IMASHEDU: Intelligent MAshups for EDUcation - Towards a Data Mining Approach

Priscila Cedillo[1,2][a], Pablo Martínez León[1][b] and Marcos Orellana[1][c]
[1]*Computer Science Research & Development Lab - LIDI, Universidad del Azuay, Cuenca, Ecuador*
[2]*Universidad de Cuenca, Cuenca, Ecuador*

Keywords: Mashup, Learning Tool, Software, Web Applications.

Abstract: Nowadays, technological tools greatly support the work of teaching-learning tasks. In this sense, there are various sources of information from which teachers and students rely on to complement their academic activities. Content is sought on the web, significantly updated and easy to understand, generally in the form of videos. As people progress in their learning, they face terms, concepts, and topics that they are not familiar with them. However, those topics are included in the video. In this context, a complex process is generated of alternating sections of the video with other sources of information that explain the related topics and contribute to the understanding of the topic discussed. In this regard, and considering the possibility of systematically consuming information from various sources, it is necessary to build a method and an application that orchestrates the contents of these sources in a convenient, fast and automatic way, according to the person's learning. This proposal contemplates the development of a Mashup. This mashup integrates different data sources in a single graphical interface. Also, it is considered the construction of a core software solution based on text mining techniques. This solution allows extracting the textual content from videos and identifying the terms that could support the knowledge of the topic. It would significantly contribute to the fact that related topics are presented unified in the same interface. At the same time, the learning experience is greatly improved, avoiding losing the common thread of the observed video. Therefore, this article presents a process of orchestrating various data sources in a Web Mashup application. It includes videos available on YouTube channels, with other sources (e.g., Wikipedia, Pinterest) that help understand the topic better, generating hypertext references based on the generation of terms through text mining techniques. A Mathematics Learning mashup has been built to show the proposal's feasibility.

1 INTRODUCTION

Nowadays, within virtual learning, several tools are used for helping teachers and learners during their classes' preparation and study. Some of the most used information platforms are YouTube (Google LLC, 2022) and Wikipedia (Wikipedia Foundation, 2022). Those solutions are comprehensive sources of knowledge for any topic that people want to study and learn about. However, there is a problem when using different platforms: the amount of related information they provide for learners. Also, the dispersion of that information makes searching and integrating the knowledge difficult. In this context, when students want to learn about a topic, they go to a specific platform to search for a video on the subject needed; once they find it, this video can generate more doubts or concerns about different topics discussed, which can be in another platform or source. This causes the student to lose the continuity of the subject since the student has to go to several sources or websites containing information on a specific topic and find its explanation.

On the other hand, mashups are Web applications whose main objective is to integrate the contents and services provided by third-party components (Insfran et al., 2012). Thus, mashups have gained popularity due to their ease of creation. Also, mashups are

[a] https://orcid.org/0000-0002-6787-0655
[b] https://orcid.org/0000-0002-9269-346X
[c] https://orcid.org/0000-0002-3671-9362

Cedillo, P., León, P. and Orellana, M.
IMASHEDU: Intelligent MAshups for EDUcation - Towards a Data Mining Approach.
DOI: 10.5220/0011087500003182
In *Proceedings of the 14th International Conference on Computer Supported Education (CSEDU 2022) - Volume 1*, pages 383-388
ISBN: 978-989-758-562-3; ISSN: 2184-5026

specific applications usually presented as a single page. Then, considering the compositional character of mashups, they can be implemented to solve the problem above.

Although mashups represent a good solution, those tools have challenges related to the source's contents (e.g., availability, replaceability). The sources format that can be integrated into mashups can be RSS, ATOM feeds, Restful, SOAP Web Services, among others. Thus, several platforms provide convenient and free access to their information in those formats. Several platforms offer video and multimedia, which are very useful when teachers and learners need to integrate information for performing their activities. That integration needs to be automatic and based on recommendations from the contents displayed on the mashup (video, audio, text) to support users when preparing information to teach or learn. Moreover, data mining techniques can be integrated into the mashup creation to perform that integration. Therefore, this paper presents a process to create mashups that combine information conveniently from different data sources. The resulting mashups will help teachers and learnings during their activities. Those mashups will present related information without searching other sources but show all the information in a unique web application. To display the feasibility of this proposal, an example and a study case will be presented for evaluating the integration process and the use of the resulting mashup.

Finally, this document is structured as follows: Section 2 presents related work. Section 3 presents the process of the mashup creation, Section 4 shows the application of the process during the development of the mashup, and Section 5 shows the conclusions and further work.

