International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 04 | April - 2023 Impact Factor: 8.176 ISSN: 2582-3930

AUTOMATIC ONLINE LECTURE HIGHLIGHTING BASED ON MULTIMEDIA ANALYSIS

E. NEHA REDDY, G. RAHUL KUMAR, M. VIJAYA(ASSISTANT PROFESSOR), B. DEEKSHIKA, G. NAVEEN PAVAN SAI

Dept. Of Computer Science And Engineering, Vidya Jyothi Institute Of Technology, Hyderabad

ABSTRACT:

The benefits of textbook highlighting for pupils are generally acknowledged. In this essay, we offer a thorough solution for highlighting the online lectures at the segment and sentence levels, much like it is done with printed books. this age of e-learning, especially with MOOCs, the answer is built on the automatic analysis of multimedia lecture materials like speeches, transcripts, and slides. Sentence-level lecture highlighting makes use of acoustic information from the audio and is incorporated in subtitle files of associated MOOC videos. The precision is over 60% when compared to expert-created ground truth, which is higher than baseline results and well-liked by users. While using statistical analysis, segment-level lecture highlighting primarily examines the speech transcripts, the lecture slides, and their linkages. A review method reveals that general accuracy can reach 70% with the ground truth produced by large numbers of users, which is pretty encouraging. Key-frame detection and automated video segmentation are used. Next, it extracts/takes out textual meta-data by using video Optical Character Recognition (OCR) technology on lecture video key-frames and Automatic Speech Recognition (ASR) on lecture audio tracks content to get audio from the video and then convert that audio track into text information. The primary goal is to provide extracted words with multimedia data, such as photos and movies. Online education is no more a sophisticated idea reserved for the ivory tower in this day of high-speed globalisation and information technology; rather, it is a fast growing sector with direct relevance to people's everyday lives.

keywords: MOOC, OCR, ASR, Video Segmentation, Key frame detection, Meta Data

I INTRODUCTION

Due to the recent rapid advancements in recording technologies, vastly enhanced video compression methsssods, and high-speed networks, digital videos are quickly becoming a very popular medium for storage and interchange. For this purpose, visual-audio recordings are being employed in e-lecturing systems more and more frequently. To give students mobility, many universities and research institutes are making use of this advantageous chance to record their lectures and then post them online for students to access at any time and from any location. in order for students to learn at any time and anyplace. This has resulted in a considerable rise in the amount of audio svisual content available online. This makes it nearly hard for users to find desired videos in video libraries without a search tool. for students to learn at any time and any place. This has resulted in a considerable rise in the amount of audio visual content available online.

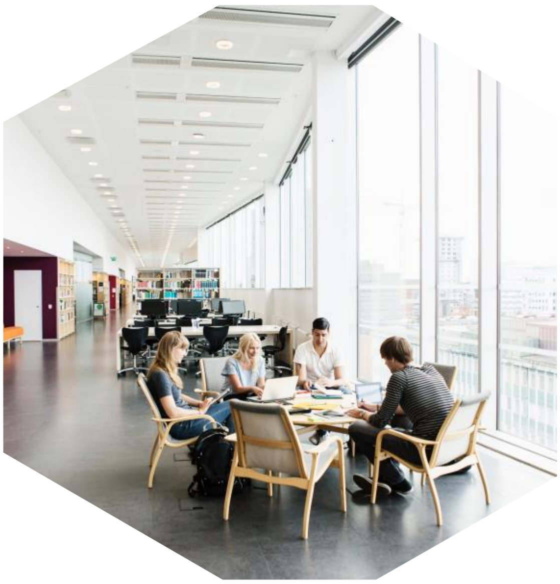
The lack of a search tool in video repositories makes it nearly difficult for users to find desired videos. Even when the user has discovered relevant video data, it is frequently challenging to compare whether a video is authentic or not by merely glancing at the title and other global meta-data offered, which are frequently brief and high-level documents. The user typically wants to acquire just the information they need without watching the entire film, thus the needed information may be covered in only a few minutes. The issue then becomes how to more effectively and conveniently find the necessary material in a huge lecture video repository. The majority of video retrieval and search engines, including those from YouTube, Bing, and Vimeo, respond based on textual meta-data like the video's title, genre, star, and detailed description that is readily available. To assure good quality, this type of metadata must often be created manually by a human, but this process takes time and money

II LITERATURE SURVEY*:*

Haojin Yang and Christoph Meinel noted in their research that e-lecturing has grown in popularity over the past few years. On the Internet, there is a sharp rise in the volume of lecture videos. So, it is imperative to develop a more effective approach for video retrieval from the internet or from sizable lecture video collections. The method for automatic video indexing and video search in sizable lecture video libraries is presented in this research. In order to provide a visual cue for navigating video content, they first apply the automated video segmentation (ASR) method and keyframe detection technique. They then employ Automated Speech Recognition (ASR) to lecture audio recordings and video Optical Character Recognition (OCR) technology to key-frames to retrieve textual metadata. In order to extract keywords for content-based video browsing and searching, the ASR and OCR transcript as well as recognised slide text line types are used, as well as video and segment-level keywords. Evaluation has shown the performance and efficacy of the suggested indexing functionalities.

