Links Related Generation to Transcribed Texts for YouTube Videos

Generación de Vínculos Relacionados a Textos Transcritos para Videos de YouTube

RTE
Revista Tecnológica
ESPOL

Pablo Martínez León
Laboratorio de Investigación y Desarrollo en Informática (LIDI), Ecuador
pdmleon@es.uazuay.edu.ec

Marcos Orellana
Laboratorio de Investigación y Desarrollo en Informática (LIDI), Ecuador
marore@uazuay.edu.ec

Abstract: The way people learn has undergone a significant transformation thanks to the advancement of technology. Various digital tools complement daily and academic activities, facilitating access to updated and diverse information. The Internet, in particular, has positioned itself as the primary source of information, offering a large amount of textual and audiovisual content, with short content in the form of videos being the most popular. However, the learning process inevitably involves acquiring new concepts and terms. When encountering unfamiliar vocabulary in videos, people often search for additional information to understand the content better. Therefore, this research seeks to develop a tool capable of analyzing video transcripts using Natural Language Processing techniques to identify key terms and relate them to other relevant information sources, thus facilitating learning. When evaluating the relevance of terms to the textual content of videos on various topics using an Artificial Intelligence model, a relevance greater than 75% was evidenced for all terms. This confirms the efficacy of this approach for analyzing and understanding the textual content of transcribed videos.

Keywords: Natural language processing, Text mining, Text classification, Web Mashup.

Resumen: La manera en que las personas aprenden ha experimentado una transformación significativa gracias al avance de la tecnología. Actualmente, una amplia variedad de herramientas digitales complementa las actividades cotidianas y académicas, facilitando el acceso a información actualizada y diversa. Internet, en particular, se ha posicionado como la principal fuente de información, ofreciendo una gran cantidad de contenidos textuales y audiovisuales, siendo los contenidos breves en forma de videos, los más populares. Sin embargo, el proceso de aprendizaje inevitablemente implica la adquisición

de nuevos conceptos y términos. Al encontrarse con vocabulario desconocido en los videos, las personas suelen buscar información adicional para comprender mejor el contenido. Por lo tanto, esta investigación busca desarrollar una herramienta capaz de analizar las transcripciones de videos utilizando técnicas de Procesamiento de Lenguaje Natural para identificar términos clave y relacionarlos con otras fuentes de información relevante, facilitando así el aprendizaje. Al evaluar la relevancia de los términos con el contenido textual en videos de distintas temáticas utilizando un modelo de Inteligencia Artificial, se evidencia una relevancia mayor al 75% en todos los términos, lo que confirma la eficacia de este enfoque para analizar y comprender el contenido textual de los videos transcritos.

Palabras clave: Categorización de texto, Minería de texto, Mashup web, Procesamiento de lenguaje natural.

Introduction

Technology has radically transformed the way we learn. Today, students and teachers have many digital resources to complement their academic activities. The Internet has become a notable source of up-to-date and accessible information, with videos being the preferred format for learning (Hsin & Cigas, 2013). However, this activity is interrupted when people encounter unfamiliar terms or concepts. This situation requires searching for additional information from other sources, which generates intermittent viewing of the video (Manasrah et al., 2021). Thus, developing a methodology and tool that integrates and organizes these contents coherently is imperative.

Thanks to advances in Natural Language Processing (NLP), analyzing and categorizing texts and videos available online is possible. This branch of Artificial Intelligence (AI) enables computer systems to understand and process human language similarly to people (Litman, 2016). NLP faces a considerable challenge due to the evolutionary nature of human language. Factors such as people's speaking and writing styles make training NLP models continuous and complex. These models improve by being trained with a large amount of data, such as Google Translate, which uses neural networks to translate full texts using the available context, achieving more accurate and fluid translations. This technique, called Zero-shot, translates directly from the original language to the target language (Vaswani et al., 2017).

Keywords serve as guides to finding relevant information in documents. They are helpful both for people searching for information and for NLP systems, such as those that generate summaries (Pal et al., 2013), categorize texts (Özgür et al., 2005), and extract information from texts (Marujo et al., 2011), among others. Additionally, automatic frameworks are proposed to extract keywords (Litvak, 2019; Medelyan et al., 2010; Turney, 2000). These systems were designed to work in ideal conditions, with structured and controlled data, such as news or texts from the Internet. However, their performance often decreases significantly when applied to more complex tasks such as translation, name recognition, or summary generation (Chang et al., 2014).