2 RELATED WORK

Several studies focused on applying different techniques and mechanisms to create tools that help people improve their study methods (Burstein, 2009). However, previous mashups that have been made have no purpose within the field of education and do not have the techniques proposed in this study.

Many studies provide several ways to create new Mashups for different needs and aspects (Ghiani et al., 2016). Various components and tools must be considered when developing a Mashup by reusing existing components from different applications and combining them within the same solution.

Methodologically, some studies apply natural language processing and text mining techniques to find new topics from a base text (Devlin et al., 2019). These proposals are based on a similar process since characteristics are extracted from the transcribed texts, or the keywords of the entered text are classified. At the same time, some studies use Natural Language Processing (NLP) and processes for Artificial Intelligence (AI) to improve teaching and learning processes.

Several practical approaches to NLP are presented, which help in educational field and other linguistic aspects (Ferreira-Mello et al., 2019). Also, several effective solutions handle patterns of grammar and other linguistic approaches. Besides, NLP is an effective technique for improving students' ability to identify relationships of different words and use such within search engines to generate new knowledge (Campos, R et al., 2019). Therefore, a practical proposal allows students to use these tools optimally. The search procedure requires entering the correct information in the text to enter the next step. NLP enables the analysis of the information gathered from students, comparing it with the content within the search carried out (Burstein, 2009). Also, studies on extracting keywords and topics from a single document were considered without a document dictionary. From the study of YAKE! (Campos, R et al., 2019), a tool was obtained, which extracts keywords from a text from a simple document without a corpus. Based on the study of the attention mechanism (Vaswani, A et al., 2017), a technique called zero-shot classification was carried out, which associates observed and unobserved classes through auxiliary information, which encodes the distinctive observable properties of objects and detects types that the model has never seen during training. It is characterized by having the human-like ability to generalize and identify new things without direct supervision. Unlike previous studies where NLP mechanisms are used to classify data and web mashups that were created for different purposes (Wong et al., 2007), our objective is to extract the keywords from the text of a video or audio by using various APIs and generate new learning topics as explained in this through a mashup. In this way, it creates a helpful tool for the student to develop a learning path fluently and without losing attention to the topics of main interests.

3 CREATING INTELLIGENT LEARNING MASHUPS

The methodology is represented with the Software & Systems Process Engineering Meta-Model 2.0 (SPEM 2.0). The entire process is divided into three main tasks: i) Video metadata extraction, ii) Natural Language Processing (NLP) task, and iii) API links task. Each of which has input and output data. The first task (i.e., extraction of metadata from the video) has as input the link of a YouTube video. The NLP task transcribes the video obtained in the first task as input. Finally, the API links task has the keywords of the text obtained through text mining techniques as input, which leads to the output of new topics of the video that the user is observing.

A. Video Metadata Extraction

In this process, the multimedia information is extracted from the video platform (e.g., YouTube); this information serves to obtain the relevant data during the learning o teaching processes. In this task, the input is a transcription of the platform video, which is enabled in several sources and will collect the text to be analyzed.

Video transcription: The first step is extracting information obtained through a platform API (e.g., YouTube API). The process should focus on the transcription of the text of the video. It is necessary to transform these data into a document to carry out the respective processing of the obtained data. Once it is performed, available data is brought to continue with all the following steps.

Platform API: Using a platform API, all the necessary information is obtained, which will help continue with the next steps of the process.

B. Natural Language Processing Task

Here, the input artifacts are i) Transcribed text of the video platform (e.g., YouTube); the transcription is used to analyze the obtained text.

1) Transcribed text of the video platform: The transcribed text will be used to apply text mining techniques to search for knowledge and provide several topics that can be interesting for final users.

2) Text mining techniques: To arrive at a coherent result and to be able to classify the main topics that each video is about, several text mining techniques are used:

Stopwords: This technique reduces the dimensionality of texts by eliminating words that are not useful for the study.

Convert the text to lowercase: This technique is used to standardize texts.

Tokenization: This technique consists of constructing a list of tokens that do not include prevalent and uninformative words from a linguistic point of view, such as conjunctions (and, or, nor, that), prepositions (a, in, for, among others) and very common verbs.

YAKE!: It is a lightweight, unsupervised automatic keyword extraction method that relies on statistical characteristics of text extracted from individual documents to select the most important keywords from a reader (Campos, R et al., 2019). This system does not need to be trained with a particular set of documents, nor does it depend on dictionaries, external corpus, text size, language, or domain.

Then users are able to get the video's keywords, which will associate the relevant topics of the video with other sources of information on the web and generate new topics of interest.

