International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-8 Issue-6, March 2020

# Automatic Annotation of Instructional Videos

**D. Bhavitha, A. Kalaivani**

***Abstract:*** ***Multimedia has a significant role in communicating the information and a large amount of multimedia repositories make the browsing, retrieval and delivery of video contents. For higher education, using video as a tool for learning and teaching through multimedia application is a considerable promise. To extract the audio information from the visual content is a very challenging because it needs the extraction of high level semantic information from low level visual data. The summarization technique for videos has been proposed to improve the browsing faster for large video collections to produce more efficient content indexing and access. The proposed video summarization system for instructional videos initially separates videos into audio track and then converted into text transcript. The text transcripts are pre-processed and extracted the important relevant keywords which can be used for indexing the video file. The automatic annotation of instructional videos reduced the summary length and also preserved the semantic of the videos. The research indicates that the students found these summarized content is very helpful for preparing and reviewing the lecture material and also for the preparation during their examination.***

***Keywords: Automatic Summarization, Instructional Videos, Summary Length, Text Transcript.***

## I. INTRODUCTION

The students who are doing their graduation will spend most of their time at their own computers, far away from the college campus, particularly students who are studying the computer subjects and related subjects. The Students from the Higher Education level should be explored more about the subjects they are learning as an addition to the subject which was given by a faculty. This active learning and studying style, and access to a different kinds of study materials, is essential for higher graduate students to succeed in their studies at college. In addition, pressure of working and travelling mean that there is no longer time to spend all the day in the campus. Support for the students learning is being provided, where faculty can upload different study related materials and documents for students to use whenever and where ever they wanted to study.

Manuscript received on February 10, 2020.
Revised Manuscript received on February 20, 2020.
Manuscript published on March 30, 2020.
* Correspondence Author
**D. Bhavitha*,** Ssaveetha School of Engineering, Saveetha Institute of Medical Sciences, Chennai, India.
**Dr. A. Kalaivani,** Associate Professor, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai, India.



We have investigated the use where we can extract information from the video to increase the student learning, and to produce the short summary notes online, as preparation notes for our students during their exams. We are looking into a new technique where the faculties prepares a 15-20 minute video which contains the summary of the main points from the lecture video and the text will generate automatically for the recognised audio from the video.

After converting audio into text we will summarize the text so that students can easily verify the notes. Text summarization can be categorized into two main parts— Extractive and Abstractive Summarization.

Extractive Summarization method depends on extracting various parts of information from the text given, like sentences, noun phrases, and words, from a bit of text and bind them together to create and generate as summary. Hence, identifying those correct and exact sentences for summarization is very important in an extractive summarization method.

Abstractive Summarization uses an advanced and updated NLP technique to produce a completely new text summary. Few sections of this summarization will not appear in the original text which we have in the text file. In our proposed system we are focusing on the extractive summarization technique because it extracts the text which is present in the main document only but abstractive summarization will generate entirely new summarization.

The rest of the paper may be prepared as follows. In section 2 affords literature survey on video summarization. Section 3 explains the block diagram of the proposed system and section 4 describes the methodology of the proposed system. Finally section 5 highlights the outcomes at each and every section of the automatic video summarization of the instructional videos. Finally, section 6 gives the conclusions of the proposed work in the form of graphs that specifies the emotion levels of the tweets.

## II. LITERATURE SURVEY

Different authors have executed their research work inside the subject of Data pre-processing on one kind of domain names and proved suitable efficiency and accuracy in getting in getting rid of noisy information with the aid of distinctive strategies. Some of the authors specially concentrated on stop phrase removal for achieving higher accuracy. Martin Malchow, Matthias Bauer, Christoph Meinel[1], introduced a video annotation approach in a HTML video player.

In comparison to other video annotation approaches, no additional software is needed and the annotation feature is available in the player and not on the website below the player. This improves the usability and supports the patience of the users. Besides annotations, the player also offers annotations marks. Annotations marks offer an easy and fast way to mark crucial parts that may be relevant for exams. The user has the opportunity to find a part of interest easily when a video is watched again.