Text classification is an essential task in the field of NLP, allowing us to sort and categorize a wide variety of content, from questions to product reviews (Minaee et al., 2022). Although various methods effectively classify texts into previously labeled categories, their performance deteriorates significantly when faced with new classes not included in the previous training (Pourpanah et al., 2019). Zero-shot classification aims to classify elements that have not been seen

during training (Wang et al., 2019). This classification technique used in this study seeks to identify new topics and determine if they are related to the main topic of the text. Zhang et al. (2022) generate features for unseen classes based on lateral information, i.e., class-level attributes or text descriptions. Ye et al. (2020) also employ reinforced self-training techniques to take advantage of unlabeled data during training. By assuming that there is no information about unknown categories, they explore different methods to classify texts without labels in a generalized way.

Based on the previous studies that mentioned techniques for extracting keywords such as YAKE! and Zero-shot classification, this research proposes analyzing transcripts of videos obtained from the Internet (YouTube) through Application Programming Interfaces (APIs). First, the keywords are extracted. Then, new related topics are generated using Zero-shot classification. Finally, these topics are linked to Wikipedia articles and presented in a web mashup. The resulting system allows a more complete video content exploration and related terms.

Finally, this document is organized as follows: Section II contains the related works; Section III is the methodology used for execution; Section IV shows the results achieved; Section V addresses the evaluation; and finally, Section VI presents the conclusions of the work.

Related Works

Several studies are focused on applying different techniques and mechanisms to create tools that help people improve their study methods (Burstein, 2009). However, some studies have not used a combination of the proposed techniques. Methodologically, some studies apply NLP and text-mining techniques to find new topics from a base text (Devlin et al., 2018). These proposals are based on a similar process since characteristics are extracted from the transcribed texts, or keywords from the entered text are classified. At the same time, some studies use NLP and AI processes to improve teaching and learning processes.

Keyword extraction is essential for analyzing texts, allowing us to identify key points quickly. Shukla and Kakkar (2016) propose a method to extract keywords from transcripts of MOOCs (Massive Open Online Courses). Using grammatical rules and term frequency-inverse document frequency or Term Frequency and Inverse Document Frequency (TF-IDF) allows this method to highlight the most important concepts and facilitate the search for information in educational materials.

In the study of Zhang et al. (2022), an automatic keyword extraction algorithm using supervised machine learning was developed. Through experiments, they significantly improved the algorithm's performance by combining the predictions of three different models. The preprocessing of the data, performed with NLP tools, was crucial to the success of this research.

NLP has revolutionized the way students interact with information. By efficiently analyzing texts, NLP facilitates learning and the generation of new knowledge (Ferreira-Mello et al., 2019). In addition, it is a powerful tool that enhances learning by helping students understand, improve language structures, and effectively use search engines (Campos et al., 2020). Therefore, a properly designed proposal would allow students to use these tools effectively. Students' responses can be analyzed and compared with the search content by entering textual information into the system and identifying their coincidences and discrepancies (Burstein, 2009).

Vaswani et al. (2017) developed a Zero-shot classification technique that links known classes to unknown classes through auxiliary information. This information encodes the distinctive properties of the objects, allowing the model to identify types that were not present in the training set.

Unlike previous research that uses NLP tools such as TF-IDF to identify keywords in organized textual data, this study explores the possibility of extracting keywords from YouTube videos and obtaining their transcriptions through specialized APIs. Additionally, new learning topics linked to these keywords are generated with this information, providing links to related topics. A web mashup is developed, where these new links are displayed. In contrast to the research above, which is limited to identifying keywords, this study extracts keywords from video transcripts. It establishes meaningful connections between these words and external information sources, such as Wikipedia articles. In this way, a complete and contextualized vision of the topics addressed in the video is provided.

Materials and Methods

The exponential growth of data from various sources, such as transactions, sensors, and multimedia content, represents an unprecedented challenge. To deal with this massive volume of information, it is necessary to develop systems capable of automatically generating reports, views, or summaries, thus facilitating data-based decision-making.

