GROUP NO-13

AJAY KUMAR GARG ENGG. COLLEGE, GHAZIABAD

Software Requirement Specification (SRS)

Video-Text Analyzer

Submitted for fulfillment of award of

Bachelor of Technology

In

Computer Science and Information Technology

By

Ishita Vashistha	2000270119004
Shivam Singh	1900270130159
Tushar Raghav	1900270110057
Umendra Mani Dwivedi	1900270110059

Under the Guidance of

Dr Anupama Sharma

AJAY KUMAR GARG ENGINEERING COLLEGE, GHAZIABAD

YEAR: 2022-2023

Table of Contents

List of Figures

1. Introduction

In the current digital era, technological development is advancing quickly, which is generating a huge volume of video data every day. Nowadays, social media is one of the most used applications on a large scale in our daily lives. One of the media types that is used most frequently on social media is video.

Video data contains beneficial textual information such as scene text and caption text. The different types of videos like movies, news videos, and TV programs video etc. are created by various video frames based on its purpose. The majority of videos are distributed via social media, and this one can be repeated or lack vital information. However, the consumer is simply left with the choice of downloading the movie or watching the entire thing. This leads to add more and more of the costs to users because the video requires a large bandwidth to download video or view it and a large space to store it and most important it is very time consuming for the user because they to watch the full video for getting only some sort of required information. Thus, it is important to be a way to enable users to watch videos in a way that helps to reduce time and costs.

Video text analyzer technique is a proposed solution to resolve different points such as power consumption and additional cost. The processing of such huge chunky videos requires high storage, high computational processing power, and consumes a lot of time. Extraction of features from the video is a time-consuming task because the user has to watch the entire video. A large number of editing tools exist that require expertise that is highly expensive.

The video text analyzer is used to overcome these issues as it produces summaries by analyzing the underlying content of a source video stream, condensing this content into abbreviated descriptive forms that represent surrogates of the original content embedded within the video. The multimodal nature of video, which conveys a wide range of semantics in multiple modes, such as sound, music, still images, moving image, and text, makes this task much more complex than analyzing text documents.

1.1 Purpose

Video text analysis deals with the extraction of metadata from raw video to be used as components for further processing in applications such as search, summarization, classification or event detection.

The purpose of video text analysis is to provide extracted features and identification of structure that constitute building blocks for video retrieval, video similarity finding, summarization and navigation.

Video text analysis transforms the audio and image stream into a set of semantically meaningful representations. The ultimate goal is to extract structural and semantic content automatically, without any human intervention, at least for limited types of video domains. Algorithms to perform content analysis including the important points discussed in video, contents that contains solutions of the particular problem, and summarization of whole video.

1.2 Product Scope:

The following aspects help to describe the scope of our project Video Text Analyzer:

- Its main aim is to facilitate large-scale video, browsing by producing short, concise summaries that are diverse and representative of original videos.

- It will analyze the content of videos and will show the best possible results and methodology to the user.

- It will also be useful in generating and highlighting the required facts and contexts from the sequence of textual data.

By making the necessary adjustments to the backend process, we can also integrate the generation of models for a wide range of videos.

- News Video: Summarizing news videos automatically allows the user to quickly look out for the important patterns shown in the news.

- E Lecture Videos: Students usually spend their large amount of time by watching long E Lecture videos which can be saved by using the Videos Text Analyzer which will create a summary of a video that comes handy in a situation when we want to just glance at the content of the video quickly.

- Research Videos: It will helpful in analyzing the research videos as it will helps in representing the required results of the research in more compact and in very handy way.

1.3 Definition, Acronyms and abbreviations

This document uses the following conventions.

SRS	Software Requirement Specification
RAM	Random Access Memory
API	Application Programming Interfaces
LSTM	Long short-term memory
GIF	Graphics interchange Format

1.4 Overview

This project of Video Text Analyzer focuses on extracting metadata from video footages. These data can be used in applications like searching, categorization, summarization and event recognition. The process tends to transform audio and images into meaningful components. These components fulfill many purposes. The text data present in video contain certain useful information for automatic annotation, indexing, and structuring of content. However, variations of the text due to differences in text style, font, size, orientation, alignment as well as low image contrast and complex background make the problem of automatic text extraction extremely difficult and challenging job for which we will perform data extraction process followed by other processes that are explained as given below:

I. Data Extraction:

This part is most important step in our project because from this step we will extract the textual-data from the video on which we have to perform the operations. Following Python Libraries are used for the Text Information Extraction from the video:

1) Speech Recognition

2) MoviePy Library: This library can read the all the most common video.

