

## A New Approach for YouTube Video Transcript Summarizer Using Python

Surbhi Kumari, Satish Kumar Alaria, Muskan Sharma, Manish Kumar Mukhija,  
Pramod Sharma

Department of Computer Science & Engineering, Arya Institute of Engineering &  
Technology, Jaipur, Rajasthan

---

### **ABSTRACT**

Throughout the day, a tremendous number of video recordings are generated and shared on YouTube. It can be tough to spend time to watching films that may run longer than anticipated, and our efforts may be for naught if we are unable to extract relevant information from them. Summarizing transcripts of such videos will help us to quickly discover important patterns in the video without having to view the entire thing. The main purpose of our project, VIDEO TRANSCRIPT SUMMARIZER USING PYTHON, is to construct a brief summary of the video content by selecting the most interesting and relevant sections.

Natural language processes are not a whole new topic. But with growing interest in human-machine communication, access to vast amounts of information, powerful computations, and improved algorithms, we are making rapid progress. You can communicate as a personality in English, Spanish or Chinese. However, most people do not understand the computer language known as computer code or machine language. At the lower level of the widget, communication is done through a number of 0s and 1s that perform logical actions. The central idea of our project, healing technology, is human language technology (natural language processing). In our facet, natural language is outlined as the process of creating fleshy phrases and sentences in the style of real language. With the help of python and its various libraries and html, CSS and JavaScript we are summarizing the video of YouTube that will be very easy to understand.

Also, we are trying to translate the language from English to Hindi and in one more language so that people comfortable in Hindi language can easily read it. When you try to translate a word or a sentence from one language to another, the Google Translate API works in the background to provide you with the necessary results. Though you can translate anything by visiting the Google Translate website, you can also use the Google Translate API to integrate Google Translate into your online or desktop applications. The best part about the API is how simple it is to set up and utilize. With the Google Translate API, you can perform everything from detect languages to simple text translation, as well as define source and destination languages and translate large lists of text phrases.

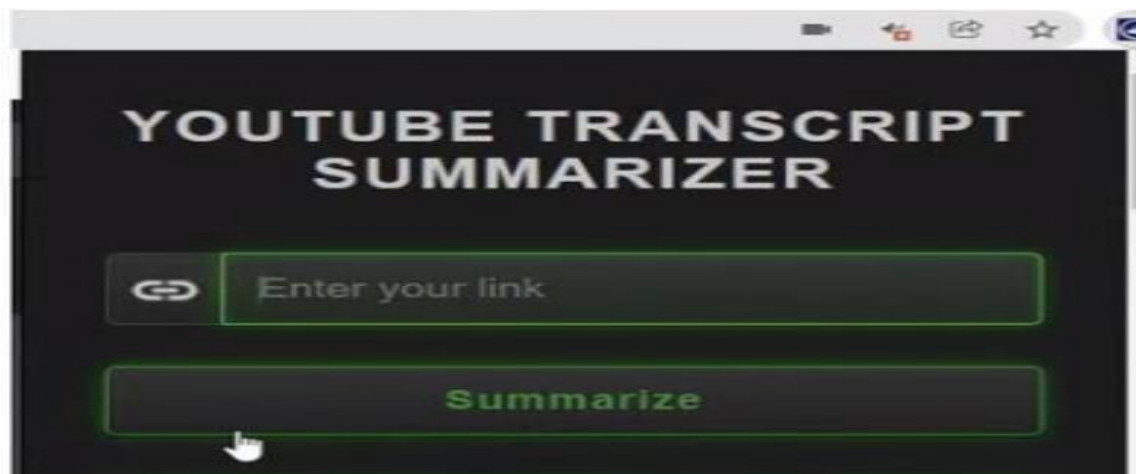
**Keywords** - YouTube, Summarizer, NLP, API, Python, Videos.

---

### **I. INTRODUCTION**

YouTube users are predicted to reach 2.3 billion by 2020, and the number is increasing every year. Every minute, 300 hours of YouTube videos are uploaded. According to Google, about a third of YouTube visitors in India view videos on their phones and spend more than 48 hours each month on the site. It takes time to find videos that contain the information we're looking for.