The Software and Systems Process Engineering Meta-Model 2.0 (SPEM 2.0) has been used to structure and represent the methodology of this research. As seen in Figure 1, each phase is

detailed here in a sequential order. Each stage is characterized by its specific inputs and outputs. Likewise, the roles of those involved in each phase are identified. The methodology was divided into five main tasks: i) Transcripts extraction; ii) Text mining techniques application; iii) Wikipedia article extraction; iv) Topics of interest classification; and v) Mashup generation. Finally, the Technology Acceptance Model (TAM) was used to evaluate the acceptance of the results obtained using the new methodology. This model, widely used in information technologies, allows us to analyze the factors that influence users' decisions to adopt and use a new technology.

Figure 1

Phases of the Research Methodology represented with SPEM 2.0

Transcript Extraction

The elements necessary to extract relevant information from the selected videos were identified. The duration of the videos, according to the study by Manasrah et al. (2021),Click or tap here to enter text. was established between two to ten minutes. The source for extracting the videos is the YouTube platform, one of the most popular and mature video hosting platforms. In addition, this platform allows integration with several programming languages through APIs with different functionalities. Furthermore, the recovered videos are limited to this language because the AI model has been trained primarily with English content. Thus, the identified components include the YouTube video link and the APIs that allow access to its content and metadata.

Initially, the audio conversion of the videos to text was performed. For this, a specialized API, such as that of YouTube, was used to obtain an accurate transcription of the audiovisual content.

Subsequently, this transcript was formatted to facilitate analysis and processing.

The video platform API obtains all the information necessary for the subsequent stages of the process. This procedure was implemented using the Python programming language in version 3.10.11 and requires the installation of the "youtube-transcript-API" library in version 0.6.2; this library recommends a version of Python more significant than or equal to version 3.8.

The transcription process is illustrated with an example of how to get the transcript of the selected video using the "youtube-transcript-api" module:

```
$pip install youtube_transcript_api

from youtube_transcript_api import YouTubeTranscriptApi

YoutubeTranscriptApi.get_trasncript(video_id)
```

As an example, two YouTube videos are selected to evaluate the process:

- "What is Programming?" from the Khan Academy channel, with 3.4 million views, available at: https://www.youtube.com/watch?v=FCMxA3m_Imc

- "Factorizing Algebraic Expressions" from the Maths Explained channel, with 1.4 million views, available at: https://www.youtube.com/watch?v=ctqviXu-mTE

Once the video transcriptions were obtained, they were stored in text variables for later analysis. From these transcripts, themes related to the content of each video are extracted.

The video transcripts serve as a starting point for this analysis. Through them and using NLP techniques, it is possible to identify the relevant terms or words and obtain a more detailed understanding of the knowledge of each video.

Text Mining Technique Application

In this task, several text mining techniques were applied to extract the most information from the transcribed videos.

Pre-processing

Before applying the computational model, it is necessary to prepare the text data to maximize its effectiveness. This preprocessing process involves several tasks:

- Elimination of irrelevant words (stopwords):
  In this technique, common words such as "the," "that," and "a," which do not contribute significantly to the analysis, were removed. A search engine configured to ignore these words was employed (Shukla & Kakkar, 2016).

- Tokenization:
  For this process, a list of tokens is created that excludes prevalent and linguistically uninformative words, such as conjunctions (and, or, nor), prepositions (a, in, for), and common verbs (Campos et al., 2020).

- Keyword extraction with YAKE!
  YAKE! It is an automatic method to identify the most critical keywords in a text based on their statistical characteristics. It does not require prior training or depend on external dictionaries or corpora. In addition, it includes a measure of relevance in the keywords, which allows for the inclusion of a threshold for selecting them (Campos et al., 2020).

Keyword extraction

In this stage, the transcription obtained from the selected YouTube video from various sources is used to analyze its content and apply data pre-processing techniques. Next, the keyword extraction technique automatically identifies the most relevant and essential words and phrases within the text. In this way, the dimensionality of the data is reduced through pre-processing techniques and keyword extraction, facilitating the classification process and subsequent evaluation with the Artificial Intelligence model.

For this, the YAKE! Library was used in version 0.4.8 in Python (3.10.11), together with the transcription of the video selected as input. Afterward, the library was configured with the desired parameters (language, n-gram size, among others.). Finally, the keywords were extracted and stored in a text list, as presented in the following implementation:

```
$pip install git+https://github.com/LIAAD/yake

import yake

custom_kw_extractor    =    yake.KeywordExtractor(lan="en",    n=2,
dedupLim=0.9,    dedupFunc='seqm',    windowsSize=1,    top=10,
features=None)
```

```
keywords = custom_kw_extractor.extract_keywords(transcription)
```

Through the application of the YAKE! Tool, extracting between eight and ten of the most relevant keywords from each video transcription was possible. This set of key terms, along with the most significant n-grams, provided an overview of the content of each video and served as a starting point for deeper analyses. The frequency with which a word appeared in the transcript was directly proportional to the value assigned by YAKE!, which allowed keywords to be ordered according to their relative importance.

