

# **1. INTRODUCTION**

## 1.1. Purpose

### 1.1.1. Introduction

This Software Requirements Specification provides a complete description of all the functions and specifications of the Chrome Extension Multi Language YouTube Transcript Summarizer.

In this project, I have created a Chrome Extension which will make a request to a backend REST API where it will perform NLP and respond with a summarized version of a YouTube Transcript.

### 1.1.2. Scope

Scope of this project is :Enormous number of video recordings are being created and shared on the Internet throughout the day. It has become really difficult to spend time watching such videos which may have a longer duration than expected and sometimes our efforts may become futile if we couldn't find relevant information out of it. Summarizing transcripts of such videos automatically allows us to quickly lookout for the important patterns in the video and helps us to save time and effort to go through the whole content of the video.

### 1.1.3. References

- I. Building RESTful APIs with Flask in Python Boiler Plate
- II. Hugging Face Transformer Python Installation
- III. YouTube Transcript API Documentation
- IV. How to Perform Text Summarization using Transformers in Python
- V. The Ultimate Guide to Building a Chrome Extension
- VI. How to Create Chrome Extensions
- VII. Design the user interface

VIII. Content Scripts

IX. Message Passing in Chrome

X. How to use XML, XMLHttpRequest to issue HTTP requests

XI. Language Translator Using Google API in Python

XII. Parsing REST API Payload and Query Parameters With Flask

#### **1.1.4. Document overview**

The remainder of this document is 8 chapters, the first providing introduction of the project. It lists all the functions performed by the system. The second chapter consists of software requirements specification and all the dependencies. The third chapter provides details about system analysis and design. The fourth chapter gives backend programming

and data dictionary information. The fifth chapter is of Chrome Extension development. The sixth chapter gives testing for the project and browsing experience without distracting from it and proper working of the project. The seventh chapter tells about the conclusion and future enhancements of the project. The final chapter concerns with the bibliography.

This document is meant for describing all the features and procedures that were followed while developing the Extension.

This document specially mentions the details of the project how it was developed, the primary requirement, as well as various features and functionalities of the project and the procedures followed in achieving these objectives.

Multi Language YouTube Transcript Summarizer is a Chrome Extension which will help you get the short summary of contents of a video so that one can save time and get the meaningful value of that video they were seeking for in a faster way possible.

For example, a student is watching a video to understand some topic for his/her study, my motto is to help the student by transcribing the audio and generate subtitle of that video and summarize that content to make the student understand the topic faster and simplest and easy way. Benefit is that this Extension can summarize the content so that the student might get help in learning and making notes.

This Chrome Extension will transcribe audio from a video and generate the short summary of the content and this is a very useful extension for people who are looking for a specific video and most useful in video conferences to make summarized notes.

### **1.1.5. Overall description**

Multi Language YouTube Transcript Summarizer Chrome Extension is designed for everyone who wants to save time on watching and searching for the content they actually need without watching every other video that pops up when you search for one particular video that will help you in providing satisfactory material.

The Chrome Extension will lead you to a page where there will be summary of a youtube video(text) or a summarized audio in whichever form they require summarization.

### **1.1.6. Functional requirements definitions**

Functional Requirements are those that refer to the functionality of the system, i.e., what services it will provide to the user. Non-functional (supplementary) requirements pertain to other information needed to produce the correct system and are detailed separately.

### **1.1.7.            Use cases**

This system will be used in different required ways of a user, as not only YouTube video summarization but also videos from websites, video conferences from different region with diverse language-based summarization to understand the content on their own language.

➤ User can do the following functions in the Chrome Extension:

- I.     Summarization in text
- II.    Summarization in audio
- III.   Make notes
- IV.    Understand the main highlights
- V.     Create questionaries for exam purpose
- VI.    Quick Revision

### **1.1.8.    User characteristics**

The user should be familiar with YouTube videos, Subtitles, Chrome Extensions to be able to use them and download them and understand its working and features to use it.

The user should be familiar with Chrome (google).

### **1.1.9. Constraints**

The video should not be long enough because of limited word limit criteria, else it might throw error and not generate summary.

Multilingual support function in audio summary is not available in Extension.

# **2. Software Requirement Specification and Dependencies**

## **Purpose**

### **2.1.1. Introduction**

This Software Requirements Specification provides a complete description of all the functions and specifications of the Chrome Extension Multi Language YouTube Transcript Summarizer.

The main objective of Chrome Extension Multi Language YouTube Transcript Summarizer is to save the useful energy and time we spend on finding a satisfactory YouTube video and to select that satisfactory YT video one summarizes its content and use it to understand.

### **2.1.2. Scope**

Scope of this project is very useful for finding the perfect content we are looking for. Few of them are: -

- I. Summarizing transcripts of such videos automatically allows us to quickly lookout for the important patterns in the video.
- II. Can be used anywhere any time as it is a web-based application.
- III. No restriction, it is free and easily used especially designed for content creators.

### **2.1.3. Glossary**

Table 2.1

Term	Definition
Admin	The only user who has the right to give permission to which URLs can be accessible in that Extension.
Entry	Video-Id of a YouTube video to differ from others and get the result on every tab they use for YouTube.
Html	Hypertext markup language
QA	Quality assurance
SCMP	Software Configuration Management Plan
SDD	Software Design Document
SQAP	Software Quality Assurance Plan
SRS	Software Requirements Specification
Web Application	A place on the world wide web which provides any accessible application.

#### **2.1.4. References**

- I. Building RESTful APIs with Flask in Python Boilerplate
- II. Hugging Face Transformer Python Installation
- III. YouTube Transcript API Documentation
- IV. How to Perform Text Summarization using Transformers in Python
- V. The Ultimate Guide to Building a Chrome Extension
- VI. How to Create Chrome Extensions
- VII. Design the user interface
- VIII. Content Scripts
- IX. Message Passing in Chrome
- X. How to use XML Http Request to issue HTTP requests
- XI. Language Translator Using Google API in Python
- XII. Parsing REST API Payload and Query Parameters with Flask

#### **2.1.5. Document overview**

The remainder of this document is two chapters, the first providing a full description of the project from the admin of the Multi Language YouTube Transcript Summarizer.

It lists all the functions performed by the system. The final chapter concerns details of each of the system functions and actions and dependencies in full for the Chrome application developers' assistance. These two sections are cross-referenced by topic; to increase understanding by developers and users involved.

## **2.2. Overall description**

Multi Language YouTube Transcript Summarizer is designed for each and every person who needs to find the satisfied content they are looking for in short amount of time.

Enormous number of video recordings are being created and shared on the Internet

throughout the day. It has become really difficult to spend time watching such videos which may have a longer duration than expected and sometimes our efforts may become futile if we couldn't find relevant information out of it. Summarizing transcripts of such videos automatically allows us to quickly look out for the important patterns in the video and helps us to save time and effort to go through the whole content of the video.

### **2.2.1. Functional requirements definitions**

Functional Requirements are those that refer to the functionality of the system, i.e., what services it will provide to the user. Nonfunctional (supplementary) requirements pertain to other information needed to produce the correct system and are detailed separately.

### **2.2.2. Use cases**

This system will be used in different required ways of a user, as not only YouTube video summarization but also videos from websites, video conferences from different region with diverse language-based summarization to understand the content on their own language.

#### **User can do the following functions in the Chrome Extension:**

- I. Summarization in text
- II. Summarization in audio
- III. Make notes
- IV. Understand the main highlights
- V. Create questionaries for exam purpose
- VI. Quick Revision for study with summarized material

### 2.2.3 Use Case: Access Chrome Extension

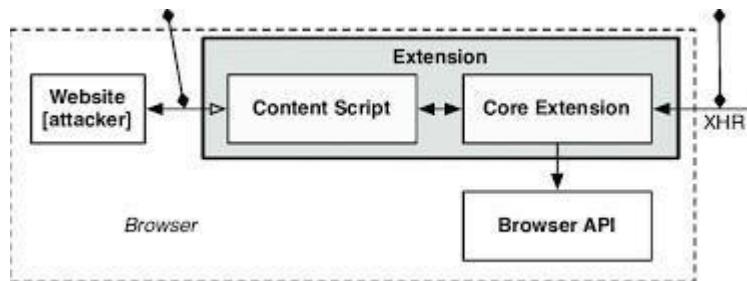


Fig. 2.1 Accessing Chrome Extension

#### Brief Description:

User uses the Chrome Extension of Multi Language YouTube Transcript Summarizer to access the Extension.

#### Initial step-by-step description:

For this use case to be initiated, the user can use the Chrome Extension Multi Language YouTube Transcript Summarizer by:

- I. The user connects to the system using a web browser compulsory Chrome browser.
- II. The user selects the Extension icon on Chrome browser at top-right corner which looks like a little greyish colored puzzle piece little icon.
- III. The system passes the user to the Chrome Extension page where their will all the Extensions are available.
- IV. The user needs to find the search button on Chrome extension page and type the name of extension that is “YTSUMMARIZER”.
- V. Then the user clicks on the search appeared Extension and clicks on Add to chrome blue colored button and the extension will be added to the users chrome browser.
- VI. Then user should pin the extension with the pinning icon present near extensions name.
- VII. Then whenever they want to access it, they can without going further inside to find that extension after pinning it.

#### 2.2.4 Use Case: User Login or Signup

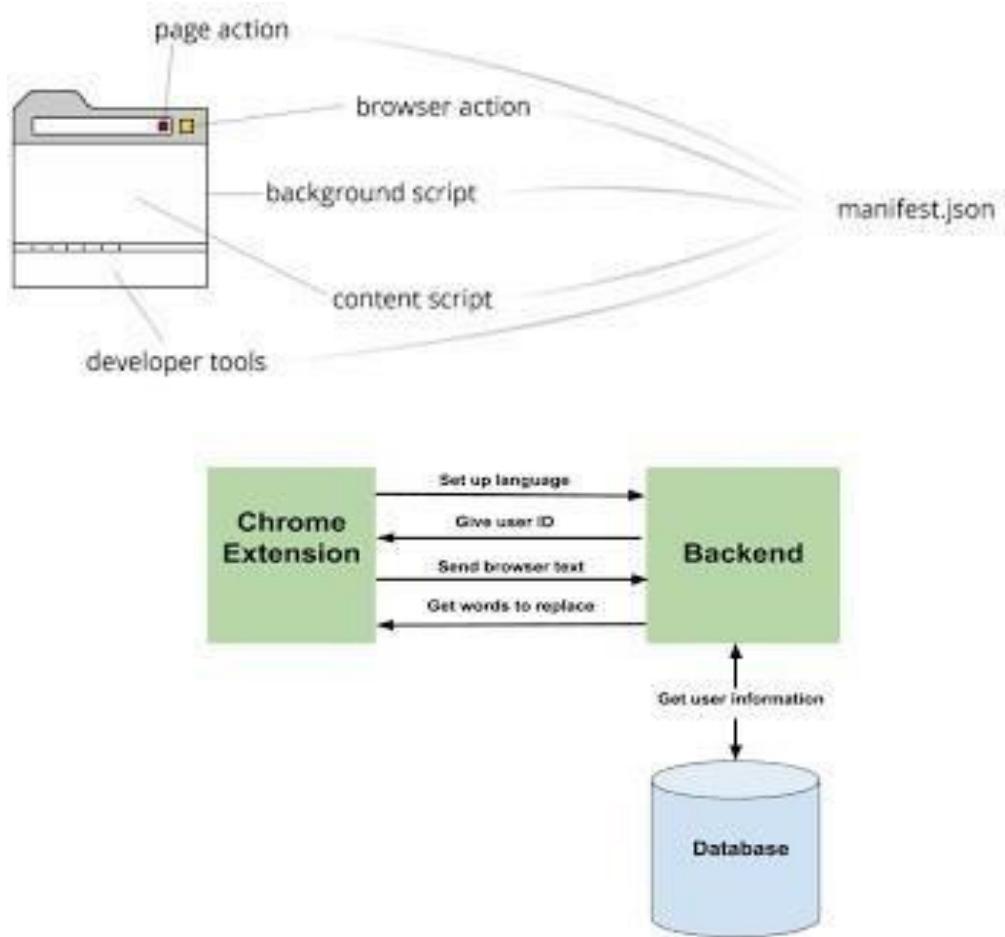


Fig. 2.2 User Login or Signup

#### Brief Description:

The user doesn't need to either log in or sign up if they already have chrome browser account to access a Chrome Extension.