Hence the video player should be integrated with the annotation function. Then, looping to a different annotation area on the website is not compulsory. Therefore, the annotated material can be transferred together with the moving slide content to a word document, and we can easily print the document. The annotated videos support and drive the students to watch and learn from the videos by an E-Learning archived video. C. Choudary and T. Liu[6], offered a robust technique to appropriately extract the textual content and images from instructional videos. They investigated that the delivery of content material pixels in instructional videos developes a statistical version for classifying the content material pixels. The given method first calculates the common and variance of the board coloration and the luminance range across the video time. Then for every sampled video frame, it should accurately divides the irrelevant regions and board regions by the way of merging photograph regions by the use of a probabilistic classifier. Finally, it extracts the content material pixels by using combining morphological processing along side a gradient-primarily based adaptive thresholding technique. The experimented consequences have proved that their algorithm achieved excessive performance on the 4 tested full duration instructional motion pictures. The implemented technique can be at once used for reinforcing textual content quality in instructional videos and for recognising the handwriting Therefore, the content extraction outcomes may aid the evaluation of video scenes and the grouping of lecture topics in instructional videos.

Lu, S., Lyu, M. R., King, I., & Sar, H. K. [8], proposed a video summarisation framework technique, for improving the skimmings in video which guarantees the visual coherence and balanced content. Firstly, they used the semi automated annotation tool for video to collect the semantic information from the video. Then, they have analyzed the video structure and determines the picture scene's target skim length from each video. Then, according to their semantic descriptions they have used mutual reinforcement principle to calculate the value of relative importance and to cluster the video shots. At last, they analyzed that the arranged pattern for the video and key shots are taken out to form the video for performing skimming for the final video, where the importance of video shot value is used as a guidance. Experiments are directed to evaluate the efficiency of their givrn approach.

Janice Whatley and Amrey Ahmad[7], proposed a video as a tool for learning and teaching during the graduation is a multimedia(video) application with a considerable promise. They Compromised video within the online documented materials for each and every module which helps students to understand the notes and to review for the assessment. They have investigated using short length videos which summarizes the lectures given in the video, as a material for students which helps them for preparation. They adopted an interpretive method which investigates how useful are these videos, throught the teaching term and during the preparation for examination. They concluded that the basic research tells that students found these summarised lectures are very useful for prepation during their assesment.

Chekuri Choudary[5], investigated a new robustic approach to summarize the visual content in instructional videos. They first provided a model which extracts the hand written content on the black board and then they computes the key frames of the videos by locating the local maxima in content fluctuation curves with a shifting window algorithm.They have Further introduced techniques based on Hausdorff-distance and connected-component-decomposition to reduce content redundancy of the key frames by matching the content and mosaicking the frames. The investigation on all the instructional videos are recorded in college classrooms which shows that their algorithm is very effective for extraction and summarization of the visual content. The comparison of our algorithm with three existing key frame selection methods demonstrated the superior performance of our method. Future work includes the analysis of the events in instructional videos and the development of indexing and retrieval systems based on the content summarization results.

## III. PROPOSED SYSTEM

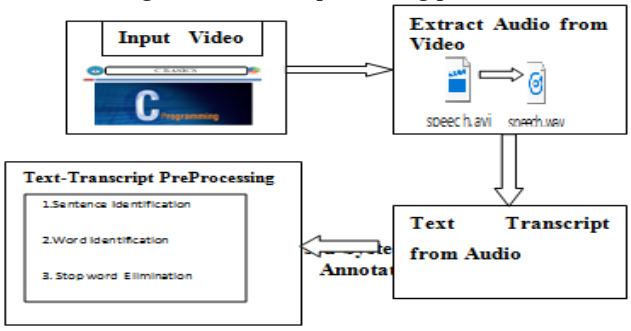
The proposed system mainly focuses on instructional videos used for educational purposes. We are looking into a new technique where the faculties prepares a 15-20 minute video which contains the summary of the main points from the lecture video and the text will generate automatically for the recognised audio from the video. After converting audio into text we will summarize the text so that students can easily verify the notes. The summarization has been done in various steps : pre-processing, feature extraction, sentence ranking, redundancy reducing and generated summary