Continuing with the themes to be evaluated for the topic of Computer Science and Mathematics, the extracted keywords are presented in Table 1:

Table 1

Keywords extracted for each topic

COMPUTING	MATHEMATICS
"programming"	"common factor"
"program"	"highest common"
"programs"	"common divisor"
"computer"	"algebraic expressions"
"Code"	"factor analysis"
"process"	"lowest common factor"
"learn"	"expression (mathematics)"
"animations"	"factor solet"
"languages"	"closed-form expression"

Wikipedia Articles Extraction

Once the critical topics in the videos were identified, they were linked to relevant Wikipedia articles. To do so, the API of this virtual encyclopedia was used, which allows searching programmatically through its large search engine. Using the Wikipedia library, in version 0.6.0 for Python 3.10.11, an iterative cycle was established for

each keyword, and queries were performed with the API, obtaining a maximum of four relevant results.

Topics of Interest Classification

Text classification within NLP is a fundamental task where a model must predict the categories to which text documents belong. However, traditional classification methods require large amounts of labeled data to train the model and cannot be generalized to new data. Unlike traditional methods, the Zero-shot technique seeks to classify text documents without needing previously labeled data. This is achieved using large AI models based on the Transformers architecture, which are trained on general language understanding tasks.

This way, the pre-trained model "bart-large-mnli" from Meta was used, and a Natural Language Inference (NLI) task was trained on it. NLI involves determining whether two sentences are related, contradictory, or unrelated. Although this pre-trained model presents high performance in NLI tasks in English, it has yet to be trained on a Spanish language corpus, which limits its applicability to this language. The video transcription was proposed as a premise for classifying a topic, and a hypothesis was built for each candidate label, such as "This text is about mathematics" if the label is "Mathematics." Then, the pre-trained model was used to evaluate the relationship between the premise and the hypothesis, obtaining probabilities of implication and contradiction. These probabilities are converted into classification scores for each candidate label.

The prediction model returns indicate the probability that the video transcript is related to a candidate tag. A high score suggests a strong relationship, while a low score indicates the relationship is unlikely. The Zero-shot technique evaluated the relationship between the themes extracted from the video transcription and the suggested Wikipedia articles. Each pair's score was calculated, indicating the relevance and correspondence between the video content and the articles. The links to the articles related to the video were used as a starting point to explore other sources of information on the web, further expanding the understanding of the topics of interest.

Mashup Generation

Through the stages described above, a complete process has been created to extract topics associated with a YouTube video selected by the user. The video transcription was used as a starting point, and using text mining and NLP techniques, a processing flow implemented in a Mashup was created. The main objective of this

Mashup is to present the user with new topics of interest related to the selected video, providing links to articles on the web where more information can be found.

This study used Django, a high-level framework for the rapid development of websites with Python, to implement this process. This way, a graphical interface was developed that allows the user to enter the URL of the desired YouTube video and obtain related articles.

The Mashup processes the transcription by entering the video URL and presenting the related topics and relevant subtopics. Each topic and subtopic includes a direct link to a Wikipedia article to complement the information. For example, in the video with a computer theme, the main topics found were "*Programming*" and subtopics were "*Computer programming,*" "*Program,*' "*Programming language,*" and "*Code.*" In contrast, with the mathematics-themed video, the main topic was "*Common Factor,*' and the subtopics were "*Greatest common divisor,*" "*Common factors theory,*" "*Lowest common factor,*" and *.Factor analysis.*' When a subtopic is selected, it is automatically redirected to the corresponding Wikipedia article, completing the process and providing detailed information on the topics of interest.

Evaluation

The TAM model is based on three primary constructs: perceived usefulness, perceived ease of use, and attitude toward use. The objective is to evaluate the acceptance of the results of the proposed methodology, using the TAM and measuring perceptions of usefulness, ease of use, and attitude towards the use of the results through a seven-question questionnaire based on the five-point Likert scale. Additionally, the Goal Question Metric (GQM) was used to define the objectives of the case study.

