



Structural review of relics tourism by text mining and machine learning

Subhankar Das
Duy Tan University, Vietnam

Subhra Mondal
Duy Tan University, Vietnam

Vikram Puri
Duy Tan University, Vietnam

Vasiliki Vrana
International Hellenic University, Greece

Abstract:

Purpose: The objective of the paper is to find trends of research in relic tourism-related topics. Specifically, this paper uncovers all published studies having latent issues with the keywords “relic tourism” from the Web of Science database.

Methods: A total of 109 published articles (2002-2021) were collected related to “relic tourism.” Machine learning tools were applied. Network analysis was used to highlight top researchers in this field, their citations, keyword clusters, and collaborative networks. Text analysis and Bidirectional Encoder Representation from Transformer (BERT) of artificial intelligence model were used to predict text or keyword-based topic reference in machine learning.

Results: All the papers are published basically on three primary keywords such as “relics,” “culture,” and “heritage.” Secondary keywords like “protection” and “development” also attract researchers to research this topic. The co-author network is highly significant for diverse authors, and geographically researchers from five countries are collaborating more on this topic.

Implications: Academically, future research can be predicated with dense keywords. Journals can bring more special issues related to the topic as relic tourism still has some unexplored areas.

Keywords: Text analysis, machine learning, artificial intelligence, topic modelling, relic tourism.

JEL Classification: Z00, Z11, Z32

Biographical notes: Subhankar Das (info.subhankardas@gmail.com) holds a Ph.D. in Social Media Advertisement & Media Planning and works as Assistant Professor in Duy Tan University, Vietnam with 14+ years of research experience. Subhra Mondal (subhra.mondal05@gmail.com) holds a Ph.D. in business management, working as assistant professor and researcher in Duy Tan University, Vietnam. She has experience with 20+ peer reviewed international publications in social media marketing, AI & Web 5.0, and digital recruitment. Vikram Puri (purivikram@duytan.edu.vn) is a Master Researcher at the Center of Simulation and Visualization, Duy Tan University, Vietnam, and Visiting Research Scholar at the University of Nevada, USA. He has published research articles in reputed indexed journals covering Machine and Deep Learning, Blockchain technologies and Cyber security. Vasiliki Vrana (vrana@ihu.gr) is a Professor at the Department of Business Administration, International Hellenic University, Greece. She has published, two books and more than 150 papers in academic journals, chapters internationally in ICT for tourism & hospitality, education assessment, Web 2.0 applications in tourism, health, and politics.

1 INTRODUCTION

Relics are historical objects that carry cultural value and relevance even in modern times despite having no current use (Cambridge Dictionary, 2021; Mugobi & Mlozi, 2021). They are the mirrors of the past for everyday people. The cultural influence of relics lies in people still believing in their importance (Alonsopérez et al., 2022). The belief and

existence in cultural folklores are the foundation of tangible heritage for humanity. Their conservation is always a topic of concern as the process is very time-consuming, and finance is always concerned. Restoration and protection of relics directly impact cultural and heritage tourism (Cerquetti & Ferrara, 2018; Manosso & Domareski Ruiz, 2021). Tourism research has progressed significantly and is now recognized as a distinct and diversified management sector (McKercher, 2018). Academic research journals on tourism



pump new life into the already crowded cauldron of management research, showing tourism research's diversity and evolution as a distinct academic discipline (Cripps, 2021; Scholtz & De Ridder, 2021). So, the topic "relics tourism" has gathered significance in the last two decades. Relics tourism is closely associated with cultural and heritage tourism in many aspects and is represented as a part of the culture in most tourism research. So, to understand the trends and explore all possible importance of different related topics in this subject, the authors investigate this topic.

Furthermore, McKercher (2018) stated that any new research targeting tourism management must enhance new avenues and explore hidden areas. Various scholars reviewed different aspects of tourism research to lend a helping guide for researchers and industry insiders (Manosso & Domareski, 2021; Mugobi & Mlozi, 2021; Sinclair-Maragh & Simpson, 2021; Daskalaki et al., 2020). There is no structural review of "relics tourism" available, which can help future researchers explore hidden areas linked with cultural and heritage tourism. So, here the researchers try to examine central databases for papers related to relic tourism.