#### Initial step-by-step description:

For this use case to be initiated the user must on the chrome browser.

- I. The system passes the user to the Chrome Extension page where their will all the Extensions are available.
- II. The user needs to find the search button on Chrome extension page

and type the name of extension that is “YTSUMMARIZER”.

- III. Then the user clicks on the search appeared Extension and clicks on Add to chrome blue colored button and the extension will be added to the users chrome browser.
- IV. Then user should pin the extension with the pinning icon present near extensions name.
- V. Then whenever they want to access it they can without going further inside to find that extension after pinning it.

## 2.3. Use Case: 2 formats of summarization

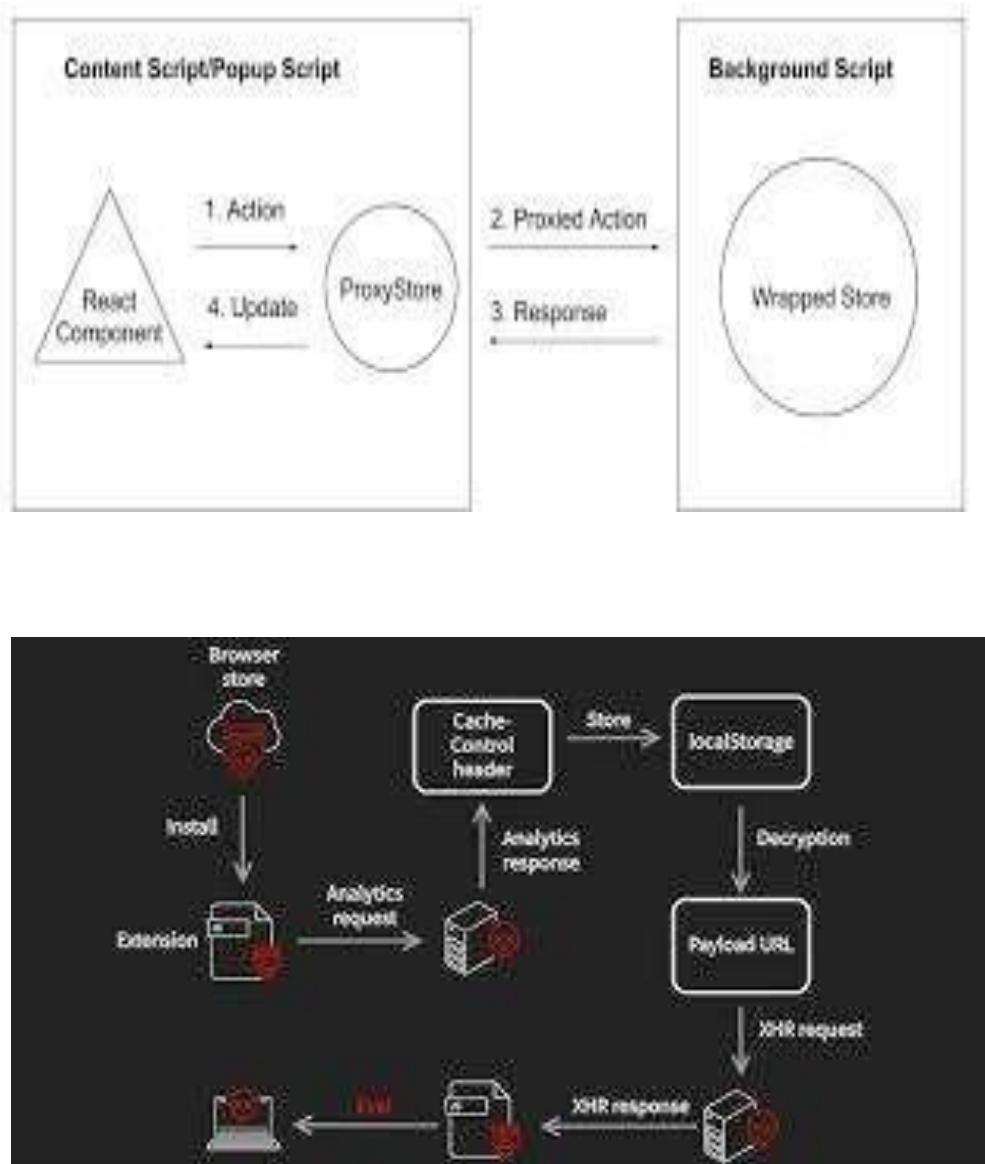


Fig. 2.3 Admin has provided two forms in Summarization Text and Audio.

Admin has provided to forms in Summarization Text and Audio for user.

### **Initial step-by-step description:**

For this use case to be initiated, the user must select option from 2 forms of summarization available

1. The user should select the button and click on it, whichever form they want summarization in.
2. The 2 buttons are text summarization and audio summarization.
3. The user when select text summarization the summary will start processing if the video is subtitle eligible and then it'll show the summary on the same page with a popup like box and you can copy paste it.
4. The user when selects audio summarization button, a summarized audio will get processed and it will play the audio.
5. Then they can make notes or record the audio on some recording device by playing it loud.

### **2.3. Requirement specifications**

Knowledge on how to use Extension or at least to know what the extensions are.

2.3.1. External interface  
specifications None

### 2.3.2. Functional Requirements

Table 2.2 Access Chrome Extension

Use Case Name:	Accessing Chrome Extension
Priority	User
Trigger	Summarization format selection
Precondition	User is on the chrome browser.
Basic Path	<ol style="list-style-type: none"> <li>1. The user connects to the system using a web browser compulsory Chrome browser.</li> <li>2. The user selects the Extension icon on Chrome browser at top-right corner which looks like a little greyish colored puzzle piece little icon.</li> </ol> <p>The system passes the user to the Chrome Extension page where their will all the Extensions are available.</p>
Alternate Path	N/A
Postcondition	The User is on the Chrome browser.
Exception Path	If there is a connection failure it'll give you the result after getting connected again to the internet.
Other	

Table 2.3User Login or Signup

Use Case Name:	User Login or Signup
Priority	Chrome browser Login is enough.
Trigger	Selects
Precondition	The User is on the chrome extension page.
Basic Path	<ol style="list-style-type: none"> <li>1. The system passes the user to the Chrome Extension page where their will all the Extensions are available.</li> <li>2. The user needs to find the search button on Chrome extension page and type the name of extension that is “YTSUMMARIZER”.</li> <li>3. Then the user clicks on the search appeared Extension and clicks on Add to chrome blue colored button and the extension will be added to the users chrome browser.</li> <li>4. Then user should pin the extension with the pinning icon present near extensions name.</li> </ol> <p>Then whenever they want to access it, they can without going further inside to find that extension after pinning it.</p>

Alternate Path	Pinning the extension is up to the user.
Postcondition	The user is on the chrome browser watching content or doing whatever work.
Exception Path	If the connection is terminated it'll later on load the summarization after the connection is back.
Other	

## 2.4. Hardware

### Specification USER

#### Side:

Chrome browser: latest version (92,93) it is ok if not the latest version available.

Processor: All

RAM: All

Hard Disk: All

#### Server Side:

Processor: All

RAM: All

Disk space: All

## 2.5. Software

### Specification

#### User Side:

Chrome Browser

Chrome Extension icon on chrome

#### browser Data Base Server:

All

## 2.6. Hardware and Software Requirements in detail

### Hardware Requirements:

Processor: All

Chrome Browser (92,93): with the updated versions if not the latest

RAM: All

Hard Disk: All

Disk space: All

## **Software Requirements:**

- a) Microsoft Visual Studio 2010
  - a. Microsoft Visual Studio is an integrated development environment (IDE) from Microsoft. It can be used to develop console and graphical user interface applications along with Windows Forms applications, web sites, web applications, and web services in both native codes together with managed code for all platforms supported by Microsoft
    - 1. Windows, Windows Mobile, Windows CE, .NET Framework, .NET Compact Framework and Microsoft Silverlight.
  - b. Python 3.10 version the latest Compiler
  - c. Pipeline
  - d. Flask, flask RESTful
  - e. JS, HTML, JSON compiler installed in VS code.
  - f. .NET Framework 4.8 or nearest below to 3.5
  - g. Natural language processing
  - h. Transformers
  - i. YouTube Transcript Api
  - j. TensorFlow
  - k. Chrome browser (92,93) versions

# **3.System Analysis and Design**

### **3.1 Study & Weaknesses of Current Working/Processing Current Working/Processing**

The Current Working/Processing of Extension is not at its best speed accuracy or satisfactory fast. Whenever a user performs summarization, the summarization is only available for the videos who already have subtitle eligibility.

- Subtitle eligible video
- Audio summary is not at its accuracy level.
- Extra short summary in text of extra small

videos. Weaknesses in Current System

The current Working/Processing is as mentioned earlier of Extension is not at its best speed accuracy or satisfactory fast. Whenever a user performs summarization, the summarization is only available for the videos who already have subtitle eligibility.

As Subtitle should be in built in a YouTube video to perform the summarization task and it will throw error when the video is large because this extension only has capacity of 1024 word limit and also larger videos can't produce summary in Audio format.

Text summarization is the task of shortening long pieces of text into a concise summary that preserves key information content and overall meaning. There are two different approaches that are widely used for text summarization:

- Extractive Summarization: This is where the model identifies the important sentences and phrases from the original text and only outputs those.
- Abstractive Summarization: The model produces a completely different text that is shorter than the original, it generates new sentences in a new form, just like humans do.

In this project, we will use transformers for this approach.

### **3.1. Requirements of Updated version**

I.Speed

II.Accuracy

III.Larger videos eligible for summarization

IV.Summarization of no-subtitle eligible videos.

#### **3.1.1. User Requirements**

The User requirements to recommend this extension to others and tell the admin by emailing the working and feedback.

- Time can be saved in text summarizations
- Feedback
- Proper internet connection clear audio.

### **3.2. Feasibility Study**

A key part of the preliminary investigation that reviews anticipated costs and benefits and recommends a course of action based on operational, technical, economic, and time factors. The purpose of the study is to determine if the systems request should proceed further.

#### **3.2.1. Does the New System Contribute to the Overall Objectives of the Extension?**

The new system would contribute to the overall objectives to of the Extension. It would provide a quick, error free and zero cost solution to the current process. It would provide a solution to many issues in the current system. As the new system is flexible and scalable it can also be upgraded and extended to meet other complex requirements which may be raised in the future.

#### **3.2.2. Can the New System be Implemented Using Current Technology?**

The Extension developer has latest and updated laptop and its all the require functions with intel core processor (i5 8<sup>th</sup> Gen) which is connected by Internet LAN

Wi-fi network 24/7 and managed by a server. It would be very easy to set up the Extension in the current environment as the application is web based. The database and the chrome browser's latest version are installed on machine, the system can be started as quick as required by the developer/User.

### **3.3. Features of the New System.**

The new system has been designed as per the user requirements so as to fulfil almost all them.

#### **1.2 Speedy Processing**

Summary can be generated very quickly as compared to the existing extension as it allows the use of previously generated summaries to see again with data stored on cloud. It saves time required to get the summarization of audio and text faster.