A brand-new system for the automatic transcription of lectures was presented by E. Leeuwis, M. Federico, and M. Cettolo. This system incorporates a number of innovative features, such as factored recurrent neural network language models and deep neural network acoustic models that use multi-level adaptive networks to include out-of-domain input. It proves the system is capable of improving on the TED lecture transcribing challenge from the 2012 IWSLT evaluation. The closest competitive system from the evaluation has a relative Word Error Rate reduction of more than 16 percent, according to the results.

A Korean-spoken document retrieval system was introduced by Lee & G. Lee for use in lecture searches. It automatically creates a general inverted index table from spoken document transcriptions, and it also extracts lecture-related data from textbooks or slide notes. . It combines these two sources to conduct a search. The produced experimental findings showed that the information gathered from the documents is only marginally helpful for retrieving lectures



The research was conducted by W.Heurst, T. Kreuzer, and M. Wiesenhutter. A common practice at many universities is to record lectures and post them online for student access. Extensive search functionality must be offered to make use of the knowledge database these papers create. This functionality should go beyond simple meta-data searches and instead do a thorough analysis of the relevant multimedia documents. This paper presents some efforts for developing a web-based search engine for presentation audio recordings. The authors assess both cutting-edge speech recognition algorithms and typical, realisable retrieval performance. To assess the worth of voice processing for lecture retrieval, they also compare the speech retrieval outcomes with a conventional, text-based strategy for searching.

III EXISTING SYSTEM:

The ability to highlight text is popular not just with traditional paper books but is equally appreciated in the age of electronic publications. It is extensively used in numerous e-book apps. In this instance, highlighting is still done using the textual resources that resemble a book even though a marker is no longer of thenecessary. On the Internet, there is a sharp rise in the volume of lecture videos. So, it is imperative to develop a more effective approach for video retrieval from the internet or from sizable lecture video collections. The method for automatic video indexing and video search in sizable lecture video libraries is presented in this research. In order to provide a visual cue for navigating video content, they first apply the automated video segmentation (ASR) method and keyframe detection technique. They then employ Automated Speech Recognition (ASR) to lecture audio recordings and video Optical Character Recognition (OCR) technology to key-frames to retrieve textual metadata. In order to extract keywords for content-based video browsing and searching, the ASR and OCR transcript as well as recognised slide text line types are used, as well as video and segment-level keywords. Evaluation has shown the performance and efficacy of the suggested indexing functionalities. The marked sentences and portions still don't correlate. Due to the difficulties of creating ground truth, the sentence-level approach's evaluation scale is highly constrained. For the voluntary users, we might additionally offer an interactive highlighting tool in order to get information from their points of view. Experiments on several courses taught by various lecturers are required to determine the robustness of the suggested approach at the segment level.

IV PROBLEM IN EXISTING SYSTEM:

The ability to highlight text is popular not just with traditional paper books but is equally appreciated in the age of electronic publications. It is extensively used in numerous e-book apps. In this instance, highlighting is still done using the textual resources that resemble a book even though a marker is no longer necessary. On the Internet, there are increasingly more lecture videos available. So, a more effective mechanism for video retrieval from the internet or from sizable lecture video collections is crucial. This study outlines a strategy for automated video indexing and video search in huge lecture video libraries. To begin with, they use the automated video segmentation (ASR) method and keyframe recognition technique to provide a visual cue for navigating through video information. They then use video Optical Character Recognition (OCR) technology on key-frames and automatic speech recognition (ASR) on lecture audio tracks to retrieve textual metadata. The ASR, OCR, and recognised slide text line types are used for keyword extraction, and both segment- and video-level keywords are extracted for content-based video browsing and search. Evaluation is used to demonstrate the effectiveness and performance of the suggested indexing functionalities. The video clip is divided into several sections by visual differences, which are strengthened by keyframes that serve as examples. A presentation video can be navigated using an interactive user interface (UI), which combines a visual representation of audio, video, text, and keyframes. Moreover, authors might investigate clustering and labelling of speaker data and publish preliminary findings. The marked sentences and portions still do not correlate. Due to the difficulties of establishing ground truth, the sentence-level approach's rating scale is highly constrained. Perhaps we could also give the voluntary users a mechanism for interactive highlighting to get information from their points of view. To evaluate the robustness of the suggested approach at the segment level, tests on various courses taught by various lecturers are required.

