Journal of Multidisciplinary Engineering Science and Technology (JMEST)
ISSN: 2458-9403
Vol. 12 Issue 11, November - 2025

Educational Video To Text Summarization And Mind Map Generation

Mohammad Izzuddin bin Ishak
Faculty of Computing and Informatics
Universiti Malaysia Sabah
Kota Kinabalu, Sabah

James Mountstephens
Faculty of Computing and Informatics
Universiti Malaysia Sabah
Kota Kinabalu, Sabah
james@ums.edu.my

***Abstract*— Students frequently struggle with the task of understanding vast amounts of knowledge learned from traditional methods. These knowledge comes in two main types of formats, which is text and videos. To tackle the overwhelming number of learning materials, a video to text summary and mind map generation system is suggested as a solution to this problem. It is to help students learn more by reading less. The system summarizes the video lectures of the course materials or other types of videos, in form of text, into clear and understandable summaries using extractive and abstractive text summarization method. The summarization is made by identifying the significant sentences of the text and add them to the summary. The technique for text summarization involves using an open-source topic and vector space modeling toolkit. The summarization algorithms used are TextRank and spaCy for extractive summarization, while using BART and its distilled version for abstractive summarization. These summaries are then transformed into graphical representations, specifically mind maps, using the 'Graphviz' library from Python. This application facilitates students to better understand the links and interactions between ideas and remember the knowledge learnt. The proposed system will not only address the difficulties faced by students in their academic pursuits, but also have a broader impact on the education sector by enabling teachers and educators to deliver more engaging and impactful lessons. The web application is built using python with Flask framework. The report will cover the problem background, project scope, objectives, methodology, and benefits of the proposed system, and provide a detailed plan for its development and implementation. The educational video text summarization and mind map generation system has the potential to revolutionize the way students learn and retain information, making education more accessible and effective for all. The current developed system has several limitations. However, overall, the system is able to achieve all the given project objectives.**

Keywords*— *educational technology; text summarisation; mind maps; video to text

I. INTRODUCTION

The challenge of managing large volumes of information from course materials is a common struggle for students. To address this issue, a text summarization and mind map generation system has been proposed for use in improving student study. The system uses advanced natural language processing (NLP) and machine learning techniques to analyze and summarize video lectures, textbooks, and other course materials into concise and easy-to-understand summaries. These summaries are then transformed into interactive, graphical representations using mind maps, allowing students to easily identify relationships and connections between concepts and retain the information for longer periods of time. Mind maps are a popular tool for generating design concepts and in general for hierarchically organizing design insights [1].

The proposed system will leverage advanced natural language processing (NLP) and machine learning techniques to analyze and summarize video lectures, textbooks, and other study materials, and convert the output into concise, easy-to-understand summaries [2]. The method that most NLP technologies use relies on breaking apart text and identifying meaningful sections [3]. The mind map component will use state-of-the-art algorithms and libraries [4] to generate interactive, graphical representations of the information, allowing students to easily identify relationships and connections between concepts, and retain the information for longer periods of time.

This project will not only address the challenges faced by students in their academic pursuits, but also have a broader impact on the education sector by enabling teachers and educators to deliver more engaging and impactful lessons [5]. The text summarization and mind map generation system will be built using the latest technologies, and will be scalable, secure, and easy to use, making it an ideal solution for universities, colleges, and other educational institutions.

Overall, the purpose of this proposal is to outline the scope, objectives, and benefits of the text summarization and mind map generation system for improving student study, and to present a detailed plan for its development and implementation.

II. LITERATURE REVIEW

A. A Review of Semantic Text Summarisation Methods for Long Video

The familiarity and excitement of recording interesting occurrence have becoming a norm in this age, which results in abundance of video data produced, with the duration of it ranges from seconds to hours. In this fast paced moving state of the world, consumers are more inclined in knowing the main points and content of the videos immediately rather that initially watching the videos, section by section, as it can be daunting [6]. This motivates the researcher to propose a novel method to summarize and annotate long videos.

The proposed method is identifying interesting segments of long videos according to consumer preference. Key frames are extracted from interesting segments where the deep visual-captioning techniques generate visual and use it as a basis for relevant textual summaries. The introduction of knobs enables to increase or decrease both video and textual summary length achieve desired outcomes of the suitable applications [6].

For the text summarization method, sumy 0.4.1 python framework and NLTK libraries were used to assess Luhn's algorithm, Latent Semantic Analysis (LSA), Edmund-son's heuristic method, LexRank, TextRank, SumBasic and KL-Sum text summarization techniques [6]. The summary length was fixed to 24 sentences for the papers, however it can be adjust to greater length that the input captions. The evaluation of summaries generated are tested onto ROUGE [7] which is a package for evaluation of automatic summaries..