#### **1.3 Accuracy**

One of the most important draw backs of the current system is that audio is not at its best accuracy and can't be generated on longer size videos because of word limit. The new system will generate the result as soon as the summarization is processed by user and will also store it in the database for future usage.

#### **1.4 High-quality Audio**

The new system makes it easy to store and retrieve information as required and does not involve storing information by the user-self it's on Auto mode from cloud. It thus saves data management problems faced in the current system as it has a Database Management System of only one-time access.

#### **1.5 Zero Cost and No Advertisements**

Unique service provider this extension as it does not show any advertisements and provide no-cost service.

### 3.4. Data Flow Diagram (DFD)

The DFD (also known as bubble chart) is a simple graphical formalism that can be used to represent a system in terms of the input data into the system, various processes carried on these data, and the output data generated by the system.

The main reason why the DFD technique is so popular is because the fact that the DFD is a very simple formalism – it is simple to understand and use. A DFD model uses a very limited number of primitive symbols to represent the functions performed by a system and the data flow among the functions. Starting with a set of high-level functions that a system performs, a DFD model hierarchy represents various sub-functions.

#### 3.1 DFD level 0

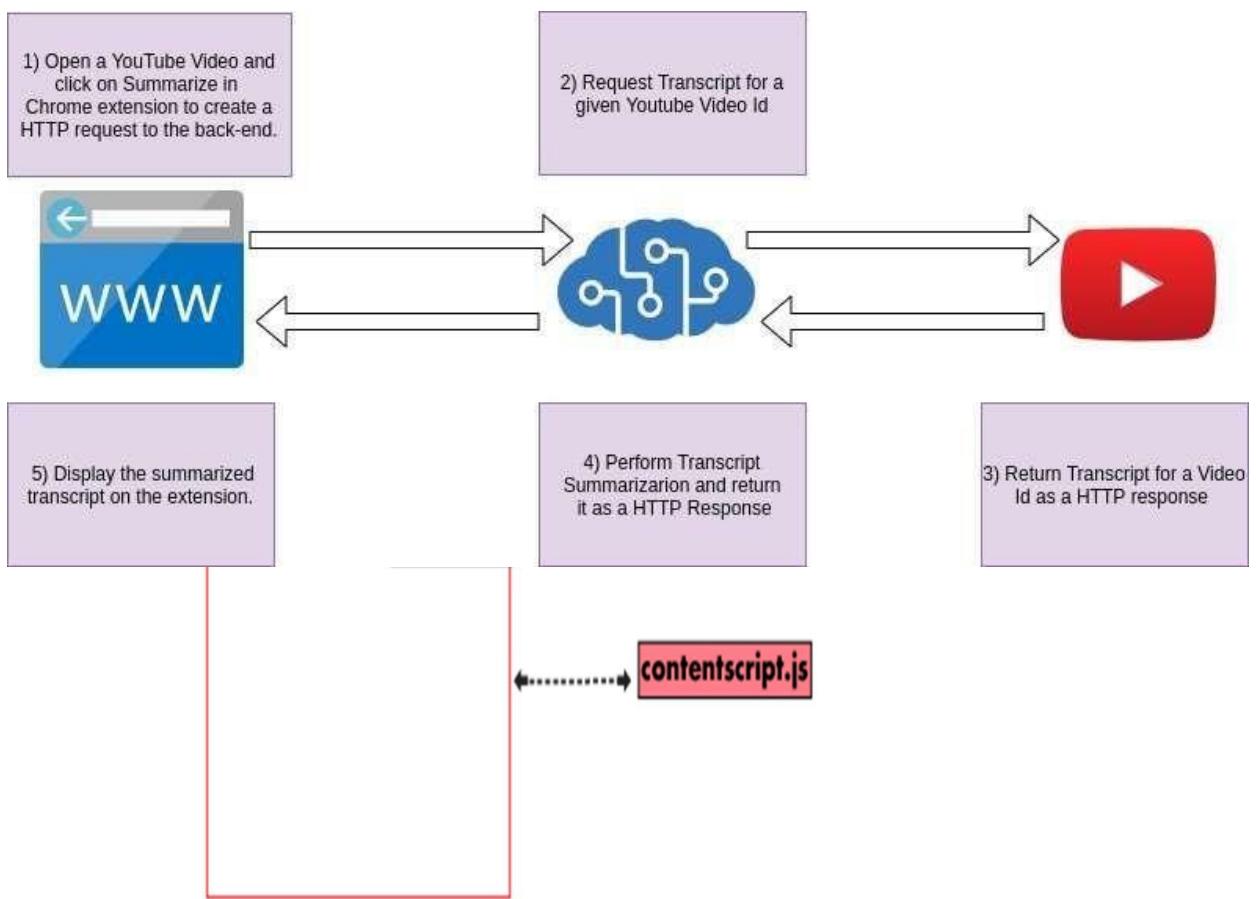


Fig. 3.2 DFD level 1

## UML Modelling

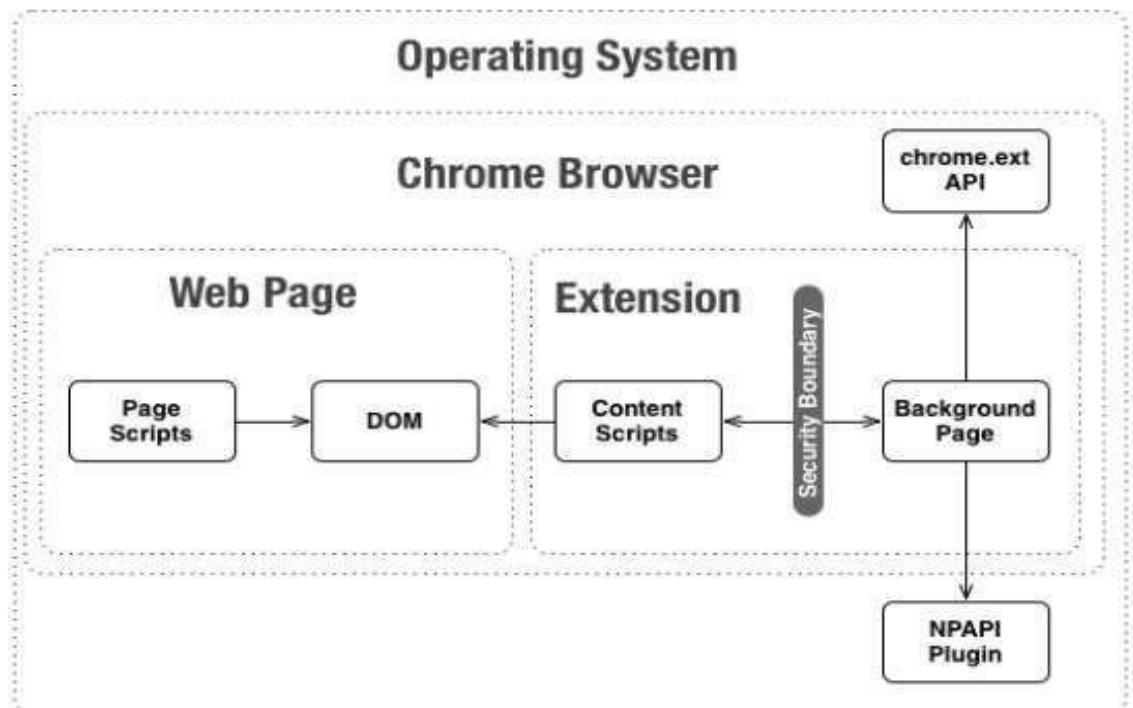


Diagram by Wade Alcorn. Thanks!

OWASP



6

Fig. 3.3 UML Modelling

### 3.6.3 Context Diagram

The context diagram is a top-level view of an information system that shows the boundaries and scope. It describes the main objective of the system and the entities involved.

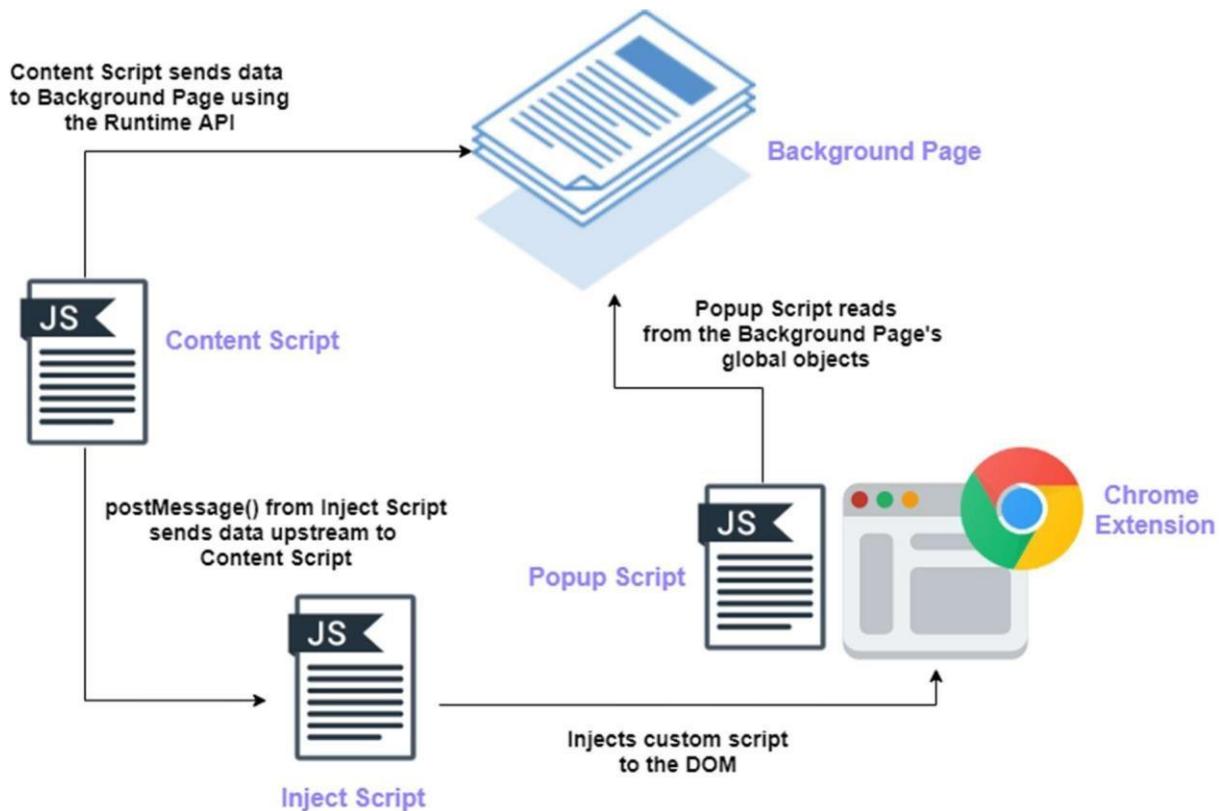


Fig. 3.4 Context Diagram

# **Data Dictionary**

A data dictionary is a catalog-a-repository of the elements in a system. As the name suggests, their elements center on data and the way they are structured to meet user requirements and organization needs. In a data dictionary you will find a list of all the elements composing the data flowing through a system. The major elements are data flows, data stores and processes. The data dictionary stores details and descriptions of these elements.

If analysis want to know characters are in a data item by what other names it is referenced in the system, or where it is referenced in the system, or where it is issued in the system, they should be able to find the answers in issued in the system, they should be able to find the answer in properly developed data dictionary.

The Dictionary contains two types of description for the data following through the system.

## **1. Data Elements**

The most fundamental data is the elements. They are building blocks for all other data in the system. Data elements are also alternatively known as fields, data item or elementary item.

## **2. Data Structure**

A data structure is a set if items that are related to one another and described a component in the system.