Over example, there are many Ted Talk videos online in which the speaker speaks for an extended period of time about a given topic, yet locating the stuff the speaker is most interested necessitates watching the entire video. And visit the site for more than 48 hours every month.



**Fig-1: UI of Chrome Extension**

Our work central concept and underlying technology is NLP (Natural Language Processing), which we define as the process of producing meaningful phrases and sentences in the form of a highly natural language. NLP has grown in importance in the field of Machine Learning due to the vital necessity to analyze text with its diverse structure, inferred meanings, feelings, and intent. YouTube transcript summarizer is based on the foundations of NLP and helps to remove objectionable and lengthy scrap or rubbish content from a video while also returning the video's summary.

NLP is concerned with creating transcript summaries because machines must understand what people have written and provide human-readable results. Nowadays, it is extremely easy to watch videos on YouTube for anything, whether instructional or entertaining. There are many videos in those genres, and because of the large number, it is critical that we identify the specific information that we want to consume. We can't always watch text-based YouTube videos in high definition because of network troubles, which causes the text to blur. There are also those long, click bait videos that are solely meant to make money and include no useful information, wasting our time.

## **II. PROBLEM STATEMENT AND METHODOLOGY**

This topic explains United States concerning the matter that is being resolved by this project with its cheap and valid answer, later it additionally explains United States concerning varied technologies that were used whereas developing this project with its even and logical options.

### **2.1 Problem Statement**

We pay a substantial quantity of your time on YouTube. That may be a primary supply of data as a result of it contains lecture videos from premier institutes. We tend to watch AN hour-long video and check out to grab the vital concepts from it. What if we are able to let a system do it? Perceive the vital concepts from the video and summarize them in a very means which will permit you to seem back and get back the key ideas without watching the video all over again.

### **2.2 Problem Answer**

The discussion of the matter Formulation of this project revolves around 2 factors that embrace the necessity and significance of the event. The necessity for development is classed in making a Chrome Extension, as we all know that Google Chrome extensions are programs that may be put in into Chrome to vary the browser's practicality. This includes adding new options to Chrome or modifying the present behavior of the program itself to create it additional convenient for the user.

Our major objective would be acting text account on obtained transcripts exploitation fondling Face transformers. We'll be building a Backend Rest API to export the summarized content of the video to the consumer. And, for enabling quick access to the user, we'll be building a Chrome Extension for the method.