We are exploring the use which enhances the student learning by extractiong information that is made out of video and providing a short online summarized lectures , as revision aids for our students. The major goal is that these summarized notes should be useful for students for preparation during their examination. This highly improves productivity as it speeds up the surfing process instead of reading full document which may contain useless information. There is a huge need to decrease the much of content from large text data, focused summaries captures the important information details, so that we can navigate it more efficiently and effectively, so that we can check whether the large documents also contains the information which we are looking for.

It makes the selection process easier, improves the effectiveness of indexing.

The structural depiction of the real input text is pre-processing. The significance of pre-processing is almost used in each and every developed system connected with natural language processing and also with text processing. This phase performs the operations like words identification, sentences identification, stop words elimination, language stemmer for nouns and proper names.

It allows the original input text in a format and eliminates the duplicate sentences or words from the given text. Reducing the size the original text is Pre-processing phase.




### A. Extract Audio from Video

In this paper we are mainly focusing on the presentational videos so we have extracted the audio content from the video. First we have to convert the mp4 formatted video into avi format. Then from that converted video file we have to extract the audio. The extracted audio content should be in wav format.

### B. Extracting Text from Audio

Here, we have extracted the text from the Audio file which we have already taken from the video file. We can easily extract the text from audio by using speech recognition module. This module first reads the audio file as source using the recognizer then prints the text from the audio.

### C. Pre – Processing

The structural representation of the original input text is applied to pre-processing module. The significance of pre-processing is almost used in each and every developed system connected with natural language processing and also with text processing. This phase performs the operations like words identification, sentences identification, stop words elimination, language stemmer for nouns and proper names. It allows the original input text in a format and eliminates the duplicate sentences or words from the given text. Reducing the size the original text is Pre-processing phase.

Words identification is splitting the sentences into words. The utmost method to extract aword from sentence is whenever a empty space or comma is recognised.

***Example :***

*Sentence – conversion of audio to text*
*Words – conversion, of, audio, to, text.*

Sentences identification is splitting the paragraph into sentences. The utmost method of extracting a sentence from the paragraph is whenever a period appears.

***Example :***

Paragraph – Recently I have bought a car from your company. The car was really in very good condition and I liked the design of your car. Two days back I met with an accident and the car was damaged very badly. I have registered a complaint two days back, but no one turned up and checked the problem.

*Sentences --*

*i) Recently I have bought a car from your company.*
*ii) The car was really in very good condition and I liked the design of your car.*
*iii) Two days back I met with an accident and the car was damaged very badly .*
*iv) I have registered a complaint two days back, but no one turned up and checked the problem.*

Stop word are words which are filtered out before processing of natural language data(text). It removes the most common words from a query in order to improve performance. Language stemmer is the process of reducing inflected words to their stem or root from a generally written word format.

Sentence which is having an entity name or a proper noun signifies as an important sentence and proper Noun are generated from this.

*Example :*
*Sentence – Bhavitha lives in Bangalore.*
*Proper Noun – Bhavitha , Bangalore*

## IV. RESULTS AND DISCUSSION

Conversion of video format is very easy. First we have to take the mp4 format video and convert it into avi format. As we are focusing on presentational videos we have to take the audio from the video. We can easily extract the audio when the video is in avi format. Then we have to extract the audio.