II. Data Cleaning:

Cleaning of the obtained data is necessary to work with them for the further process, because the working with the data that are not cleaned will produce the inappropriate and inaccurate result for the user.

Following Process are involved in Data Cleaning:

- Removing of Punctuation Marks from the data.
- Identification of Stop words and removes them because these words always create a noise into the data that's why it is better to remove them.

III. Data Preprocessing: Real-world texts are incomplete and they cannot be sent directly to the model that will cause certain errors. So, we clean all our texts and convert them into a presentable form for prediction tasks

IV. Creation of Model:

- For creation of model, we will use the stacked Long Short-Term Memory (LSTM) layers in which we will create at most 3 layers stacked on the top of each other. This will make our model prediction much better.
- LSTM Layer captures all the contextual information present into the input sequence and also return the hidden states output and cell state after execution of every LSTM layer.

System Architecture for Video Text Analyzer is shown in Figure1.

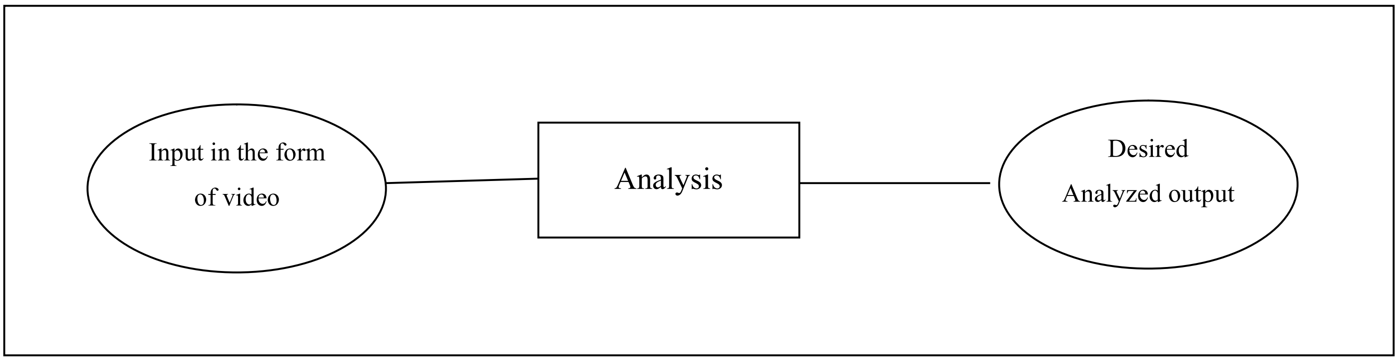


Figure 1: Proposed System

2. Overall Description

2.1 Product Perspective

Video Text Analyzer is a platform concerned with the extraction of metadata from raw video for use as components in applications such as multilingual video information access, semantic video summarization and indexing, video security and surveillance, etc. Audio and visual elements are often converted into meaningful parts by the procedure. These parts serve a variety of functions and we can also use the same model to produce the analyzed result on wide variety of videos.

The generated findings can also be translated into another format, such as an audio summary, which will be useful for those who wish to view only the most important parts of a video rather than the entire thing.

2.2 Product Functions

Video Textual Analyzer is a highly handy project that may assist with a wide range of tasks. Video Text Analyzer will perform the following key functions:

- Producing textual summary from analyzed data:

  After completing the backend procedures, we will be able to produce textual summaries of the uploaded video, which will be useful for emphasizing significant aspects. Extracted text serves as a significant indicator of the video's content, and it is relatively simple to classify videos.

- Making the audio version of the video text analysis findings:

  The video text analyzer will have the added capability of being able to turn the retrieved text into a sound signal to help the blind.

  With the use of this added feature, a person who is blind may access the video's valuable information that they would otherwise be unable to see.

2.3 User Classes and Characteristics

- Video Text Analyzer should be useable by any user.
- The user should know how to operate mobile phones and PC.
- He/she should also be able to understand English.

- To begin analyzing the video text, they are not need to master a specific set of commands.

2.4 Constraints

- The project will be developed using HTML, CSS, JavaScript, Python etc.

- It will use android app for coding the Video text analyzer.

- Work product will be in compliance with IEEE standards.