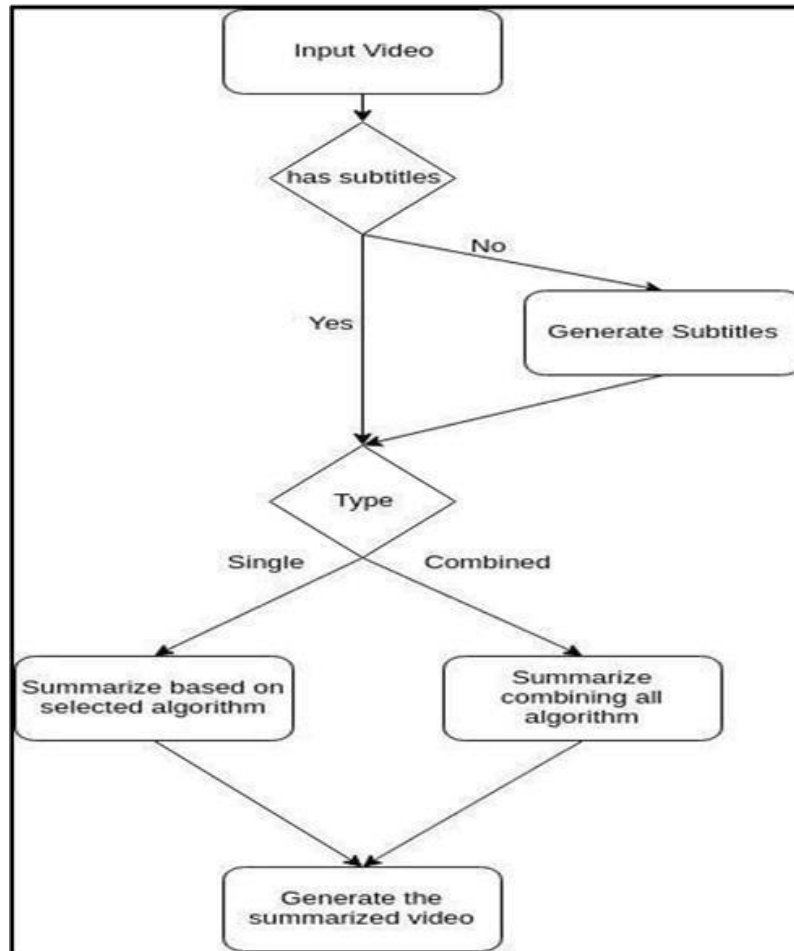


Fig 2: flowchart where various steps are being displayed with its fundamental functions and specific outcomes.

Automatic NLP based summarization algorithms is being used here which helps in generating the summary of the given YouTube Video. Basically, the video content is converted into a text document through subtitles, and then summarization algorithm is being applied to gain the desired output.

### 2.3 Features

- Conduct a large-scale investigation. The language process assists robots in mechanically comprehending and analyzing massive amounts of unstructured text data, such as social media comments, client support tickets, online reviews, news headlines, and more.
- In a short amount of time, automate procedures. Machines will be able to learn to type and route data with little to no human intervention - swiftly, effectively, precisely, and around the clock.
- Customize IP tools for your company. Terminology processing algorithms are frequently adjusted to your preferences and criteria, such as complex, industry-specific language – even witticism and misspelt terms.
- Machine learning algorithms are fed coaching information and expected outcomes (tags) to train machines to make associations between a certain input and its related output using IP tools to convert text into something a machine can understand.
- Grammar examination Parsing, also known as syntax analysis, detects the grammar structure of a text and, as a result, the dependence relationships between words, as shown in Figure 3.
- Linguistics research focuses on the distinctive features of language However, because language is polysomic and imprecise, linguistics is regarded as one of the most difficult areas of IP.
- Adding a region of speech category to each token in a text is known as part-of-speech tagging. Pos tags include, among others, unit verb, adjective, noun, pronoun, conjunction, preposition, and intersection.

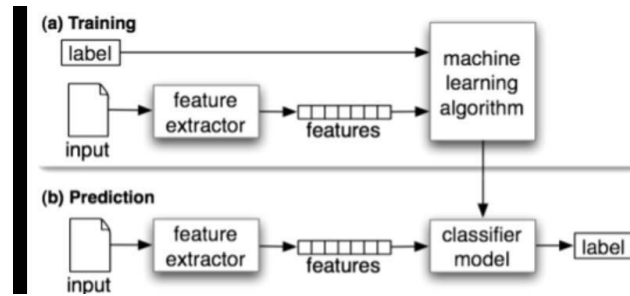


Fig 3: Training and prediction of a paragraph

**Several options of Chrome Extension:**

- **Architecture of a browser extension.** Chrome extensions work with three distinct areas that Communicate: Background, content and inner scripts. Every of them has their space of expertise inside the page and serves many functions.
- **Content script.** It acts within the space that the user visits and may serve to speaking regarding the present state of the page. Any data will be obtained from it and sent to the extension so as to alter its behavior.
- **Background script.** It acts within the browser area; it's the extension's event handler. This Is where all event listeners that area unit necessary to the extension area unit keep. It remains in active until an occasion is triggered and executes the logic that has been allotted there to.
- **Inner script.** (in the below Fig two.3.1, the popup.js) is answerable for the visual functionalities of the extension, interacting directly with the hypertext markup language of the extension and in conjunction with the background script, it will perform, for instance, calls to associate degree API and information come back one thing.