V PROPOSED SYSTEM:

Users find it difficult to read the huge amounts of text available online since it takes so much time. Automated Text Summarization, the suggested system, is considerably more useful and applicable in the present. Natural language processing is used to process the input text, and word embedding is used to turn the processed input into a vector format. Word embeddings provide us a highly in-depth understanding of a document, which improves the outcomes of any NLP activity. To extract higher-ranked sentences, sentences are scored against each other. This creates the extractive summary of the input. Then, polarity and subjectivity criteria are applied to the condensed text for analysis. Moreover, speech conversion is applied to the condensed text.

VI ARCHITECTURE:

The architecture shows the relationship and overview of the project like how the components are connects in which flow the project is running, by this we can understand the project at which component what process is going on.

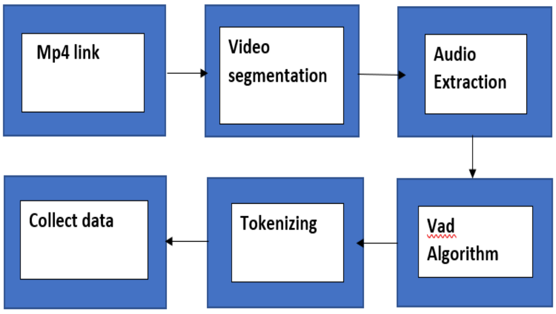


Fig : Architecture

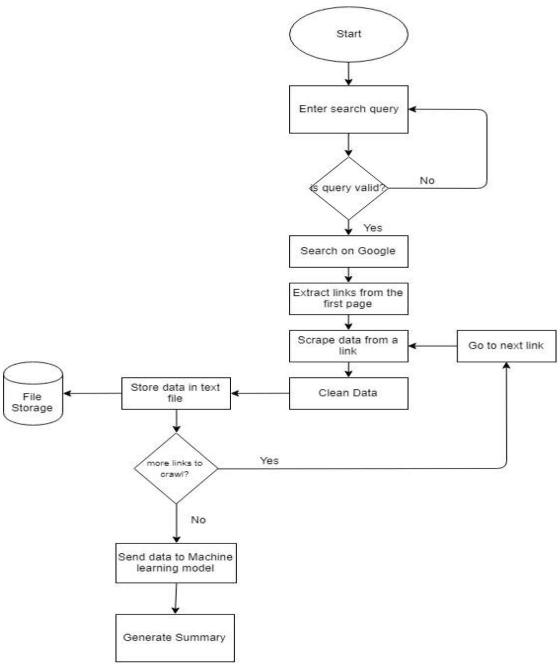


Fig: Flow process

.VII RESULTS AND DISCUSSIONS :

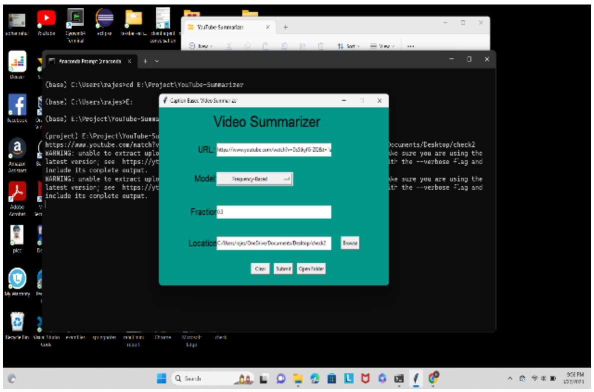


Fig: Output Screen

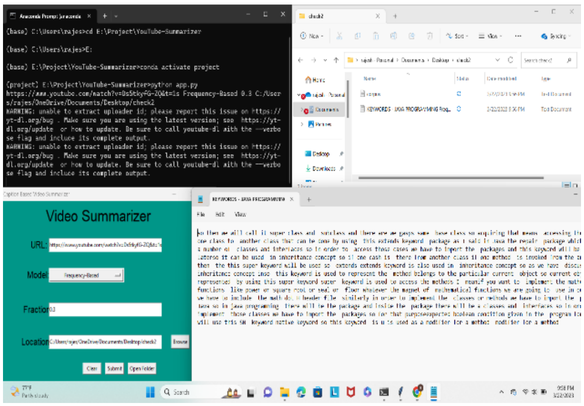


Fig : Highlighted Text

VIII ADVANTAGES:

The primary goal is to provide extracted words with multimedia data, such as photos and movies.

It is useful to children if they missed any classes.

It is useful to teachers to easily explain the topics and to highlight the important sentence.

Flexible to every video for segmentation.

Easy to translate words from voice.