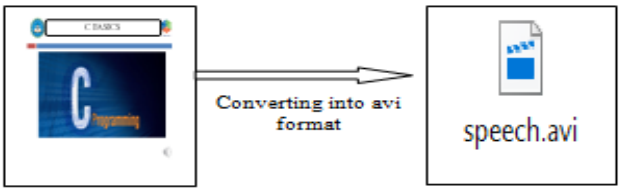



**Figure 2 : Conversion from mp4 video format to avi format**

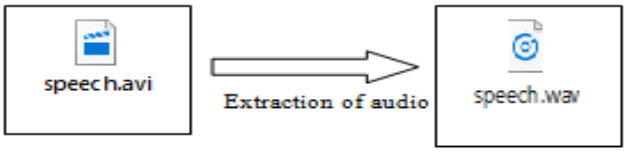



**Figure 3 : Extraction of audio file from video file**

welcome to Saveetha University lecture cast on data reduction we all know very well that computer is an electronic device which includes both hardware and software hardware is a peripheral devices software comprises of data and programs program
is a set of instruction was just returned to do a specific task data data nothing but a distant piece of information that has been given to carry on the process data from text includes uppercase and lowercase letters numbers and taken from 0 to 9 bits is a combination of binary digits it takes values of 0.15 the combinations of data of text on numbers a bi LED Technology advances different data places are being used in the information community big data at the volume of birth structured and unstructured data in was terabytes of data Big Data Analytics at the process of collecting organising and analysing a big data in a Discover Patan some useful information Data integrity it's used to check the validity of the data validity can be checked in number of please identify errors are the data entry point for identifying errors are the data transmission thank you students by patient listening will meet you in the next class

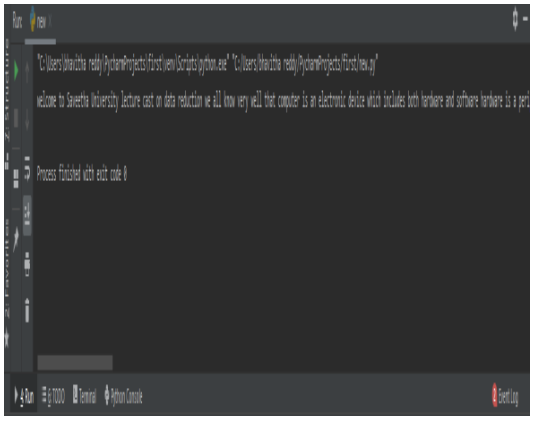


**Figure 4 : Extraction of text transcript from the audio file**

The PunktSentenceTokenizer module is used to identify the sentences from the paragraphs.

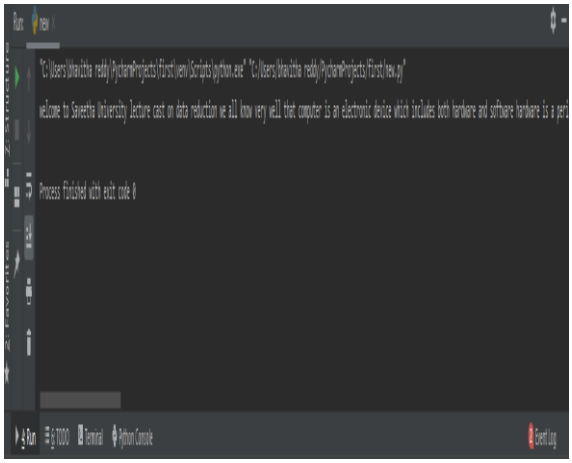


**Figure 5 : Identification of sentences from the extracted text**

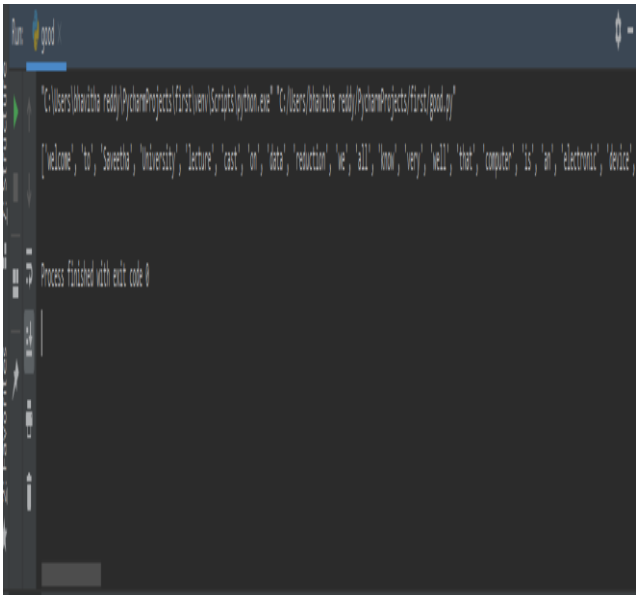


**Figure 6: Identification of words from the sentence identified**

The idea of sorting the text to normalizing method is called as stemming. Different forma of words will have the same meaning, when tense is involved in the text. This is the reason why we stem and shorten the sentences.