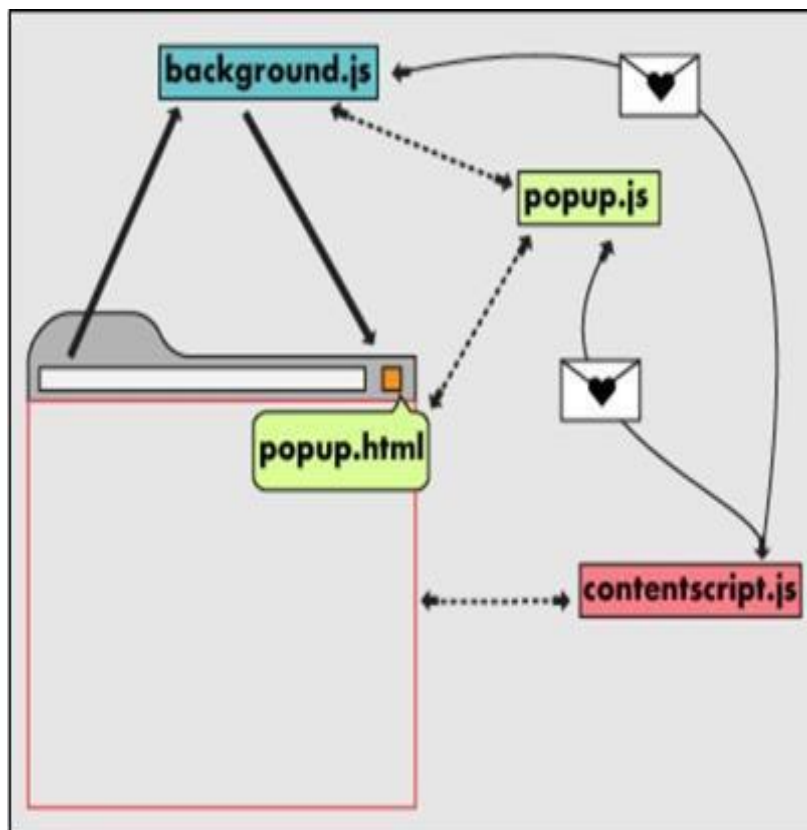
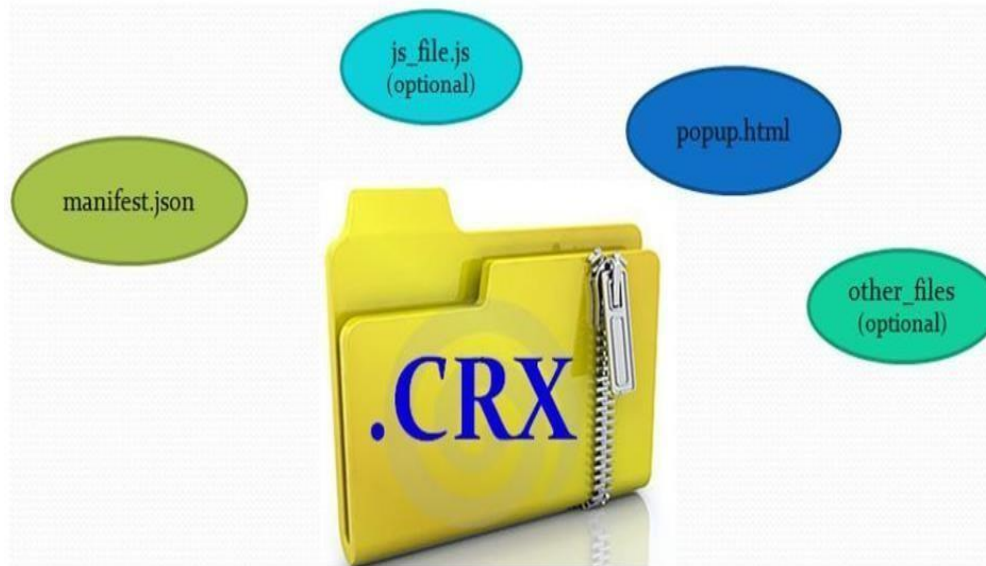


Fig 4: Chrome Extensions working process

**CRX File.** May be a Chrome Extension file want to extend the practicality of the Google Chrome application through very little programs that add further options to the default browsing expertise

Brief regarding CRX. It's saved during JS, JSON, and other files, such as photos and executable programmers, will be compressed into this format. CRX files are used to install browser extensions such as games, ad blockers, and news readers.