In this way, a case study was performed in which the methodology was applied to a YouTube video about mathematics, and the results were evaluated manually and through a survey of experts in different areas. The GQM template defines the case study's objectives, as seen in Table 2.

Table 2

Definition of objectives of the case study

	QUESTIONS	ANSWERS
EVALUATE	What is being studied?	The result of applying the methodology to identify related topics from a mathematics video on YouTube through implementing and experimenting with text mining and NLP techniques and methods.
FOR THE PURPOSE OF	What is the intention of the study?	Evaluate the validity of the results obtained from the method.
WITH RESPECT TO	What effect is studied?	Related topics obtained from a video in the mathematics area.
FROM THE POINT OF VIEW OF	Who is affected? Refers to the study group	Systems and Data Science Engineer.
IN THE CONTEXT OF	Where, how, and by whom is the study performed?	Researchers from the Computer Science Research and Development Laboratory (LIDI) of the University of Azuay.

Subsequently, the variables of perception of usefulness, perception of ease of use, and attitude towards use are defined. Hypotheses were established, and expert surveys were conducted to evaluate these variables. The survey included questions about the effectiveness of the methodology results, the precision of the results, the ease of implementation, and the simplicity of the user interface, where a five-point Likert scale was used, the same as in indicator one, it refers to "Strongly disagree" and five "Totally agree".

Hypothesis

The hypotheses proposed for the evaluation are presented below:

- H1: The perception of the usefulness of the results is adequate for productivity or efficiency when finding related topics.

- H2: The results are easy to understand and use.

- H3: The results have significant value for the user.

The questionnaire for measuring variables with TAM included questions about the effectiveness, precision, ease of implementation, and simplicity of the user interface of the results.

Evaluation Execution

For the evaluation to be performed successfully, the experts were introduced to the methodology used and how the *Mashup generated* the results. Afterward, a demonstration and interaction were performed with each of the experts, who, based on the results obtained, performed the provided survey.

Analysis and Interpretation of Collected Data

The survey results included the variables' minimum values, maximums, and averages and the scale's reliability through Cronbach's Alpha coefficient, which varies between zero and one. The higher the value, the higher the internal reliability of the items.

Results and Discussion

The links to related Wikipedia articles were analyzed for their relevance to the video content. For this, the *Zero-shot* classification technique was used. This process was divided into three stages: i) data preparation, the video transcription is sent as a premise and the related topics as candidate labels to the classifier; ii) relationship analysis, the classifier evaluates the relationship between the premise and each candidate label, assigning a probability score; and iii) interpretation of the results, the scores are interpreted as percentages of relevance. A value of 100% indicates a high relationship between the topic and the video.
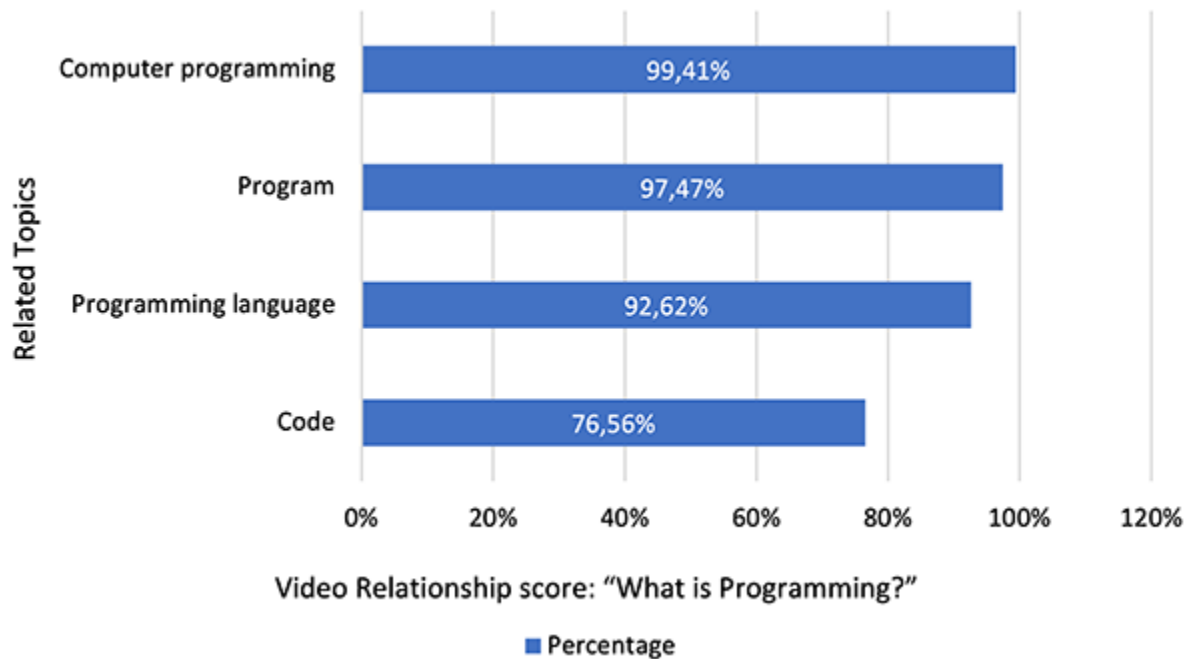
Figure 2

