

**COMPUTER GENETRATED SUMMARIES AND KEYWORD EXTRACTION FROM VIDEO CONTENT USING NLP TECHNIQUES**

**A MINI PROJECT REPORT**

***Submitted by***

# SANTHOSH N J(1920110041)

**GOKUL K P (1920110012)**

**THANUSRI S (1920110054)**

***in partial fulfilment for the award of the degree***

***of***

**BACHELOR OF TECHNOLOGY**

**IN**

ARTIFICIAL INTELLIGENCE AND DATA SCIENCE

**SONA COLLEGE OF TECHNOLOGY, SALEM-5**

**(Autonomous)**

**ANNA UNIVERSITY: CHENNAI 600 025**

November 2023

**ANNA UNIVERSITY: CHENNAI 600 025**

**BONAFIDE CERTIFICATE**

Certified that this project report “**COMPUTER GENETRATED SUMMARIES AND KEYWORD EXTRACTION FROM VIDEO CONTENT USING NLP TECHNIQUES”** is the bonafide work of **“SANTHOSH N J (1920110041), GOKUL K P (1920110012),**

**THANUSRI S (1920110054)”** who carried out the project work under my supervision

**SIGNATURE**

**SIGNATURE**

Dr. J. Akilandeswari

Ms. M. Sasikala

**HEAD OF THE DEPARTMENT**

**SUPERVISOR**

Professor

Assistant Professor

Department Of Information Technology

Department Of Information Technology

Sona College of Technology

Sona College of Technology

Salem- 636 005.

Salem- 636 005.

Submitted for Project viva voce examination held on 16.11.2023 to 18.11.2023

|  |  |
| --- | --- |
| **INTERNAL EXAMINER** | **EXTERNAL EXAMINER** |

i

**ACKNOWLEDGEMENT**

First and foremost, we thank to **power of almighty** for showing us inner peace and for all blessings. Special gratitude to our parents, for showing their support and love always.

We express our sincere thanks to chairman **Sri. C. VALLIAPPA** and principal **Dr. S.R.R. SENTHILKUMAR** for providing adequate facilities to complete the project.

We are immensely grateful to our Head of Information Technology Department, **Dr. J. AKILANDESWARI** forthe continue encouragement to complete theproject.

We express our heartfelt thanks to our project supervisor

**Ms. M. SASIKALA** for her valuable guidance and faithful discussions throughoutthe course of the project work.

We feel proud in sharing this success with all our department Faculty, Staff members and friends who helped directly and indirectly to completing this project successfully.

ii

**ABSTRACT**

Computer generated summaries using machine learning and natural language processing to transcribe spoken words from video recordings. In the current fast-paced environment, individuals and Organization’s regularly having to manage enormous amounts of data in the form of audio recordings, video recordings, and text documents. The manual summarization and transcription of such data can be laborious and time-consuming. Keyword extraction is a crucial step in the information retrieval and categorization of text. This problem has been resolved by computer generated note-taking software for video recordings. Natural language processing (NLP) techniques are used to transform spoken words into text, which is subsequently used to build a summary or set of notes and analyze the transcribed text to identify and extract key terms using spacy. Keywords are identified based on their frequency or importance within the transcribed text. Keyword extraction is a vital component of the project aim to efficiently manage and utilize video recordings in an educational context. It facilitates the creation of meaningful summaries, enables precise information retrieval, ultimately benefiting both educators and learners.

iii

**TABLE OF CONTENTS**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **CHAPTE NO.** | |  | | | **TITLE** | | **PAGE NO.** |
|  | | **ABSTRACT** | | | | | **III** |
|  | | **LIST OF FIGURES** | | | | | **V** |
| **1.** | | **INTRODUCTION** | | | | | 07 |
|  | | 1.1. | Objective the project | | | | 09 |
|  | | 1.2. | Scope of the project | | | | 10 |
|  | | 1.3. | Literature Review | | | | 11 |
|  | | 1.4. | Existing system | | | | 14 |
| **2.** | | **SYSTEM DESIGN** | | | | |  |
|  | | 2.1. | Proposed System | | | | 15 |
|  | | 2.2. | Proposed Block Diagram | | | | 17 |
|  | | 2.3. | List of Modules | | | | 18 |
|  | | 2.4. | Module Description | | | | 18 |
|  | | | 2.4.1 | | | Video Collection | 18 |
|  | | | 2.4.2 | | | Video to Audio Extraction | 19 |
|  | | | 2.4.3 | | | Audio to Text Conversion | 20 |
|  | | | 2.4.4 | | | Text Summarization | 20 |
|  | | | 2.4.5 | | | Keyword Extraction | 21 |
|  | | | 2.4.6 | | | Document Compilation | 23 |
|  | 2.5. | | | Hardware/Software | | | 24 |
|  | 2.6. | | | Results / Screenshots | | | 25 |
| **3.** | | **CONCLUSION** | | | | |  |
|  | | 3.1. | Conclusion | | | | 30 |
|  | | 3.2. | Future Enhancement | | | | 31 |
|  | | **REFERENCES** | | | | | 32 |

**APPENDIX** 34

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |

**LIST OF FIGURES**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **FIG NO.** | | **TITLE** | **PAGE NO.** | |
| 2.1 | Proposed Block Diagram | | 17 | |
| 2**.**2 | Obtaining Video Files | | 18 |
| 2.3 | Transferring from mp4 to mp3 | | 19 |
| 2.4 | Transcribing spoken Words to Textual Content | | 20 |
| 2.5 | Selection of Keywords | | 22 |
| 2.6 | Keyword Retrieval from Textual Content | | 22 |
| 2.7 | Document Creation with The Text and Keywords | | 23 |
| 2.8 | Collecting Video Files | | 27 |
| 2.9 | Video to Audio Conversion | | 28 |
| 2.10 | Summary Generation | | 28 |
| 2.11 | Keyword Extraction | | 28 |
| 2.12 | Document Creation | | 29 |
| 2.13 | Inside the Document | | 29 |
|  |  | |  |

**LIST OF ABBREVIATIONS**

|  |  |
| --- | --- |
|  |  |
| **NLP** | Natural Language Processing |
| **TF-IDF** | Term Frequency – Inverse Document Frequency |
| **BERT** | Bidirectional Encoder Representations from Transformers |
| **RAKE** | Rapid Automatic Keyword Extraction |
| **AMIS** | Access Multilingual Information opinions |
| **AWS** | Amazon Web Service |
| **IBM** | International Business Machine |
| **ASR** | Automatic Speech Recognition |
| **WAV** | Wave from Audio file format |
| **ROUGE** | Recall Oriented Understudy for Giving Evaluation |
| **BLUE** | Bilingual Evaluation Understudy |
| **SEO**  **LSA**  **LDA** | Search Engine Optimization  Genism excels in Latent Semantic Analysis  Latent Dirichlet Allocation |

**CHAPTER 1**

**INTRODUCTION**

In today's fast-paced digital landscape, the sheer volume of data in the form of audio recordings, video content, and textual documents has become overwhelming for both individuals and organizations. The manual processes of summarizing and transcribing this wealth of information can be incredibly labour-intensive and time-consuming. However, the emergence of cutting-edge technologies that harness the power of machine learning NLP has ushered in a new era of data management and accessibility.