**Fig 5: CRX Image**

The on top of figure 5 displays the total file structure of CRX which incorporates numerous JS and JSON files, as we have a tendency to all understand a file with the CRX file extension may be a Google Chrome Browser Extension Archive file. Basically, these styles of files may be put in on Google Chrome to feature additional options to the browser. CRX files may be opened mistreatment WinZip/WinRAR or the other archive tools to view the contents of the pre packed plugin/extension (.CRX file).

### **III. Software and Tools Used**

This section covers the enhance technologies that has been utilized in this project with its supported platforms and its interface.

#### **3.1 Supported Platforms:**

Visual Studio Code is an ASCII text editor that supports Java, JavaScript, Go, Node.js, Python, and C++, among other programming languages. It's compatible with the Lepton framework, which is used to construct Blink-based Node.js web apps.

Visual Studio Code has a similar editing widget in Azure DevOps Instead of a project management system, it allows users to access one or more directories, which can then be saved in workspaces and processed later. As a result, it can be used as a code editor for any language.

#### **3.1.2 VS Code Interface**

##### **VS Code UI (User Interface)**

Visual Studio Code could be a code editor at its core. VS Code, like many other code editors, has a standard programming associate degreed layout with an individual on the left showing all of the files and directories you have access to, and an editor on the right showing the content of the files you've accessed.

VS Code has a simple and intuitive structure that maximizes the editor's space while still allowing you to browse and access the entire context of your folder or project.

The UI is divided into five areas: Editors, Sidebar, Status bar, Activity bar, and Panel. Figure 3 shows the many aspects of VS Code, such as Activity bar. Editor Groups, aspect bar and standing Bar.