IX CONCLUSION AND FUTURE SCOPE:

As distance learning technology quickly advances, more and more lectures are professionally filmed and published for public viewing. Nonetheless, lecture videos are always the main source of content, whether in traditional te-le-teaching or more current, well-liked MOOCs. In this thesis, several technical improvements to the learning resources—outside of videos—are suggested to better serve online learners. Online learning is no longer only about watching lecture videos with the provided subtitles, summaries, and highlights. Students may more easily locate desired lectures, comprehend lectures given in other languages, and better plan their study schedules. This system uses an algorithm to retrieve videos based on their content and provide learners with relevant search results. The employed method is more effective than the current one. The ASR algorithm is used to extract audio from video, which is the key component of this procedure. We hope to employ more effective and practical algorithms in the future to detach audio from video, extract text from audio and video segments, and split video into several segments so that keyframes can be selected from among them. We can add more nodes and assess how well they work.The quality of automatically generated subtitles, outlines, and highlights needs to be improved in the future. It might be difficult to improve MT output, especially for subtitles in the target language. Also, it is feasible to create interactive user interfaces and ask users of e-learning portals to contribute to the modification of these learning resources, increasing the learning experiences from the perspective of the student. It's never too late to learn something new or to get better

X REFERENCES :

[1] "Content Based Lecture Video Retrieval Using Speech and Video Text Information," by Haojin Yang and Christoph Meinel, published in IEEE Transactions on Learning Technologies, Vol. 7, No. 2, April-June.

[2] TED corpus talks' language modelling and transcription; Proc. IEEE Int.Conf. Acoust.,Speech Signal Process, 2003, pp. 232–235. E. Leeuwis, M. Federico, and M. Cettolo.

[3] D. Lee and G. G. Lee described a spoken document retrieval system for lectures in the 2008 Proceedings of the ACM Special Interest Group Spontaneous Conversational Speech Workshop . At Proc. IADIS Int. Conf.WWW/Internet, pp. 135143, 2002, a qualitative study on the use of big vocabulary automatic voice recognition to index recorded presentations for search and access over the web was published.

[4] "Augmented segmentation and visualisation for presentation videos," Proc. 13th Annual ACM International Conference on Multimedia, pp. 5160, Haubold and J. R. Kender, 2005.4

[5] "Analysis and Processing of Lecture Audio Data: Early Investigations," in Proc. HLT-NAACL Workshop Multidisciplinary Approaches Speech Indexing Retrieval, edited by James Glass, Timothy J. Hazen, Lee Hetherington, Chao Wang, and others, p. 912. (2004). Information Process Management, vol. 24, no. 5, pp. 513-523, "Term-weighting techniques in automatic text retrieval" (1988).

[6] Using the stroke width transform to recognise text in natural environments, Proc. Int. Conf. Comput. Vis. PatStern Recog., pp. 2963-2970, B. Epshtein, E. Ofek, and Y. Wexler, 2010.

[7] In the proceedings of the 2011 2nd International Conference on e-Learning, C. Meinel, F. Moritz, and M. Siebert discuss community tagging in teleteaching scenarios.

[8] In 2009, M. Grcar, D. Mladenic, and P. Kese published Semi-automatic classification of movies on videolectures.net in the proceedings of the European Conference on Machine Learning and Knowledge Discovery Databases, pp. 730733.

[9] A structure for more accurate video text detection and recognition, Multimedia Tools Appl., pages. 129, 2012, by H. Yang, B. Quehl, and H. Sack.

[10] Using linguistically salient gestures to transform lectures into comic books, J. Eisenstein, R. Barzilay, and R. Davis, Proc. 22nd Nat.Conf.Artif.Intell., 1, pp. 877882, 2007.

[11] Teaching Sociology, vol. 40, no. 3, pp. 242-256, 2012, ""I'm ambivalent about it" the problems of PowerPoint." A. Hill, T. Arford, A. Lubitow, and L. M. Smollin.

[12] "Structuring lecture films by automatic projection screen localization and analysis," IEEE Transactions on Pattern Analysis & Machine Intelligence, vol. 37, no. 6, 2015, pp. 1233–1246; K. Li, J. Wang, H. Wang, and Q. Dai

[13] "Pedagogy meets PowerPoint: A study review of the effects of computer-generated slides in the classroom," D. G. Levasseur and J. Kanan Sawyer, The Review of Communication, vol. 6, no. 1-2, pp. 101-123, 2006.
[14] An effort at mooc localization for Chinese-speaking Users by X. Che, S. Luo, C. Wang, and C. Meinel was published in the International Journal of Information and Education Technology in 2016, issue 2, page 90
[15] Punctuation forecast for an unsegmented transcript using a word vector by X. Che, C. Wang, H. Yang, and C. Meinel is published in the 2016 proceedings of the tenth international conference on language resources and evaluation (LREC 2016)