3. Specific Requirements

- Hardware Requirements:

  I. OS- Windows 8 or above Architecture: 32- or 64-bit operating system x64 i.e., 86 bits, x32 i.e., 86 bits

  II. Minimum 4GB RAM

  III. Minimum 5GB of disk space to install Anaconda

- Software Requirements:

  I. Anaconda Version 2.1 or Latest Version of Anaconda

  II. Jupyter Notebook

  III. Python 3.5 or above

IV. Latest Version of HTML and CSS installed over machine

V. Browser: Chrome or Internet Explorer.

- Network requirements:

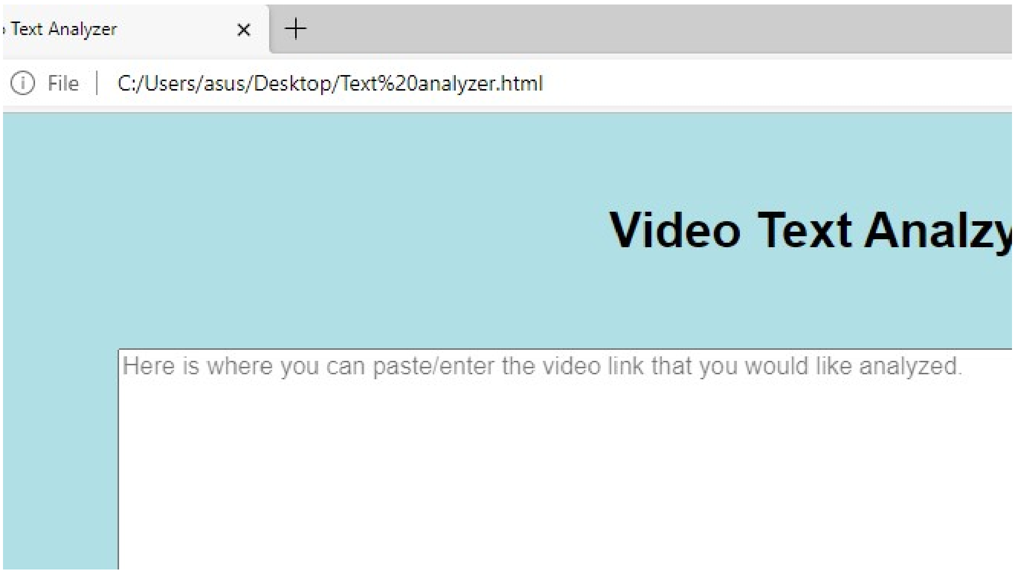
I. The Jupyter notebook runs on a local server on your computer, so there is no need of internet connection.

II. Internet Connection with 500+kbps

III. Bandwidth:3-5Mbps

3.1 External Interfaces

The first screen of Video Text Analyzer will have a space for the video's URL as shown in Figure2, after which, when we click Start analyzing, it will begin analyzing the video's content. From there, we can choose to make a textual summary or convert the written output into audio, which is useful for people who are blind as shown in Figure 3.



Text Analyzer
File
C:/Users/asus/Desktop/Text%20analyzer.html
Video Text Analzy
Here is where you can paste/enter the video link that you would like analyzed.

Figure 2 Mainframe of Video Text Analyzer

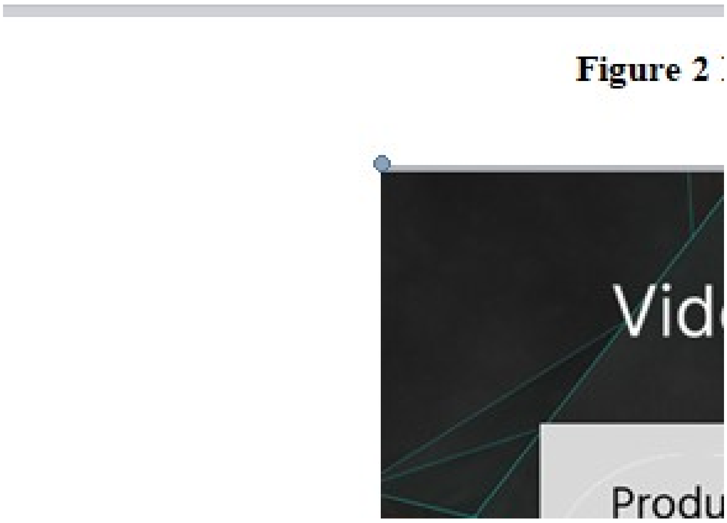


Figure 3 Options available for user

3.2 Performance requirements

- The performance of the functions and every module must be good.
- The overall performance of the software will enable the users to work efficiently.
- It should produce the result in a single click of a button, and video will upload easily that will make it less time consuming and reliable.
- This project will be available for the users only in the condition where the bandwidth of internet will be greater than 2.5Mbps.

3.3 Software System attributes

- *User Friendly*- It will be user friendly for the users as it will provide the user interface which will be easily handle able and understandable to user without any hassle.