B. A Critical Review of Hybrid Video-to-Text Summarization Frameworks and Algorithms

This paper proposed a hybrid video-to-text summarization (VTS) framework, as it utilizes both extractive and abstractive summarization algorithms [8]. This hybrid method is highly desirable as it eases the navigation process to determine the informative and important theme, idea and plot by taking account the text or caption from the long video.

The main approach of state-of-the-art text summarization generally divided into three categories: extractive, abstractive and hybrid [9]. However, by combining both approaches, extractive and abstractive summarization, it does not take advantages of each other to generate a reliable synopsis for the video [8].

The proposed framework consists of five main modules, which is PyScene Detector (PD), Audio Extractor (AE), Human Dialog Text Extractor(HDTE), Extractive-based Text Summarizer (ETS), and Abstractive-based Split-and-Merged Text Summarizer (ASMTS). The python libraries used are PD and AE, which is the MoviePy library. PD is used to analyze and detect scene changes in the video, and automatically split the video into a number of key content-aware scene clips, while the AE is used to extract their audio portions in the waveform audio (WAV) format. The Google Speech-to-Text API, HDTE, takes the WAV file as input from the AE, to transcribe the human dialogues into a text transcript. It is then passed onto ETS to create the initial summary. The initial summary is then fed onto ASMTS to create new phrases and terms, by still keeping the content and plot the same, as it generates one final video synopsis based on the inputs.

In this paper, researchers propose a new way of summarizing videos called the hybrid VTS framework. The framework automatically generates text transcripts from human dialogues and summarizes them into a short video synopsis. The approach uses an advanced algorithm to split and merge long transcripts to produce a final semantic video synopsis. The researchers tested the method on three different videos and found it outperforms most individual methods by 75% [8]. However, there is still a need to figure out how to integrate video background sound and key image frames. Overall, the researchers' approach offers a promising solution for summarizing videos.

C. A Comparative Study of Ensemble Methods for Extractive Text Summarization

The high volume of articles, links and video produced in this digital age has push the importance of semantic density, as it made us struggle to make informed decision quickly [10]. This paper discusses the experimentation and result based on Neural Network, Logistic Regression, Decision Tree, Random Forest, SVM models, XGBoost and Naïve Bayes. The results between these algorithms are compared and finally, an ensemble approach was proposed. Metric used to validate this research claim are called Recall-Oriented Understudy for Gisting Evaluation (ROUGE). It showcased convincing improvements by utilizing 1-grams, bigrams and longest common subsequence-based statistics.

The summarization of text is approached commonly using these two techniques, which is extraction and abstraction based technique [10]. Extraction based fetch key phrases from source document and aggregate them to make a summary, while abstraction based needs to paraphrase and shorten parts of the source documents. Deep learning method is applied for text summarization to overcome grammar disparities. Five based approaches are used, which is statistical based, machine learning based, topic based, graph based and lastly discourse based.

Results drawn from this paper are the process of evaluating both reference summaries and system-generated summaries has been completed, and comparisons have been made between the ROUGE scores of existing base models and a proposed model..

III. METHODOLOGY

The methodology used in this project is Extreme Programming (XP). It is a methodology that emphasizes agile software development life cycle and practices.

A. *Selection of Development Methodology*

Extreme programming (XP) methodology is selected because of its agility, flexibility and rapid development. It is lightweight to developer as it has the ability to respond to changing and unexpected requirements, making it an ideal methodology to be chosen for this product development [11], as important factor or key points of project might have been overlooked. There is multiple advantage on using this methodology such as improved software quality, increased flexibility of the project development phase, prioritize user satisfaction, reduced time and cost and helps developers and potential users to align expectation of the system or the application.

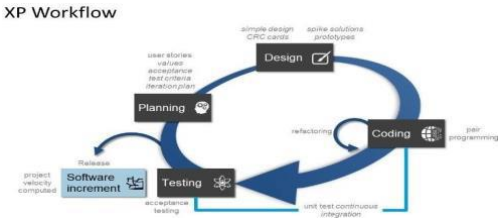

Fig. 1. *Workflow of XP Methodology*

B. *Workflow of XP Methodology*

Plan: The planning phase involves breaking down the project into small and manageable chunks of functions and modules. The requirement gathering and analysis via multiple resources such as journals, articles, product reviews and user's stories need also be done in this phase of the project. The requirements will then be evaluated and determine its priority to move to the next phase.