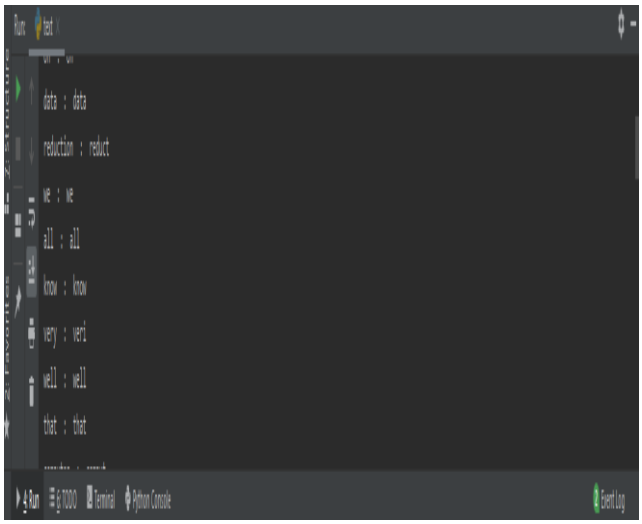


**Figure 7: Identification of stemming words from the extracted text**

Stopwords are words that do not contribute to the meaning of a sentence. Hence, they can safely be removed without causing any change in the meaning of the sentence. The NLTK library has a set of stopwords and we can use these to remove stopwords from our text and return a list of word tokens.

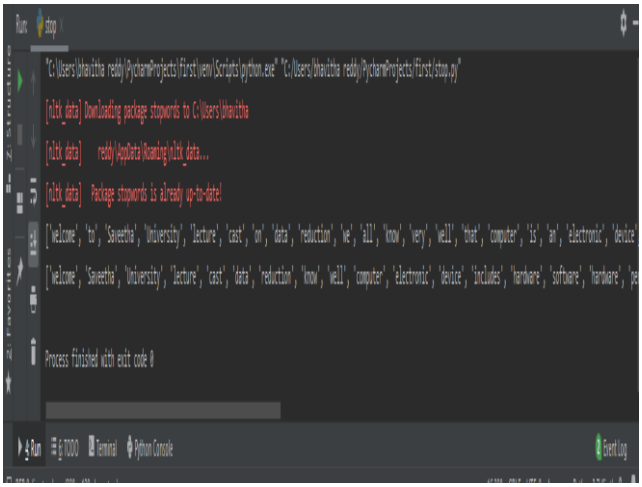


**Figure 8: Elimination of stop words from the extracted text**

Chunking all proper nouns (tagged with NNP) is a very simple way to perform named entity extraction. A simple grammar that combines all proper nouns into a NAME chunk can be created using the RegexpParser class.

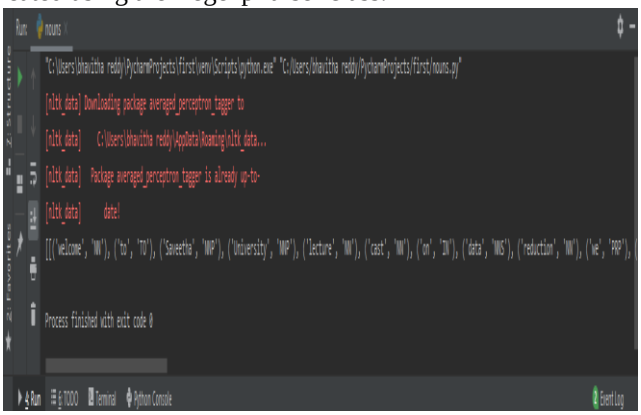


**Figure 9: Identification of proper nouns from the extracted text**

## V. CONCLUSION

In this paper, we investigated a new method to summarize the audio content from the instructional videos. We have first provided a method to extract the audio content into text format, then we applied extractive techniques to summarize the transcript text. Our future scope is to provide a summarized text for the students which helps them to revise easily during their examination. In future we are going to extract the features and give sentence ranking and will generate a summarized text.