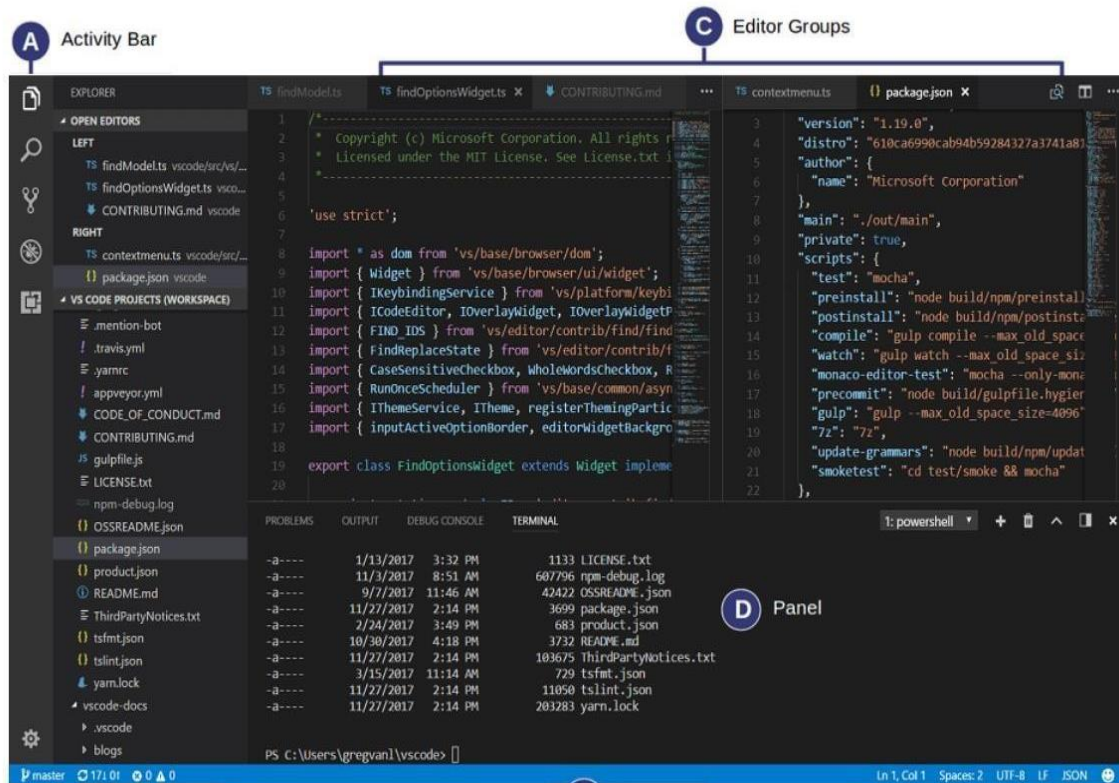


Fig 6: Home Page of VS Studio

### 3.2 Editors Group in VS Code

When you split Associate in Nursing editor, a new editor zone is created that can house a variety of items. You can open as many editor regions as you want, vertically and horizontally, aspect by aspect. You can see these plainly in the OPEN EDITORS column at the top of the person view: you can drag and Drop editor teams onto the worktable, move individual Tabs across teams, and fast close entire teams. Whether or not you have Tabs enabled, VS Code employs editor teams. Editor teams, unlike Tabs, are a stack of open items with the most recently selected item shown in the receiver pane.

### 3.2. Introduction to Google Colab

#### 3.2.1 What is Google Colab?

Google Colab could be a product of Google, from the name itself you'll be able to perceive. It'sprimarily a free notebook setting that runs absolutely within the cloud. Its options that assist

#### 3.2.2 Need of Google Colab in our project

Google is kind of concerned within the AI analysis and development half. One of the most popular AI frameworks was created by Google known as Tensor Flow and additionally a development tool known as Collaborator. Nowadays, Tensor Flow is one among the foremost used open- supply tools since 2017

### 3.3 Libraries Used

#### 3.3.1 Tensor Flow

Tensor Flow is a deep learning ASCII text file library with an associate's degree. It was originally designed by the Google Brain Team to do massive computations. It wasn't designed with deep learning in mind. They quickly realized, however, that Tensor Flow was useful for deep learning implementations, and they've subsequently implemented it as an ASCII text file resolution. Tensor Flow is a collection of deep learning and machine learning algorithms and models.

You can easily train and run deep neural networks for a variety of metric capacity unit applications using Tensor Flow. Word embedding's, written digit classification, continuous neural networks, picture recognition, tongue process, and partial equation simulations are all examples of these. You can run those programs with C++ with Tensor Flow and get exceptional performance.



Fig 7: Tensor Flow

### 3.3.2 YouTube Transcript API Library.

Python has an enormous number of API general from which to choose. Google has an API for each and every service it offers. YouTube Transcript API is one of them, and it's quite easy to use and has a lot of features. This module is used to extract captions and subtitles from YouTube videos. If YouTube video has the following uniform resource locator, <https://youtu.be/SW14tOdaKi>, for example, the video id for this film would then be "SW14tOda Ki," which would include all phrases after the? v=counts reason the video id. We'll start with the basics now. We tried to collect the transcript of the video ID mistreatment the in the beginning. () obtain a transcript

```
[{'text': 'hello friends and welcome to geeks for', 'start': 0.0, 'duration': 1.92}, {'text': 'geeks in this tutorial we will learn how', 'start': 1.92, 'duration': 2.64}, {'text': 'to reverse a stack using recursion so', 'start': 4.56, 'duration': 2.279}, {'text': 'here is the problem we have to reverse a', 'start': 6.839, 'duration': 2.131}, {'text': 'stack and we are not allowed to use', 'start': 8.97, 'duration': 1.65}, {'text': 'loops like while four etc we can use', 'start': 10.62, 'duration': 4.05}, {'text': 'functions of stacks like is empty push', 'start': 14.67, 'duration': 3.42}, {'text': 'and pop the idea of the solution is to', 'start': 18.09, 'duration': 3.51}, {'text': 'hold all the values and function call', 'start': 21.6, 'duration': 2.46}, {'text': 'stack until the start becomes empty when', 'start': 24.06, 'duration': 3.059}, {'text': 'the start becomes empty we insert all', 'start': 27.119, 'duration': 2.4}, {'text': 'the hell values one by one at the bottom', 'start': 29.519, 'duration': 3.031}, {'text': 'of the start let us now look at how it', 'start': 32.55, 'duration': 2.22}, {'text': 'works', 'start': 34.77, 'duration': 0.51}, {'text': 'let us assume our stack as 1 2 3 4 with 1', 'start': 35.28, 'duration': 4.92}, {'text': 'at top we pop all the values 1 by 1 and', 'start': 40.2, 'duration': 4.41}, {'text': 'hold all of them in function call stack', 'start': 44.61, 'duration': 2.73}, {'text': 'until this side becomes empty so here we', 'start': 47.34, 'duration': 3.059}, {'text': 'go we pop all the values until the side', 'start': 50.399, 'duration': 4.171}, {'text': 'becomes empty so a function called stack', 'start': 54.57, 'duration': 2.94}, {'text': 'in order is 4 3 2 and 1', 'start': 57.51, 'duration': 4.639}, {'text': 'since the stack is empty now we need to', 'start': 62.149, 'duration': 3.19}, {'text': 'insert all the hello elements one by one', 'start': 65.339, 'duration': 2.491}, {'text': 'at the bottom of stack now how to insert', 'start': 67.83, 'duration': 3.42}, {'text': 'elements at the bottom of stack to do so', 'start': 71.25, 'duration': 3.24},
```