#### 4.1. Table Details

Table 4.1. User Accessing Extension

Field Name	Description	Constraints	Size	Data Type
Text Summary	Unique video - id /URL		1024	Varchar, int
Audio Summary	Unique video - id /URL		1024	Varchar, int

Table 4.3. Answer

Field Name	Description	Constraints	Size	Data Type
YTSUMMARIZE R	Chrome Extension for video summarization		1024	varchar, int

## E-R Diagram

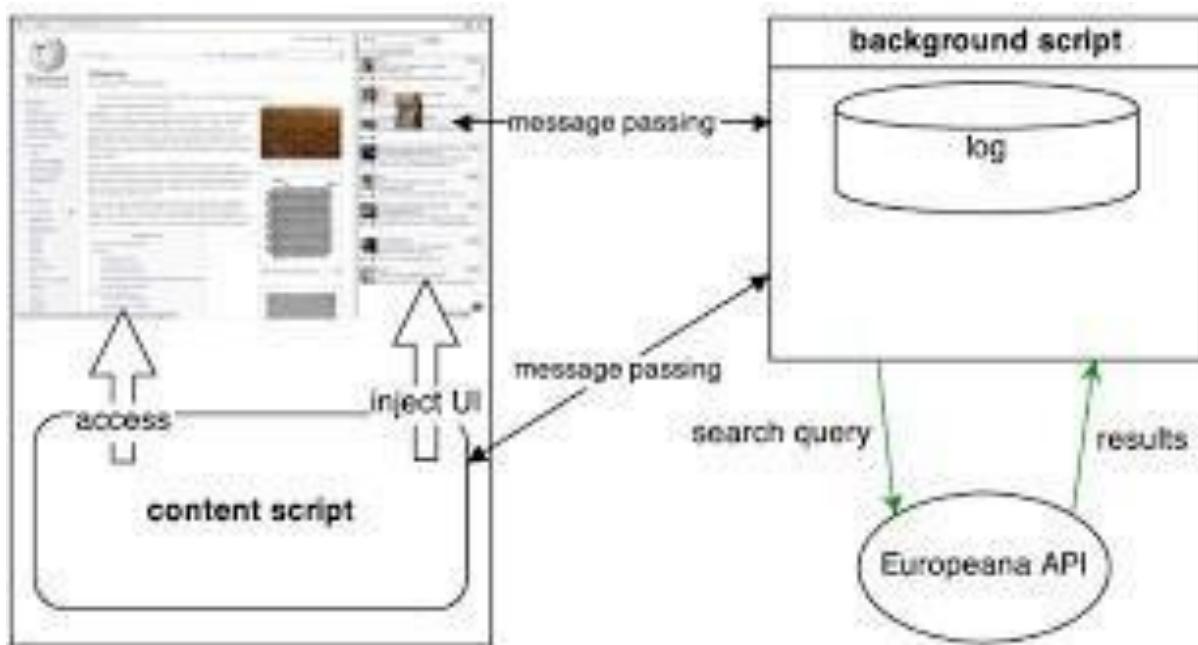


Fig. 4.1 E-R Diagram

## **5.SCREENSHOTS**

## 5.1. Google Chrome Browser

### 5.1.1. Browser home page

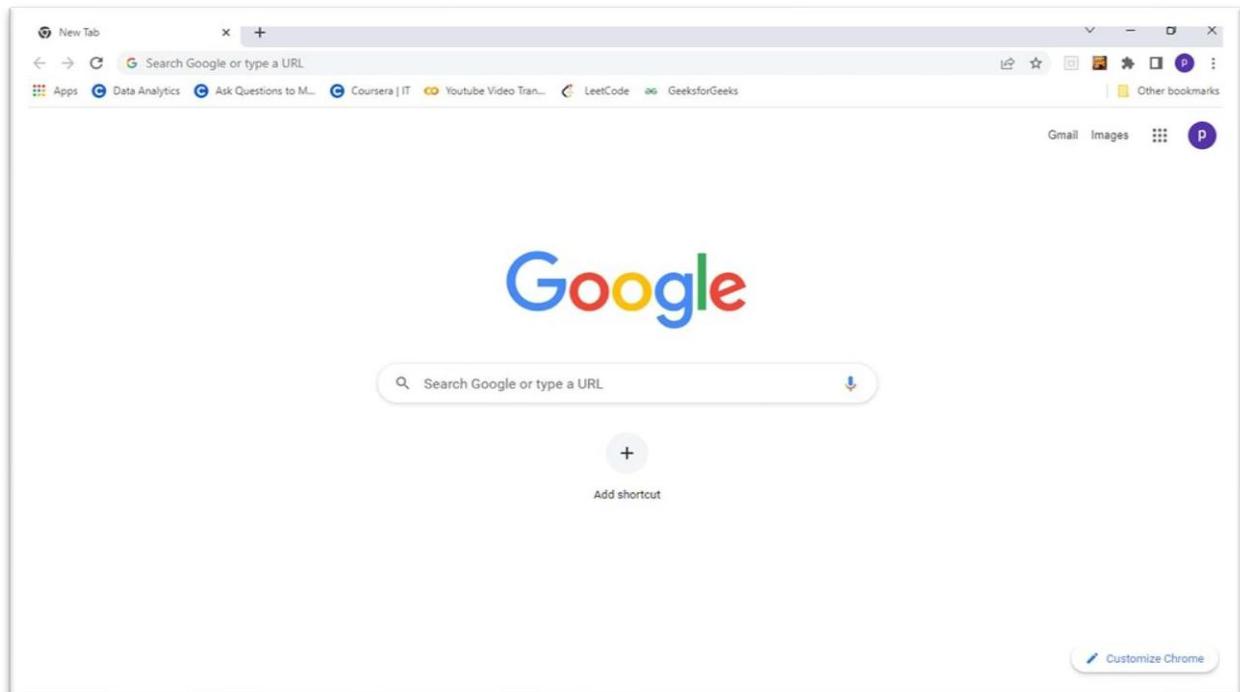


Fig-5.1 Browser home page

### 5.1.2. Chrome Extensions icon

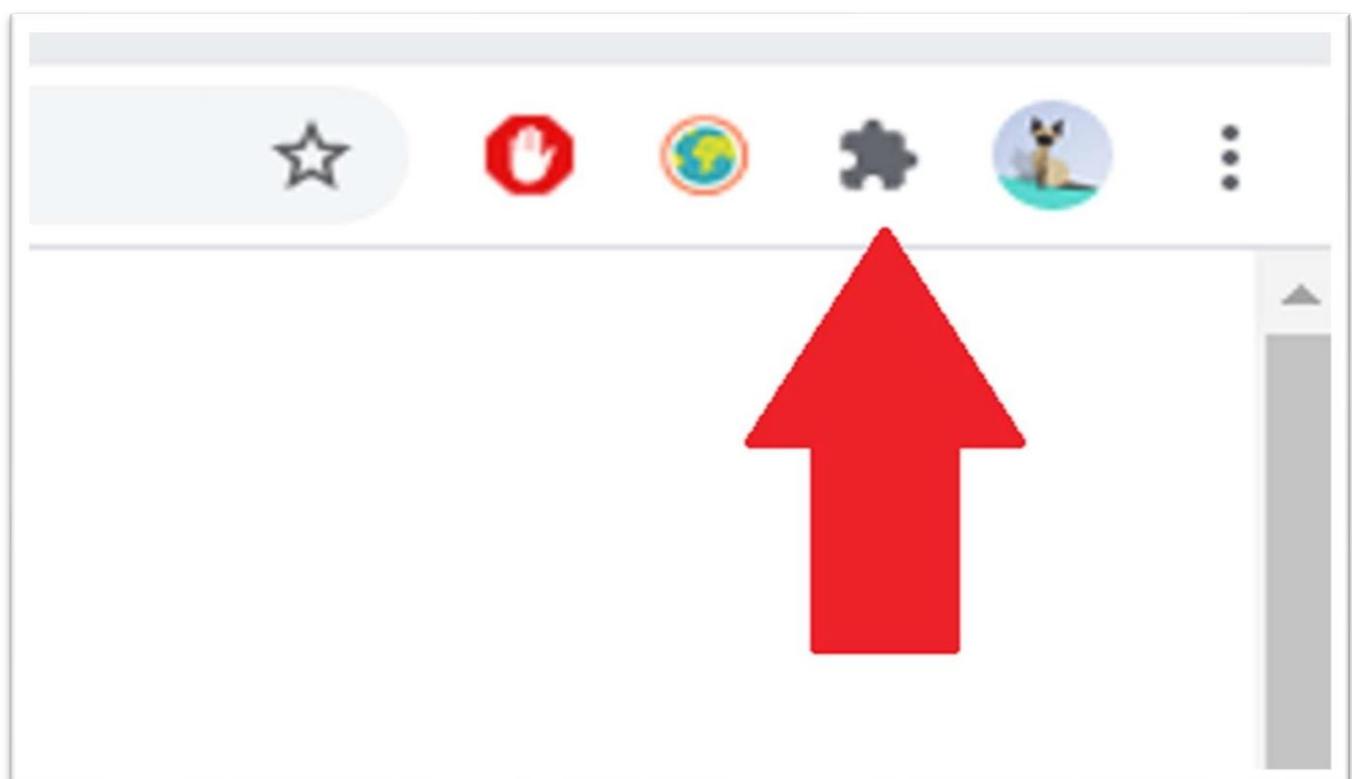


Fig-5.2 Chrome Extensions icon

### 5.1.3. Chrome Extension store

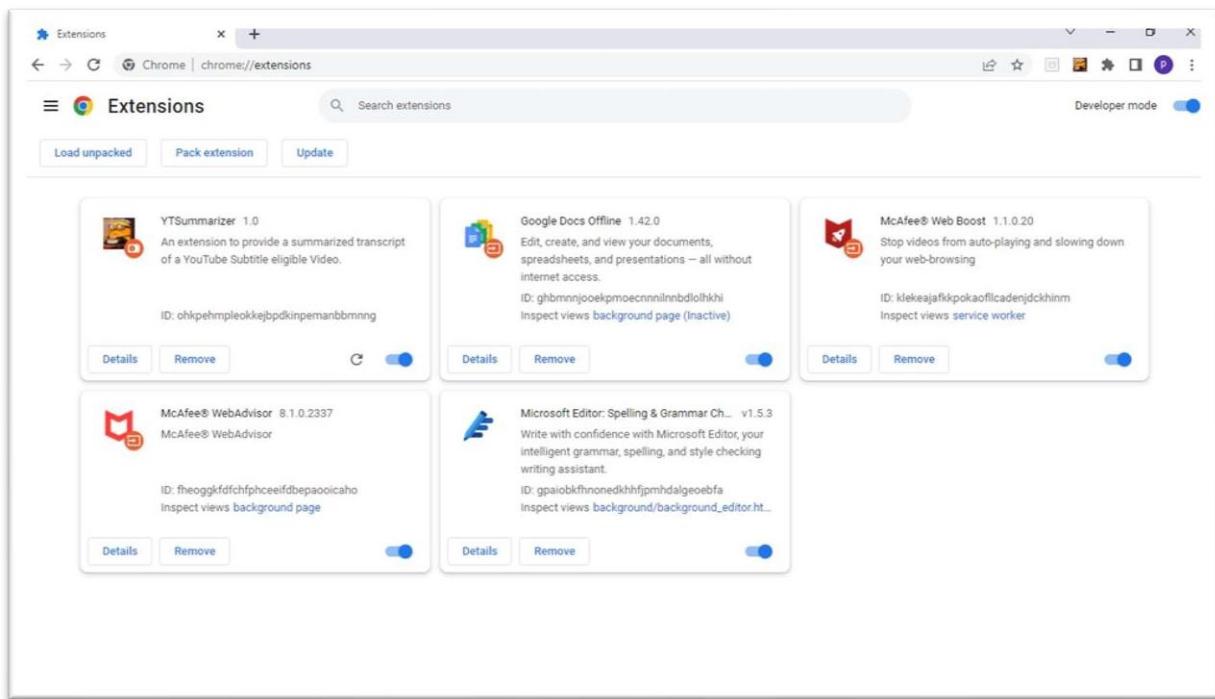


Fig -5.3 Chrome Extension store

#### 5.1.4. Searching the chrome extension YT Summarizer

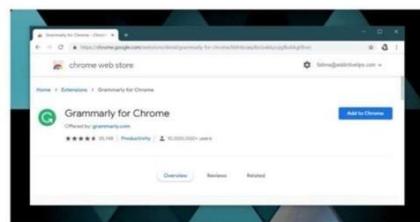


Fig -5.4 Searching the chrome extension YT Summarizer

### 5.1.5. YTSummarizer chrome extension add to chrome blue button

#### Install extensions

Chrome extensions are mostly available in the Chrome Web Store. You can search the store, click an extension, and click the Add to Chrome button.