C. API Links

The next step is to associate them with Wikipedia articles through the API provided by this virtual encyclopedia, which we will use to get associated topics regarding the topics of interest which were extracted in the previous steps via your search engine. After finding and obtaining similar topics within the Wikipedia platform, the zero-shot-classification technique is used based on the attention mechanism (Vaswani, A et al., 2017), which will help us determine the effectiveness of the extracted results. In this way, it will be possible to determine if the similar topics extracted through keywords and the Wikipedia API have a special relationship with the issues spoken in the main video that the user enters.

All this was joined by a pipeline to build a workflow for extracting topics associated with the YouTube video that the user chose. The text transcription of the video was inserted as a parameter. This flow will be implemented through a Mashup, which has the objective of showing the user all the new and suggested topics that may be of interest to them according to the video they enter, directing them to links to other websites where they will find more information about these recommended topics and thus improve the users' study flow.

Figure 1 shows the entire process towards developing a Mashup which includes data mining techniques to accomplish the objective of this study. Thus, the mashup development begins with selecting the components integrated into the Web application. It is necessary to explain that the central component should be a media container API; once the components that integrate the Mashup have been selected, they maintain an orchestration or choreography that reaches the main objective of this proposal.

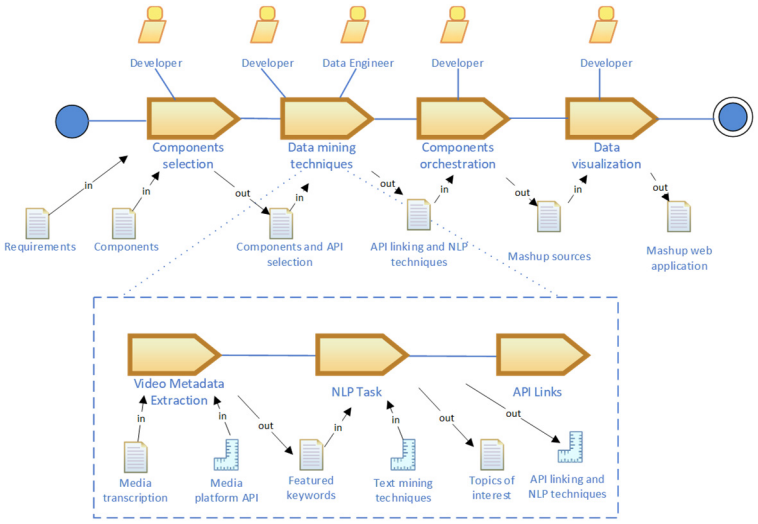

Figure 1: The process to be followed during the intelligent mashup's development.

4 APPLYING THE PROCESS

With the different techniques of text mining, NLP, and APIs used in this study, a Mashup was obtained. This application allows users to enter a video of interest from YouTube and generate topics according to the displayed video. Thus, watching a video with support while it is being played improves the flow of study and provides you with more information on these topics by directing you to Wikipedia articles to improve the way you study and answer the questions that arise when looking at new issues.

The use of text mining and NLP were critical aspects for verifying if the recommended topics were relevant to the main video. Therefore, it is possible to avoid showing issues that are not relevant or do not correspond to the theme of the video. Therefore, it is need a tool that provides more knowledge of various topics for learning. A proof of concept was carried out with different stages to demonstrate the usefulness and benefits of the Mashup obtained, which is presented below.

A. Linking the processes of the methodology in the Mashup. This study seeks to generate a Mashup that obtains new topics of interest related to a YouTube video that a person is watching. Here, it is expected to shorten study periods and create a linear study flow when searching for information within videos on YouTube.

With the Mashup made, it is possible to create this study flow and at the same time generate new topics from a related topic. This will be reflected in the next step from the workflow tests. The following shows the anchoring of the process designed in Figure 1 through the Mashup interface to reach the objectives and thus visualize a friendly and straightforward approach to show the results of the flow process to the user.

1) As shown in Figure 2, an interface was designed in which the user is allowed to enter the link to the YouTube video of interest. In this way, the flow enters the Video Metadata Extraction process, where the output is the transcribed text of the video entered by the user. This process is done by linking the YouTube API, where the transcription of the video and its different components are extracted. In this way, the flow is directed to the following process.

2) After obtaining the transcript as the output in the previous process, text mining and natural language processing techniques are performed to get the main topics related to the video entered by the user. These are linked through the Wikipedia API to obtain articles associated with the different topics extracted and thus orchestrate the different APIs. Therefore, the process flow shown in Figure 1 is completed, and the results are shown below.