Figure 8: is a sample image which shows the working of YouTube Transcript API

### 3.3.3 Hugging Face and Transformers

Hugging Face might be a big ASCII text file community that soon became a popular destination for pre-trained deep learning models geared mostly at data processing. The employment of Transformers is at the heart of their natural language processing strategy. The Python-based Transformers library presents a well-designed API for leveraging a large number of deep learning architectures for cutting-edge data processing applications such as those previously stated. One central startup, you may have imagined.

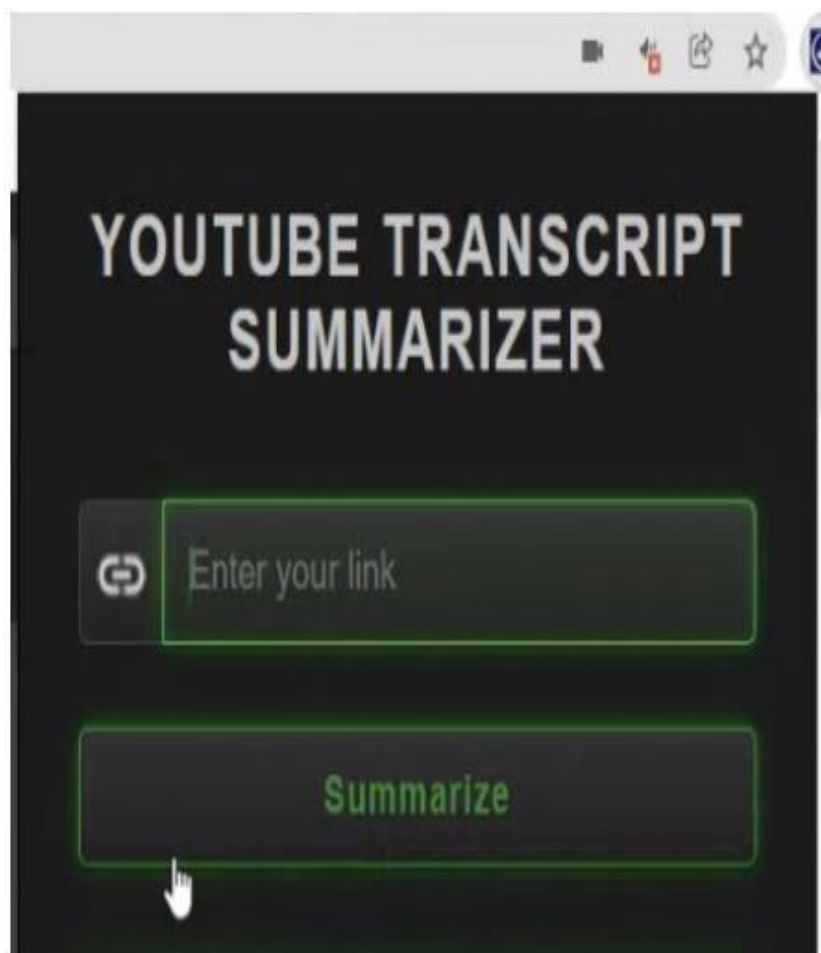
Reusability is important—all of their models come with a set of pre-trained weights that you may fine-tune for your individual needs. Transformers is a self-opinionated library for NLP researchers and educators who want to utilize, examine, and extend large-scale transformers models.