Design: Developer develops a high-level design for the product based on the modules that has been identified in planning phase. The proper design is needed to ensure that in the long run, the system does not become too complicated which results in inevitable halt. It is critical to create a design structure which don't rely on too many dependencies as the integration of it might rises another problem. It is the highest priorities to ensure that the system implementation is easy to be implemented as long as it meets all the required functionality.

Coding: Developing the product features based on the design developed in the previous phase. XP emphasizes the use of test-driven development, where tests are written before the code is developed. Series of test is made up from user's stories or tasks, where it will be coded and tested to ensure it is working. Main programming language used for this project is in Python language

Test: Involves testing the product features or modules developed in the previous phase. XP emphasizes the testing in every stage of iteration process to ensure that the product value meets the user's requirements.

Release: The product features or modules are released to the user for feedback. This feedback is used to refine the product for the next iteration, if there is a need to do so.

C. *Project Flow*

To ensure the project development runs smoothly and complete within the proposed timeframe, the project is broken into five stages, which started with project planning, design, coding, testing and product release.

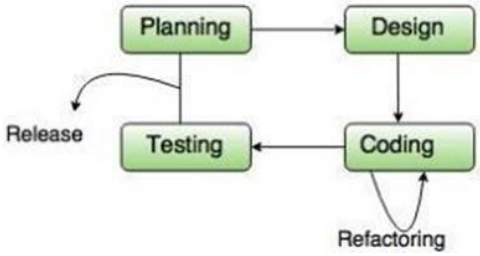

Fig. 2. *Project Flow*

IV. TESTING AND RESULTS

The testing is done using a bottom-up approach, where the components of the systems are being tested individually and isolated, to ensure it is working as expected before moving onto the integration phase. The integration testing is conducted, once all the components tested on unit testing passed their test and evaluation. The integration testing will test the interaction between the components, to ensure it is compatible and able to integrate successfully. Next, system testing is carried out to test the system as a whole. Lastly, user acceptance test (UAT) is carried out to determine whether the system is functional and acceptable by the end user.

A. *Unit Testing*

Unit testing is where software units or components are tested, with the goal in mind to ensure that each individual unit of software code or parts operates as planned [12]. In this part, each unit is tested to check the usability and functionality for each components or criterion. Unit testing is performed by in-house developer throughout the development phase to ensure its functionality and its validity. Unit tests are used to isolate a part of code and verify its correctness [12]. There is 10 units needed to be test, such as uploading video, transcribe function, text summarization, summary customization, word counts, mind map generation, word cloud, visualization tools, feature extraction and lastly export functionality.

B. *Integration Testing*

Integration testing is carried out once all the modules are completed. All modules combined and compiled are then tested to validate the functionality of the system, as if all the modules can work well and integrate without any complication with each other. As the integration is done, it should be working as expected without any errors or bugs.

C. *System Testing*

System testing validate the fully integrated and the entirety of the web application. A system test is carried out to validate and verify the end-to-end system

specification. System testing is done to carry out a series of different tests with the sole purpose is to utilize the full web application system.

TABLE I. SYSTEM TESTING BY USERS

Module	Tasks	Result
Upload video	Uploads video	Pass
Transcribe video	Get video transcribe	Pass
Text summarization	Summarize transcribe	Pass
Word count	Ensure number of transcribed and summary words are correct	Pass
Summarization customization	Customize summary	Pass
Mind Map generation	Generate mind map off summary	Pass
Word Cloud	Generate word cloud	Pass
Visualization tools	Generate bar chart	Pass
Feature extraction	Get keyword and NER	Pass
Export functionality	Export summary and images generated	Pass

D. User Acceptance Test and Results

User acceptance test can be carried out with multiple variety of users to test their acceptance and understanding toward the system. The test conduct follows the structure of Technology Acceptance Model (TAM) questionnaire to gather feedback and user acceptance level on the system.

The Technology Acceptance Model (TAM) questionnaire created was divides into 4 sections which are perceived usefulness, perceive ease of use, attitudes toward usage of the system and behavioral intentions when using the system. It is created that way to gain information on users when they are testing the system as a whole. A number of 10 students was picked to test the system. Each of them were then asked to fill up the TAM questionnaire on their opinion of the use of the “Educational Video to Text Summarization and Mind Map Generation” system.