### 5.1.6. Chrome Extension for YouTube Transcript Summarizer YT Summarizer

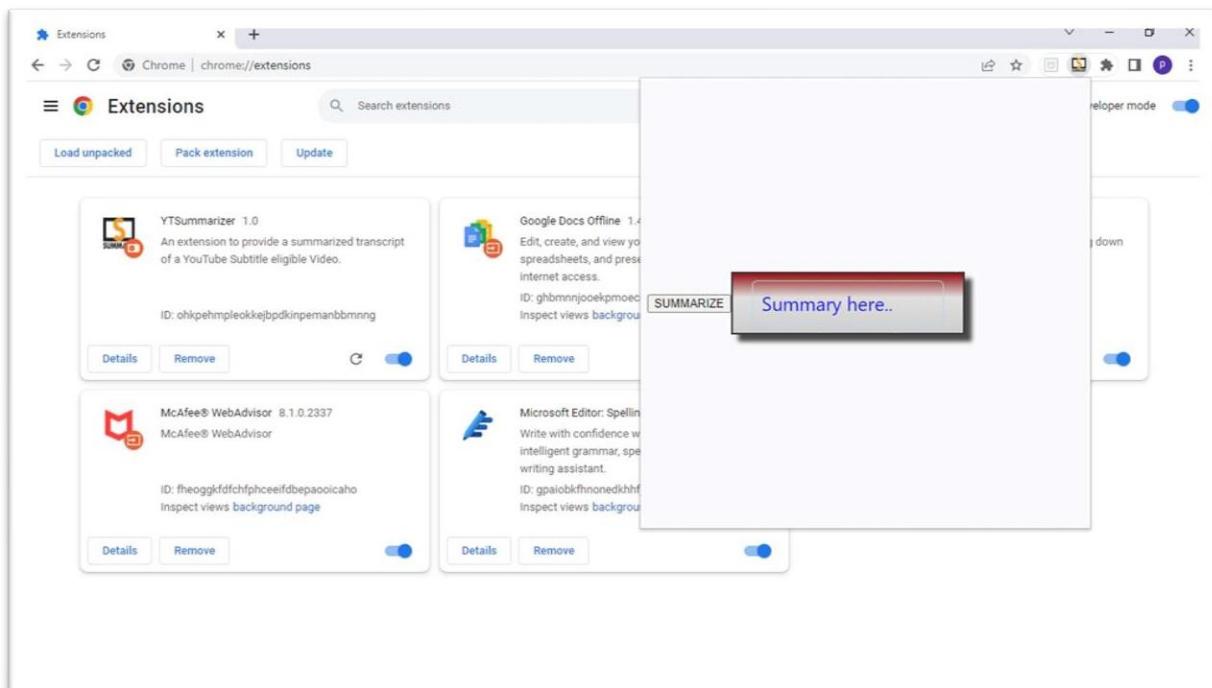


Fig -5.5 Summarizer YT Summarizer

#### 5.2.14. Results

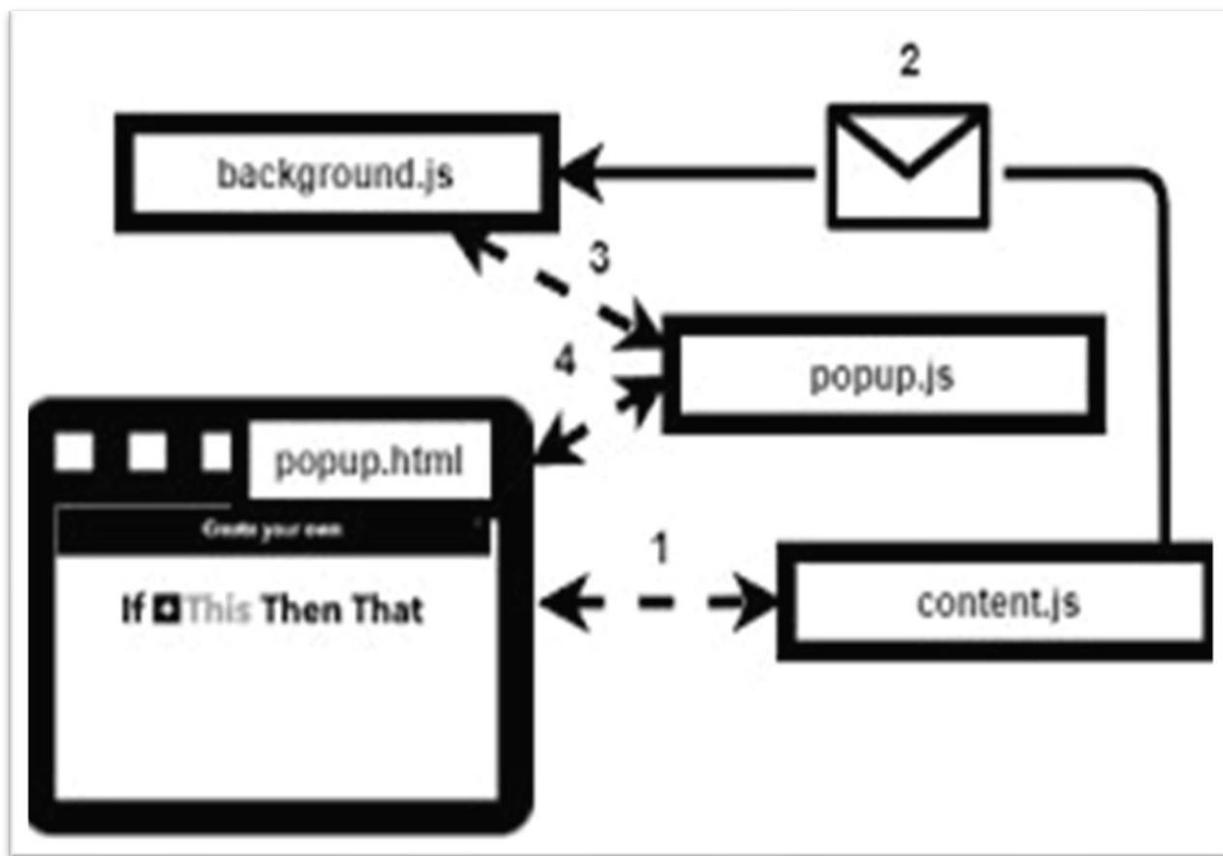


Fig -5.6 Result

# **6. Testing**

## Testing Methodology

World now is not about books for reference or help to find anything relatable or something we are searching for or research about, it has changed and for everything people want to research on they search on browsers, to help and make their searches easy this YT Summarizer extension is created to generate summary on text or audio format of a YouTube video to save time and get the required satisfactory content. A proper backend leads to proper development and testing is the most important part of that.

Rising customer expectations for fault-free, requirements-exact system have increased awareness of the importance of software testing as a critical activity.

We begin the testing process by developing a comprehensive plan to test the general functionality and special features on a variety of platform combinations. Strict quality control procedures are used. The process very files that the application meets the requirements specified in the system requirements document and is bug free. At the end of each testing day, we prepare a summary of completed and failed tests. Applications are not allowed to launch until all identified problems are fixed. A report is prepared at the end of testing to show exactly what was tested and to list the final outcomes.

Our software testing methodology is applied in three distinct phases: unit testing, system testing, and acceptance were testing.

**Unit Testing:** The programmers conduct unit testing during the development phase. Programmers can test their specific functionality individually or with other units. However, unit testing is designed to test small pieces of functionality rather than the system as a whole. This allows the programmers to conduct the first round of testing to eliminate bugs before they reach the testing staff. In unit testing the analyst tests the programs making up a system.

For this reason, unit testing is sometimes called program testing. Unit testing gives stress on the modules independently of one another, to find errors. This helps the tester in detecting errors in coding and logic that are contained within that module alone.

The errors resulting from the interaction between modules are initially avoided.

For example, a hotel information system consists of modules to handle reservations; guest checking and checkout; restaurant, room service and miscellaneous charges; convention

activities; and accounts receivable billing. For each, it provides the ability to enter, modify or retrieve data and respond to different types of inquiries or print reports. The test cases needed for unit testing should exercise each condition and option.

Unit testing can be performed from the bottom up, starting with smallest and lowest-level modules and proceeding one at a time. For each module in bottom-up testing a short program is used to execute the module and provides the needed data, so that the module is asked to perform the way it will when embedded within the larger system.

**System Testing:** The objective of system testing is to ensure that all individual programs are working as expected, that the programs link together to meet the requirements specified and to ensure that the computer system and the associated clerical and other procedures work together.

The initial phase of system testing is the responsibility of the analyst who determines what conditions are to be tested, generates test data, produced a schedule of expected results, runs the tests and compares the computer produced results with the expected results with the expected results.

The analyst may also be involved in procedures testing. When the analyst is satisfied that the system is working properly, he hands it over to the users for testing. The importance of system testing by the user must be stressed. Ultimately it is the user must verify the system and give the go-ahead.

During testing, the system is used experimentally to ensure that the software does not fail, i.e., that it will run according to its specifications and in the way, users expect it to. Special test data is input for processing (test plan) and the results are examined to locate unexpected results.

A limited number of users may also be allowed to use the system so analysts can see whether they try to use it in unexpected ways. It is preferably to find these

surprises before the organization implements the system and depends on it. In many organizations, testing is performed by persons other than those who write the original programs. Using persons who do not know how certain parts were designed or programmed ensures more complete and unbiased testing and more reliable software.

The system is tested as a complete, integrated system. System testing first occurs in the development environment but eventually is conducted in the production environment. Functionality and performance testing are designed to catch bugs in the system, unexpected results, or other ways in which the system does not meet the stated requirements.

The testers create detailed scenarios to test the strength and limits of the system, trying to break it if possible. Editorial reviews not only correct typographical and grammatical errors, but also improve the system's overall usability by ensuring that on-screen language is clear and helpful to users. Accessibility reviews ensure that the system is accessible to users with disabilities.

