Rohit_V_Shastry_MARK2

by Rohit V Shastry

Submission date: 02-May-2023 04:34PM (UTC+0530)
Submission ID: 2081942318
File name: Rohit_V_Shastry_MARK2.pdf (894.74K)
Word count: 6325
Character count: 33429

Dissertation on

"News on the Go" – A Video and Text Summarization and Translation Service

Submitted in partial fulfilment of the requirements for the award of degree of

Bachelor of Technology

in

Computer Science & Engineering

UE20CS390A – Capstone Project Phase - 1

Submitted by:

Niranjan Rao SS	**PES2UG20CS226**
Rohit V Shastry	**PES2UG20CS282**
Rushil Ranjan	**PES2UG20CS288**
Sai Hardik Sriram Talluru	**PES2UG20CS296**

Under the guidance of

Prof. Shilpa S
Assistant Professor
PES University

January - May 2023

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
FACULTY OF ENGINEERING
PES UNIVERSITY
(Established under Karnataka Act No. 16 of 2013)
Electronic City, Hosur Road, Bengaluru – 560 100, Karnataka, India

PES UNIVERSITY

(Established under Karnataka Act No. 16 of 2013)
Electronic City, Hosur Road, Bengaluru – 560 100, Karnataka, India

FACULTY OF ENGINEERING

CERTIFICATE

This is to certify that the dissertation entitled

'News on the Go' – A Video and Text Summarization and Translation Service

is a bonafide work carried out by

Niranjan Rao SS	**PES2UG20CS226**
Rohit V Shastry	**PES2UG20CS282**
Rushil Ranjan	**PES2UG20CS288**
Sai Hardik Sriram Talluru	**PES2UG20CS296**

In partial fulfilment for the completion of sixth semester Capstone Project Phase - 1(UE20CS390A) in the Program of Study -Bachelor of Technology in Computer Science andEngineering under rules and regulations of PES University, Bengaluru during the period Jan. 2023 – May. 2023. It is certified that all corrections / suggestions indicated for internal assessment have been incorporated in the report. The dissertation has been approved as itsatisfies the 6th semester academic requirements in respect of project work.

Signature Prof. Shilpa S Assistant Professor	Signature Dr. Sandesh B J Chairperson	Signature Dr. B K Keshavan Dean of Faculty

External Viva

Name of the Examiners	**Signature with Date**
1. _____	_____
2. _____	_____

DECLARATION

We hereby declare that the Capstone Project Phase - 1 entitled **"News on the Go – A Video and Text Summarization and Translation Service"** has been carried out by us under the guidance of Prof. Shilpa S, Assistant Professor and submitted in partial fulfilment of the course requirements for the award of degree of **Bachelor of Technology** in **Computer Science and Engineering** of **PES University, Bengaluru** during the academic semester January – May 2023. The matter embodied in this report has not been submitted to any other university or institution for the award of any degree.

PES2UG20CS226	**Niranjan Rao SS**
PES2UG20CS282	**Rohit V Shastry**
PES2UG20CS288	**Rushil Ranjan**
PES2UG20CS296	**Sai Hardik Sriram Talluru**

ACKNOWLEDGEMENT

I would like to express my gratitude to Prof. Shilpa S, Department of Computer Science and Engineering, PES University, for her continuous guidance, assistance, and encouragement throughout the development of this UE20CS390A - Capstone Project Phase – 1.

I am grateful to the Capstone Project Coordinators, Dr.Sarasvathi V, Professor and Dr. Sudeepa Roy Dey, Associate Professor, for organizing, managing, and helping with the entire process.

I take this opportunity to thank Dr. Sandesh B J, Professor & Chairperson, Department of Computer Science and Engineering, PES University, for all the knowledge and support I have received from the department. I would like to thank Dr. B.K. Keshavan, Dean of Faculty, PES University for his help.

I am deeply grateful to Dr. M. R. Doreswamy, Chancellor, PES University, Prof. Jawahar Doreswamy, Pro Chancellor – PES University, Dr. Suryaprasad J, Vice-Chancellor, PES University for providing to me various opportunities and enlightenment every step of the way. Finally, this Capstone Project could not have been completed without the continual support and encouragement Ihave received from my family and friends.

ABSTRACT

News on the Go is a mobile application that aims to provide a revolutionary new way for people to stay informed with the latest news. With the fast-paced nature of the world today, it can be difficult to keep up with all the news and information that is being produced every day. This app offers a unique solution to this problem by retrieving news videos from YouTube and summarizing them into concise and accurate textual and visual summaries.

One of the key features of News on the Go is its ability to translate both the video and textual summaries into multiple languages, making it accessible to people all around the world. This feature is especially valuable in a world where language barriers can often be a hindrance to accessing important news and information.

In summary, News on the Go is an innovative and user-friendly mobile application that offers a unique solution to the problem of staying informed with the latest news. Its advanced algorithms and artificial intelligence ensure that the summaries provided are accurate and informative, while its translation feature makes it accessible to people all around the world.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

In today's fast-paced world, staying informed with the latest news is more important than ever. With the rise of social media and the prevalence of smartphones, it's easier than ever to access news content. However, with so much information available, it can be challenging to keep up with everything that's happening. That's where 'News on the Go' comes in.

'News on the Go' is a mobile application that provides users with a unique and convenient way to stay up to date with the latest news. The app retrieves news videos from YouTube using the YouTube data API and uses advanced algorithms and artificial intelligence to summarize the videos into concise textual and visual summaries.

One of the key features of 'News on the Go' is its translation service. The app can translate both the video and textual summaries into multiple languages, making it accessible to a global audience. This is particularly useful for people who speak languages other than English, as it allows them to stay informed with news from around the world.

The app is designed to be accessible to as many people as possible, regardless of their technological expertise.