According to McKercher (2018), researchers of different universities aim to publish approved journals in SSCI (social science citation index) to get more reach and enhance their reputation in the research fraternity. So, here we also believe that publications indexed in SSCI can give prominence to the published study and enrich the importance of authors. So here, we believe that indexing a journal in SSCI can significantly alter the course of published research in the journal and the authorship structure. Web of Science is the best source for SSCI journals. Thus, the Web of Science database is explored, and the authors collected papers mainly focused on relic tourism. So, here the researchers intend to

- a. Identify the topics and their unexplored relations in all publications by text classification modeling using ML
- b. Highlight prominent researchers and their network by bibliometrics analysis using Vos viewer.

The enormous amount of research papers available on the Web of Science is not the only problem; the more significant obstacle is a lack of other methodological rigor in the systematic review. Too many replication studies are available with different representations (Pratt et al., 2019). Most authors use Vos viewer software and bibliometrics to study citation, authorship, keywords, and geographical distribution of authorship. But besides all these, there is a strong need to explore the trends and patterns that text mining and ML applications can explore. ML tools help give more accuracy to information analysis as they have no biases in human judgment. The output of AI and ML-based text mining for reviews of research articles will have fewer critical errors. Here, we don't intend to say that text mining based on ML is devoid of all limitations or the best way to analyze. Here, we tried to give a more precise, unbiased output that will differ from all the existing tools that can be more explored. The AI tools use all the organic inputs which humans entered. ML analyzes these inputs to give a more unbiased output. AI took human entered data for learning and repeated the prejudiced organic thought that we popularly call ML analysis (LaGrandeur, 2020).

Recent developments in ML applications capture researchers' attention. In recent years scholars have started to use ML to review the literature and explore hidden trends ((Cabitza et

al., 2018; Sabahi & Parast, 2020). With AI and ML, researchers can conduct many text analyses and explore the complexities of information (Hamel et al., 2020). We can use classical bibliometrics analysis to explore the authorship, citation, keywords, geographical distribution, and co-authorship. But here, the authors want to explore the available research of "relics tourism" by text mining methods with the help of various tools of artificial intelligence (AI) and machine learning (ML). Here, the authors do not suggest that all the systematic review research should be done by AI and its ML tools replacing the human element of research. Instead, researchers of this study emphasize exploring AI and ML tools to explore text mining and bring out hidden patterns of social science studies.

Since social science, especially management and tourism research papers, are very diverse, future researchers should try AI and ML tools to explore and get output with less ambiguity. These AI and ML tools are predominantly used in medical imaging research to produce more specific results (Hamel et al., 2020). So, social science research will open new avenues if these tools are used to clarify topic identification. Our manuscript intends to explore this research gap of not having more structural and systematic reviews of papers in relic tourism and no previous use of text mining in such subjects by AI and ML tools. Generally, researchers explored heritage tourism by cultural aspect where they studied religious places, monuments, arts, history, dark events, and placed of cultural importance. But significantly relics tourism is not studied so much as compared to other aspects of heritage tourism.

We may say this area of heritage tourism is subsided under dark tourism or cultural heritage when this aspect has its own identity and significance in overall field of heritage tourism. Here in this study, we argue that tourism scholars will benefit from the evidence-based explanation of current keywords, patterns of topics, and network analysis for relics tourism. Our research results will help new and experienced researchers get more insights into using text mining techniques to identify trends and topics. This study will add more valid and reliable output to relic tourism studies. Researchers of cultural and heritage tourism who have less knowledge of ML and its use for text mining can also learn how to use the tools to explore patterns, topics, and keyword clusters in future studies. Here we review 109 articles from tourism indexed by the Web of science to demonstrate the text analysis. This analysis is done with the help of AI's ML tools, which are slowly capturing the researchers' logical minds rapidly. We are sure they will contribute significantly to the existing studies in the future. Our research will add more food for thought to tourism research and bring more opportunities to researchers in various related fields.

