Generation

Once the transcript and translations are ready, the application provides options for output:

•The user can view the transcript, translations, and highlighted keywords in a well-organized format.

•Additionally, the application offers the functionality to generate and download a PDF version of the transcript using the reportlab library, ensuring offline access and easy sharing.

Flowchart

CompSci Journal
Double Blind Peer Reviewed, Open Access International Research Journal.
ISSN: 3067-3089 | Website: https://csjour.com | Volume-1, Issue-1 | February - 2025

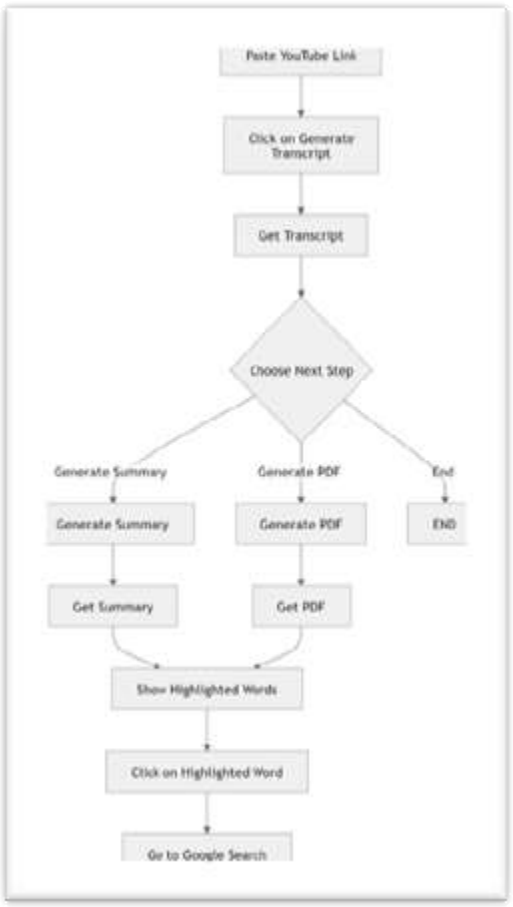
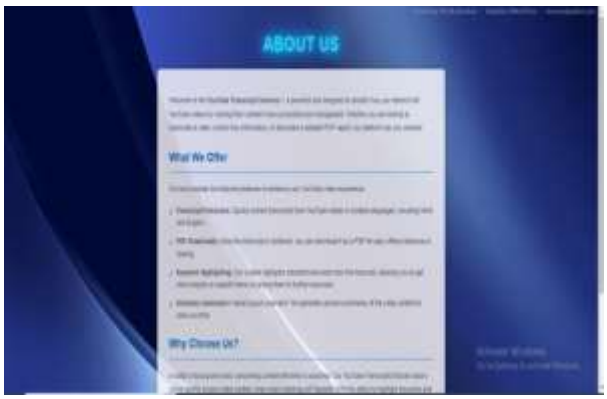


Fig. Flowchart of our project

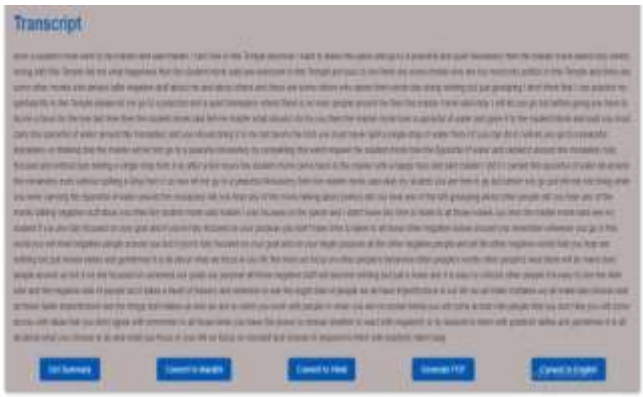
Result:

1. Home Page:

2. About Us:



3. Transcript:

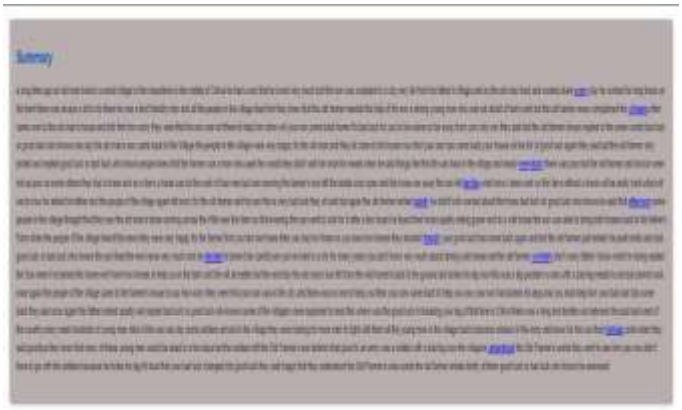


Transcript Translation:

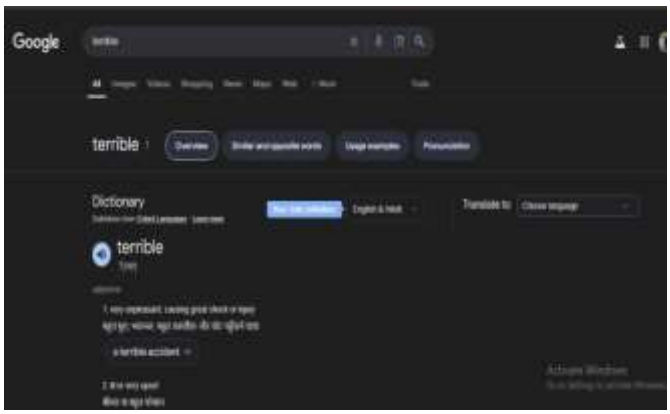
a. Marathi:

b. Hindi:

Highlighting Keywords:



More Information About:



Conclusion:

The YouTube Video Transcripter represents a significant advancement in enhancing the accessibility and comprehension of video content. By integrating automated transcription, multilingual translation, keyword extraction, and contextual information, the application effectively addresses the challenges faced by users when engaging with multimedia content. Throughout the development process, a systematic approach was adopted, ensuring that each component functioned harmoniously to deliver a seamless user experience. The use of libraries such as YouTube Transcript Api, translate, and nltk facilitated the efficient extraction of subtitles, accurate translations, and insightful keyword analysis. The feedback received during user acceptance testing confirmed the application's efficacy in meeting user needs and improving understanding of complex topics.

This project not only contributes to the existing body of knowledge in the fields of natural language processing and educational technology but also promotes inclusivity by making video content accessible to diverse audiences, including non-native speakers and individuals with hearing impairments.

Future work may involve further enhancements, such as incorporating additional languages, refining the contextual understanding of translations, and improving the overall user interface. By continuously evolving, the YouTube Video Transcripter has the potential to become an invaluable resource for learners and content consumers alike, fostering a deeper engagement with educational and informational videos across various platforms.

Acknowledgment:

We would like to express our deep sense of gratitude to our guide Mr. M.S. Koli for his invaluable help and guidance for the duration of project. We are highly indebted him for constantly encouraging us by giving critics on our work. We would like to thank our HOD Dr. S. V. Pingale for his encouragement throughout the semester which gave us confidence that we can complete our project successfully. We express gratitude towards, Project Coordinator Mr. R. S. Yevale for providing the support and giving his valuable time, indispensable support and his priceless suggestions. We also express gratitude towards our all teaching & non-teaching staff, family members and our Friends for encouraging us with their valuable suggestions and motivating us from time to time.

Financial support and sponsorship

Nil.

Conflicts of interest

There are no conflicts of interest.

References:

1. V. Mane, K. Dupare, P. Dupare, T. Ganvir, R. Ghode, "YouTube Transcript Summarizer using HuggingFace Transformer," Vishwakarma Institute of Technology, Pune, Maharashtra, India, vol. 2023, pp. 145–153.
2. P. V. Kumari, M. C. Keshava, C. Narendra, P. Akanksha, K. Sravan, "Youtube Transcript Summarizer Using Flask And NLP," vol. X, pp. 2022, Year.
3. A. Rice, "Using YouTube as the Primary Transcription and Translation Platform for Remote Corpus Work," NFLRC, vol. 15, pp. 514-550, 2021.
4. G. Olasina, "An evaluation of educational values of YouTube videos for academic writing," The African Journal of Information Systems, vol. 9, no. 4, pp. 1-15, October 2017.
5. L. Mangal, D. Aggarwal, G. Kumar, J. Kumar, and M. Aggarwal, "YouTube Transcript Summarizer," 2022, Year