Overall, this application is a revolutionary way to stay informed in the present day. By providing users with a convenient way to access news content, the app makes it easier than ever to keep up with the latest developments.

CHAPTER 2

PROBLEM DEFINITION

The problem we aim to solve with this project is the difficulty people face in keeping up with the latest news in a fast-paced world. With the rise of social media and online news, there is a wealth of information available, but it can be overwhelming to sift through it all. Many people find themselves unable to keep up with the constant stream of news, and may miss important events or updates. Additionally, language barriers can prevent individuals from accessing news in languages other than their own. Our project seeks to address these challenges by creating a mobile application that uses advanced algorithms and artificial intelligence to provide concise summaries of news videos, and a translation service that can make these summaries accessible to people in multiple languages.

CHAPTER 3

LITERATURE SURVEY

3.1. Robust Speech Recognition via Large-Scale Weak Supervision

3.1.1. Introduction

Proposes a new approach to training speech recognition models that relies on weak supervision and large-scale data to improve model performance and scalability. The paper proposes to improve speech recognition. It uses large amounts of speech and textual data which is poorly labelled. They use this data to train the neural network.

3.1.2. Advantage

The paper depends on self-supervision while learning. This allows to learn from poorly labelled data or text and helps give a better model.

3.1.3. Disadvantage

The approach needs manual effort to create unsupervised data.

3.2. Abstractive Document Summarization

3.2.1. Introduction

The team proposes a graph-based model to provide abstractive text summaries. They make use of Graph CNNs, attention mechanisms and coverage techniques.

3.2.2. Advantage

The paper gives the best results on diverse datasets. This shows the method'saccuracy and effectiveness. The method can also be generalized for abstractive summaries.

3.2.3. Disadvantage

Evaluation is performed on a variety of datasets and diverse datasets. However, the evaluation is performed on small datasets. These small datasets cannot portray the model's efficiency and performance. When used for large datasets, the model might underperform.

3.3. Unsupervised Extractive Text Summarization

3.3.1. Introduction

The paper puts forth a proposal for an unsupervised method to perform extractive summarization for a transcript. This paper employs a method called sentence graph-based summarization. This method uses sentence graphs that are created automatically by the model and selects the key sentences or key phrases in the transcript for summarization.

3.3.2. Advantage

This method gives satisfying text summaries for unsupervised data. The results are comparable to supervised summaries with strong baselines.

3.3.3. Disadvantage

The paper achieves decent results for weakly labelled data using unsupervised methods but its still not as efficient as supervised methods in generating summaries.

3.4. Unsupervised Usage of Multi-source Features

3.4.1. Introduction

This paper proposes the use of the MCSF where frames are sampled at a rate of 2 FPS and the frames are retrieved using an encoder.

3.4.2. Advantage

The use of chunk and stride fusion allows the algorithm to observe local information along with universal information. Results in more accurate and representative video summaries.

3.4.3. Disadvantage

The effectiveness depends on the quality of the feature extraction process.

Doesn't work well with complex videos with more plots and movements and people.

3.5. Multi-Task Learning

3.5.1. Introduction

The objective is to propose a multi-task learning approach that learns recommendation along with giving ranks to improve efficiency of web-search and provide more accurate recommendations.

3.5.2. Advantage

The use of this improves the effectiveness and efficiency in web search.

3.5.3. Disadvantage

Cannot handle new users.

3.6. GPT2MVS

3.6.1. Introduction

The aim of this paper was to create a summary using an explicit request in a text format and then make use of attention schemes.

3.6.2. Advantage

The tests showed that the approach was useful and showed an increase in accuracy by quite a few percentage points.

3.6.3. Disadvantage

The proposed approach requires significant computational resources, making it difficult to implement in resource-constrained environments. It may not be as effective for certain types of videos, such as those with complex visual content or those with non-standard language use.

3.7. Time-Aware Transformers

3.7.1. Introduction

The paper puts forth a TAMT to create visual summaries. To explore the time information, they implement an attention mechanism in the encoder of the TAMT.

3.7.2. Advantage

The method manages to provide visual summaries that capture both objective and subjective summaries. It captures the creator's intention behind the video.

3.7.3. Disadvantage

Summary generated is not in chronological order and some detailed sequences are not well summarized.

CHAPTER 4

DATA

4.1. Overview

For this project, we are going to develop a system that retrieves news videos from YouTube. We shall pull these videos and their meta-data from YouTube and store them in a cloud-based management system – Google Cloud Storage being our initial choice. We design a web-crawler that continuously runs on a remote server. This web-crawler waits for the respective YouTube Channels to upload news videos which it can then extract. We use these videos for further processing.

4.2. Dataset

The data we will be using for this project will be the news videos of the news channels. These are uploaded by the news channels on YouTube and we retrieve them using the YouTube Data API which is an Open-source API. We extract videos that cover a wide range of news like politics, weather, business, finance, sports and entertainment. We also make use of the additional data of the video to give more personalized summary recommendations to users based on their search history and watch history. The videos undergo pre-processing to extract relevant information such as transcripts, video frames, timestamps and text overlays. This information will then be used to generate textual and visual summaries of the news videos using advanced algorithms and artificial intelligence techniques. The data will be processed and utilized in different ways to give the users an efficient way to view their daily news.

CHAPTER 5

SYSTEM REQUIREMENTS SPECIFICATION

5.1. Product Features

Video Summarization:

- The app summarizes news videos into concise visual summaries.
- It utilizes artificial intelligence to ensure accurate reporting.
- Captures objective and subjective data in a video leading to better summaries.

Text Summarization:

- Provides users with a textual summary of the news video.
- Ensures that the summary is concise and easy to understand.

Translation:

- Provides users with an option to translate summaries in multiple languages.
- Helps users to stay informed about global news, regardless of language barriers.