- *Reliability*-It will produce the result in a single click of a button, and video can be uploaded easily that makes it less time consuming and reliable.

- *Portability*- As the project is developed using the open-source technologies like HTML, Python and on open-source platforms like Jupyter Notebook hence it will work on both windows as well as the android. Hence portability problem will not arise.

- *Availability*- This project will be available for the users only in the condition where the bandwidth of internet will be greater than 2.5Mbps.

3.4 Diagrams with detailed explanation

3.4.1 Use Case Diagram

- In UML, use-case diagrams model the behavior of a system and help to capture the requirements of the system. Use-case diagrams describe the high-level functions and scope of a system. These diagrams also identify the interactions between the system and its actors. The use cases and actors in use-case diagrams describe what the system does and how the actors use it, but not how the system operates internally.

- For our project, a video text analyzer, two actors—a user and a system—are needed as shown in Figure4.

- The user can upload the video that will be used as input for the video text analyzer and access the summarized data. Additionally, the textual output can also be converted to audio format.

- The system of Video Text Analyzer will be in charge of conducting process and segmentation, feature extraction, data cleaning and preprocessing, model creation, then training and managing the Data, and lastly creating written or audio summary of text taken from video. Process and Segmentation will include captioning of key frames.

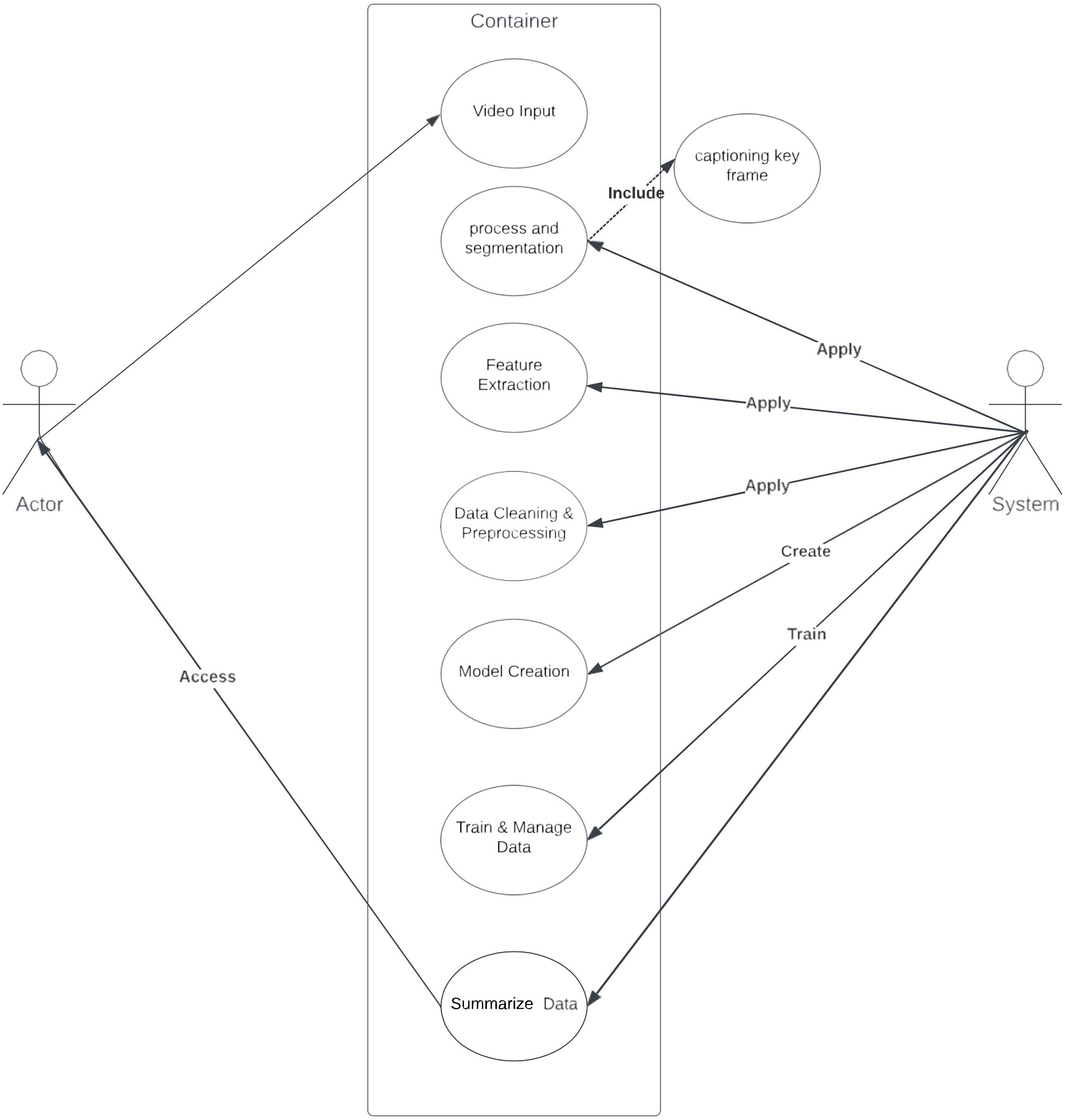


Figure 4 Use Case Diagram for Video Text Analyzer

3.4.2 Sequence Diagram

- The sequence diagram represents the flow of messages in the system and is also termed as an event diagram. It helps in envisioning several dynamic scenarios. It portrays the communication between any two lifelines as a time-ordered sequence of events, such that these lifelines took part at the run time.

- In UML, the lifeline is represented by a vertical bar, whereas the message flow is represented by a vertical dotted line that extends across the bottom of the page. It incorporates the iterations as well as branching.

- As shown in the Figure 5 the user is providing the input as a video URL to the system and along the timeline of System all the processes will be performed and, in the end, model will be formed with the help of LSTM.

- LSTM is a kind of recurrent neural network. LSTM was proposed to specifically address this issue of learning long-term dependencies. The LSTM maintains a separate memory cell inside it that updates and exposes its content only deemed necessary. The final output will be in the form of textual summary we can further convert it into audio form.