One of the most significant challenges in dealing with this data deluge is the efficient extraction of meaningful insights and information. This is where the innovative project we are about to embark upon comes into play. Through the NLP techniques, this project aims to revolutionize the way we handle video recordings. By transcribing spoken words from these recordings into text, we lay the foundation for a host of remarkable capabilities, including automated summarization and the extraction of crucial keywords.

The process is both sophisticated and powerful. As spoken words are transformed into textual data, our system leverages the method to identify and extract keywords that hold particular significance within the transcribed content. These keywords can be identified based on their frequency, relevance, or importance, giving us a profound understanding of the underlying content.

Crucially, this project has far-reaching implications, particularly in the context of education. It promises to be a game-changer, making the management and utilization of video recordings in educational settings significantly more efficient.

By facilitating the creation of meaningful summaries and enabling precise information retrieval, it empowers both educators and learners to access knowledge in an unprecedented manner. In essence, it unlocks the potential for more effective, engaging, and informative educational experiences.

In this project, we embark on a journey to harness the power of technology to streamline data management, making it more accessible and efficient for a broad spectrum of applications. The goal is not only to ease the burden of data processing but also to enrich the way we interact with information, ultimately contributing to a brighter future in education and beyond.

**1.1. OBJECTIVE OF THE PROJECT**

In digital landscape, where the influx of data in various formats, including audio recordings, video content, and textual documents, has reached overwhelming proportions, the need for efficient data management and accessibility has never been more apparent. Manual processes for summarizing and transcribing this wealth of information are not only labour-intensive but also time-consuming, making it a formidable challenge for individuals and organizations alike. However, we are on the cusp of a transformative solution, poised to harness the power of cutting-edge technologies that leverage NLP to tackle this challenge head-on.

Our innovative project aims to redefine the way we handle video recordings, setting the stage for an array of groundbreaking capabilities. At its core, the project focuses on transcribing spoken words from video content into text, laying the foundation for automated summarization and the extraction of crucial keywords. This process, while sophisticated, is a powerful one, as it seamlessly converts spoken words into textual data and employs advanced NLP techniques to identify and extract keywords of particular significance within the transcribed content. These keywords are selected based on factors such as frequency, relevance, and importance, providing us with a profound understanding of the underlying content.

Crucially, this project holds far-reaching implications, particularly in the field of education. It has the potential to be a game-changer, greatly enhancing the management and utilization of video recordings in educational settings.

By enabling the creation of meaningful summaries and facilitating precise information retrieval, it empowers both educators and learners to access knowledge in an entirely unprecedented manner.

**1.2. SCOPE OF THE PROJECT**

The project's core scope includes the development of an advanced system for processing video content. This system will be designed to investigate a diverse range of video formats, including limited to MP4. The system will also accommodate videos of varying lengths, from short clips to feature-length films. The scope includes the development of a user-friendly interface or API that allows users to interact with the system easily.

Users should be able to upload video content, initiate the summarization and keyword extraction processes, and generate document files without significant technical barriers. We make use of Assembly AI API to extraction summaries from audio content. It will utilize audio recognition techniques to transcribe spoken words and identify sound patterns. The project will develop an advanced video summarization system that watches videos, understands their content, and generates coherent textual summaries.

The summaries will aim to capture the essential information, context, and key insights from the audio. The scope here involves not only automatic summarization but also ensuring readability of the summaries. Keyword extraction from video content is a vital component of this project's scope. The system will extract relevant keywords from summaries. Extracted keywords will enhance content categorization and search optimization. It is important to note that the system aims to provide relevant keywords but may not capture every possible keyword.

The project goals are to integrate the generated summaries and extracted keywords into structured document files. The documents will be in form of Word documents. This aspect of the project will enable users to create organized, shareable, and searchable documents based on the video content.

Performance considerations involve processing speed, accuracy, and resource requirements, ensuring the system can efficiently process videos while maintaining high-quality results. This project has the potential to significantly impact various domains, from education to entertainment, by making video content more accessible and actionable.

**1.3. LITERATURE REVIEW**

**SURVEY 1:** Automatic Video Summarization with Timestamps using Natural Language Processing Text Fusion, IEEE 11th Annual Computing and Communication Workshop and Conference, 2021.

The author aimed to save time for users by watching videos and saves their time using timestamps. One of the main goals of that project is to keyword extraction. Extracted keywords help finding videos with the significant video's keywords. The summarizing of the video depends on frames, emotions, and speech. Techniques such as tokenization, sentence segmentation and lemmatization & stemming, and then abstractive summarization. Video summarization occurs to get a meaningful accurate description of the video. Having an accurate description helps finding the inquired content matching the description. The implemented experiment showed that on average 87% of the participants found generated text well representing the video. The evaluation is conducted on a small dataset of only 20 videos, which may not be representative of the real-world distribution of videos.

**SURVEY 2:** Video Summarization Using Deep Neural Networks: A Survey, arXiv preprint in 2021.

Video summarization technologies aimed to create a concise and complete synopsis by selecting the most informative parts of the video content. Several approaches have been developed over the last couple of decades and the current state of the art is represented by methods that rely on modern deep neural network architectures. They reported on protocols for the objective evaluation of video summarization algorithms and compare the performance of several deep-learning-based approaches. Technology used are Computer Vision and Pattern Recognition, Machine Learning and Multimedia. The paper does not discuss the limitations of deep learning-based video summarization methods.

Another drawback is that the paper does not focus on a specific aspect of video summarization. Instead, it provides a broad overview of the topic.

**SURVEY 3:** Video Summarization using NLP, International Research Journal of Engineering and Technology (IRJET) in 2021.