Accessibility:

- Designed such that disabled people can also use it.
- Has TTS and brightness and lighting modes for people with visual impairments.

User Experience:

- Ensures a clean and easy to interact UI.
- Fast and reliable performance, with minimal loading times.
- Constantly updates in the background based on user feedback and app performance.

5.2. Operating Environment

The system will operate on smart phones and other handheld devices. It will use APIs to access news videos and articles, and pre-trained machine learning models for video summarization, transcription, and translation. No additional hardware will be required to access this app.

5.3. General Constraints, Assumptions and Dependencies

5.3.1. Legal Implications

- Intellectual property infringement: This project should not infringe on any patents or copyrighted material when using APIs, pre-trained models, or Python libraries.
- Licensing: Some APIs, pre-trained models, and Python libraries such as OpenAI have specific licensing agreements that need to be followed.
- Usage limits: Some APIs and pre-trained models may have usage limits that need to be followed to avoid legal issues.
- Privacy laws: This project will need to comply with data privacy regulations since it collects and stores user data.

- Translation accuracy: The translations need to be accurate and not infringe on any intellectual property or trademark laws.

- Data security: The software tool may need to ensure that any data accessed or stored is secure and not subject to unauthorized access or use.

5.3.2. Usage Limitations

- The accuracy may depend on the complexity of the video, the transcript and the quality of the translation algorithms used.
- The project may be limited to only a certain number of languages.
- The project may have certain restrictions due to the usage limit of certain Partner APIs.

5.3.3. Dependencies

- **APIs**: The project is dependent on third-party and Partner APIs to access news videos and perform transcription and translations.
- **Pre-trained models**: The project may depend on pre-trained models for natural language processing, video summarization, and machine translation. For this project , we shall be employing the services of BERT and OpenCV.
- **Python libraries**: The project may depend on various Python libraries for data processing, machine learning, and natural language processing. Examples of such libraries include NumPy, Pandas, Scikit-Learn, and NLTK.

- **Software development frameworks**: The project may be dependent on specific software development frameworks to build and deploy the app on different mobile operating systems, such as React Native or Flutter.
- **User feedback**: The project may depend on user feedback to improve the accuracy and quality of the video summarization, translation, and transcription. It may be necessary to collect and analyse user feedback to identify areas for improvement and adjust the algorithms accordingly.

5.3.4. Assumptions

- We assume that the videos used in the project are in a format that can be processed by the algorithms such as MP4.
- We assume that the quality and accuracy of the video summarization, translation, and transcription algorithms used are sufficient for the purposes of the project.
- We assume that the text summarization and translation algorithms used are effective in generating accurate summaries and translations.
- We assume that the video and audio quality of the news videos used is sufficiently high to be processed accurately.
- We assume that users have an internet connection to access the app.
- We assume that users are able to understand and use the app's interface without significant difficulty.
- We assume that users are willing to provide feedback on the app's performance and functionality in order to improve its accuracy and effectiveness.

- We assume that the project will not infringe on any intellectual property or copyright laws.
- We assume that the project will comply with all relevant privacy laws and regulations, and that user data will be collected and stored in a secure and ethical manner.

5.3.5. Risks

Technology failures: Algorithms and technologies used may not work as expected, leading to errors in the app.

Hardware failures: Hardware used to develop or run the app may fail or malfunction, resulting in delays or loss of progress.

Version compatibility problems: Different versions of libraries, frameworks, and technologies may not be compatible, resulting in errors or other issues that could delay project completion.

Security threats: The project could be prone to different security threats, such as DoS- denial of service, hacking, data breach and malware attacks leading to loss or theft of user data or sensitive information.

Legal issues: The project may violate laws and regulations related to intellectual property, privacy, or other areas, leading to legal action that could delay or halt project delivery.

Resource constraints: The project may require significant resources, such as time which is limited, leading to delays or reduced functionality in the app.

5.4. Functional Requirements

Video Search:

- The app should provide a search functionality for users to find news videos based on news channels, keywords, genre, and other parameters. The functionality will also have a recommendation algorithm working to provide better recommendations of news videos to users based on their search and watch history.
- The search results should be a list of already summarized videos that match the user's search criteria.

Video Playback:

- The app will allow users to select and play the summarized video from the search results and recommendations.

Transcription and Translation:

- The app will transcribe the audio to text along with timestamps which will help in the summarization process.
- The app will have the option to translate the textual and visual summary into different languages.

User Interface:

- The app should have a user-friendly interface for searching and playing summarized videos.
- The app should allow users to interact with the summary text by highlighting or selecting specific sections.
- The app should provide the option to save or share the summary text and video.

- The app should allow users to listen to the summarized news as an audio by plugging in earphones.

Error Handling:

- The app should be able to handle errors or issues that may arise during transcription or translation.
- The app should notify the user if there is an error or issue with the video or summary generation.

5.5. External Interface Requirements

5.5.1. Hardware Requirements

- The application will require a mobile device with internet connectivity, such as a smartphone or tablet. The device should be handheld.
- The device should have enough storage space to install the app.
- The app should be compatible with some components such as the microphone. In order for this to work, the user should give permission to use the microphone of the device.
- The app will utilize cloud-based storage for video and text summaries and all the related video meta-data including keywords, phrases, people, news channels etc.
- The app will be designed such that it doesn't drain the device's battery life.

5.5.2. Software Requirements

Primary Programming Language - Python It is widely used in machine learning and natural language processing.

Backend Server - Python Language will be used for the backend. Python has a large standard library that includes many modules for app development, database connectivity, and more.

App Development Framework - FLUTTER will be used to develop the app.

Database Management Systems - We will be using Google Cloud SQL.

Cloud Services - Google Cloud will be used to host the application, store data, which would allow for scalability and cost-effectiveness.

GitHub: the source code for the app will be stored on GitHub for version control and collaboration.