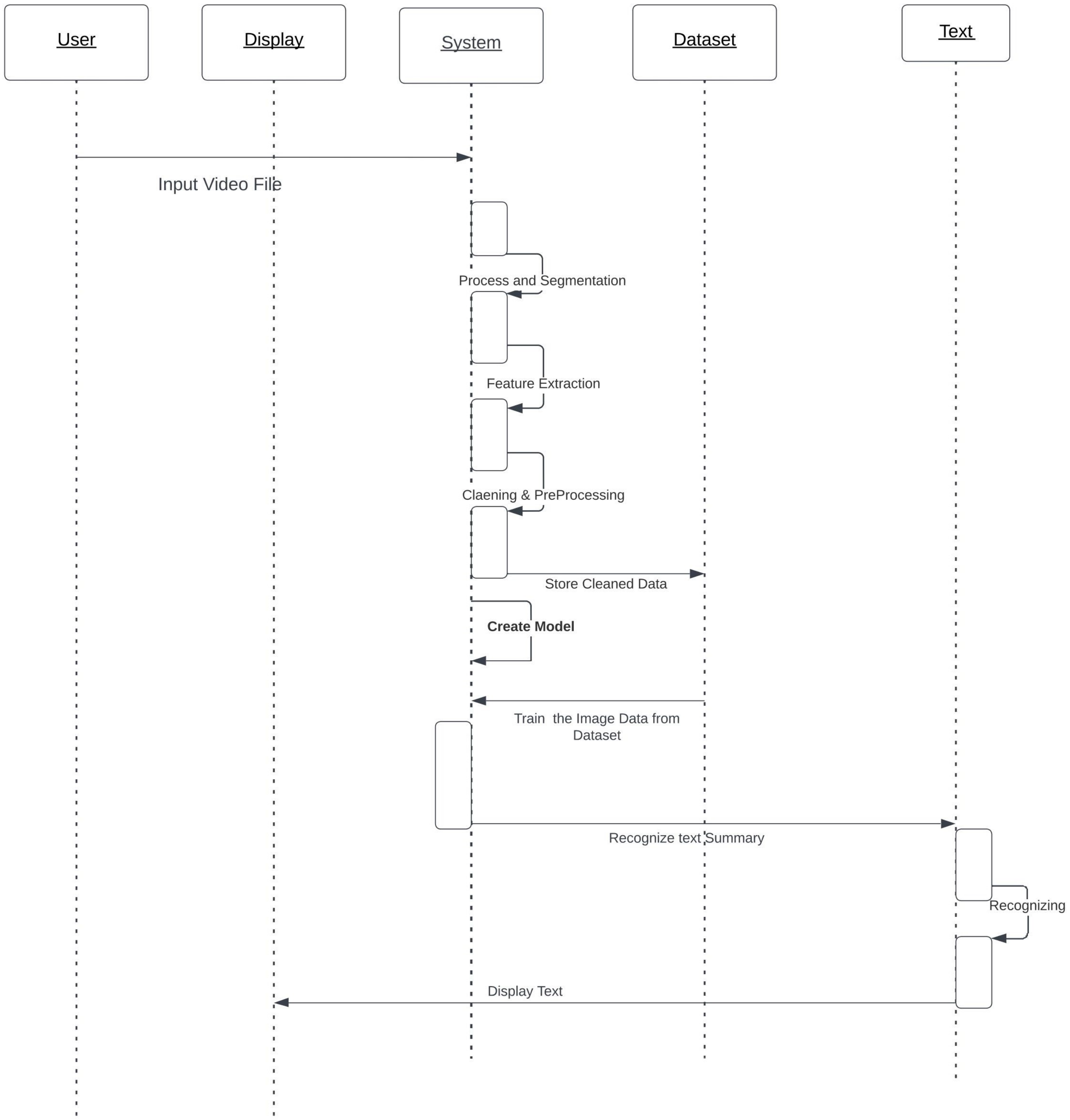


Figure 5 Sequence Diagram

3.4.3 Zero Level and one level DFD

DFD Level 0 is also called a Context Diagram. It's a basic overview of the whole system or process being analysed or modelled. It's designed to be an at-a-glance view, showing the system as a single high-level process, with its relationship to external entities. It should be easily understood by a wide audience, including stakeholders, business analysts, data analysts and developers. Zero level DFD for Video Text Analyser is shown in figure below:

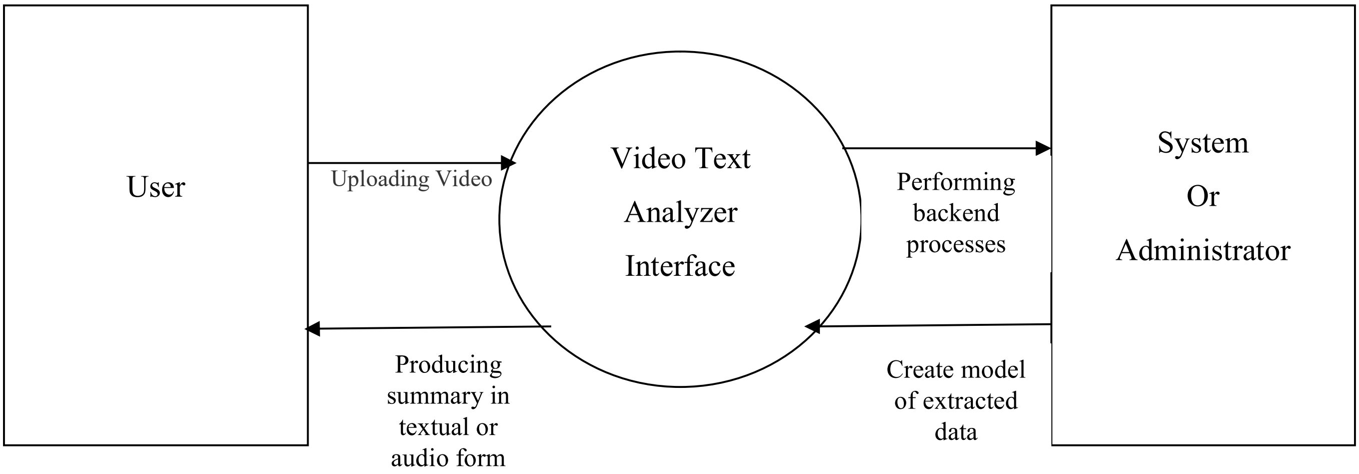

User
Uploading Video
Video Text Analyzer Interface
Performing backend processes
System Or Administrator
Producing summary in textual or audio form
Create model of extracted data

Figure 6 Zero Level DFD for Video Text Analyzer

With the help of this Figure 6, we can explain how the user will provide input by uploading the URL of the video through the video text analyzer interface, and how the system will create a model of the extracted data after completing the backend tasks and producing the summary in textual or audio format.

The context diagram is divided into many bubbles and processes in 1-level DFD. As shown in Figure 7, we have divided the Video Text Analyser processes into two modules, one for producing textual summaries and the other for converting textual output into audio, along with all the processing being done in the backend. In this level, we highlight the main functions of the system and break down the high-level process of 0-level DFD into sub processes.