An automatic video summarization algorithm using NLP based algorithms has been proposed to handle an increase in internet videos on the video repository platforms like YouTube, Instagram etc. there is an increase in demand for good summarization algorithms to summarize various videos. The paper aimed to produce short and concise video summary that summarizes various YouTube videos. This method lacks in the field of limited coverage and generalization. Many machines learning-based video summarization techniques are present but they require devices with large processing powers, this is because each video contains thousands of frames and processing all frames takes a very long time. In this paper we propose to use the LSA Natural Language Processing algorithm, which requires less processing power and no training data required to train the algorithm. The proposed system is that it requires a lot of processing power and memory.

**SURVEY 4:** A survey on NLP based text summarization for summarizing product reviews, in Proceedings of the 2nd International Conference on Inventive Research in Computing Applications, ICIRCA in 2020.

Text summarization is the process of extracting the main idea of the context or the text and briefly explaining about the context. This process is not only to extract key idea and phrases from the text sources but also generating meaningful summary in a concise and crisp way. Depending on the NLP techniques used, some nuances or subtleties in the text may be overlooked.

The effectiveness of text summarization heavily depends on the NLP techniques applied, which may inadvertently overlook nuances and subtleties within the text. Though text summarization aims to provide concise summaries, it results in a loss of context or important details present in the original text.

**SURVEY 5:** A Comparative Study on different Keyword Extraction Algorithms, 3rd International Conference on Computing Methodologies and Communication (ICCMC) in 2019

Extraction of keywords manually is a tedious task, The author has developed an Automatic keyword extraction techniques, helps in overcoming this challenging task. This paper is a comparative study of unsupervised keyword extraction algorithms without using corpus. It compares the performance of Position Rank which considers the position of all word’s occurrences in the document with Text Rank and RAKE.The study's focus on comparing unsupervised keyword extraction algorithms without using a corpus may offer valuable insights into those specific algorithms, but it might not provide a comprehensive overview of key extraction methods.

Supervised or domain-specific techniques and the use of a corpus may yield different results, which are not covered in this study.

**SURVEY 6:** A first summarization system of a video in a target language”, International Conference on Multimedia and Network Information System in 2018.

The main goal of this work is to understand the content of a video in a foreign language. In this work, let consider the understanding process, such as the aptitude to capture the most important ideas contained in a media expressed in a foreign language. Mainly understanding media content in a foreign language is not only about language translation but also requires an understanding of cultural nuances, idioms, and context. The system struggle to capture these aspects accurately, potentially resulting in incomplete or inaccurate summaries.

**SURVEY 7:** Automated video summarization using speech transcripts, Storage and Retrieval for Media Databases in 2002

The paper focused on automated video summarization using speech transcripts. They proposed a method to automatically generate video summaries for long videos. video summarization approach involves mainly two tasks: segmenting the video into small, coherent segments, ranking the resulting segments.

The proposed algorithm scores segments based on word frequency analysis of speech transcripts. Then a summary is generated by selecting the segments with the highest score to duration ratios and these are concatenating them.

The quality of the video summarization heavily relies on the accuracy of the speech recognition system. The speech transcripts are not accurately transcribed, it can lead to inaccuracies in the summarization, potentially missing important content or including irrelevant segments.

**1.4. EXISTING SYSTE****M**

The existing system leverages a variety of powerful tools and technologies for video content processing, transcription, and keyword extraction, there is an opportunity to further enhance the project's capabilities with the integration of the Genism library.

It specializes in topic modelling and document similarity analysis, offering an innovative approach to understanding textual content. By implementing Genism alongside the current technology stack, we can unlock new dimensions of knowledge extraction and representation. Genism excels in Latent Semantic Analysis (LSA) and Latent Dirichlet Allocation (LDA), two popular techniques for uncovering latent topics in a collection of documents. By applying Genism’s topic modelling capabilities, we can categorize and organize textual data from video transcripts into coherent themes and topics, further facilitating content analysis. Genism also provides Word2Vec and Doc2Vec models for word and document embeddings, respectively.

This is particularly valuable for understanding semantic relationships between keywords and phrases, enabling a more nuanced and context-aware keyword extraction process. Several powerful tools are available for video transcription and speech-to-text services. Google's Speech-to-Text is a top solution known for its multilingual support. IBM's Speech to Text service offers customization for specific industries, delivering high accuracy and flexibility. Amazon Web Services provides Amazon Transcribe, a cloud-based tool for converting audio from video into machine-readable text, useful for various applications like keyword extraction .

**CHAPTER 2**

**2.1. PROPOSED SYSTEM**

In today's digital age, the consumption of video content is on the rise. From educational materials to entertainment, videos have become a primary medium for sharing information and stories. However, with this surge in video content, there arises a need for efficient ways to extract, process, and document valuable insights from these videos. This is where the convergence of cutting-edge technologies and tools, combined with the power of Natural Language Processing (NLP) techniques, plays a pivotal role.

This article delves into the technologies and tools that collectively enable the extraction, processing, and documentation of summaries and keywords from video content, with a specific focus on NLP techniques. These solutions have the potential to revolutionize the way we engage with video content, transforming it into a goldmine of information waiting to be explored.

The process begins with the crucial phase of video content preprocessing, where the first tool on our journey comes into play—Movie PY. Movie PY is a remarkable asset in this initial phase, as it facilitates the extraction of audio content from video files. By doing so, it allows for the separation of spoken words and soundtracks from the visual components of a video, thereby creating a distinct audio file. The importance of this cannot be overstated, as it paves the way for subsequent steps in the analysis.

Movie PY simplifies what can be a complex task, providing a user-friendly interface that allows users to specify the video file as input, after which it generates a separate audio file in their preferred format, be it MP3 or WAV. Beyond its user-friendly features, Movie PY also caters to those who require advanced customization options, granting them a high degree of control over video and audio manipulation. This initial phase sets the stage for a more in-depth analysis of the video's audio content, which is crucial for understanding and summarizing the spoken words within.

Transcriptions serving as the bridge that connects the audio content to the world of textual data. To convert spoken audio content into text, Assembly AI transcription services come to the forefront. These services are designed to provide an efficient and highly accurate solution for the task of transcribing spoken words and audio content into machine-readable text format. Such a conversion is not merely a formality but is fundamental for any subsequent NLP tasks, such as summarization and keyword extraction.

Assembly AI's prowess in this area ensures that the spoken content from videos becomes accessible, analysable, and most importantly meaningful. Without this transformation, the wealth of information within the video content would remain locked away, inaccessible to the advanced NLP techniques we intend to employ.

The next integral component of this journey is the use of NLP libraries, with spacy being a standout example. NLP libraries are the engines that drive the text analysis and keyword extraction processes. In the context of this project, NLP libraries play a critical role in analysing the textual content, particularly when it comes to extracting keywords and key phrases from the generated summaries. Spacy stands out as a modern and efficient NLP library designed for production use. Known for its speed and accuracy, spacy provides a user-friendly experience, making it accessible to a wide range of users. Moreover, it comes equipped with pre-trained models and data resources, simplifying many NLP tasks, including keyword extraction. These tools ensures that the analysis of video content is not just thorough but also highly efficient.