5.5.3. Communication Interfaces

- Internet connection for retrieving and storing news videos.
- APIs for Transcription and Translation.
- LAN for communication between the app and the server.
- HTTPS for secure communication and data transfer.
- Serial ports or USB for connecting to external devices such as headphones or speakers.

Communication Standards

- High-speed internet connection for faster retrieval of news videos.

- Standard protocols such as TCP/IP for reliable and efficient data transfer.

5.6. Non-Functional Requirements

5.6.1. Performance Requirements

- The application should be able to handle significant number of people at any given time.
- The backend system should be able to process and transcribe the video into text within 30 seconds. This is achieved using OpenAI.
- The system should be able to summarize the text into a shorter summary.
- The system should have an accuracy rate of at least 90% for summarization and 95% for transcription.
- The system should be able to translate the video summary and text summary into at least 2 different languages – Kannada and Hindi being our initial choice.
- The system should have a response time of no more than 2 seconds for user interactions.

Quality Attributes

- The system should be reliable and not result in errors or malfunctions frequently.
- The system should be available 24/7 and shouldn't have a downtime of more than 2 hours in a month. These 2 hours will be marked for maintenance and updates.

5.6.2. Safety Requirements

- The system should not produce any harmful outputs or content.
- The system should not store any user information or data without their consent.

5.6.3. Security Requirements

- The system should have secure authentication methods to ensure user privacy and prevent unauthorized access.
- The system should use encryption methods to protect user data and prevent data breaches.
- The system should comply with GDPR rules and other relevant data protection laws.

CHAPTER 6

SYSTEM DESIGN

6.1 High Level Design

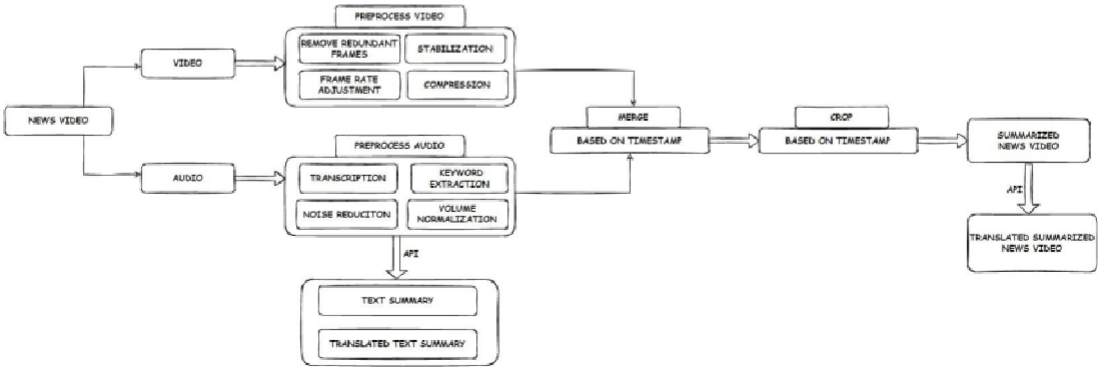


Figure 1

The raw, unprocessed news video is firstfed into the model and is then split into separate audio and video pipelines.

Video Pipeline - The video pipeline processes the video by compressing it, stabilizing it, adjusting the frame rates, and removing redundant frames.

Audio Pipeline - The audio pipeline normalizes the audio, transcribes it and extracts important keywords and phrases.

In the video pipeline, the key frames are obtained from the from the timestamps using the important keywords obtained during pre-processing, and the frames are labelled with numbers.

In the audio pipeline, the news video is converted to an mp3 file and then processed by the summarization module.
The module generates chapters and highlights containing the summary and time-intervals of the key topics.
Some of the time intervals are extracted from the chapters and highlights and merged with the video timestamps.

Text Summarization - The transcripts obtained from the video will be given to the text summarization model. First, the text will be tokenized or broken down into words. The tokenizer then encodes the text as a sequence of numbers that can be input into the transformer.
The transformer has an encoder and a decoder. The encoder captures the important part of the text and the decoder decodes the input from the encoder and gives a summary out of this.

Transcription and Translation will be performed with the use of an API.

Use of BERT – BERT is a pre-trained transformer which will be used in the text summarization model of this project. We will be using it to tokenize the words and assign them into vectors by giving integer IDs. The vector will be fed to the transformer blocks of BERT which consist of 2 layers. The mechanism is used to focus on various parts of the input at the same time to learn the relationships between different pairs of words and derive patterns in the text. We compute attention scores and give this as input to the feed forward network which transforms this by computing the weighted sum and ranking them. This output is given to the decoder which extracts the text from the vectors. Hence, we get our textual summaries. In order to obtain more abstract summaries, we need BERT to learn the nuances of our language. So, we fine tune BERT by feeding it news related data and videos. By doing this, it learns the language and can distinguish

between objective and subjective data, and can also achieve more abstractive summaries which are much more human-like compared to extractive summaries.

Video Summarization – before performing summarization, we pre-process the videos by adjusting the frame rates, stabilizing the video and by compressing it to acceptable sizes using libraries. We then remove the redundant frames by finding the key frames in the video. Then, we use open-source python libraries to perform object detection. For videos containing factual information in the form of text overlays, we will be using Optical Character Recognition – OCR, which will extract those facts from the text on screen. We will also use the transcripts of the videos to find the key words or phrases and extract those scenes using the timestamps.