2 OVERVIEW OF TEXT MINING WITH ML AND AI

The recent developments in various research databases, web technologies, and digital documentation techniques due to AI and ML advancements attract the attention of researchers. Various researchers use text mining (TM) to explore hidden research trends in a particular area (Indurkha, 2015). But TM techniques face the biggest obstacle in meeting the standards of natural language processing (NLP). Researchers

consider text mining with the ever-growing demand to extract information from texts stored in digital resources (Haq et al., 2019). Manually it isn't easy to remove and analyze the texts. The availability of textual information from various research databases makes the analysis difficult manually or with software like Vosviewer. Though Vosviewer helps in network analysis for bibliometrics, it cannot determine the trend of research and keywords pattern (Salloum et al., 2017). It is now common to assume that AI can be the best alternate replacement for all human activities (Pi & Fan, 2021). Theoretically, researchers argue that a considerable amount of data can be analyzed by ML tools, but the complexity of interpretation can only be productive with human orientation (Zanzotto, 2019). Machines in the near past showed that they are catching up with the complex cognitive process of humans, like playing chess and making routing decisions. So, soon, machines may catch up with the decision-making ability of humans and more effectively perform a better analysis of massive data (Zuluaga et al., 2016). Due to rapid advancements in computation and data analysis, AI technologies with ML tools are better equipped to analyze, understand, and predict human textual language (Blanco-Ruiz et al., 2020). Despite a dynamic application in medical imaging literature studies, AI still finds itself in nascent social science and tourism (Watanabe et al., 2020). NLP and ML can significantly apply to systematic review studies (Porciello et al., 2020). NLP helps extract relevant academic studies from databases, and ML uses its tools for topic trend analysis. So, AI always helps study the trends of topics and keywords by text mining with tools like NLP and ML.

2.1 Text mining literature tools

a. Extraction of Information

It is the first step where researchers extract relevant information from the unstructured research articles (Linn, 1996). This step follows pre-defined text arrangements by matching the pattern. Here, researchers will select appropriate articles and link them with cooperative association with the terms (Waqas et al., 2018). Here researchers discover and collect information by differentiating relevant text, extracting the pertinent data, and converting them into functional forms (Krótkiewicz, 2018).

b. Discovery of Suitable Text

The Discovery of suitable text (DostoText) is the most critical aspect of text mining. It gathers structured data from unstructured texts (Clifton et al., 2004). Authors use tools of knowledge discovery from databases of research (KDD) to get structured, relevant data that can help in analysis ("Information Extraction from Text Using Text Mining," 2020). Authors use keyword extraction to classify text, make cluster terms, summarize topics, and give a pattern of the used terms ("Text Mining Technique for Driving Potentially Valuable Information from Text," 2020).

c. Text Analytics

Text analytics is an automated process that helps to interpret a large amount of unstructured text into qualitative data, especially for uncovering insights, trends, and patterns (Lakshmi & Baskar, 2021). The amalgam of text analytics and visualization tools better understands the problem and accommodates fruitful decision-making (Ren & Han, 2018). In this work, we apply text analytics to generate the word cloud and plot the most frequent terms used in the abstract

and title (Yusuf et al., 2017). The word cloud is the visual representation of the words associated with the text data. It is also used to highlight the words as per their frequency and relevance (Shivakumara et al., 2021). The steps of text analysis follow the specific steps mentioned below.

- I. Fetch word count from title and abstract: After successfully importing the dataset, both the words in the abstract and title must be counted. A lambda function is used to count the number of words. A lambda function is a small function that contains single expressions and can also act as an anonymous function where it doesn't need any name (Porter et al., 2020). The Lambda function consists of three parts: 1) keyword, 2) bound variable 3) expression. In this study, this function is deployed to fetch the number of words in every abstract and title (Lehr, 2019).
- II. Stopwords: Stopwords are a filter-out technique for natural language processing (Rüdiger et al., 2017). It is a technique that is used to pre-process the collected text. It helps to remove the most common words, such as prepositions and articles that do not have significant meaning in the text (Antons et al., 2020). In this step, we created a list of customs stopwords such as using, show, large, one, one, two, new, previous, and shown to set up a dictionary (Sarica & Luo, 2021).
- III. Remove the punctuations and special characters: The next step of text pre-processing initially converts all lower cases and removes the punctuations and characters. Afterward, tokenization is applied to convert strings into a list and remove the stopwords (Jatani, 2020). Stemming and lemmatization help remove the -ing, -ly and reclaim the word into root words. Again, in the final step, convert lists into strings (Sohrabi & Khaliljafarabad, 2018).
- IV. Word Cloud: It is a technique of visual representation of keywords depending on their structured frequency (Jayashankar & Sridaran, 2016). It is a prevalent method and is primarily acceptable by researchers. It is used to extract words from texts and help in detailed analysis. It also provides text and structured information (DePaolo & Wilkinson, 2014). It is more statistical than linguistic, allowing for summarization and providing little correlation among data (Sinclair & Cardew-Hall, 2007).