The culmination of this intricate process is the creation of structured documents. This final step brings together all the results of the various processes, including summaries and keywords, and compiles them into a format that is not just organized but also logical.

The structured document serves as a treasure trove of insights extracted from the video content, offering a comprehensive view of its content and key elements. It is this document that becomes a valuable resource for users seeking to glean knowledge from the video or for further analysis by researchers and experts in various fields.

**2.2. PROPOSED BLOCK DIAGRAM**

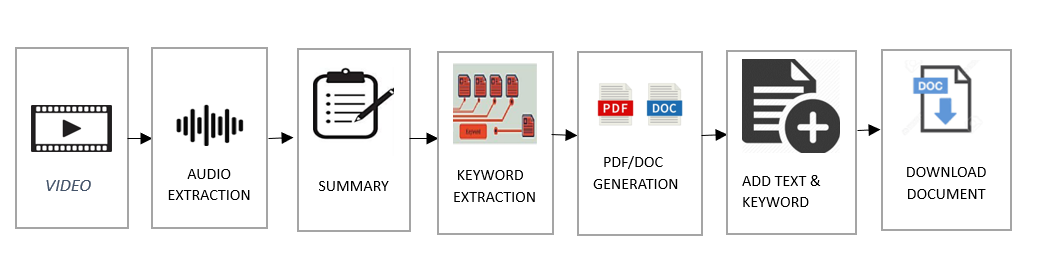


Fig.2.1 Proposed Block Diagram

In the process of harnessing insights from video content, a carefully orchestrated sequence of tools and technologies were employed. The video collection was facilitated by Movie PY, ensuring that a diverse range of videos were readily available for analysis. Movie PY was also instrumental in the extraction of audio from these videos, effectively separating spoken words and soundtracks from the visual components.

The critical bridge between audio and text was established using Assembly AI's audio-to-text conversion service, which transcribed the spoken content into a machine-readable text format. This transcription was pivotal in making the audio content accessible for further Natural Language Processing (NLP) tasks.

The NLP engine of choice, Spacy, played a key role in extracting meaningful keywords from the transcribed content. Its speed, accuracy, and user-friendly features made the process efficient and effective.

Lastly, the compiled insights, summaries, and keywords were organized into structured documents using FPDF, providing a valuable resource for knowledge seekers and researchers. This seamless integration of tools and technologies transformed video content into a rich source of information and knowledge

**2.3. LIST OF MODULES**

1. VIDEO COLLECTION
2. VIDEO TO AUDIO EXTRACTION
3. AUDIO TO TEXT CONVERSION
4. TEXT SUMMARIZATION
5. KEYWORD EXTRACTION
6. DOCUMENT COMPILATION

**2.4. MODULE DESCRIPTION**

**2.4.1. VIDEO COLLECTION**

The "Video Collection" module serves as the initial step in the data processing, responsible for aggregating video content from various sources. This module can employ multiple methods such as class recordings, YouTube videos, or manual input of video Paths. Its primary role is to accumulate a wide range of video content, making it accessible for further analysis and summarization within the system. Essentially, it acts as the system's data intake mechanism, ensuring a diverse pool of video content is available for subsequent processing.

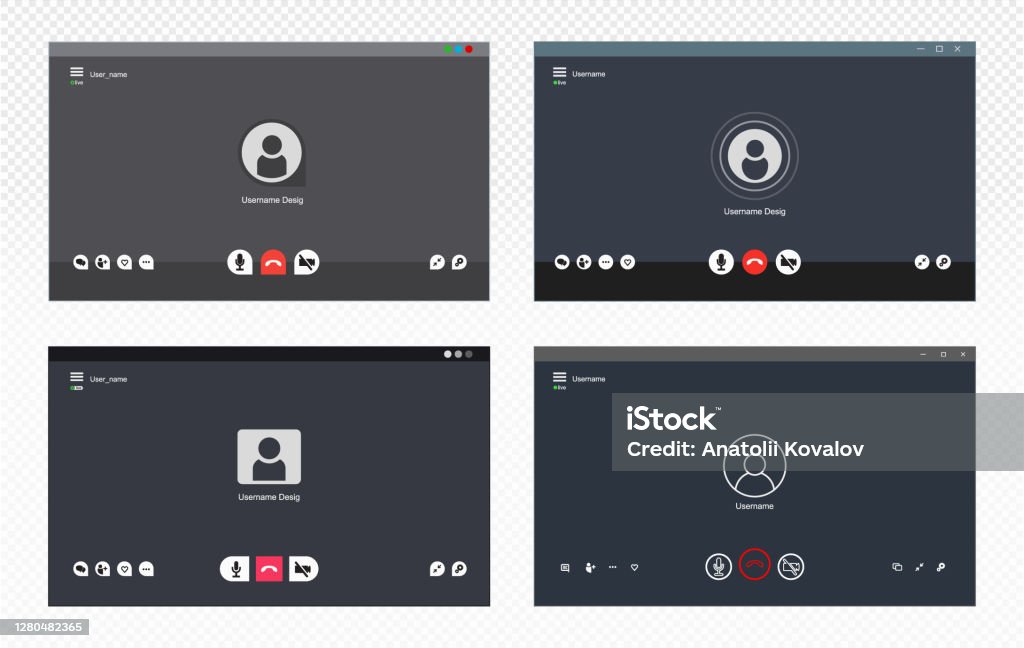


Fig 2.2. Obtaining Video Files

**2.4.2. VIDEO TO AUDIO EXTRACTION**

This module serves as the initial step in the process by focusing on the extraction of the audio track from video content. The conversion from audio to text is made possible through the utilization of multimedia processing libraries such as FFMPEG and movie PY. Specifically, within these libraries, movie PY is employed to extract the audio track from the video sources.

This extraction of audio content is a pivotal and necessary step in the data processing as it lays the groundwork for the subsequent conversion of audio data into textual content. By bridging the gap between visual and auditory content and the subsequent text-based processing stages, this module ensures that the extracted audio becomes a crucial component in the analysis and summarization of video content.

