

LectYS: A System for Summarizing Lecture Videos on YouTube

Taewon Yoo
twyoo@khu.ac.kr
Kyung Hee University
Yongin, Republic of Korea

Donghwan Lee
derick_lee@khu.ac.kr
Kyung Hee University
Yongin, Republic of Korea

Hyewon Jeong
woni980911@khu.ac.kr
Kyung Hee University
Yongin, Republic of Korea

Hyunggu Jung*
hjung@uos.ac.kr
University of Seoul
Seoul, Republic of Korea

ABSTRACT

Students leverage online resources such as online classes and YouTube is increasing. Still, there remain challenges for students to easily find the right lecture video online at the right time. Multiple video search methods have been proposed, but to our knowledge, no previous study has proposed a system that summarize YouTube lecture videos using subtitles. This demo proposes LectYS, a system for summarizing lecture videos on YouTube to support students search for lecture video content on YouTube. The key features of our proposed system are: (1) to summarize the lecture video using the subtitle of the video, (2) to access to the specific parts of the video using the start time of video subtitle, and (3) to search for the video with keyword. Using LectYS, students are allowed to search for lecture videos on YouTube faster and more accurately.

CCS CONCEPTS

• **Information systems** → *Information extraction; Document filtering.*

KEYWORDS

YouTube, lecture video, text summarization

ACM Reference Format:

Taewon Yoo, Hyewon Jeong, Donghwan Lee, and Hyunggu Jung. 2021. LectYS: A System for Summarizing Lecture Videos on YouTube. In *26th International Conference on Intelligent User Interfaces (IUI '21 Companion)*, April 14–17, 2021, College Station, TX, USA. ACM, New York, NY, USA, 3 pages. <https://doi.org/10.1145/3397482.3450722>

1 INTRODUCTION

Recently, students who use online resources for learning are increasing. According to Snyder et al., the percentage of students who take online classes is increased from 15.6% in 2004 to 43.1% in 2016 [15]. The percentage of undergraduate and graduate students who fully take online degree programs have increased from 3.8%

*Corresponding author.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

IUI '21 Companion, April 14–17, 2021, College Station, TX, USA

© 2021 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-8018-8/21/04.

<https://doi.org/10.1145/3397482.3450722>

in 2008 to 10.8% in 2016, and from 6.1% in 2008 to 27.3% in 2016, respectively [15]. Through surveys, researchers discovered that a number of high school and college students use online resources on the Internet to achieve their educational purposes [1, 4, 9]. Nevertheless, students faced challenges when searching for educational resources online [2, 5, 6, 10, 14, 20]. For instance, students have difficulty choosing a search query before they look for the resources online. In addition to struggling to create search queries and to spell them correctly [20], students had difficulty in searching for unfamiliar research areas online [5]. Likewise, prior studies revealed that students struggle while searching the resources online. For example, students were concerned about selecting an appropriate result on the search list, such as judging which information to choose due to too much information available [2], selecting a website from a search list [20], and finding information they want in a reasonable time [6, 14]. Moreover, Patel and Dave revealed that students had difficulty in finding information online because information is too scattered [14]. After searching for resources online due to the limitations of online resources retrieval services, students struggled with performing multiple tasks, such as viewing search results due to inaccessible websites [2, 6, 10] and understanding contents in search results [20].

Despite these challenges, there exist three benefits to searching for educational resources on YouTube, such as obtaining a large amount of educational resources in the form of streaming services [21], providing additional supplement for students in their school classes [17], and enabling students to exchange and learn information through social interaction among students [13]. However, YouTube has its limitations when searching for educational resources. In YouTube, it is not possible to know what is in the video without watching the video. Even if a description of the video is provided, students are not able to find a specific part of the video without the time stamp. Also, it is hard to know the difficulty of the lecture video without watching the video. Still, little is known about how to summarize the lecture videos on YouTube to make it easier for students to find the videos. To address this gap, we developed a system that summarizes the lecture videos on YouTube to improve students' learning efficiency. This study aims to develop a YouTube lecture summarizing system by answering the following research question (RQ): *How can we develop a system that summarizes lecture videos on YouTube?* In the next section, we present related work our methods for summarizing YouTube lecture. We then discuss about the benefits and limitations of our study.