### **System testing consists of the following five steps:**

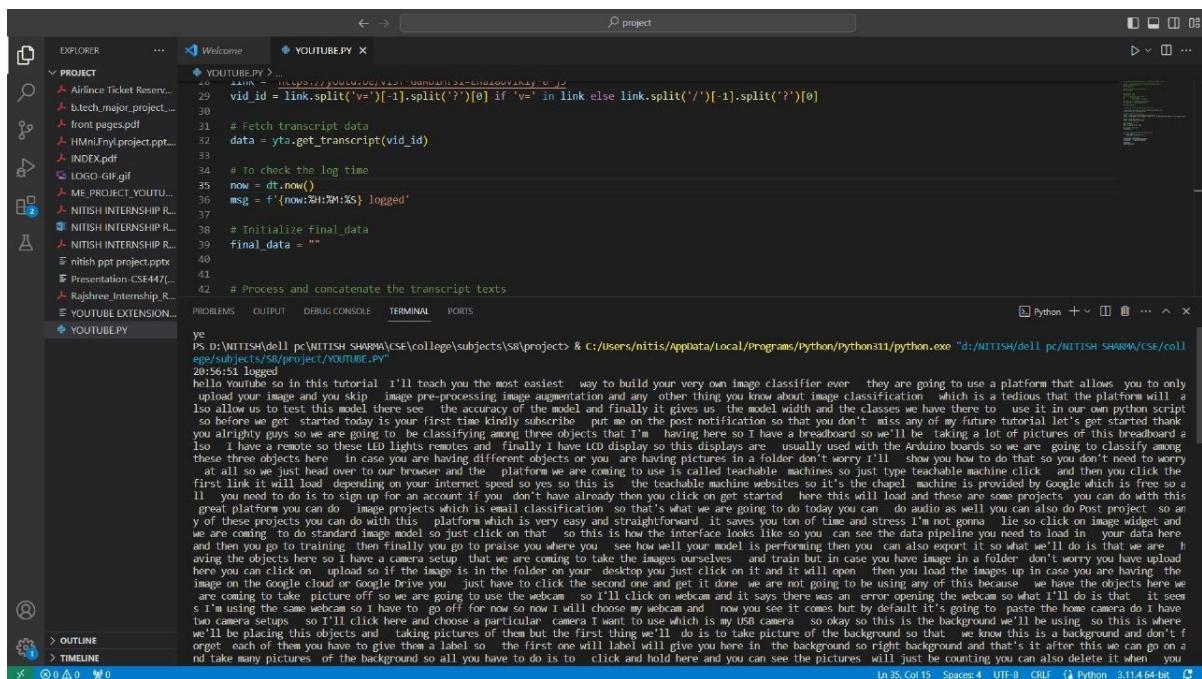
- i. Program testing
- ii. String testing
- iii. System testing
- iv. System documentation
- v. User acceptance testing

## Program Testing

A program represents the logical elements of a system. For a program to run satisfactorily, it must compile and test data correctly and tie in properly with other programs. It is the responsibility of a programmer to have an error free program. At the time of testing the system, there exists two types of errors that should be checked. These errors are syntax and logic.

A syntax error is a program statement that violates one or more rules of the language in which it is written. An improperly defined field dimension or omitted key words are common syntax errors. These errors are shown through error messages generated by the computer. A logic error, on the other hand, deals with incorrect data fields out of range items, and invalid combinations.

Since the logical errors are not detected by compiler, the programmer must examine the output carefully to detect them. When a program is tested, the actual output is compared with the expected output. When there is a discrepancy, the sequence of the instructions, must be traced to determine the problem. The process is facilitated by breaking the program down into self-contained portions, each of which can be checked at certain key points.



The screenshot shows a code editor interface with the following details:

- Explorer:** Shows a project structure with files like "Airlines Ticket Reserv...", "b.tech\_major.project...", "front pages.pdf", "HMnIProject.ppt...", "INDEX.pdf", "LOGO-GIF.gif", "ME\_PROJECT\_YOUTU...", "NITISH INTERNSHIP R...", "NITISH INTERNSHIP R...", "nitish ppt project.pptx", "Presentation-CS447(...", "Rajdhree Internship\_R...", "YOUTUBE EXTENSION...", and "YOUTUBE.PY".
- Terminal:** Displays the command: `ps D:\NITISH\dell pc\NITISH SHARMA\CSE\college\subjects\S6\project> & C:/Users/nitish/AppData/Local/Programs/Python/Python311/python.exe "d:/NITISH/dell pc\NITISH SHARMA\CSE\college\subjects\S6\project\YOUTUBE.PY"`.
- Code Editor:** Shows Python code for a YouTube project. The code includes imports for requests, BeautifulSoup, and re. It defines functions for fetching transcript data and processing transcript texts. It also includes a section for logging current time and a main loop for processing and concatenating transcript texts.
- Output:** Shows the terminal output of the executed Python script, which includes a welcome message from the video and some log information.

FIG-6.1 ENGLISH VIDEO SUMMERIZATION

## SONG SUMMERIZATION

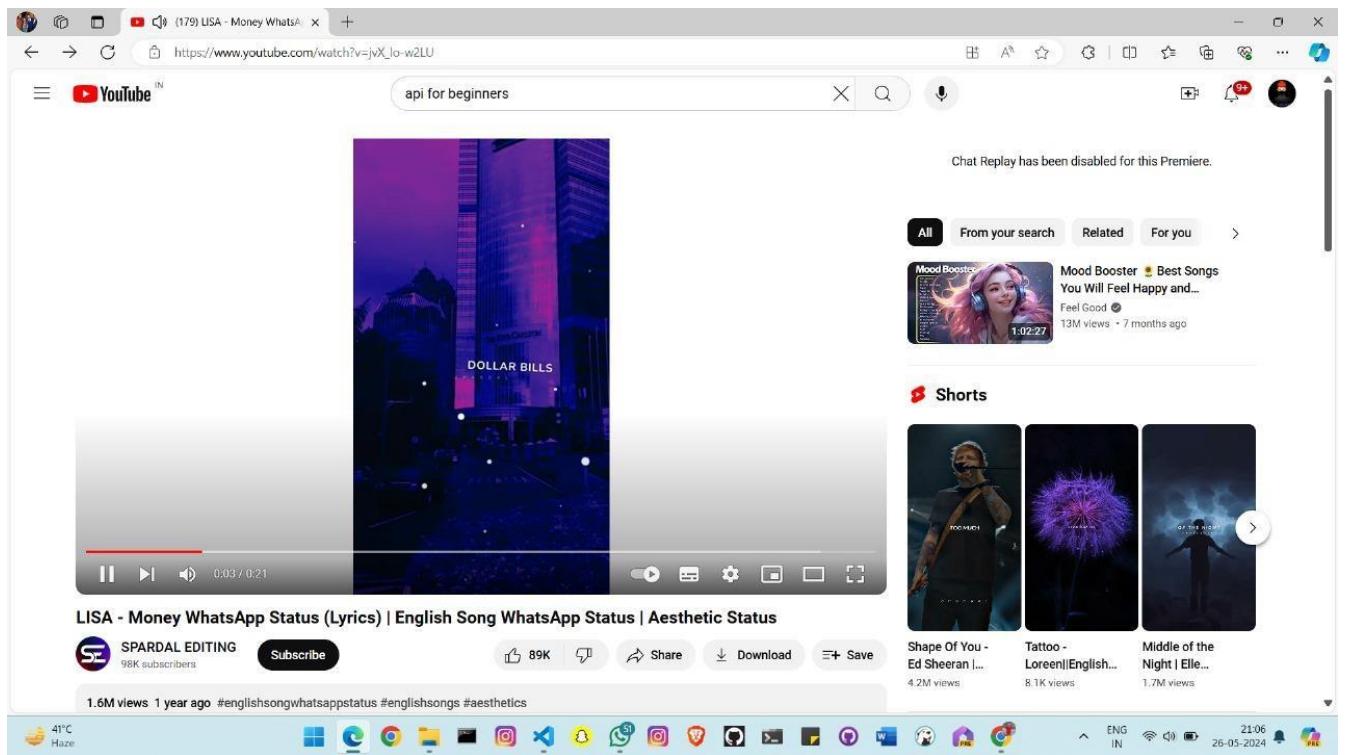
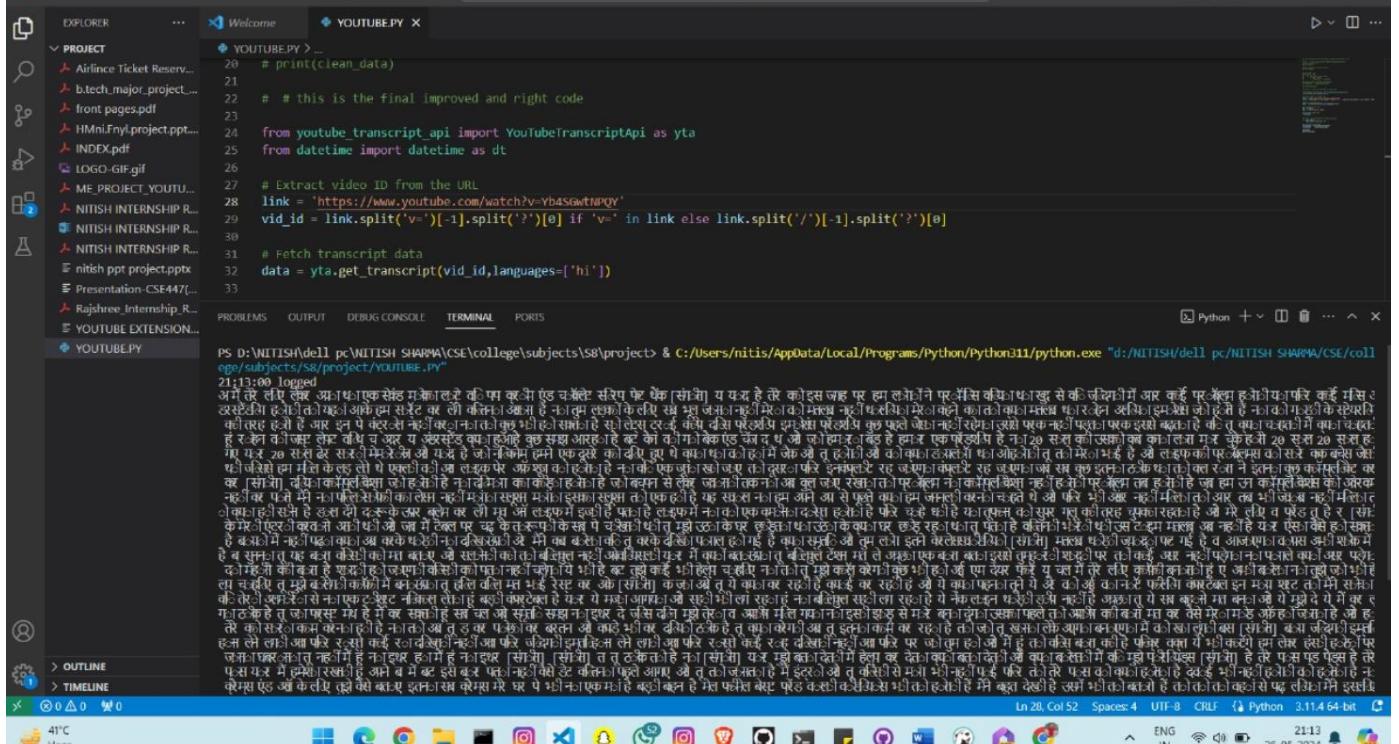