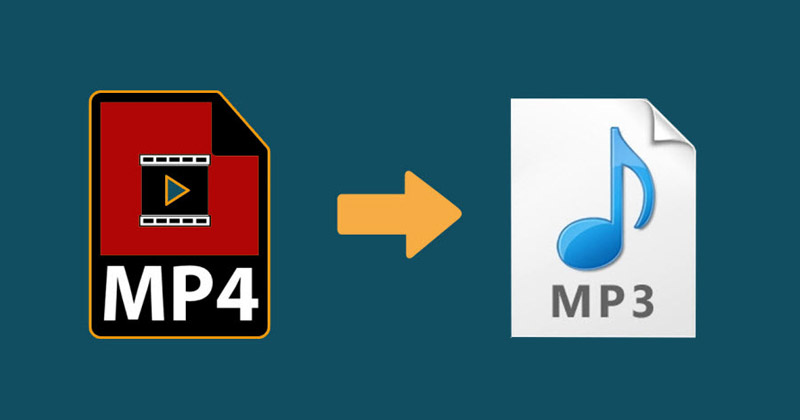


Fig 2.3. Transferring from mp4 to mp3

**2.4.3. AUDIO TO TEXT CONVERSION**

In this module, the extracted audio track is transformed into written text through the application of an Automatic Speech Recognition (ASR) algorithm. Assembly AI is a specialized expert in the field of ASR technology, and they play a pivotal role in ensuring the successful transcription of spoken language audio into written text. ASR technology is instrumental in various applications, including transcription services, voice assistants, and more.

Assembly Ai’s ASR services are designed to convert spoken words and phrases into textual data with remarkable accuracy. The converted text serves as the cornerstone for subsequent analysis and processing steps.

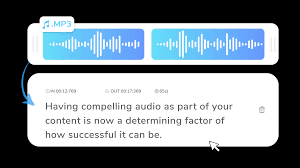


Fig .2. 4. Transcribing spoken words into textual form

**2.4.4. TEXT SUMMARIZATION**

"Text Summarization" is a module that transcribed textual content from videos into concise and manageable summaries. This process can be accomplished using a range of techniques, notably extractive and abstractive methods, each with its unique approach to summarizing text.

Extractive summarization identifies and extracts sentences or phrases directly from the original text, selecting them based on their relevance and importance.

Abstractive summarization, on the other hand, generates new sentences that capture the core ideas of the content in a more human-like manner. Various libraries and tools, Genism Transformers facilitate the summarization process. They help distil the content down to its most vital and informative elements. Text summarization is particularly valuable in scenarios where time and attention are limited, such as in news articles, research papers, and educational content.

**2.4.5. KEYWORD EXTRACTION**

The module plays a vital role in the text processing pipeline. Once text summaries are generated, this module is responsible for identifying and extracting keywords or key phrases from those contents.

To accomplish this, the module leverages Natural Language Processing (NLP) techniques. spacy is the powerful techniques available for the extraction of keywords from the summaries.

These extracted keywords serve as critical metadata for understanding the primary topics and subjects covered in the videos. They provide a quick and efficient way to grasp the essential focus of the content. By aiming to providing valuable insights into the subject matter of the videos. Keyword extraction enhances the accessibility and utility of the summarized content, making it easier for users to identify and navigate the most relevant information.

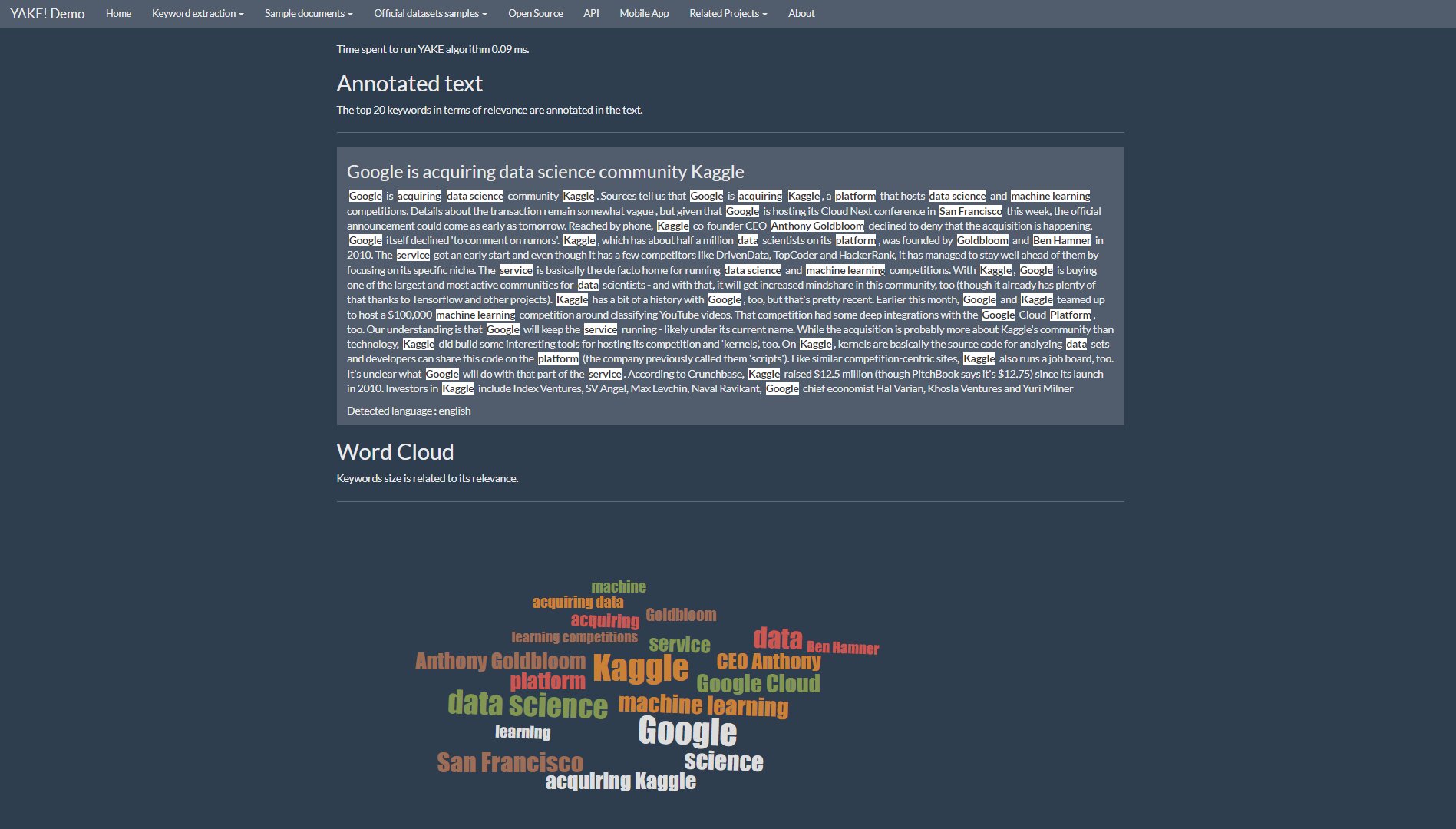


Fig .2.5. Selection of keywords



Fig.2.6. Keyword retrieval from textual content

**2.4.6. DOCUMENT COMPILATION**

This module serves as the concluding step in the data processing process. In this stage, the module combines the video summaries, which provide concise representations of the video content, with the extracted keywords, which encapsulate the central themes and topics, into a structured document.