Here in this study, the authors observed that collecting and extracting information from the database has a limited scope with Vos Viewer-based literature review research papers. So, here we tried to employ text mining and some ML tools to extract patterns in keywords and project a holistic output (Albashir et al., 2020). Tourism and hospitality research has shown many enhancements in the last two decades. Many research articles are published, and journals have an upward journey for their impact and popularity. But in the scope of tourism, relics tourism has minimal research. Mostly they are related or embedded in culture or dark tourism. So, here we focus on relics tourism to extract more unexplored areas where researchers and journals can explore more. After carefully analyzing the available articles and literature in the Web of Science database for this study, the researchers assume that extraction of information and text mining tools were not applied to tourism. It demands a requisite effort where all the articles need to undergo text mining to explore

published topics and trends which can help industrial practitioners. Researchers want to explore the following research questions:

RQ1: What is the Word Cloud for titles & abstracts of all extracted publications?

RQ2: What high-frequency keywords are used in all extracted publications collected from Web of Science in their titles and abstracts?

RQ3: What are the MLM (Masked language modeling) & Next Sentence Prediction (NSP) in relics tourism?

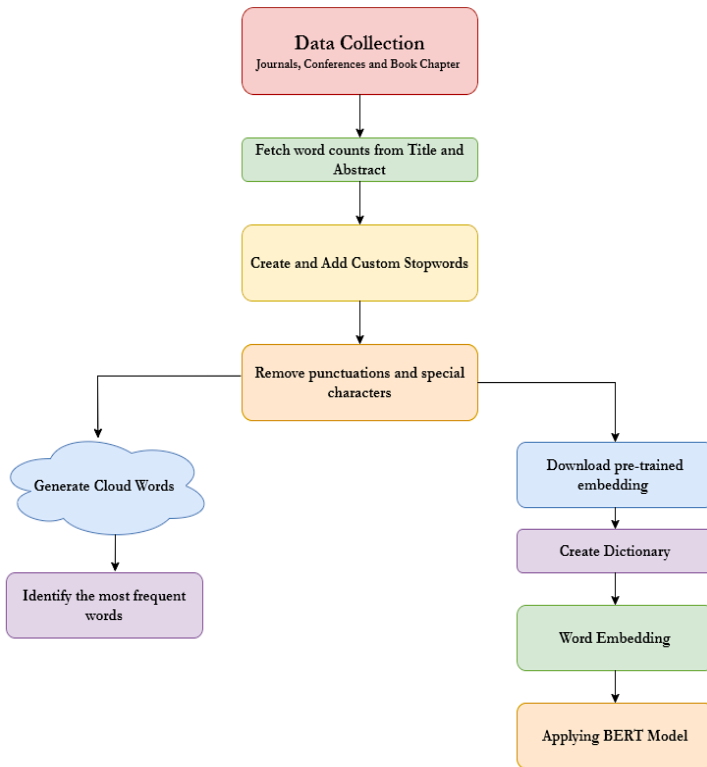
RQ4: Who are the prominent authors, and how does network analysis relate to them?

RQ5: What are the co-citation and citation networks among prominent researchers?

3 METHODOLOGY

For this study, researchers adopted the following architecture to identify the most frequent words used in the published relics tourism indexed in the Web of Science database. We use a combination of text analysis, text extraction, and text classification using AI models (Kabir et al., 2018). Two different approaches are used in this study: 1) Text Analytics and 2) BERT-based Text Classification.

Figure 1: Architecture of the proposed text miningsSolution



Source: Authors' conception

3.1 Data collection

Researchers accumulated all the research publications from the Web of Science. We found 143 articles with the keyword "relics tourism." Out of 143 papers, we exclude 34 from physical science backgrounds to reduce linguistic noise using the RapidMiner tool. Most search tools use text classification, an unsupervised learning technique that enables them to sort

documents into different categories. For instance, they can create a vector representation of a document that contains a set of unique words. For example, if a paper has a value of 0.5, it has a better chance of being categorized into a specific category. Through its algorithm, RobotAnalyst learns to classify each article according to its coefficients.