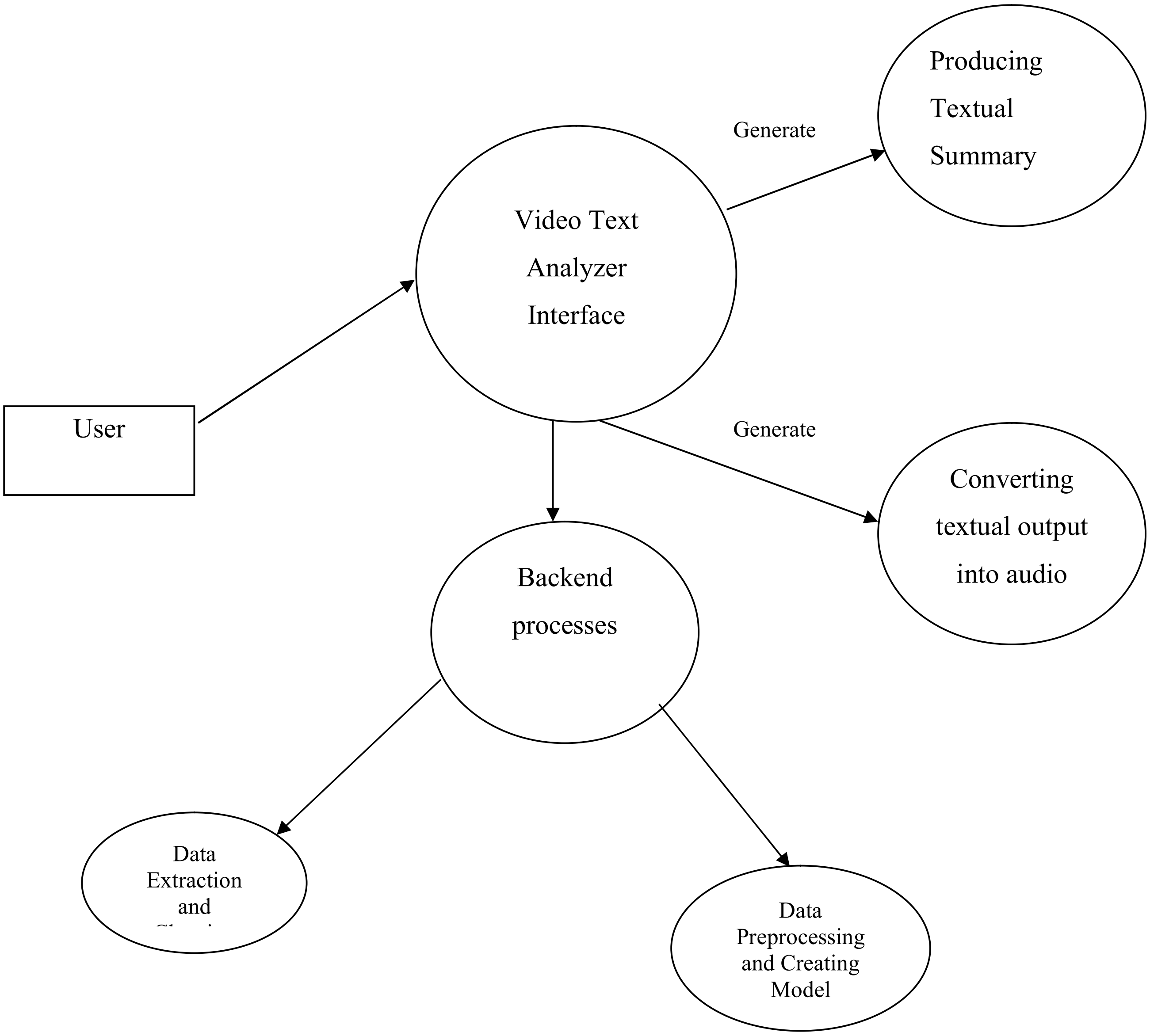

Producing Textual Summary
Generate
Video Text Analyzer Interface
User
Generate
Converting textual output into audio
Backend processes
Data Extraction and
Data Preprocessing and Creating Model

Figure 7 One Level DFD for Video Text Analyzer

3.5 Gantt chart

- Gantt charts are useful for planning and scheduling projects. They help you assess how long a project should take, determine the resources needed, and plan the order in which you'll complete tasks. They're also helpful for managing the dependencies between tasks.