6.2. UML Design

6.2.1. Master Class Diagram

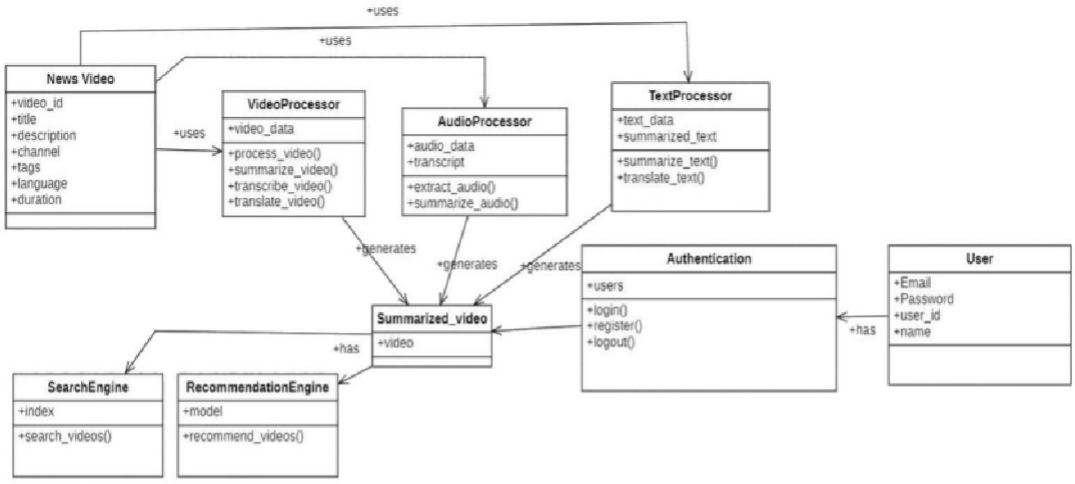


Figure 2

The Class diagram consists of a News Video class, VideoProcessor class, AudioProcessor class, TextProcessor class, Summarized_video class, SearchEngine class, Recommendation class, Authentication class, User class.

The News Video class will have the all the video title, id, description, tags, duration etc.
The VideoProcessor will perform functions like summarize video, translate video, and transcribe video. It receives the news video from the News Video class.

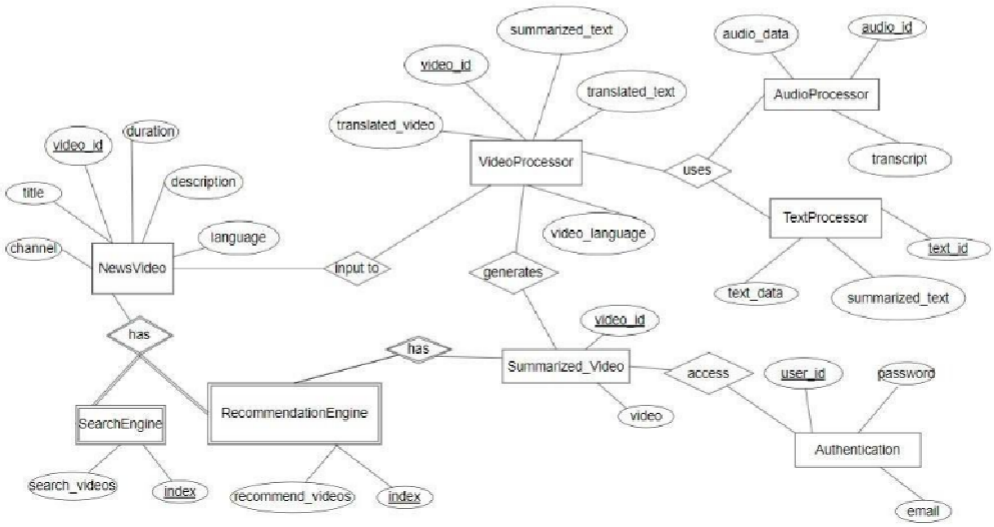
AudioProcessor Class has extract_audio function and summarize_audio function that produces highlights, chapters and sentence transcripts, generate timestamps takes these parameters into consideration and generates audio timestamps, speech intervals.

TextProcessor class has summarize_text function and translate_text function that will summarize the text and translate the text into other languages.

Summarized_video class will contain the summarized video.
User class will have the all the users id, email, name and the password.

6.2.2. Entity Relationship (ER) Diagram

Figure 3 – ER Diagram



The News Video entity holds a dataset of videos and includes various attributes such as video ID, duration, title, channel name, description, and language. This entity serves as an input to the Video Processor entity.

The Video Processor entity has several attributes, including the translated video, video ID, summarized text, translated text, and language. This entity utilizes two other entities, the Audio Processor and the Text Processor, to process the audio file and generate summarized text, respectively.

The Audio Processor entity has attributes such as audio data, audio ID, and transcript, while the Text Processor entity has attributes like text ID, text data, and summarized text.

The Video Processor entity generates the Summarized Video entity, which includes attributes such as video ID and summarized video. To access this entity, an Authentication entity is used, which has attributes such as user ID, password, and email.

Furthermore, the Summarized Video entity is linked to two weak entities, the Search Engine and the Recommendation Engine. The Search Engine allows users to search for summarized videos and has attributes such as searched videos and ID. The Recommendation Engine recommends summarized videos to users and has attributes such as recommended videos and ID.

#	Entity Name	Definition	Type
1.	NewsVideos	Contains the input data for Processing i.e.; the Original Video	Strong Entity
2.	VideoProcessor	Processes the Video to generate a Summarized video version of it	Strong Entity
3.	AudioProcessor	Processes the Audio of the Video file.	Strong Entity
4.	TextProcessor	Processes the transcription of the Video to generate a summarized version of it	Strong Entity
5.	Summarized_Video	Contains the Summarized Video	Strong Entity
6.	Authentication	Access to Users	Strong Entity