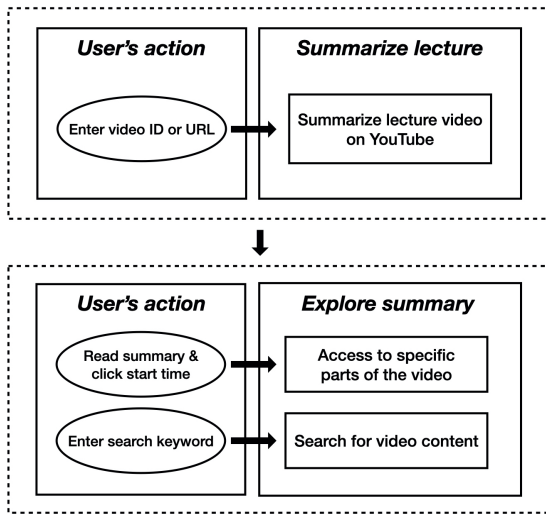


Figure 1: Overview of LectYS. When users enter the video ID or URL of the lecture video, LectYS creates a summary of the lecture video. Then the users are allowed to read the summary and click the start time of the summary to go to the specific part of the video. The users are also allowed to enter search keyword so that the summary that includes the keyword is displayed. While the keyword in the summary is highlighted.

2 RELATED WORK

To help students find lecture videos faster and more accurately, eight studies were proposed [3, 7, 8, 11, 12, 18, 19, 22]. Two methods are suggested in the studies. The first method is to search video contents with video metadata and keywords [3, 7, 11, 12, 18, 22]. The second method is to access to the specific part of the video [18, 19]. Prior studies showed how their method helps students find lecture videos faster and more accurately. Studies suggested search functions to find the contents of the video using metadata and keywords [3, 7, 11, 12, 18, 22]. Also, several studies made students to access to the specific part of the video [18, 19]. In addition, one study recommended lecture videos for students to search for lecture videos that students want to find on YouTube [8]. Multiple studies suggested systems that support searching for the content in lecture videos using metadata such as speech text, mouse cursor movements, annotation, and time information [7, 11, 12, 22]. Similarly, other prior studies demonstrated their systems that allow students to search for educational videos using video keywords [3, 18].

On the other hand, previous studies provided indexed information of videos that help readers find the specific parts of the video quickly and easily [18, 19]. Nonetheless, prior studies have limitations. First, to our knowledge, little is known about how to summarize the contents of the lecture videos available on YouTube. Second, none of the prior studies supported YouTube lecture videos for content searching or navigating to the specific part of the video. To address these limitations, our suggested method is to create a summary of the lecture video on YouTube so that the students are able to find the content of the lecture video easily.

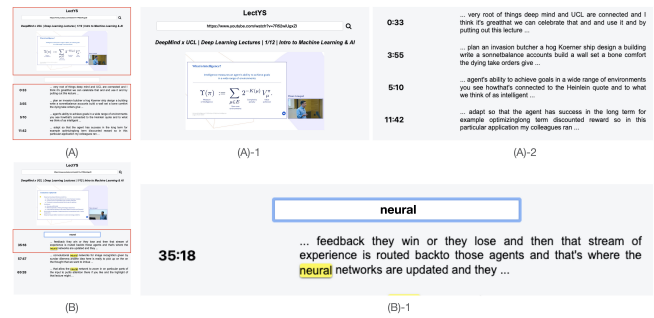


Figure 2: Screenshots of the LectYS user interface. In (A)-1, users enter the ID or URL of the YouTube lecture video in order to search for the lecture video. Then, LectYS returns a list of important subtitles in the video and the start time of each subtitle as illustrated in (A)-2. Then, users are allowed to search for the videos with subtitles including the keyword. (B)-1 illustrates the result of the search process after entering “neural” as a keyword. The result shows a set of subtitles, each of which contains the highlighted keyword.

3 USER WALKTHROUGH AND FEATURES

3.1 System overview

To answer our RQ, we propose LectYS, a YouTube lecture summarizing system that uses video subtitles, so that students who use YouTube for educational purposes are allowed to search the content of the lecture quickly and accurately. As illustrated in Fig. 1, LectYS provides key three features: (1) summarizing the lecture video, (2) accessing to specific parts of the video, and (3) searching for video content. We demo this system by developing a web application using Python, JavaScript and Django web framework (see Fig. 2).