FIG-6.2 SONG SUMMARIZATION

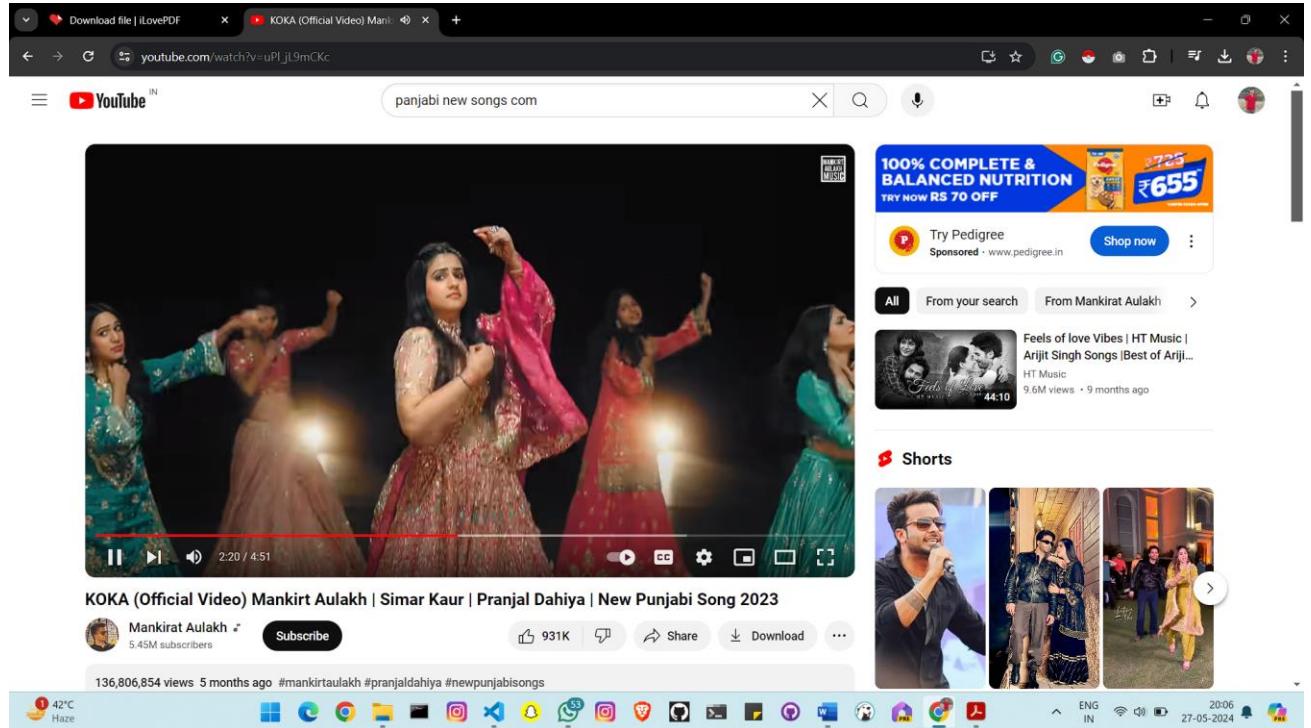
## HINDI VIDEO SUMMERIZATION



```

PROJECT
YOUTUBE.PY > ...
20 # print(clean_data)
21
22 # # this is the final improved and right code
23
24 from youtube_transcript_api import YouTubeTranscriptApi as yta
25 from datetime import datetime as dt
26
27 # Extract video ID from the URL
28 link = 'https://www.youtube.com/watch?v=yb4sawtnpQY'
29 vid_id = link.split('v')[1].split('?')[0] if 'v=' in link else link.split('/')[1].split('?')[0]
30
31 # Fetch transcript data
32 data = yta.get_transcript(vid_id, languages=['hi'])
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
279
280
281
282
283
284
285
286
287
288
289
289
290
291
292
293
294
295
296
297
298
299
299
300
301
302
303
304
305
306
307
308
309
309
310
311
312
313
314
315
316
317
318
319
319
320
321
322
323
324
325
326
327
328
329
329
330
331
332
333
334
335
336
337
338
339
339
340
341
342
343
344
345
346
347
348
349
349
350
351
352
353
354
355
356
357
358
359
359
360
361
362
363
364
365
366
367
368
369
369
370
371
372
373
374
375
376
377
378
379
379
380
381
382
383
384
385
386
387
388
389
389
390
391
392
393
394
395
396
397
398
399
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
419
420
421
422
423
424
425
426
427
428
429
429
430
431
432
433
434
435
436
437
438
439
439
440
441
442
443
444
445
446
447
448
449
449
450
451
452
453
454
455
456
457
458
459
459
460
461
462
463
464
465
466
467
468
469
469
470
471
472
473
474
475
476
477
478
479
479
480
481
482
483
484
485
486
487
488
489
489
490
491
492
493
494
495
496
497
498
499
499
500
501
502
503
504
505
506
507
508
509
509
510
511
512
513
514
515
516
517
518
519
519
520
521
522
523
524
525
526
527
528
529
529
530
531
532
533
534
535
536
537
538
539
539
540
541
542
543
544
545
546
547
548
549
549
550
551
552
553
554
555
556
557
558
559
559
560
561
562
563
564
565
566
567
568
569
569
570
571
572
573
574
575
576
577
578
579
579
580
581
582
583
584
585
586
587
588
589
589
590
591
592
593
594
595
596
597
598
599
599
600
601
602
603
604
605
606
607
608
609
609
610
611
612
613
614
615
616
617
618
619
619
620
621
622
623
624
625
626
627
628
629
629
630
631
632
633
634
635
636
637
638
639
639
640
641
642
643
644
645
646
647
648
649
649
650
651
652
653
654
655
656
657
658
659
659
660
661
662
663
664
665
666
667
668
669
669
670
671
672
673
674
675
676
677
678
679
679
680
681
682
683
684
685
686
687
688
689
689
690
691
692
693
694
695
696
697
698
699
699
700
701
702
703
704
705
706
707
708
709
709
710
711
712
713
714
715
716
717
718
719
719
720
721
722
723
724
725
726
727
728
729
729
730
731
732
733
734
735
736
737
738
739
739
740
741
742
743
744
745
746
747
748
749
749
750
751
752
753
754
755
756
757
758
759
759
760
761
762
763
764
765
766
767
768
769
769
770
771
772
773
774
775
776
777
778
779
779
780
781
782
783
784
785
786
787
788
789
789
790
791
792
793
794
795
796
797
797
798
799
799
800
801
802
803
804
805
806
807
808
809
809
810
811
812
813
814
815
816
817
817
818
819
819
820
821
822
823
824
825
826
827
828
829
829
830
831
832
833
834
835
836
837
838
839
839
840
841
842
843
844
845
846
847
848
849
849
850
851
852
853
854
855
856
857
858
859
859
860
861
862
863
864
865
866
867
868
869
869
870
871
872
873
874
875
876
877
878
879
879
880
881
882
883
884
885
886
887
888
889
889
890
891
892
893
894
895
895
896
897
897
898
899
899
900
901
902
903
904
905
906
907
908
909
909
910
911
912
913
914
915
915
916
917
918
919
919
920
921
922
923
924
925
926
927
928
929
929
930
931
932
933
934
935
936
937
938
939
939
940
941
942
943
944
945
946
947
948
949
949
950
951
952
953
954
955
956
957
958
959
959
960
961
962
963
964
965
966
967
968
969
969
970
971
972
973
974
975
976
977
978
979
979
980
981
982
983
984
985
986
987
987
988
989
989
990
991
992
993
994
995
996
997
997
998
999
999
1000
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1039
1040
1041
1042
1043
1044
1045
1046
1047
1048
1049
1049
1050
1051
1052
1053
1054
1055
1056
1057
1058
1059
1059
1060
1061
1062
1063
1064
1065
1066
1067
1068
1069
1069
1070
1071
1072
1073
1074
1075
1076
1077
1078
1079
1079
1080
1081
1082
1083
1084
1085
1086
1087
1088
1089
1089
1090
1091
1092
1093
1094
1095
1096
1096
1097
1098
1099
1099
1100
1101
1102
1103
1104
1105
1106
1107
1108
1109
1109
1110
1111
1112
1113
1114
1115
1116
1117
1118
1119
1119
1120
1121
1122
1123
1124
1125
1126
1127
1128
1129
1129
1130
1131
1132
1133
1134
1135
1136
1137
1138
1139
1139
1140
1141
1142
1143
1144
1145
1146
1147
1148
1149
1149
1150
1151
1152
1153
1154
1155
1156
1157
1158
1159
1159
1160
1161
1162
1163
1164
1165
1166
1167
1168
1169
1169
1170
1171
1172
1173
1174
1175
1176
1177
1178
1179
1179
1180
1181
1182
1183
1184
1185
1186
1187
1188
1189
1189
1190
1191
1192
1193
1194
1195
1195
1196
1197
1198
1199
1199
1200
1201
1202
1203
1204
1205
1206
1207
1208
1209
1209
1210
1211
1212
1213
1214
1215
1216
1217
1218
1219
1219
1220
1221
1222
1223
1224
1225
1226
1227
1228
1229
1229
1230
1231
1232
1233
1234
1235
1236
1237
1238
1239
1239
1240
1241
1242
1243
1244
1245
1246
1247
1248
1249
1249
1250
1251
1252
1253
1254
1255
1256
1257
1258
1259
1259
1260
1261
1262
1263
1264
1265
1266
1267
1268
1269
1269
1270
1271
1272
1273
1274
1275
1276
1277
1278
1279
1279
1280
1281
1282
1283
1284
1285
1286
1287
1288
1289
1289
1290
1291
1292
1293
1294
1295
1296
1297
1297
1298
1299
1299
1300
1301
1302
1303
1304
1305
1306
1307
1308
1309
1309
1310
1311
1312
1313
1314
1315
1316
1317
1318
1319
1319
1320
1321
1322
1323
1324
1325
1326
1327
1328
1329
1329
1330
1331
1332
1333
1334
1335
1336
1337
1338
1339
1339
1340
1341
1342
1343
1344
1345
1346
1347
1348
1349
1349
1350
1351
1352
1353
1354
1355
1356
1357
1358
1359
1359
1360
1361
1362
1363
1364
1365
1366
1367
1368
1369
1369
1370
1371
1372
1373
1374
1375
1376
1377
1378
1379
1379
1380
1381
1382
1383
1384
1385
1386
1387
1388
1389
1389
1390
1391
1392
1393
1394
1395
1395
1396
1397
1398
1399
1399
1400
1401
1402
1403
1404
1405
1406
1407
1408
1409
1409
1410
1411
1412
1413
1414
1415
1416
1417
1418
1419
1419
1420
1421
1422
1423
1424
1425
1426
1427
1428
1429
1429
1430
1431
1432
1433
1434
1435
1436
1437
1438
1439
1439
1440
1441
1442
1443
1444
1445
1446
1447
1448
1449
1449
1450
1451
1452
1453
1454
1455
1456
1457
1458
1459
1459
1460
1461
1462
1463
1464
1465
1466
1467
1468
1469
1469
1470
1471
1472
1473
1474
1475
1476
1477
1478
1479
1479
1480
1481
1482
1483
1484
1485
1486
1487
1488
1489
1489
1490
1491
1492
1493
1494
1495
1495
1496
1497
1498
1499
1499
1500
1501
1502
1503
1504
1505
1506
1507
1508
1509
1509
1510
1511
1512
1513
1514
1515
1516
1517
1518
1519
1519
1520
1521
1522
1523
1524
1525
1526
1527
1528
1529
1529
1530
1531
1532
1533
1534
1535
1536
1537
1538
1539
1539
1540
1541
1542
1543
1544
1545
1546
1547
1548
1549
1549
1550
1551
1552
1553
1554
1555
1556
1557
1558
1559
1559
1560
1561
1562
1563
1564
1565
1566
1567
1568
1569
1569
1570
1571
1572
1573
1574
1575
1576
1577
1578
1579
1579
1580
1581
1582
1583
1584
1585
1586
1587
1588
1589
1589
1590
1591
1592
1593
1594
1595
1595
1596
1597
1598
1599
1599
1600
1601
1602
1603
1604
1605
1606
1607
1608
1609
1609
1610
1611
1612
1613
1614
1615
1616
1617
1618
1619
1619
1620
1621
1622
1623
1624
1625
1626
1627
1628
1629
1629
1630
1631
1632
1633
1634
1635
1636
1637
1638
1639
1639
1640
1641
1642
1643
1644
1645
1646
1647
1648
1649
1649
1650
1651
1652
1653
1654
1655
1656
1657
1658
1659
1659
1660
1661
1662
1663
1664
1665
1666
1667
1668
1669
1669
1670
1671
1672
1673
1674
1675
1676
1677
1678
1679
1679
1680
1681
1682
1683
1684
1685
1686
1687
1688
1689
1689
1690
1691
1692
1693
1694
1695
1695
1696
1697
1698
1699
1699
1700
1701
1702
1703
1704
1705
1706
1707
1708
1709
1709
1710
1711
1712
1713
1714
1715
1716
1717
1718
1719
1719
1720
1721
1722
1723
1724
1725
1726
1727
1728
1729
1729
1730
1731
1732
1733
1734
1735
1736
1737
1738
1739
1739
1740
1741
1742
1743
1744
1745
1746
1747
1748
1749
1749
1750
1751
1752
1753
1754
1755
1756
1757
1758
1759
1759
1760
1761
1762
1763
1764
1765
1766
1767
1768
1769
1769
1770
1771
1772
1773
1774
1775
1776
1777
1778
1779
1779
1780
1781
1782
1783
1784
1785
1786
1787
1788
1789
1789
1790
1791
1792
1793
1794
1795
1795
1796
1797
1798
1799
1799
1800
1801
1802
1803
1804
1805
1806
1807
1808
1809
1809
1810
1811
1812
1813
1814
1815
1816
1817
1818
1819
1819
1820
1821
1822
1823
1824
1825
1826
1827
1828
1829
1829
1830
1831
1832
1833
1834
1835
1836
1837
1838
1839
1839
1840
1841
1842
1843
1844
1845
1846
1847
1848
1849
1849
1850
1851
1852
1853
1854
1855
1856
1857
1858
1859
1859
1860
1861
1862
1863
1864
1865
1866
1867
1868
1869
1869
1870
1871
1872
1873
1874
1875
1876
1877
1878
1879
1879
1880
1881
1882
1883
1884
1885
1886
1887
1888
1889
1889
1890
1891
1892
1893
1894
1895
1895
1896
1897
1898
1899
1899
1900
1901
1902
1903
1904
1905
1906
1907
1908
1909
1909
1910
1911
1912
1913
1914
1915
1916
1917
1918
1919
1919
1920
1921
1922
1923
1924
1925
1926
1927
1928
1929
1929
1930
1931
1932
1933
1934
1935
1936
1937
1938
1939
1939
1940
1941
1942
1943
1944
1945
1946
1947
1948
1949
1949
1950
1951
1952
1953
1954
1955
1956
1957
1958
1959
1959
1960
1961
1962
1963
1964
1965
1966
1967
1968
1969
1969
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1989
1990
1991
1992
1993
1994
1995
1995
1996
1997
1998
1999
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2019
2020
2021
2022
2023
2024
2025
2026
2027
2028
2029
2029
2030
2031
2032
2033
2034
2035
2036
2037
2038
2039
2039
2040
2041
2042
2043
2044
2045
2046
2047
2048
2049
2049
2050
2051
2052
2053
2054
2055
2056
2057
2058
2059
2059
2060
2061
2062
2063
2064
2065
2066
2067
2068
2069
2069
2070
2071
2072
2073
2074
2075
2076
2077
2078
2079
2079
2080
2081
2082
2083
2084
2085
2086
2087
2088
2089
2089
2090
2091
2092
2093
2094
2095
2095
2096
2097
2098
2099
2099
2100
2101
2102
2103
2104
2105
2106
2107
2108
2109
2109
2110
2111
2112
2113
2114
2115
2116
2117
2118
2119
2119
2120
2121
2122
2123
2124
2125
2126
2127
2128
2129
2129
2130
2131
2132
2133
2134
2135
2136
2137
2138
2139
2139
2140
2141
2142
2143
2144
2145
2146
2147
2148
2149
2149
2150
2151
2152
2153
2154
2155
2156
2157
2158
2159
2159
2160
2161
2162
2163
2164
2165
2166
2167
2168
2169
2169
2170
2171
2172
2173
2174
2175
2176
2177
2178
2179
2179
2180
2181
2182
2183
2184
2185
2186
2187
2188
2189
2189
2190
2191
2192
2193
2194
2195
2195
2196
2197
2198
2199
2199
2200
2201
2202
2203
2204
2205
2206
2207
2208
2209
2209
2210
2211
2212
2213
2214
2215
2216
2217
2218
2219
2219
2220
2221
2222
2223
22
```