7.	RecommendationEngine	Provides recommended videos based on the user.	Weak Entity
8.	SearchEngine	Provides a search option for the user	Weak Entity
#	**Attribute Name**	**Definition**	**Type (size)**
1a.	Channel	The Channel Type of the News Video	String(50)
1b.	Title	The Title of the News Video	String(50)
1c.	Video Id	Unique Id given to the News Video	Int(5000)
1d.	Description	Description of the News Video	String(500)
1e.	Duration	Time period of the News Video	Time(1:00:00)
1f.	Language	Language used in the News Video	String(20)
2a.	Translated Video	The Translated Language of the News Video	String(1000)
2b.	Video Id	Unique Id given to the News Video	Int(5000)
2c.	Summarized Text	Summary of the Transcript	String(1000)
2d.	Translated Text	The Translated Transcript	String(1000)
2e.	Video Language	Language of the Video	String(50)
3a.	Audio data	Audio of the News Video	String(100)
3b.	Audio Id	Unique Id given to	Int(5000)

		the News Video for the audio file	
3c.	Transcript	The written version of the News video	String(5000)
4a.	Text Id	Unique Id given to the text	Int(5000)
4b.	Summarized text	Brief Statements of the text.	String(1000)
4c.	Text data	Description of the text	String(1000)
5a.	Video Id	Unique Id given to the News Video	Int(5000)
5b.	Video	Summarized video of the original one	String(100)
6a.	User Id	Unique Id given to the User	Int(5000)
6b.	Password	Password of the User	String(100)
6c.	Email	Email of the User	String(100)
7a.	Index	Unique Id given to the News Video	Int(5000)
7b.	Recommend Videos	Recommended Videos are suggested to the user	String(100)
8a.	Index	Unique Id given to the News Video	Int(5000)
8b.	Search Videos	The videos can be searched by the user	String(100)

6.2.3. Use Case Diagrams

A) Search

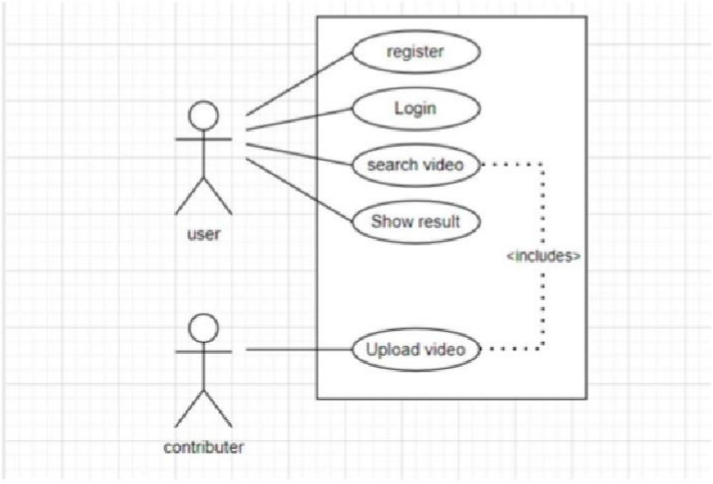


figure 4 – Search

B) View

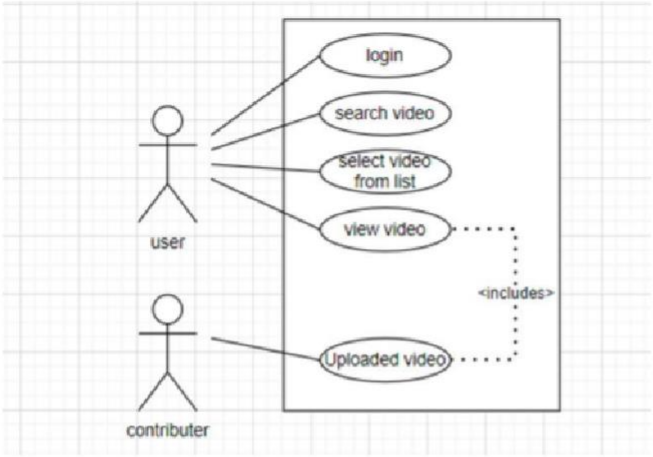


figure 5 – View

C) Translate

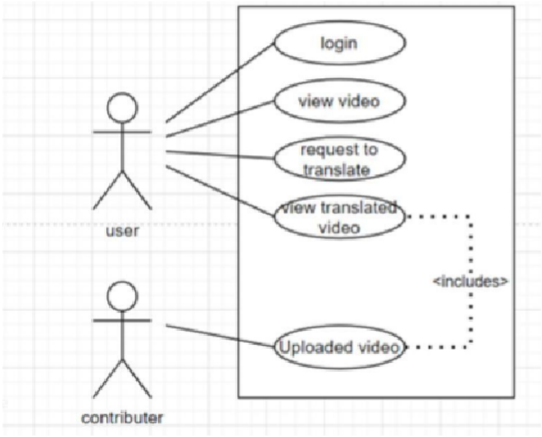


figure 6 – Translate

6.2.4. Logical User Groups

Logical User Groups can be widely classified into two - **Viewers and Administrators.**

Viewers

- Casual users
- News enthusiasts
- Researchers
- Journalists
- Business professionals
- Students
- Travelers

Administrators

Administrators are responsible for maintaining the system and ensuring that it is running smoothly. They monitor system performance from time to time and observe usage patterns and handle technical issues if any happen to occur.