3.2 Summarize lecture videos on YouTube

On the first page of the demo (see Fig. 2 (A)-1), users are required to enter the YouTube video ID or URL. The videos with English or auto-generated English subtitles are available for the summarization. When the users enter the appropriate video ID or URL, the system downloads the video subtitles using Python *youtube-transcript-api* library. If the video has both English and auto-generated English subtitles, the system will only download English subtitles. The downloaded data includes subtitle text, subtitle start time and subtitle duration. After downloading the subtitle data, we created a summary of the video. First, the subtitle text data was preprocessed. For example, we stemmed subtitle text and removed stop words such as “and”, “a”, and “or”. Stop words in subtitles were only removed during the summarization process. When the summary is displayed, users are able to view the summary of the original subtitles. Then we summarize the subtitle text using Python *sumy* library. We adopted Latent Semantic Analysis (LSA) methods for text summarization algorithm to identify semantically important sentences in the lecture video [16]. The result of the summarizing process includes the list of important subtitles in the video and the start time of each subtitle (e.g., “3:55” in Fig. 2 (A)-2 means that the start time corresponding to the subtitle in the video is 3 minutes and 55 seconds.).

3.3 Explore summary

3.3.1 Access to specific parts of the video. After the summarizing process, the lecture video and its summary are displayed on the web browser together with the start time of the subtitle. We used the YouTube IFrame Player API to embed a YouTube video in the system. Then, the summary will be displayed below the video in chronological order of start time of subtitle (see Fig. 2 (A)-2). The users are able to read the lecture video summary and click on the corresponding time stamp to access to a specific part of the video.

3.3.2 Search for video content. The users are allowed to search for the summary using the search function while reading the summary. There is a search bar for the users to find a summary that contains the keyword. When the users enter a keyword, LectYS returns the list of subtitles that contains the keyword and the start time of each subtitle. As a result, the subtitles and start time are displayed. The keywords in each subtitle are highlighted as illustrated in Fig. 2 (B)-1.

4 CONCLUSION AND FUTURE WORK

The ultimate goal of this study is to help students identify lecture videos on YouTube faster and more accurately. To achieve this goal, we developed LectYS, a system that summarizes YouTube lecture videos. LectYS summarizes the subtitle text of the lecture video on YouTube. Therefore, LectYS allows students to explore the overall content of lecture videos by reading a video summary. The start time provided with the summary gives students immediate access to what they want to learn, making it easier for students who are having difficulty finding the specific content in the video. In addition, keyword search makes it easier to verify whether the content that the students are looking for is included in the lecture video. Using LectYS, students are able to read the summary of the lecture video, while students are allowed to click the start time of the subtitle corresponding to the summary to go to specific parts of the video. Also, students are able to search for summaries that include a specific keyword. Therefore, students are allowed to search for the content of the lecture video on YouTube faster and more accurately when using our system. Future work still remains to evaluate the performance of the summarizing process of LectYS and its usability with potential target users. Moreover, it would be beneficial for students to recommend lecture videos on YouTube with video summary provided by this system and other features such as keywords, topic tags, and key frames as Jung et al. suggested [8]. Besides, developing a system that addresses the challenges such as difficulty in choosing search queries that Vanderschantz et al. reported [20] would be helpful for students searching for lecture videos. Additionally, in order to enable users to view the summary, we plan on applying LectYS to the public using web extensions, such as Mozilla WebExtensions API rather than providing a website.

ACKNOWLEDGMENTS

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. 2020R1G1A1009133).