### **3.3.4 Working of Chrome Extension**

In this section the operating of Chrome Extension is mentioned with each major technicalside.

#### **UI of Chrome Extension.**

The UI of Chrome Extension is developed enterprise its friendly usage and ease, it's coming upwith is done with the assistance of JavaScript, HTML, and CSS Alternative technologies such as CSS and JavaScript augment and change the basic structure of websites. CSS is used in order to control the look, feel, and arrangement of your presentation. JavaScript is used to control the behavior of different elements.



**Fig 9: UI of Chrome Extension**

## **IV. Conclusion and Future Scope**

### **4.1 Conclusion:**

There is continually an area for enhancements in any computer code package, but smart and economical it may be done. However, the foremost vital factor ought to be flexibility to just acceptany modification. Right now, we have a tendency to area unit simply coping with video transcriptsummarization, whereas in future this computer code may be extended to be delivered as an internet application with its chrome extension. Posting a transcript aboard a speaker video, for example, makes it potential for users with disabilities to scan the text or hear it.

### **4.2 Future Scope**

Here are a unit some Future Scope of our comes

It is often utilized in analysis work, it'll cut back the time consumption, someone rather than observation a video will directly scan the transcript outline. Posting a transcript aboard a speaker video, for instance, makes it potential for users with disabilities to scan the text or hear it.



**REFERENCES:**

- [1]. <https://docs.python.org/3/>
- [2]. <https://developer.chrome.com/docs/extensions/>
- [3]. Natural Language Toolkit <https://www.nltk.org/>
- [4]. Christian Bokhove, Christopher Downey, “Automated generation of “good enough” transcripts as first step to transcription of audio-recorded data” (Published in Jul’18) [https://www.researchgate.net/publication/326405221\\_Automated\\_generation\\_of\\_good\\_enough\\_transcripts\\_as\\_a\\_first\\_step\\_to\\_transcription\\_of\\_audio-recorded\\_data](https://www.researchgate.net/publication/326405221_Automated_generation_of_good_enough_transcripts_as_a_first_step_to_transcription_of_audio-recorded_data)
- [5]. <https://developer.mozilla.org/en-US/docs/web/javascript>
- [6]. Morgan Persson “JavaScript DOM Manipulation Performance Comparing Vanilla JavaScript and JavaScript Front-end Frameworks “(Published in May’20)<https://www.diva-portal.org/smash/get/diva2:1436661/FULLTEXT01.pdf>
- [7]. Nidhi Mittal, Manish Kumar Mukhija, Satish Kumar Alaria, “Design Implementation and Assessment of Efficient Brain Tumor Detection and Classification System Using Improved Machine Learning Techniques”, Turkish online Journal of Qualitative Inquiry (TOJQI), pp. 4124 -4140, Vol. 12, Issue:8, July 2021, e-ISSN 1309-6591.
- [8]. Ravi Khandelwal, Manish Kumar Mukhija, Satish Alaria, “Numerical Simulation and Performance Assessment of Improved Particle Swarm Optimization Based Request Scheduling in Edge Computing for IOT Applications”, International Journal of Contemporary Architecture "The NewARCH", ISSN 2198-7688, pp. 155-169, Vol. 8, No. 2, 2021.
- [9]. Manish Kumar, Dr. Sunil Kumar, Dr. Harish Nagar, “Comparative Analysis of Different Cryptography Technique for image or Data Security”, in Wesleyan Journal of Research, Vol. 13, No. 69, pp.9-20, March, 2021, UGC, Care-listed, ISSN:0975-1386.
- [10]. Davis David, Published in Towards Data Science, Jan 10 2021. <https://towardsdatascience.com/how-to-detect-and-translate-languages-for-nlp-project-dfd52af0c3b5>
- [11]. Training and Predication of a paragraph source- Shivam Bansal, Published in Analytics Vidhya, January 12, 2017.
- [12]. Chrome Extension working Process source- Published on Tuesday, September 18, 2012 • Updated on Thursday, June 7, 2018 <https://developer.chrome.com/docs/extensions/mv2/architecture-overview/>
- [13]. Fig 3.3.1 Tensor flow Image source-<https://github.com/tensorflow/tensorflow>