This document can take on various formats, including PDFs, Word documents, depending on the specific needs and preferences of the system or users. The primary goal of this module is to present the summarized content in an organized and readable format, with the associated keywords providing a helpful reference. By structuring the information in this way, the module facilitates easy access and navigation, allowing users to quickly locate and understand the essential information contained in the video contents.



Fig.2.7. Document creation with the text and keywords

**2.5. HARDWARE/SOFTWARE**

**Hardware:**

Computing Equipment: High-performance computers with multi-core processors and ample RAM are essential for handling the intensive processing of video content. The ability to swiftly transcode and process video files is greatly influenced by the processing power of the hardware.

Storage Solutions: Large-scale video collections require substantial storage capacity. High-capacity hard drives or cloud storage solutions are vital for storing video files, as well as the derived audio and textual data.

Audio Equipment: High-quality microphones and audio recording equipment are crucial for capturing clear and accurate audio from videos, especially when the audio quality in the source video is not optimal.

**Software:**

Movie PY: This software tool, built on top of the MoviePy library, enables video collection, video-to-audio extraction, and assists in separating audio from visuals. It simplifies the preprocessing of video content.

Assembly AI: This software service offers an audio-to-text conversion solution, effectively transcribing spoken content into machine-readable text format. It's an indispensable bridge between audio and text, pivotal for Natural Language Processing (NLP) tasks.

Spacy: A powerful NLP library, Spacy aids in extracting meaningful keywords and key phrases from the transcribed content. Its speed, accuracy, and user-friendly features make it an efficient choice for NLP-based analysis.

FPDF: A software tool for creating structured documents, FPDF is used to compile and organize insights, summaries, and keywords into user-friendly documents. It transforms the extracted data into a valuable resource for knowledge seekers and researchers.

**2.5. RESULTS / SCREENSHOT**

**RESULTS**

The section presents the results of our research on the application of NLP for generating computer-generated summaries and keywords from video content. We delve into the key findings and engage in a comprehensive discussion regarding the implications and prospects of technology.

In the pursuit of extracting audio from video, we delved into several libraries, exploring the options available to achieve this crucial step. While there are several libraries to consider, including PyDub, FFMPEG, and OpenCV, our ultimate choice for this task was the versatile Movie PY library Movie PY, with its user-friendly interface and robust capabilities, proved to be the standout choice for extracting audio from our video content.

In summarization techniques, our journey took us on a path that began with audio-to-text extraction. Along this trail, we explored various libraries and APIs, giants like Google, IBM, and AWS among them, each offering their unique merits and capabilities.However, which truly captured our attention and earned the starring role in our project was Assembly AI.

Its precision and performance were crucial in helping us achieve our project's goals with finesse and effectiveness. In Assembly AI, we found the key to unlocking the power of audio-to-text summarization, enabling us to craft compelling narratives and insights from our video content.

Keyword Extraction Algorithms: Various algorithms were tested for keyword extraction, with TF-IDF, keyword frequency analysis, and Named Entity Recognition showing promising results. Indeed, Spacy is known for providing accurate keywords and demonstrating efficient performance, making it an excellent choice for our project's specific requirements.

**Discussion**

When it came to extracting audio from video content, our team conducted a thorough review of available libraries. PyDub, FFMPEG, and OpenCV were all considered, but Movie PY ultimately won us over. Even though, PyDub is a popular choice for audio extraction but for more complex audio extraction needs, it might fall short in terms of features and flexibility.

FFMPEG is a powerful multimedia framework, but it can be complex to work with, especially for users who are not familiar with its command-line interface.

OpenCV is primarily designed for computer vision tasks and image processing. While it can handle video extraction, it may not be as intuitive and straightforward for audio extraction as specialized libraries like PyDub and Movie PY.Considering the disadvantages of the above-mentioned libraries, Movie PY's user-friendly interface, extensive functionality and comprehensive documentation make it an excellent choice for audio extraction.

It simplifies the process, ensuring that even users with limited multimedia experience can perform the task efficiently. Additionally, Movie PY's integration with Python is an advantage for those working within the Python ecosystem.

In summarization techniques, our journey began with the critical step of audio-to-text extraction. We diligently explored various libraries and APIs, with industry giants like Google, IBM, and AWS. While these are reputable options, they often come with complex pricing structures and may be expensive for projects with a large volume of audio-to-text conversions.

Additionally, the private API key feature, which enhances data security, might not be straightforward. Here comes the Assembly AI with its user-friendly interface and streamlined integration, making it accessible to a broader range of users, even those without extensive technical expertise. Assembly AI's private API key is a significant advantage. It ensures data security and privacy, which is crucial when handling sensitive or proprietary content.

Our project also involved keyword extraction, a vital component in understanding and categorizing video content. Various algorithms were put to the test, including TF-IDF, keyword frequency analysis, and Named Entity Recognition.

TF-IDF is a method for keyword extraction, it might not capture the context and relationships between words effectively.

Keyword Frequency Analysis limited in its ability to discern the importance of specific words or phrases within a document. It may result in a list of frequent words, but it doesn't necessarily provide meaningful insights into the content. NER is excellent for identifying proper nouns and entities in text but might not be as effective in extracting general keywords or terms.

It focuses on named entities like names of people, organizations, or locations. Whereas in spacy known for its accuracy in keyword extraction. It employs advanced natural language processing techniques to identify and extract relevant keywords, considering both frequency and context. This leads to more precise results that are better aligned with the content's meaning.