The screening process aims to recommend relevant academic literature based on the interaction between humans and the documents they encounter. This process uses text classification techniques to classify the articles according to their coefficients. A logistic regression model analyzes the results to predict new papers' relevance with the information collected. An example of this process is Rayyan, a tool researchers use to find relevant documents. Rayyan learns by analyzing the data it has collected. Once it has enough data to learn about the topic, it can recommend the most relevant papers based on the researcher's initial selection. Data extraction is a process utilized by machine learning programs to find and understand the text in documents. ML techniques then add tags to the words extracted from the papers. This process is referred to as data extraction. Aside from searching, robotreviewer also provides various tools for data extraction.

Table 1: All categories of publications considered for this study

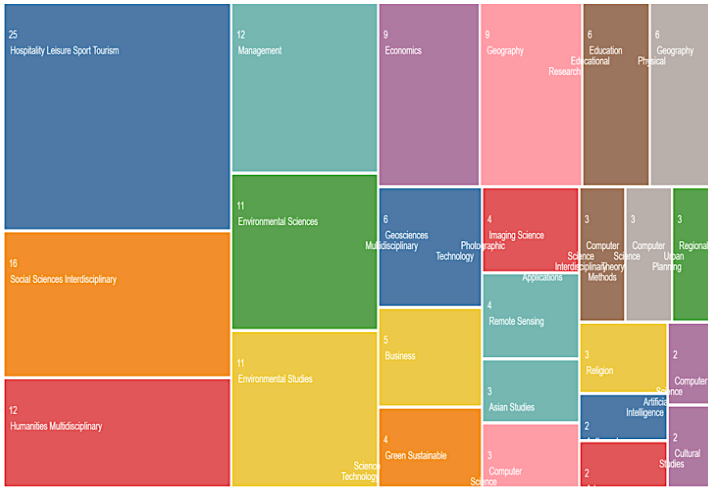
Serial No.	Web of Science Categories	Record Count	% Of 109
1	Hospitality Leisure Sport Tourism	25	22.936
2	Social Sciences Interdisciplinary	16	14.679
3	Humanities Multidisciplinary	12	11.009
4	Management	12	11.009
5	Environmental Sciences	11	10.092
6	Environmental Studies	11	10.092
7	Economics	9	8.257
8	Geography	9	8.257
9	Education Educational Research	6	5.505
10	Geography Physical	6	5.505
11	Geosciences Multidisciplinary	6	5.505
12	Business	5	4.587
13	Green Sustainable Science Technology	4	3.67
14	Imaging Science Photographic Technology	4	3.67
15	Remote Sensing	4	3.67
16	Asian Studies	3	2.752
17	Computer Science Information Systems	3	2.752
18	Computer Science Interdisciplinary Applications	3	2.752
19	Computer Science Theory Methods	3	2.752
20	Regional Urban Planning	3	2.752
21	Religion	3	2.752
22	Anthropology	2	1.835
23	Art	2	1.835
24	Computer Science Artificial Intelligence	2	1.835
25	Cultural Studies	2	1.835

Source: Web of Science

Finally, after cleaning the data, we got 109 articles. Here, researchers limit all the irrelevant characters to increase data quality. In this study, the authors first used tokenization to divide documents into tokens. In the next step, researchers apply the transformation process by creating the title of articles in lowercase. In the third step, authors filter stopwords. Each token signifies a single English word. Authors remove similarities between token and stopwords with the operator's help by arranging one-stop word per

article line. In the final step of the text mining process, researchers filter tokens as per length, where the minimum characters for a token are four and the maximum is twenty-five. Table 1 and figure 2 represent all the publication categories considered for this study.

Figure 2: Tree diagram of considered publications



Source: Web of Science

4 EXPERIMENTAL FINDINGS

After applying various text mining tools to all the extracted research publications, we got our desired results. Nevertheless, their tools have not yet been used in the publications concerned with relics tourism. So this reason makes this study and its findings unique and gives much value to the tourism research fraternity.