3) As shown in Figure 3, the results obtained through the process flow of the methodology are displayed so that each of the main topics of the video entered by the user contains several Wikipedia articles related to it, where the user can automatically redirect to each of them by clicking on any of the displayed topics. Through all these steps, the process flow is linked through a Mashup, which finalizes the process of the methodology and achieves the objectives.

B. Testing and Process Flow

Next, the integration and flow of the Mashup processes as a final product will be demonstrated.

As shown in Figure 4, the first step is to extract the information from the URL of the YouTube video of the user's choice is inserted. Then, it has been linked the Video Metadata Extraction process, which will obtain the transcript of the video to connect with the NLP Task process.

Figure 2: Home page and YouTube link text field.

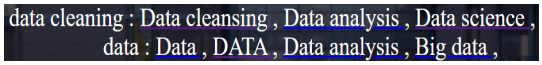

Figure 3: Results obtained within the Mashup.

Also, it is possible to extract the topics related to the video and their respective associated Wikipedia articles. After entering the YouTube video URL, it automatically redirects to the results page, where the keywords of the video are displayed with their related topics that may be of interest to the user. These topics have associated articles that automatically redirect to Wikipedia. Therefore, it is possible to connect the YouTube API with the Wikipedia API through text mining techniques and natural language processing, obtaining expected results and a Mashup that helps both the student and the teacher generate more sources of information on a topic they need. Finally achieved this process, all the explained steps are linked to complete the objective of the Mashup developing the desired tool. Therefore, the API Links process is finished, and this study's process flow culminates. Following this, the respective tests and analyzes that were carried out to verify the validity and usefulness of this study are explained, which demonstrates that the objectives of the study are met. Optimal construction of the mashup was reached; in this way, a tool for a better flow of study is available for both students and teachers.

C. Feedback

The participants belong to a group of Software and Data Engineers to collect their arguments and experiences with the final product. Different functional tests of the Mashup were carried out. It was determined that the results presented within the methodology obtained were consistent and accurate according to the entered search topics. As a general experience, it was argued that the reviewed product meets its primary objectives, both for integration and its use. The usability section contends that the Mashup generates a determined and fast study flow for direct purposes.

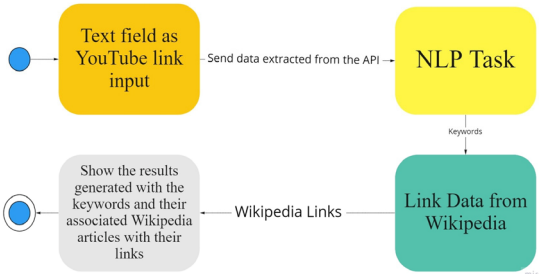

Figure 4: Flow Process.

D. Final Product

Once the Mashup meets the objectives and purposes of this study, a tool is provided to the community to help them search for information on the Web.

5 CONCLUSIONS

It was concluded that when using text mining techniques within the Mashups functions, beneficial information can be linked, which supports learners. Using the YAKE! Method: The extracted data analysis facilitated the analyzed transcripts' different topics, and the keywords can be identified with this technique. Also, NLP is essential when determining topics that will be taught as a final result of the study. With the use of the transcripts, new topics associated with the video were obtained from what was discussed. Once the result has been obtained, it is possible to expand the field of study and knowledge of subjects so that the flow of analysis of a person is improved. It is recommended that for future work, through the data collected with the use of the Mashup, methodologies can be applied to analyze the behavior of users when using the tool to improve the flow and results according to each type of search and topic. At the same time, a tool helps stakeholders to have a better understanding of the interest topics. Finally, it was possible to verify that, using techniques and methods of text mining and NLP, the obtained mashup combines those mentioned above with an orchestration of APIs, generating valuable topics for the user. Based on the results obtained, it was concluded that it is a great option to reach the objective and purpose of this study.

ACKNOWLEDGEMENTS

The authors wish to thank the Vice rector for Research of the Universidad del Azuay for the

financial and academic support and all the Laboratory for Research and Development in Informatics (LIDI) staff and the Department of Computer Science, Universidad de Cuenca. Specifically to the research project “Fog Computing applied to monitor devices used in assisted living environments.”, DIUC for its academic and financial support.