**SCREENSHOT**

**A screenshot of a video

Description automatically generated**

Fig.2.8. Collecting video file

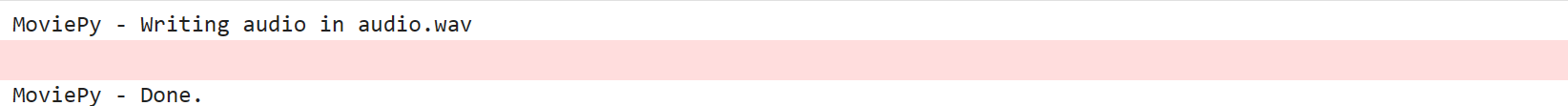


Fig.2. 9. Video to audio conversion

A close up of a text

Description automatically generated

Fig.2.10. Summary generation

A close-up of a computer code

Description automatically generated

Fig.2.11. Keyword extraction

A close-up of a computer screen

Description automatically generated

Fig.2.12. Document conversion

A screenshot of a computer

Description automatically generated

Fig.2.13. Inside the document

**CHAPTER 3**

**CONCLUSION**

**3.1. CONCLUSION**

In our exploration of audio extraction, summarization techniques, and keyword extraction, the significance of technology and tool selection in the effective processing of video content has become abundantly clear. These steps are pivotal in unlocking the wealth of information concealed within videos, and the choices we make in terms of tools and technologies can profoundly influence the outcomes.

For audio extraction, we evaluated multiple libraries, including PyDub, FFMPEG, and OpenCV. Movie PY emerged as the superior choice due to its user-friendly interface, extensive functionality, and comprehensive documentation. Its seamless integration with Python adds a layer of convenience for users within the Python ecosystem, making it accessible to those with varying levels of multimedia expertise.

Summarization techniques led us to explore various libraries and APIs for audio-to-text conversion. While industry giants like Google, IBM, and AWS offer reputable options, their complex pricing structures can be prohibitive for projects with substantial audio-to-text needs. Assembly AI shone as a frontrunner with its user-friendly interface and streamlined integration, making it accessible to a broader range of users, even those lacking technical expertise. The private API key feature ensured data security, a critical concern when handling sensitive or proprietary content.

In the realm of keyword extraction, we rigorously tested various algorithms, including TF-IDF, keyword frequency analysis, Named Entity Recognition (NER), and spacy. Spacy emerged as a standout for its precision in keyword extraction, effectively considering both frequency and context to provide more insightful and meaningful keywords.

This journey underscores the importance of selecting the right tools for each phase of video content processing. Movie PY simplifies audio extraction, Assembly AI streamlines audio-to-text conversion, and spacy elevates keyword extraction.

These tools not only simplify the intricate process of video content analysis but also make it more accessible to a wider range of users, regardless of their technical background. In the ever-evolving landscape of multimedia content, these tools serve as essential companions for those seeking to harness the full potential of video data.

**3.2. FUTURE ENHANCEMENT**

Real-time Processing: Developing real-time processing capabilities for video content would be a significant advancement. This would enable live transcription and analysis of video streams, making it valuable for applications like live broadcasts, webinars, and conferences.

Customizable Summarization: Future enhancements could include the development of customizable summarization techniques. Users could have the ability to define specific criteria for generating summaries, such as the length of the summary, the level of detail, or the inclusion of key points.

Integration with Machine Learning: Leveraging machine learning models for more advanced keyword extraction and content analysis could significantly improve the accuracy and relevance of the extracted keywords and summaries. This would involve training models on a diverse dataset of video content to enhance understanding and context-based extraction.

These enhancements would not only advance the state of the art in video content processing but also further democratize the field, making it more accessible and valuable to a broader range of users. As technology continues to evolve, these improvements will play a pivotal role in unlocking even more insights and knowledge from the ever-growing pool of video content available.

**REFERENCES**

1. Emad A, Bassel F, Refaat M, Abdelhamed M, Shorim N and AbdelRaouf A ( 2021, January) “Automatic Video summarization with Timestamps using natural language processing text fusion”, IEEE 11th Annual Computing and Communication Workshop and Conference (CCWC), pp. 0060-0066 .
2. Apostolidis Evlampios et al., (2021) “Video Summarization Using Deep Neural Networks: A Survey”, arXiv preprint arXiv:2101.06072.
3. Sanjana et al. R, (2021) “Video Summarization using NLP”, International Research Journal of Engineering and Technology (IRJET).
4. Boorugu R and Ramesh G (2020, Jul) “A survey on NLP based text summarization for summarizing product reviews,” in Proceedings of the 2nd International Conference on Inventive Research in Computing Applications, ICIRCA 2020 , pp. 352–356, Doi: 10.1109/ICIRCA48905.2020.9183355.
5. PRIYANKA G and PRASHA MEENA M (2020) “Survey and Evaluation on Video Summarization Techniques”, Journal of Critical Reviews 7.8.
6. Holly Smaïli, Kamel et al. (2018) “A first summarization system of a video in a target language”, International Conference on Multimedia and Network Information System.
7. Jayanta Basak, Varun Luthra and Santanu Chaudhury (2008) “Video Summarization with Supervised Learning”.
8. ZHU Wei REN Yusheng, (2008) “A Video Summarization Approach based on Machine Learning”, International Conference on Intelligent Information Hiding and Multimedia Signal Processing.
9. Takira, M. Cuneyt et al., (2008) "Automated video summarization using speech transcripts", Storage and Retrieval for Media Databases 2002, vol. 4676.
10. Holly Smaïli, Kamel et al. (2018) “A first summarization system of a video in a target language”, International Conference on Multimedia and Network Information System.
11. Moses T M and Balachandran K (2017) “A classified study on semantic analysis of video summarization”, 2017 International Conference on Algorithms Methodology Models and Applications in Emerging Technologies (ICAMMAET), pp. 1-6.
12. Sah et al S (2017) “Semantic Text Summarization of Long Videos”, IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 989-997.
13. Albawi S, Mohammed T A and Al-Zawi S (2017) “Understanding of a convolutional neural network”, International Conference on Engineering and Technology (ICET), pp. 1-6.
14. Qin Jin and Junwei Liang (2016) “Video Description Generation Using Audio and Visual Cues”, Proceedings of the ACM on International Conference on Multimedia Retrieval. ICMR '16, pp. 239-242, ISBN 9781450343596.
15. Jayanta Basak, Varun Luthra and Santanu Chaudhury (2008) “Video Summarization with Supervised Learning”.
16. ZHU Wei REN Yusheng, (2008) “A Video Summarization Approach based on Machine Learning”, International Conference on Intelligent Information Hiding and Multimedia Signal Processing.
17. Takira, M. Cuneyt et al., (2008) "Automated video summarization using speech transcripts", Storage and Retrieval for Media Databases 2002, vol. 4676.
18. Guang Li, Shubo Ma and Yahong Han (2015) “Summarization-Based Video Caption via Deep Neural Networks”, Proceedings of the 23rd ACM International Conference on Multimedia. MM '15, pp. 1191-1194, ISBN 9781450334594.
19. Vasconcelos N and Lippman A (1998) “Bayesian modelling of video editing and structure: Semantic features for video summarization and browsing,” Proceedings of IEEE International Conference on Image Processing, Chicago, IL. 28
20. Jing H, Barzilay R, McKeown K, and Elhadad M (1998) “Summarization evaluation methods: Experiments and analysis,” Proceedings of the AAAI Symposium on Intelligent Summarization, March 23 - 25 , Palo Alto,CA.
21. <http://www.java2s.com/example/android/android.media/get-real-video-path-from-uri.html>
22. <https://elgg.org/discussion/view/2705321/how-can-i-get-the-real-video-file-path>
23. <https://biteable.com/tools/extract-audio-from-video/>
24. <https://www.steadylearner.com/blog/how-to-extract-an-audio-file-from-a-video-with-python-3pja>
25. <https://towardsdatascience.com/extracting-audio-from-video-using-python-58856a940fd>
26. <https://copyassignment.com/extract-audio-from-video-using-python/>
27. <https://www.assemblyai.com/blog/conformer2/?utm_source=google&utm_medium=cpc&utm_campaign=retargeting_search_speechtotext&utm_term=voice%20to%20text&gclid=Cj0KCQiAuqKqBhDxARIsAFZELmI7hRsiBOsDUKFfZa7dqXDFzUfYeYGB6TamcsS-qZGp29N-fV8qdfgaAiGHEALw_wcB>
28. <https://techchannel.com/SMB/01/2022/speech-to-text-python#:~:text=Figure %201 %20Code%3A,Microphone()%20as%20UserVoiceInputSource%3A%20UserVoiceRecognizer>.
29. <https://www.analyticsvidhya.com/blog/2022/01/speech-to-text-conversion-in-python-a-step-by-step-tutorial/>
30. <https://www.geeksforgeeks.org/python-convert-speech-to-text-and-text-to-speech/>
31. <https://www.analyticsvidhya.com/blog/2022/01/four-of-the-easiest-and-most-effective-methods-of-keyword-extraction-from-a-single-text-using-python/>
32. <https://towardsdatascience.com/keyword-extraction-process-in-python-with-natural-language-processing-nlp-d769a9069d5c>
33. <https://www.analyticsvidhya.com/blog/2022/03/keyword-extraction-methods-from-documents-in-nlp/>
34. <https://monkeylearn.com/keyword-extraction/>
35. <https://www.geeksforgeeks.org/keyphrase-extraction-in-nlp/>
36. <https://www.geeksforgeeks.org/python-program-to-extract-keywords-from-a-list/>
37. <https://testbook.com/python-programming/how-to-automatically-extract-keywords-from-sentences-in-python>
38. <https://www.johnsnowlabs.com/the-experts-guide-to-keyword-extraction-from-texts-with-spark-nlp-and-python/>
39. <https://www.kaggle.com/code/akhatova/extract-keywords>
40. <https://www.vennify.ai/keybert-keyword-extraction/>