RQ1: What is the Word Cloud for titles & abstracts of all extracted publications?

A cloud is a collection of numerous words based on the frequency or importance of every expression referred to as a word cloud. Some literature points mainly to tourism, relics, and culture are more significant than the other words in the cloud. We can generate the word cloud for the "Abstract" and "Title" based on affiliation. Figure 3 and Figure 4 represent the word cloud for the "Title" and "Abstract," respectively. The keywords cloud is substantial because the frequency of the word in the text is either in "Title" or "Abstract." It is directly proportional to the size of the word in the cloud.

Figure 3: "Title" word cloud



Source: Analysis of data

Figure 4: "Abstract" word cloud

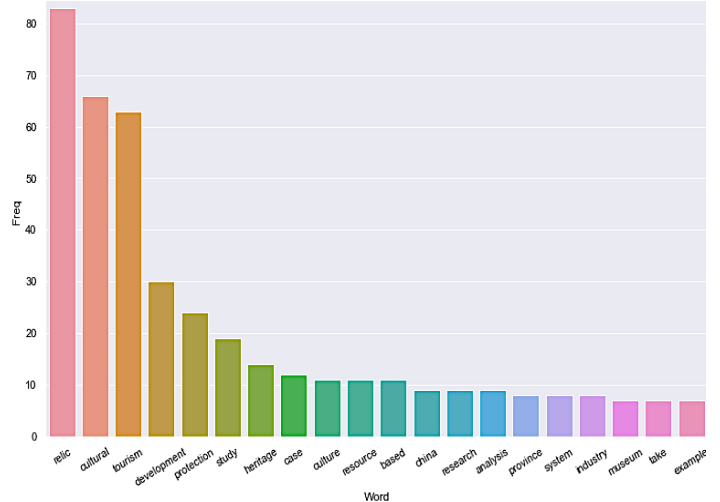


Source: Analysis of data

RQ2: What high-frequency keywords are used in all extracted publications collected from Web of Science in their titles and abstracts?

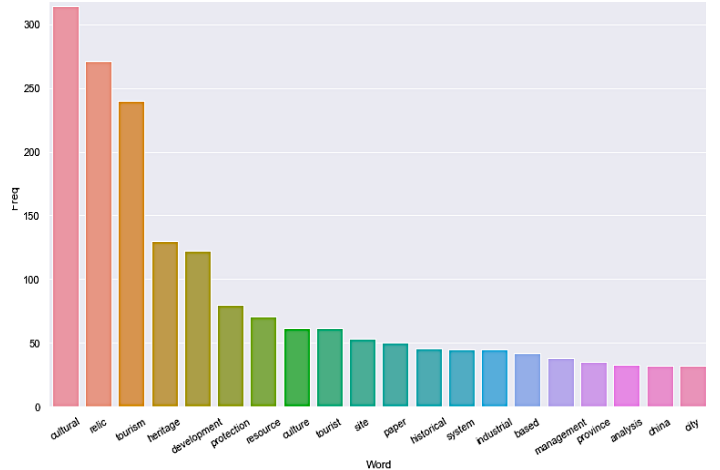
To identify the most frequent words, researchers used N-gram for this study. The N-gram model predicts the occurrence of a word based on its N-1 previous work (García et al., 2021). Figures 5 and 6 represent the word frequency analysis. Figure 5 illustrates the most frequent words in the "Title" section. The most frequent terms in the Titles are relic, culture, and tourism, and then is heritage and development.

Figure 5: Most frequently, words occur in the "Title" section



Source: Analysis of data

Figure 6: Most frequently, words occur in the "Abstract" section



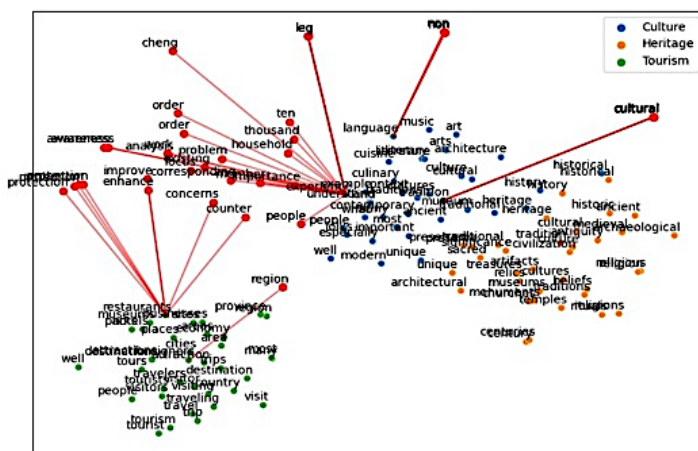
Source: Analysis of data

highest similarity score.