REFERENCES

- [1] Tamara Almarabeh, Yousef Majdalawi, and Hiba Mohammad. 2016. Internet Usage, Challenges, and Attitudes among University Students: Case Study of the University of Jordan. *Journal of Software Engineering and Applications* 09 (01 2016), 577–587. <https://doi.org/10.4236/jsea.2016.912039>
- [2] Swiga Alphonse and Kelefa Mwantimwa. 2019. Students' use of digital learning resources: diversity, motivations and challenges. *Information and Learning Sciences* (2019).
- [3] D Anderson, L Harvel, M Hayes, Yoshihide Ishiguro, J Jackson, and M Pimentel. 2000. Internet course delivery-making it easier and more effective. In *2000 IEEE International Conference on Multimedia and Expo. ICME2000. Proceedings. Latest Advances in the Fast Changing World of Multimedia (Cat. No. 00TH8532)*, Vol. 1. IEEE, 84–87.
- [4] Oberiri Destiny Apuke and Timothy Onosahwo Iyendo. 2018. University students' usage of the internet resources for research and learning: forms of access and perceptions of utility. *Heliyon* 4, 12 (2018), e01052.
- [5] Kumaripaba Athukorala, Eve Hoggan, Anu Lehtiö, Tuukka Ruotsalo, and Giulio Jacucci. 2013. Information-seeking behaviors of computer scientists: Challenges for electronic literature search tools. *Proceedings of the American Society for Information Science and Technology* 50, 1 (2013), 1–11.
- [6] Raya Fidel, Rachel K Davies, Mary H Douglass, Jenny K Holder, Carla J Hopkins, Elisabeth J Kushner, Bryan K Miyagishima, and Christina D Toney. 1999. A visit to the information mall: Web searching behavior of high school students. *Journal of the American society for Information Science* 50, 1 (1999), 24–37.
- [7] Martin Gutbrod, Christian Werner, and Stefan Fischer. 2006. Platform-independent authoring and production of rapid e-learning content. *Interactive Technology and Smart Education* 3, 3 (2006), 197–206.
- [8] Seungwon Jung, Minjae Son, Chung-il Kim, Jehyeok Rew, and Eenjun Hwang. 2019. Video-based learning assistant scheme for sustainable education. *New Review of Hypermedia and Multimedia* 25, 3 (2019), 161–181.
- [9] Mpubane Matlala and Siyanda Kheswa. 2019. The perceived usefulness of the Internet to high school learners in Makhuduthamaga Local Municipality, Limpopo province, South Africa. *SA Journal of Information Management* 21 (12 2019). <https://doi.org/10.4102/sajim.v21i1.1098>
- [10] Victor Ngessa. 2018. User Challenges and Top used Online Databases: A Survey of Higher Education Institutions in Tanzania. *Sciences* 8, 5 (2018), 381–395.
- [11] Bernardo Pereira Nunes, Marco A Casanova, and Ricardo Kawase. 2012. Automatically generating multilingual, semantically enhanced, descriptions of digital audio and video objects on the Web.
- [12] Bernardo Pereira Nunes, Alexander Mera, Marco A Casanova, and Ricardo Kawase. 2012. Boosting retrieval of digital spoken content. In *International Conference on Knowledge-Based and Intelligent Information and Engineering Systems*. Springer, 153–162.
- [13] Carlos Orús, María José Barlés, Daniel Belanche, Luis Casaló, Elena Fraj, and Raquel Gurrea. 2016. The effects of learner-generated videos for YouTube on learning outcomes and satisfaction. *Computers & Education* 95 (2016), 254–269.
- [14] Pramod Patel and Govind B Dave. 2016. A Study Of Difficulties Faced During Information Serach Pattern For E-Resources Among Students Of Management. *Voice of Research* (2016), 10.
- [15] TD Snyder, C de Brey, and SA Dillow. 2019. Digest of education statistics 2018 (NCES 2020-009). Washington, DC: National Center for Education Statistics, Institute of Education Sciences, US Department of Education.
- [16] Josef Steinberger and Karel Jezek. 2004. Using latent semantic analysis in text summarization and summary evaluation. *Proc. ISIM* 4 (2004), 93–100.
- [17] Elaine Tan and Nick Pearce. 2011. Open education videos in the classroom: exploring the opportunities and barriers to the use of YouTube in teaching introductory sociology. *Research in Learning Technology* 19 (2011).
- [18] Tayfun Tuna, Jaspal Subhlok, Lecia Barker, Shishir Shah, Olin Johnson, and Christopher Hovey. 2017. Indexed captioned searchable videos: A learning companion for STEM coursework. *Journal of Science Education and Technology* 26, 1 (2017), 82–99.
- [19] Tayfun Tuna, Jaspal Subhlok, and Shishir Shah. 2011. Indexing and keyword search to ease navigation in lecture videos. In *2011 IEEE Applied Imagery Pattern Recognition Workshop (AIPR)*. IEEE, 1–8.
- [20] Nicholas Vanderschantz, Annika Hinze, and Sally Jo Cunningham. 2014. "Sometimes the internet reads the question wrong": Children's search strategies & difficulties. *Proceedings of the American Society for Information Science and Technology* 51, 1 (2014), 1–10.
- [21] Jon Watkins and Michael Wilkins. 2011. Using YouTube in the EFL classroom. *Language Education in Asia* 2, 1 (2011), 113–119.
- [22] Baoquan Zhao, Songhua Xu, Shujin Lin, Ruomei Wang, and Xiaonan Luo. 2019. A New Visual Interface for Searching and Navigating Slide-Based Lecture Videos. In *2019 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, 928–933.