Search Results with Computing Topic

As we can observe in Figure 2, four different themes were analyzed: "*Computer Programming,*" "*Program,*" "*Programming Language,*" and "*Code,*" *classified in order of most significant relevance for the analyzed video on "What is Programming."* The topic "*Computer Programming*" obtained the highest relevance with 99.41%, followed by "*Program*" with 97.47%, the topic "*Programming Language*" with 92.62%, and "*Code*" had the slightest relevance with 76.56%, which suggests that its connection with the video is less significant. These results determined the most relevant and associated articles for the video. Subsequently, these results were transferred to the *Mashup as* a direct link to each Wikipedia article.
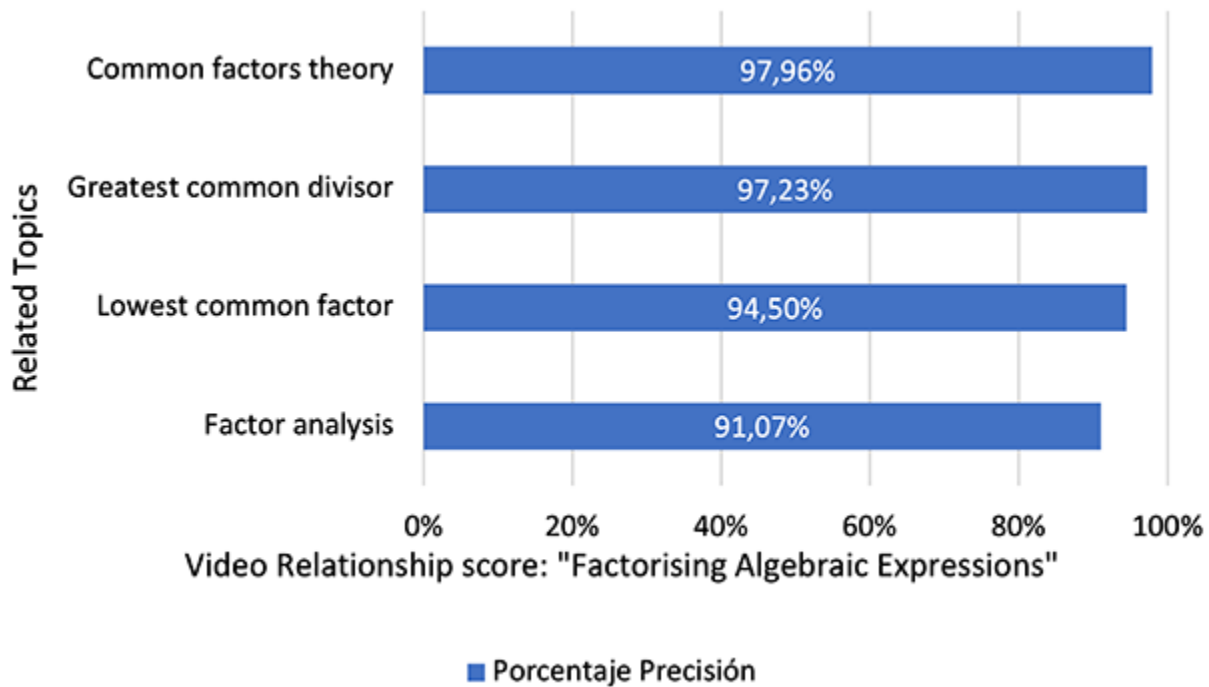
Figure 3

Search Results with Mathematics Topic

Figure 3 presents the analysis of four different topics: "Common Factors Theory," "Greatest Common Divisor," "Lowest Common Factor," and "Factor Analysis," classified in order of most significant relationship to the "Factorizing Algebraic Expressions." The most relevant topic was "*Common Factor Theory*" with 97.96%, followed by "*Greatest Common Divisor*" with 97.23%, and the topic "*Lowest Common Factor*" with 94.50%. *Moreover, "Factor Analysis'* presented a lower relevance of 91.07%. With these ratings, the topics most related to the video analyzed were determined to be viewed as part of the *Mashup*.