Based on the results of questionnaire given to participant, these are the following results. There are 60% of the students that participate in the testing of the system strongly agrees that by using this system, it would enhance their effectiveness in performing tasks. The other 40% also agrees as well that the system would improve their effectiveness in performing tasks. Next, 80% of the participants strongly agrees that using this system, would improve their productivity. The other 20% also agrees as well that the system would improve their productivity. 90% of the testers strongly agrees that using this system, would make their job or task easier. The other 10% also agrees as well with the statement. About the web application easiness of learning, 50% of the testers strongly agrees, and the other 50% only agrees that using this system is easy for them. According to the questionnaire result, the system is easy to use, since 70% of the testers strongly agrees, 20% agrees and the other 10% is neutral on the easiness of usage of the system. Majority of the students found this system is easy to use and does not requires a lot of mental effort on their part in using the system. For overall easiness of system usage, 60% of the students strongly agrees, 20% agrees and the other 20% is neutral. For attitude toward use, 70% of the students strongly agrees and 20% agrees on the statement given about the enjoyment of using the system. The other 10% reacts neutral to the statement given. Using the system is a positive experience for the participants, results in 80% of the students strongly agrees and other 20% agrees.

V. CONCLUSION

In summary, the website system has effectively realized its primary objectives, offering an intuitive interface for educational video text summarization and mind map generation. The front-end design, implemented with HTML, CSS, and JavaScript, ensures user-friendly navigation. Python, integrated into the backend through the Flask framework, facilitates video uploads, transcribing with Whisper OpenAI, and employs various summarization methods such as TextRank, SpaCY, BART, and distill-BART. The system seamlessly enables users to upload content, transcribe it, and generate both text summaries and mind maps.

Moreover, the implementation phase showcases the system's robustness by connecting the designed interface with backend processes. The choice of Whisper OpenAI for speech recognition, with its capability to handle extended inputs, underscores the system's adaptability. The incorporation of two extractive and two abstractive summarization methods caters to diverse user needs. Additionally, the use of an OpenAI plugin for mind map generation, albeit incurring a cost of $0.02 per token and API request, enriches the system's capabilities.

In the testing phase, the system undergoes comprehensive validation, encompassing unit testing, integration testing, system testing, and user acceptance testing (UAT). This exhaustive testing

process affirms the system's usability, functionality, and overall viability as a practical solution for educational video analysis and summarization..

REFERENCES

[1] Camburn, B., Arlitt, R., Anderson, D., Sanaei, R., Raviselam, S., Jensen, D., & Wood, K. L. (2020). Computer-aided mind map generation via crowdsourcing and machine learning. Research in Engineering Design (pp. 383-409). SpringerLink.

[2] Shimada, A., Okubo, F., Yin, C., & Ogata, H. (2018). Automatic Summarization of Lecture Slides for Enhanced Student Preview Technical Report and User Study. IEEE Transactions on Learning Technologies, 165-178.

[3] McIntyre, S. (2021). Mind Map Automation: Using Natural Language Processing to Graphically Represent a Portion of a U.S. History Textbook. University Honors CollegeMiddle Tennessee State University.

[4] Vimalaksha, A., Vinay, S., & Kumar, N. S. (2019). Hierarchical Mind Map Generation from Video Lectures. 2019 IEEE Tenth International Conference on Technology for Education (T4E) (pp. 110-113). IEEE.

[5] R. Hill, J. R., & Hannafin, M. J. (2001). Teaching and learning in digital environments: The resurgence of resource-based learning. Educational Technology Research and Development, 37–52

[6] Sah, S., Kulhare, S., & Allison Gray, S. V. (2017). Semantic Text Summarization of Long Videos. 2017 IEEE Winter Conference on Applications of Computer Vision (WACV)(pp. 989-997). Santa Rosa, CA, USA: IEEE

[7] Lin, C.-Y. (2004). ROUGE: A Package for Automatic Evaluation of Summaries. Barcelona: Association for Computational Linguistics.

[8] Ramakrishnan, A., & Ngan, C.-K. (2022). A Hybrid Video-to-Text Summarization Framework and Algorithm on Cascading Advanced Extractive- and Abstractive-based Approaches for Supporting Viewers' Video Navigation and Understanding. 2022 IEEE Fifth International Conference on Artificial Intelligence and Knowledge Engineering (AIKE)(pp. 36-39). Laguna Hills: IEEE.

[9] Awasthi, I., Gupta, K., Bhogal, P. S., Anand, S. S., & Soni, P. K. (2021). Natural Language Processing (NLP) based Text Summarization - A Survey. 2021 6th International Conference on Inventive Computation Technologies (ICICT) (pp. 1310-1317). Coimbatore: IEEE.

[10] Singh, P., Chhikara, P., & Singh, J. (2020). An Ensemble Approach for Extractive Text Summarization. 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE) (pp. 1-7). Vellore: IEEE.

[11] Arkbauer. (n.d.). What is Extreme Programming (XP). Retrieved from Arkbauer: https://arkbauer.com/blog/what-is-extreme-programming-xp/

[12] Hamilton, T. (2023, December 9). What is System Testing ? Retrieved from Guru99: https://www.guru99.com/system-testing.html