- Gantt chart for Video Text Analyzer is shown below:

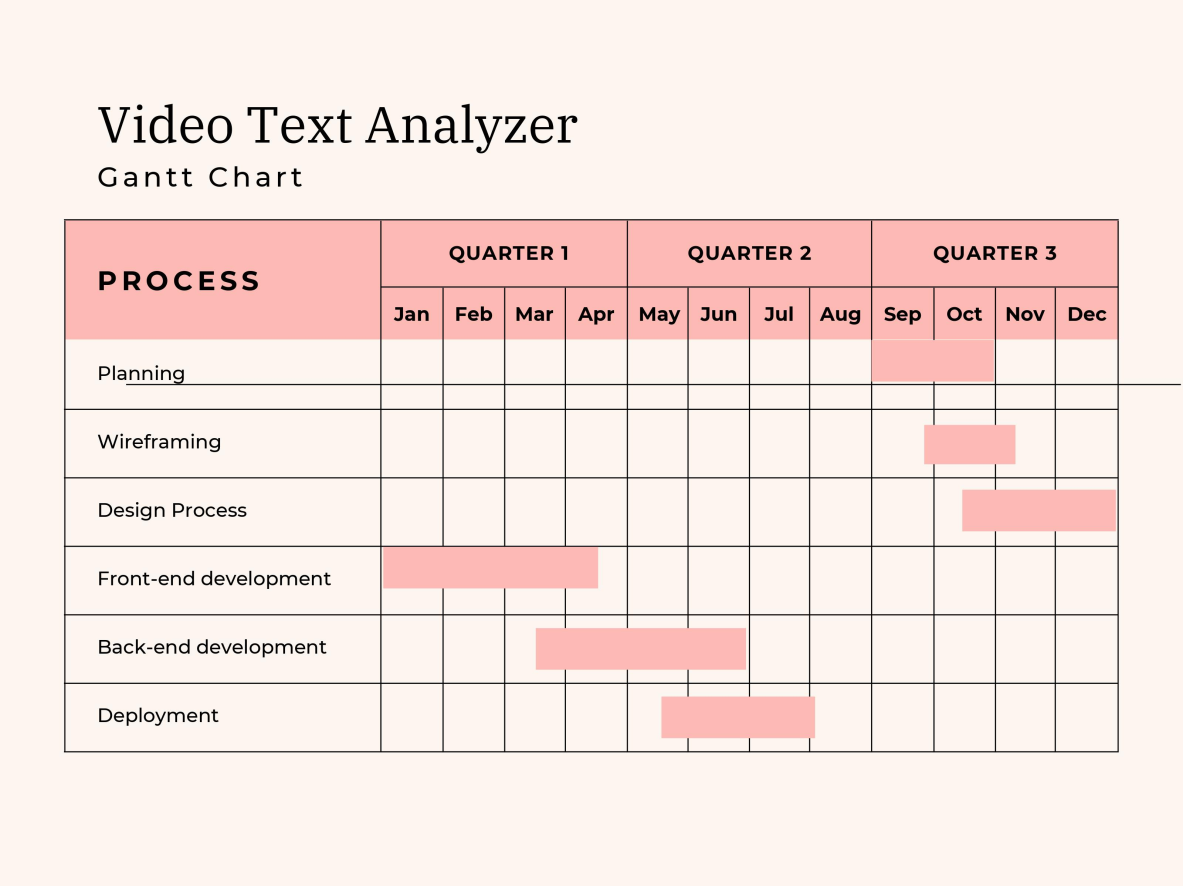


Figure 8 Gantt chart for Video Text Analyzer

4. Glossary

URL- URL stands for Uniform Resource Locator. It is the address of a resource, which can be a specific webpage or a file, on the internet.

IEEE-IEEE is the world's largest technical professional organization dedicated to advancing technology for the benefit of humanity. IEEE and its members inspire a global community through its highly cited publications, conferences, technology standards, and professional and educational activities.

5. References

[1]. https://www.citefactor.org/journal/pdf/Video-To-Text-Analysis-Deep-Learning.pdf

[2]. https://www.hindawi.com/journals/mpe/2016/2187647/

[3]. https://www.repustate.com/video-analysis/

[4]. https://arxiv.org/abs/2210.02399

[5] G. G. Rajput and Anita H.B., "Handwritten Script Recognition using DCT and wavelet Features at Block Level", IJCA, Special Issue on RTIPPR (3):158-163, 2010.

[6] G. G. Rajput and Anita H.B., "Kannada, English, and Hindi Handwritten Script Recognition using multiple features", Proc. Of National Seminar on Recent trends in Image Processing and Pattern Recognition, ISBN: 93- 80043-74-0, pp 149-152, 2010.

[7] Lajish V.L and Anoop K, “Mathematical Morphology and Region Clustering Based Text Information Extraction from Malayalam News Videos”, Springer International publishing, pp.431-442, 2015.

[8] Thika Ali H. Subber and Abbas H. AL-Asadi, “Arabic Text Extraction from Video Images”, Journal of Basrah Researches ((Sciences)), Vol. 39, No. 4, pp. 120-136, 2013.