REFERENCES

Burstein, J. (2009). Opportunities for natural language processing research in education. Lecture Notes in Computer Science, 5449 LNCS, 6–27. https://doi.org/10.1007/978-3-642-00382-0_2

Campos, R., Mangaravite, V., Pasquali, A., Jorge, A., Nunes, C., & Jatowt, A. (2020). YAKE! Keyword extraction from single documents using multiple local features. Information Sciences, 509, 257–289. https://doi.org/10.1016/j.ins.2019.09.013

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). Bert: Pre-training of deep bidirectional transformers for language understanding. Conf. of the North American Association for Computational Linguistics, 2019: Human Language Technologies, 4171–4186.

Ennals, R., & Gay, D. (2007). User-friendly functional programming for web mashups. ACM SIGPLAN Notices, 42(9), 223–233. https://doi.org/10.1145/1291220.1291187

Ferreira-Mello, R., André, M., Pinheiro, A., Costa, E., & Romero, C. (2019). Text mining in education. Wiley Int. Reviews: Data Mining and Knowledge Discovery, 9(6). https://doi.org/10.1002/widm.1332

Ghiani, G., Paternò, F., Spano, L. D., & Pintori, G. (2016). An environment for End-User Dev. of Web mashups. Int. Journal of Human Comp. Studies, 87, 38–64. https://doi.org/10.1016/j.ijhcs.2015.10.008

Grammel, L., & Storey, M. A. (2010). A survey of mashup development environments. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 6400, 137–151. https://doi.org/10.1007/978-3-642-16599-3_10

Tuchinda, R., Knoblock, C. A., & Szekely, P. (2011). Building mashups by demonstration. ACM Transactions on the Web, 5(3). https://doi.org/10.1145/1993053.19 93058

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 2017-December (Nips), 5999–6009.

Wong, J., & Hong, J. I. (2007). Making mashups with marmite. https://doi.org/10.1145/1240624.1240842

Belleau, F., Nolin, M. A., Tourigny, N., Rigault, P., & Morissette, J. (2008). Bio2RDF: Towards a mashup to build bioinformatics knowledge systems. Journal of Biomedical Informatics, 41(5), 706–716. https://doi.org/10.1016/j.jbi.2008.03.004

Chary, M., Parikh, S., Manini, A. F., Boyer, E. W., & Radeos, M. (2019). A review of natural language processing in medical education. Western Journal of Emergency Medicine, 20(1), 78–86. https://doi.org/10.5811/westjem.2018.11.39725

Google LLC. (2022). *YouTube*. www.youtube.com

Grobelnik, M., Mladenic, D., & Jermol, M. (2002). Exploiting text mining in publishing and education. ICML-2002 Workshop on Data Mining Lessons Learned, 34–39, 2002.

Insfran, E., Cedillo, P., Fernández, A., Abrahão, S., & Matera, M. (2012). Evaluating the usability of mashups applications. *Proceedings - 2012 8th Int. Conference on the Quality of Information and Communications Technology, QUATIC*, 323–326. https://doi.org/10.1109/QUATIC.2012.28

Karami, A., White, C. N., Ford, K., Swan, S., & Yildiz Spinel, M. (2020). Unwanted advances in higher education:Uncovering sexual harassment experiences in academia with text mining. Information Processing and Management, 57(2), 102167. https://doi.org/10.1016/j.ipm.2019.102167

Litman, D. (2016). Natural language processing for enhancing teaching and learning. 30th Conference on Artificial Intelligence, AAAI 2016, 4170–4176.

Odden, T. O. B., Marin, A., & Caballero, M. D. (2020). Thematic analysis of 18 years of physics education research conference proceedings using natural language processing Thematic Analysis of 18 Years ... Odden, Marin, And Caballero. Physical Review Physics Education Research, 16(1).

Petersen, S. (2007). Natural Language Processing Tools for Reading Level Assessment and Text Simplification for Bilingual Education. 145.

Romero, C., & Ventura, S. (2013). Data mining in education. Data Mining and Knowledge Discovery, 3(1), 12–27. https://doi.org/10.1002/widm.1075

Wikipedia Foundation. (2022). Wikipedia. https://es.wikipedia.org/wiki/Wikipedia:Portada

Yunanto, A. A., Herumurti, D., Rochimah, S., & Kuswardayan, I. (2019). English education game using non-player characters based on natural language processing. Procedia Computer Science, 161, 502–508. https://doi.org/10.1016/j.procs. 2019.11.158

Ngu, A. H. H., Carlson, M. P., Sheng, Q. Z., & Paik, H. Y. (2010). Semantic-Based Mashup of Composite Applications. IEEE Transactions on Services Computing, 3(1), 2–15. https://doi.org/10.1109/TSC.2010.8