6.2.5. External Interfaces

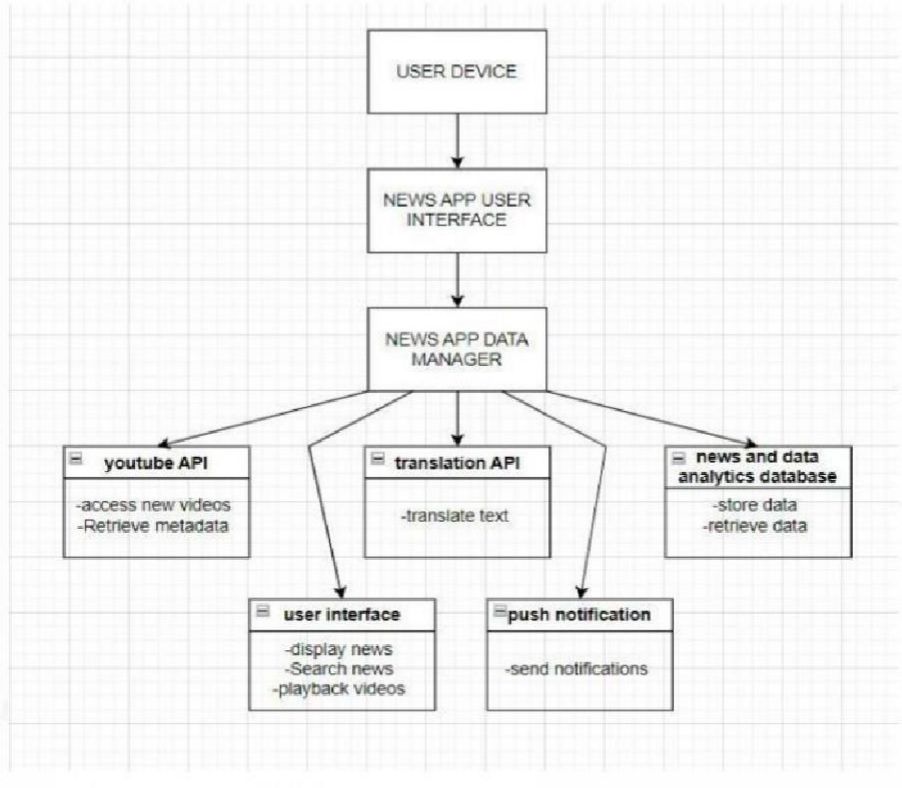


figure 7 – External Interfaces

External interfaces are the connections between the system and the outside world, including other systems and devices. For this app, the following are the external interfaces –

YouTube API: This is the interface that the system uses to access and extract news videos from the respective channels on YouTube.

Translation API: This is the interface that the news app uses to translate the summaries of news videos into different languages.

Database API: This is the interface that the system uses to store and retrieve data from the database.

User interface: This is the interface that the users use to interact with the app.

Push notifications: This is the interface that the system uses to send push notifications to users about new videos, personalized recommendations, or other app-related updates.

6.2.6. Packaging and Deployment Diagram

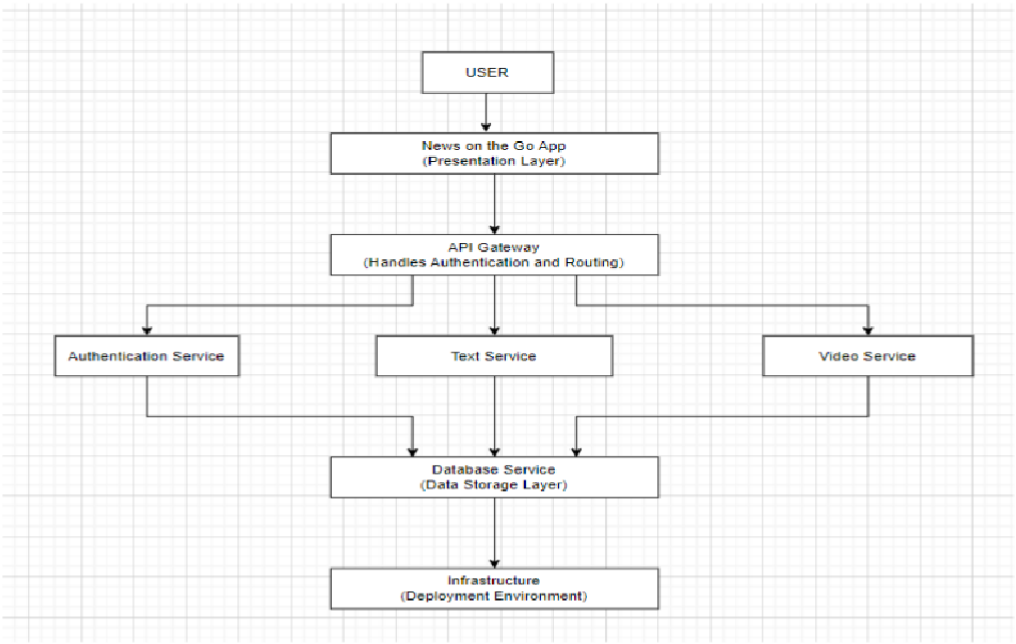


Figure 8 – Packaging and Deployment Diagram

CHAPTER 7

IMPLEMENTATION

The implementation of the project follows the following steps:

Data Retrieval: we use the YouTube Data API which is open-source and we retrieve the news videos from a specified number of news channels as and when they upload their videos on YouTube. We design a Web Crawler that runs continuously on a remote server and it watches these respective channels and pulls the videos and any meta-data available and stores them in a database which is setup using the Google Cloud Storage Platform.

Transcription: In order to transcribe our videos, we make use of the Whisper AI which is a Partner API from OpenAI. WhisperAI transcribes videos with an accuracy of over 96%, with a word error rate of under 4 words per 1000 words. It timestamps the transcripts. These transcripts will be useful for the summarization part.

Text Summarization: we will make use BERT and develop a new model by fine-tuning this highly pre-trained language model in order to come up with more abstractive summaries. We will be addressing both subjective and objective data in the transcripts while generating summaries. We shall make use of self-attention mechanisms and feed forward networks to achieve best results.

Video Summarization: we first pre-process the video. We adjust the frame rate to 30 fps if it isn't already at 30 fps. This is done since 30 fps is the international standard for news videos. Then, we stabilize the video in case of any harsh movements of the camera and we compress the video to an acceptable size so that it is easier on the storage side of the system. We perform key-frame extraction using python libraries and then we perform object detection either by using OpenCV or by using Detectron2AI which is made by Facebook. We perform OCR- Optical Character

Recognition in case of text overlays in the video. Using these methods, we come up with a concise visual summary of the news video.