## REFERENCES

1. Martin Malchow, Matthias Bauer, Christoph Meinel, "Enhance Learning in a Video Lecture Archive with Annotations" , IEEE Global Engineering Education Conference (EDUCON), 17-20 April, 2018, Santa Cruz de Tenerife, Canary Islands, Spain. 2018
2. Shashi Narayan , Shay B. Cohen, Mirella Lapata "Ranking Sentences for Extractive Summarization with Reinforcement Learning ", Proceedings of NAACL-HLT 2018, pages 1747–1759 New Orleans, Louisiana, June 1 - 6, 2018.
3. Laxmi B. Rananavare, P. Venkata Subba Reddy, "An Overview of Text Summarization ", International Journal of Computer Applications (0975 - 8887) Volume 171 - No.10, September 2017.
4. Hitoshi Nishikawa , Tsutomu Hirao, "Text Summarization Model based on RedundancyConstrained Knapsack Problem " , Proceedings of COLING 2012: Posters, pages 893–902, COLING 2012, Mumbai, December 2012.
5. Chekuri Choudary, " Summarization of Visual Content in Instructional Videos ", IEEE TRANSACTIONS ON MULTIMEDIA, 2008
6. C. Choudary and T. Liu, "Extracting content from instructional videos by statistical modeling and classification," Int. J. Pattern Anal. Applicat., vol. 10, no. 2, pp. 69–81, 2007
7. Janice Whatley and Amrey Ahmad, "Using Video to Record Summary Lectures to Aid Students' Revision" Interdisciplinary Journal of Knowledge and Learning Objects Volume 3, 2007.
8. Lu, S., Lyu, M. R., King, I., & Sar, H. K, "Semantic Video Summarization Using Mutual Reinforcement Principle and Shot Arrangement Patterns.", 2005.

## AUTHORS PROFILE

**D. Bhavitha,** Persuing my BE final year in Ssaveetha School Of Engineering, Saveetha Institute of Medical Sciences.

**Dr. A. Kalaivani,** Associate Professor, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences. She is an Academician and Researcher received her M.C.A., M.Phil., M.E., Ph.D. She has gained her 18+ years of teaching experience from well reputed educational institution in Chennai. She received South Indian "Best Non-Circuit Faculty Award", Best Teacher award , Women Researcher Award. She is acting as a Trustees for P.T.LeeChengalvarayaNaicker Trust by Honourable High Court of Madras for the period of three years (2017-2020). She is appointed as a Governing Council Member and School Committee Member for CNT Trust Schools and Colleges. She has contributed her research in the field of Data Mining and Image Processing. She has published 21 research papers in highly reputed journals indexed by Scopus, ThomsonReuters and Google Scholar Indexed Journal. She published her 9 research papers in IEEE and Springer Conferences.She has published 4 book chapters in highly reputed Book Chapter Series. She also published an editorial article "Automatic Semantic VideoAnnotation"in ICSES Transactions on Data Science, Engineering and Technology. She published a Book on "Data Structures" for Anna University Engineering Students. She is an editor and working on research book on "AI Innovations in Medical Imaging ". She is a Reviewer for ISA Transactions, International Journal of Machine Learning and Computing, International Journal of Intelligent Information Technologies , ICSES Transaction on Image Processing and Pattern Recognition. She is also honored as a reviewer for all International Conference organized by Association of Scientist, Developers and Faculties (ASDF). She is a Program Committee Member, National Advisory Board Member, Outreach Committee Member and Organizing Committee Member for various International Conferences. She has membership in IAENG,IFERP, ICSES,CSTA and ASDF.