TAM Evaluation Results

Through the results obtained in the survey, Cronbach's alpha coefficient was used to evaluate the internal reliability of the questionnaire and its results. The consistency and correlation between the items that measure each variable are indicated to evaluate its validity. Below are the scales used to measure the level of reliability:

- 0.90 or higher: Excellent reliability. The items are highly correlated, and the variable is consistently measured.

- 0.80 to 0.89: Good reliability. The items are moderately correlated, and the variable is measured consistently.

- 0.70 to 0.79: Acceptable reliability. The items are moderately correlated, and the variable is measured acceptably.

- 0.60 to 0.69: Reliable in certain circumstances. The items are weakly correlated, and the variable is measured moderately.

- Less than 0.60: Low reliability. The items are weakly correlated, and the variable needs to be measured consistently.

The results indicated that participants perceived the results as practical, easy to use, and valuable. Cronbach's Alpha coefficient was 0.85, indicating solid variables and responses' internal consistency.

Validation of Hypotheses

In this section, the hypotheses raised at the beginning of the study are examined using the results obtained from the survey.

- Hypothesis 1: Are the results helpful in finding related topics?
The analysis of the average value for the variable "perception of the usefulness of the results" was 4.16, indicating that the participants consider the results obtained helpful in terms of productivity and efficiency, validating hypothesis 1.

- Hypothesis 2: Are the results easy to understand and use?
The average value for the variable "perceived ease of use of the results" was 4.5, indicating that participants consider the results easy to understand and use, supporting hypothesis 2.

- Hypothesis 3: Do the results generate trust in the user?
The average value for the variable "attitude toward use" was 4.13, indicating that participants perceived significant value in the results and trust in their operation, validating hypothesis 3.

The average results of the survey supported the three hypotheses stated. The participants perceived that the proposed methodology was practical, easy to use, and generated trust, which suggests positive acceptance by users. These results are a positive indicator that the proposed methodology has the potential to be a valuable tool for effectively identifying related topics in YouTube videos. However, several video themes could hinder the developed system, and the keyword extraction technique may not be the most effective in some cases. Therefore, it is necessary to investigate new techniques and

more advanced algorithms. Finally, the results of this research highlight the marked disparity between the quantity and quality of pre-trained language models available in Spanish and English. This gap shows the urgent need to develop more models in Spanish to democratize access to artificial intelligence technologies in the Spanish-speaking world.

Conclusions

This study implemented a process of identifying and extracting related topics from YouTube videos. Text mining and NLP techniques, such as YAKE! and Zero-shot classification, were combined to perform the task. Two specific subject areas were selected to evaluate the effectiveness of this methodology: computer science and mathematics. Experiments were performed in both fields to analyze the results obtained and compare the performance of the different techniques.

The results were obtained through the application of the YAKE! The tool accurately revealed the relevant themes in the analyzed videos' transcripts. This tool proved to be effective in extracting relevant keywords and concepts from poorly structured text. On the other hand, NLP was essential in preparing the data for all processes, reducing dimensionality, and improving its quality. The Zero-shot classification technique allowed us to objectively evaluate whether the extracted themes aligned with the video's general content. By assigning a score to each topic, it was possible to establish a relevance ranking and discard those unrelated to the original video. Visualizing results through a web mashup, which connected the identified topics with Wikipedia articles, provided a powerful tool for exploring and understanding the data, enriching its analysis. In conclusion, combining YAKE! NLP techniques, Zero-shot classification, and visualization through a web mashup allowed us to achieve the objectives proposed in this study. It was possible to accurately identify the main themes of the videos, evaluate their relevance, and provide an interactive tool for exploring related information.

Acknowledgments

References

Burstein, J. (2009). Opportunities for natural language processing research in education. *Computational Linguistics and Intelligent Text Processing: 10th International Conference, CICLing 2009, Mexico City, Mexico, March 1-7, 2009. Proceedings 10,* 6-27.