CHAPTER 8

CONCLUSION OF CAPSTONE PROJECT PHASE-1

- Identifying the key challenges in creating Visual Summaries, Textual Summaries of News Videos and Translating them into multiple languages.

- Conducted an extensive Literature Survey on the topics like transcription, translation, BERT usage, WhisperAI, text summarization, video summarization, UI/UX and Recommendation Systems.

- The implementation section will involve the use of Python and its libraries like OpenCV, MoviePy, ffmpeg etc. We shall also make use of the YouTube API to retrieve videos. Whisper AI will be used for transcription and translation. BERT will be used for NLP.

- An application will be created to bring the project to the public.

- Data is acquired using a web-crawler that accesses news videos.

- Decided on the novelty component of the project.

- Creation of high-level design diagram, master-class diagram, Use-Case Diagram, ER Diagram, Packaging and Deployment Diagram completed.

CHAPTER 9

PLAN OF WORK FOR CAPSTONE PROJECT PHASE-2

- Data retrieval and storage.
- Acquiring Whisper AI API key and Implementing Transcription.
- Implement the Video Summarization, Text Summarization and Translation Models.
- Implement the Recommendation Model for the app.
- Design the backend of the App
- Design the app and integrate with the backend server.
- Review, alpha testing, beta testing, and deployment.
- Authoring the Research Paper.

APPENDIX A – DEFINITIONS, ACRONYMS AND ABBREVIATIONS

- BERT- Bi-Directional Encoder Representational Transformer
- API- Application Programming Interface
- MHSAM- Multi-Head Self Attention Mechanism
- OpenCV – is a python computer vision library that performs object detection, key-frame extraction and many other tasks.
- MoviePy- is another computer vision and video editing library of python.
- WhisperAI- is an API that can transcribe and translate videos in real-time.
- Flutter – is an application development framework
- detectron2: A library used mainly for object detection and other computer vision tasks.

REFERENCES

- **Multimodal Video Summarization via Time-Aware Transformers** –Xindi Shang, Zehuan Yuan, Anran Wang, and Changhu Wang. 2021. Multimodal Video Summarization via Time-Aware Transformers. In Proceedings of the 29th ACM International Conference on Multimedia (MM '21), October 20–24, 2021, Virtual Event, China. ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3474085.3475321

- **Supervised Video Summarization via Multiple Feature Sets with Parallel Attention** Junaid Ahmed Ghauri, Sherzod Hakimov, Ralph Ewerth https://arxiv.org/abs/2104.11530

- **Unsupervised Video Summarization via Multisource Features** Hussain Kanafani, Junaid Ahmed Ghauri, Sherzod Hakimov, Ralph Ewerth https://arxiv.org/abs/2105.12532

- **GPT2MVS: Generative Pre-trained Transformer-2 for Multi-modal Video Summarization** ia-Hong Huang, Luka Murn, Marta Mrak, Marcel Worring1. 2021. GPT2MVS: Generative Pre-trained Transformer-2 for Multi-modal Video Summarization. In Proceedings of the 2021 International Conference on Multimedia Retrieval (ICMR '21), August 21–24, 2021, Taipei, Taiwan. ACM,
  New York, NY, USA, 10 pages. https://doi.org/10.1145/

- Chen, G., Chai, S., Wang, G., Du, J., Zhang, W.-Q., Weng, C., Su, D., Povey, D., Trmal, J., Zhang, J., et al. Gigaspeech: An evolving, multi-domain asr corpus with 10,000 hours of transcribed audio. arXiv preprint arXiv:2106.06909, 2021.

- **Unsupervised Extractive Text Summarization with Distance-Augmented Sentence Graphs - 44th ACM SIGIR CONFERENCE**
  Jingzhou Liu, Dominic J. D. Hughes, and Yiming Yang. 2021. Unsupervised Extractive Text Summarization with Distance-Augmented Sentence Graphs. In Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '21), July 11–15, 2021, Virtual Event, Canada. ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3404835.3463111

- Chan, W., Park, D., Lee, C., Zhang, Y., Le, Q., and Norouzi, M. SpeechStew: Simply mix all available speech recognition data to train one large neural network. arXiv preprint arXiv:2104.02133, 2021. "Robust Speech Recognition via Large-Scale Weak Supervision" by Alec Radford, Jong Wook Kim, Tao Xu, Greg Brockman, Christine McLeavey, and Ilya Sutskever. OpenAI – 2022

- Jizhou Huang, Haifeng Wang, Wei Zhang, and Ting Liu. 2020. Multi-Task Learning for Entity Recommendation and Document Ranking in Web Search. ACM Trans. Intell. Syst. Technol. 11, 5, Article 54 (July 2021)

- V. K. Jeevitha and M. Hemalatha. "Natural Language Description for Videos Using NetVLAD and Attentional LSTM". In: 2020 International Conference for Emerging Technology (INCET). 2020, pp. 1–6. DOI: 10. 1109/INCET49848.2020.9154103.

- Julie Beth Lovins. "Development of a stemming algorithm". In: Translation and Computational Linguistic 11.1 (1968), pp. 22–31.

Rohit_V_Shastry_MARK2

ORIGINALITY REPORT

3%	3%	1%	0%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	coursesnew.iitm.ac.in Internet Source	2%
2	jntuhceh.ac.in Internet Source	1%
3	www.coursehero.com Internet Source	<1%
4	static.tti.tamu.edu Internet Source	<1%
5	internshipreports.yolasite.com Internet Source	<1%
6	cdn.openai.com Internet Source	<1%

Exclude quotes	On	Exclude matches	< 5 words
Exclude bibliography	On		