# PANJABI SONG SUMMERIZATION



The screenshot shows a Windows desktop environment. In the center is a code editor window titled 'YOUTUBE.PY'. The code is as follows:

```
from youtube_transcript_api import YouTubeTranscriptApi as yta
from datetime import datetime as dt

# Extract video ID from the URL
link = 'https://www.youtube.com/watch?v=uP1_jt9nCKC'
vid_id = link.split('=')[1].split('?')[0] if 'v=' in link else link.split('/')[1].split('?')[0]

# Fetch transcript data
# data = yta.get_transcript(vid_id)
data = yta.get_transcript(vid_id, languages=['hi', 'ja', 'pa', 'en'])

# To check the log time
now = dt.now()
```

The code editor has tabs for 'EXPLORER', 'PROJECT', 'PROBLEMS', 'OUTPUT', 'DEBUG CONSOLE', 'TERMINAL', and 'PORTS'. The 'TERMINAL' tab is selected. Below the code editor, there are two command-line windows labeled 'Python' and 'Python'. The status bar at the bottom shows 'Ln 28, Col 52' and 'Spaces: 4'. The taskbar at the bottom has icons for File, Edit, Selection, View, Go, Run, Terminal, Help, and several pinned applications.

## **String Testing**

Programs are invariably related to one another and interact in a total system. Each program is tested to see whether it conforms to related programs in the system. Each part of the system is tested against the entire module with both test and live data before the whole system is ready to be tested.

## **System Testing**

System testing is designed to uncover weaknesses that were not found in earlier tests. This includes forced system failure and validation of total system as it will be implemented by its user in the operational environment. Under this testing, generally we take low volumes of transactions based on live data. This volume is increased until the maximum level for each transaction type is reached.

The total system is also tested for recovery and fallback after various major failures to ensure that no data are lost during the emergency.

All this is done with the old system still in operation. When we see that the proposed system is successful in the test, the old system is discontinued.

## **System Documentation**

All design and test documentation should be well prepared and kept in the library for future reference. The library is the central location for maintenance of the new system.

## **User Acceptance Testing**

An acceptance test has the objective of selling the user on the validity and reliability of the system. It verifies that the system's procedures operate to system specifications and that the integrity of important data is maintained. Performance of an acceptance test is actually the user's show. User motivation is very important for the successful performance of the system. After that a comprehensive test report is prepared. This report shows the system's tolerance, performance range, error rate and accuracy.

Table 6.1 Test Report with test data

TEST REPORT WITH TEST DATA		
Project Name: Multi Language YouTube Transcript Summarizer		
S No.	Testing Parameter	Observations
A.	<b>INTERFACE TESTING</b> <ul style="list-style-type: none"> <li>1) User-friendliness</li> <li>2) Consistent menus</li> </ul>	OK NA
B.	<b>CONTROL FLOW TESTING</b> <ul style="list-style-type: none"> <li>1) IF-THEN-ELSE</li> <li>2) DO WHILE</li> <li>3) CASE-SWITCH</li> </ul>	OK OK OK
C.	<b>VALIDATION TESTING</b> <ul style="list-style-type: none"> <li>1) Check for improper or inconsistent typing</li> <li>2) Check for erroneous initialization or default values</li> <li>3) Check for incorrect variable names</li> <li>4) Check for inconsistent Data Types</li> <li>5) Check for relational/arithmetic operators</li> </ul>	OK OK OK OK OK

D.	<b>DATA INTEGRITY/SECURITY TESTING</b> <ul style="list-style-type: none"> <li>1) Data Insertion/ Deletion/ Updating</li> <li>2) Boundary condition (Underflow, Overflow Exception)</li> <li>3) Check for unauthorized access of data</li> <li>4) Check for data availability</li> </ul>	OK OK OK OK OK
E.	<b>EFFICIENCY TESTING</b> <ul style="list-style-type: none"> <li>1) Throughput of the system</li> <li>2) Response time of the system</li> <li>3) Online disk storage required by the system</li> <li>4) Primary memory required by the system</li> </ul>	OK OK OK OK
F.	<b>ERROR HANDLING ROUTINES</b> <ul style="list-style-type: none"> <li>1) Error description are intelligent/understandable</li> <li>2) Error recovery is smooth</li> <li>3) All error handling routines are tested and executed at least once</li> </ul>	OK OK OK

# **7.CONCLUSION AND FUTURE ENHANCEMENTS**

## **7.1. Limitations**

The new system has been designed to meet almost all of the user requirements but this too has certain limitations some of which can be enhanced in the future enhancements or updates

### **7.1.1. Supports only 1024 words for summarization**

The existing system modules and models are only able to generate summary of up to 1024 words and less sometimes because of less accuracy. The new system will be able to overcome this in future with the updated models being used in backend.

### **7.1.2 Accuracy of summary**

The existing system does not provide high rate of accuracy and its hard to be only dependent on this model for future works because this might lead to make us miss some good video content if summary is not all accurate enough.

### **7.1.3 Audio accuracy and quality**

The system currently does not provide audio summary of larger videos because it's hard for low-level modules to generate huge amount of summary and process it, it does not have the capability to process large transcription.

### **7.1.4 Subtitle eligibility in videos**

The existing system has the eligibility criteria for a YouTube video to generate summarization in whatever required format because it cannot transcribe the improper language audio to transcribe summary of a video, this leads to one single criteria to generate only subtitle in-built videos to generate summarization.

## **7.2 Future Enhancements**

The new system would contribute to the overall objectives to of the Extension. It would provide a quick, error free and zero cost solution to the current process. It would provide a solution to many issues in the current system. As the new system is flexible and scalable it can also be upgraded and extended to meet other complex requirements which may be raised in the future.

### **7.2.1 Features of the New System.**

The new system has been designed as per the user requirements so as to fulfill almost all of them below:

#### **7.2.2 Speedy Processing**

Summary can be generated very quickly as compared to the existing extension as it allows the use of previously generated summaries to see again with data stored on cloud. It saves time required to get the summarization of audio and text faster.

#### **7.2.3 Accuracy**

One of the most important draw backs of the current system is that audio is not at its best accuracy and can't be generated on longer size videos because of word limit. The new system will generate the result as soon as the summarization is processed by user and will also store it in the database for future usage.

#### **7.2.4 High-quality Audio**

The new system makes it easy to store and retrieve information as required and does not involve storing information by the user-self it's on Auto mode from cloud. It thus saves data management problems faced in the current system as it has a Database Management System of only one-time access.

#### **7.2.5 Zero Cost and No Advertisements**

Unique service provider this extension as it does not show any advertisements and provide no-cost service.

### 7.3 Conclusion

The development of software includes so many people like user system developer, user of system and the management, It is important to identify the system requirements by properly collecting required data to interact with supplier and customer of the system.

Proper design builds upon this foundation to give a blue print, which is actually implemented by the developers.

On realizing the importance of systematic documentation all the processes are implemented using a software engineering approach. Working in a live environment enables one to appreciate the intricacies involved in the System Development Life Cycle (SDLC).

I have gained a lot of practical knowledge from this project, which i think, shall make me stand in a good state in the future.

## **8. REFERENCES**

## REFERENCES

### List of useful References:

- <https://atmamani.github.io/blog/building-restful-apis-with-flask-in-python/>
- <https://huggingface.co/transformers/installation.html>
- <https://pypi.org/project/youtube-transcript-api/>
- <https://www.thepythoncode.com/article/text-summarization-using-huggingface-transformers-python>
- <https://medium.com/swlh/parsing-rest-api-payload-and-query-parameters-with-flask- better-than-marshmallow-aa79c889e3ca>
- <https://medium.com/better-programming/the-ultimate-guide-to-building-a-chrome-extension-4c01834c63ec>
- [https://developer.chrome.com/docs/extensions/mv2/user\\_interface/](https://developer.chrome.com/docs/extensions/mv2/user_interface/)
- <https://www.thepythoncode.com/article/translate-text-in-python>