**APPENDIX**

from tkinter import \*

from moviepy.editor import VideoFileClip

from moviepy.editor import AudioFileClip

from tkinter import filedialog

from tkinter import messagebox

from tkinter import ttk

from fpdf import FPDF

import assemblyai as aai

import threading

import speech\_recognition as sr

import spacy

from collections import Counter

from string import punctuation

import os

from docx import Document

#variables

video\_clip = ''

audio\_clip = ''

#api key

aai.settings.api\_key = f"21f1e0326c544c6081112fb5d6091b89"

#function to get video

def get\_video():

global video\_filepath, video\_clip

try:

video\_filepath.set(filedialog.askopenfilename(title="Select your video file"))

video\_clip = VideoFileClip(str(video\_filepath.get()))

except:

messagebox.showerror("Error", "No video selected")

#function to convert audio to pdf

def audio\_to\_pdf():

global audio\_clip

try :

#extract audio

audio\_clip = video\_clip.audio.write\_audiofile(r"audio.wav")

transcriber = aai.Transcriber()

transcript = transcriber.transcribe("audio.wav")

t=transcript.text

print(t)

# Load the English language model

# Load the English language model

nlp = spacy.load("en\_core\_web\_sm")

# Define the list of part-of-speech tags you want to include

pos\_tag = ['PROPN', 'NOUN']

# Process the text with spaCy

doc = nlp(t.lower())

# Initialize an empty list to store the keywords

re = []

# Iterate through tokens in the processed text

for token in doc:

# Check if the token is not a stop word and not in punctuation

if token.text not in nlp.Defaults.stop\_words and token.text not in punctuation:

# Check if the token's part-of-speech tag is in the specified list

if token.pos\_ in pos\_tag:

re.append(token.text)

o = ' - '.join(set(re))

print(o)

messagebox.showinfo("Message", "Conversion Successfull")

except :

messagebox.showerror( "Error", "Conversion not performed")

d = Document()

# Add content to the document

d.add\_paragraph(t)

d.add\_paragraph(o)

# Save the document as a .docx file

d.save('document.docx')

#function to run the script

def run():

global progress\_bar

t1 = threading.Thread(target = progress\_bar.start)

t2 = threading.Thread(target = audio\_to\_pdf)

t2.start()

t1.start()

# GUI CODE starts

# Intializing main program settings

root = Tk()

root.title("Summaries Converter")

# Variables for file paths

video\_filepath = StringVar()

# Creating UI Frame

UI\_frame = Frame(root, width=500, height=500, relief = "raised")

UI\_frame.grid(row=0, column=0)

convert\_frame = Frame(root, width=500, height=500, relief="raised")

convert\_frame.grid(row=1, column=0)

# Labels and buttons

select = Label(UI\_frame, text="Select Video : ", font = ("Arial", 12))

select.grid(row=1, column=1, padx=5, pady=5, sticky=W)

browse = Button(UI\_frame, text="Browse", command = get\_video, font = ("Arial", 12))

browse.grid(row=1, column=2, padx=5, pady=5)

video\_selected = Label(UI\_frame, text = "Selected video : ", font = ("Arial", 12))

video\_selected.grid(row = 2, column = 1, padx = 5, pady = 5, sticky = E)

video\_path = Label(UI\_frame, textvariable=video\_filepath)

video\_path.grid(row=2, column=2, padx=2, pady=5, sticky=W)

convert = Button(convert\_frame, text="Convert", command = run, font = ("Arial", 12))

convert.grid(row=3, column=1, pady=5)

progress\_bar = ttk.Progressbar(root, orient=HORIZONTAL, mode='indeterminate', length=500)

progress\_bar.grid(padx=25, pady=25)

# Calling main program

root.mainloop()