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

Chang, A., Savva, M., & Manning, C. D. (2014). Semantic parsing for text to 3d scene generation. *Proceedings of the ACL 2014 workshop on semantic parsing,* 17-21.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805.*

Ferreira-Mello, R., André, M., Pinheiro, A., Costa, E., & Romero, C. (2019). Text mining in education. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 9*(6), e1332.

Hsin, W.-J., & Cigas, J. (2013). Short videos improve student learning in online education. *J. Comput. Sci. Coll., 28*(5), 253-259.

Litman, D. (2016). Natural Language Processing for Enhancing Teaching and Learning. *Proceedings of the AAAI Conference on Artificial Intelligence, 30*(1). https://doi.org/10.1609/aaai.v30i1.9879

Litvak, M. (2019). *Deep Dive into Authorship Verification of Email Messages with Convolutional Neural Network* (pp. 129-136). https://doi.org/10.1007/978-3-030-11680-4_14

Manasrah, A. M., Masoud, M. Z., & Jaradat, Y. (2021). Short Videos, or Long Videos? A Study on the Ideal Video Length in Online Learning. *2021 International Conference on Information Technology (ICIT)*, 366-370. https://api.semanticscholar.org/CorpusID:236482423

Marujo, L., Grazina, N., Luis, T., Ling, W., Coheur, L., & Trancoso, I. (2011). BP2EP-adaptation of Brazilian Portuguese texts to European Portuguese. *Proceedings of the 15th Annual conference of the European Association for Machine Translation.*

Medelyan, O., Perrone, V., & Witten, I. H. (2010). Subject metadata support powered by Maui. *Proceedings of the 10th annual joint conference on Digital libraries*, 407-408. https://doi.org/10.1145/1816123.1816204

Minaee, S., Kalchbrenner, N., Cambria, E., Nikzad, N., Chenaghlu, M., & Gao, J. (2022). Deep Learning--based Text Classification. *ACM Computing Surveys, 54*(3), 1-40. https://doi.org/10.1145/3439726

Özgür, A., Özgür, L., & Güngör, T. (2005). *Text Categorization with Class-Based and Corpus-Based Keyword Selection* (pp. 606-615). https://doi.org/10.1007/11569596_63

Pal, S. K., Banerjee, R., Dutta, S., & Sarma, S. Sen. (2013). An Insight Into The Z-number Approach To CWW. *Fundamenta Informaticae, 124*(1-2), 197-229. https://doi.org/10.3233/FI-2013-831

Pourpanah, F., Lim, C. P., Wang, X., Tan, C. J., Seera, M., & Shi, Y. (2019). A hybrid model of fuzzy min–max and brain storm optimization for feature selection and data classification. *Neurocomputing, 333*, 440-451. https://doi.org/10.1016/j.neucom.2019.01.011

Shukla, H., & Kakkar, M. (2016). Keyword extraction from Educational Video transcripts using NLP techniques. *2016 6th International Conference - Cloud System and Big Data Engineering (Confluence)*, 105-108. https://doi.org/10.1109/CONFLUENCE.2016.7508096

Turney, P. D. (2000). Learning Algorithms for Keyphrase Extraction. *Information Retrieval, 2*(4), 303-336. https://doi.org/10.1023/A:1009976227802

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł. ukasz, & Polosukhin, I. (2017). Attention is All you Need. En I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), *Advances in Neural Information Processing Systems* (Vol. 30). Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf

Wang, W., Zheng, V. W., Yu, H., & Miao, C. (2019). A Survey of Zero-Shot Learning. *ACM Transactions on Intelligent Systems and Technology, 10*(2), 1-37. https://doi.org/10.1145/3293318

Ye, Z., Geng, Y., Chen, J., Chen, J., Xu, X., Zheng, S., Wang, F., Zhang, J., & Chen, H. (2020). Zero-shot Text Classification via Reinforced Self-training. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 3014-3024. https://doi.org/10.18653/v1/2020.acl-main.272

Zhang, Y., Yuan, C., Wang, X., Bai, Z., & Liu, Y. (2022). Learn to Adapt for Generalized Zero-Shot Text Classification. *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics*

(Volume 1: Long Papers), 517-527. https://doi.org/10.18653/v1/2022.acl-long.39

Pablo Martínez León, Marcos Orellana
**Links Related Generation to Transcribed Texts for
YouTube Videos**
Generación de Vínculos Relacionados a Textos Transcritos
para Videos de YouTube