4. Output Prediction: The final step predicts the final model through the similarity check.

Figures 9 and 10 represent the model's similarity computation with "Abstract" and "Title." Figure 9 shows a certain similarity in using particular words by authors in the abstract. In this study, researchers found that the words in abstracts and titles have three clusters: culture, heritage, and tourism. Authors have used words like "cultural," "history," and "architecture" under the cluster culture. Researchers have used words like "protection," "religion," "belief," and "place" under heritage. Similarly, the cluster tourism has "place", "ancient", "sites", "travel", and "destination". These words have greater similarity and high frequency in their use inside the abstracts by various authors.

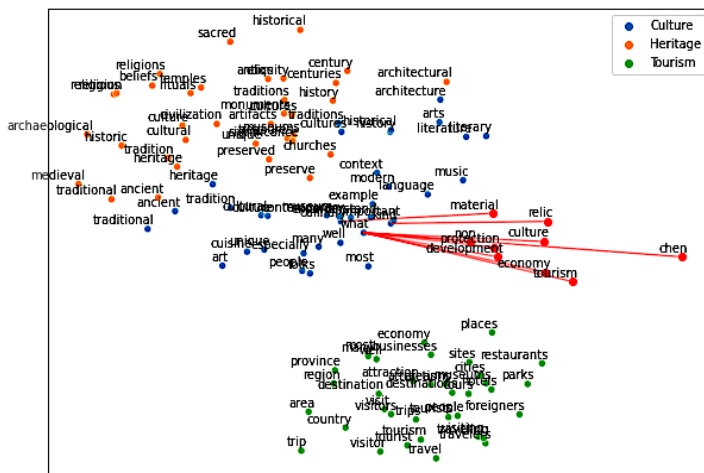
Figure 9: Similarity nodes of words used by authors in "Abstract"



Source: Analysis of data

Figure 10 represents the three clusters of words in all the titles. Authors used these words or similar ones primarily in their titles and restrained themselves from using words from the tourism cluster. Researchers prefer to use words from culture clusters more than heritage. So, we can say authors prefer to use words from culture as they might think the words are more related to relics.

Figure 10: Similarity nodes of words used by authors in "Title"

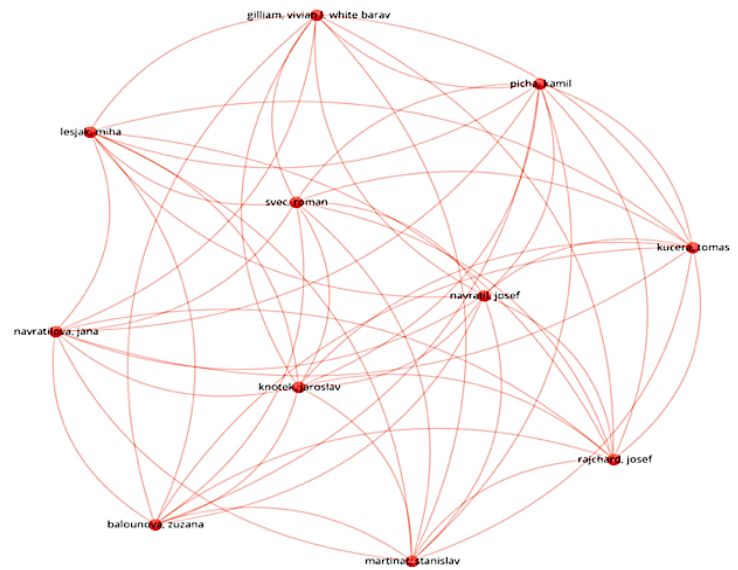


Source: Analysis of data

RQ4: Who are the prominent authors, and how does network analysis relate to them?

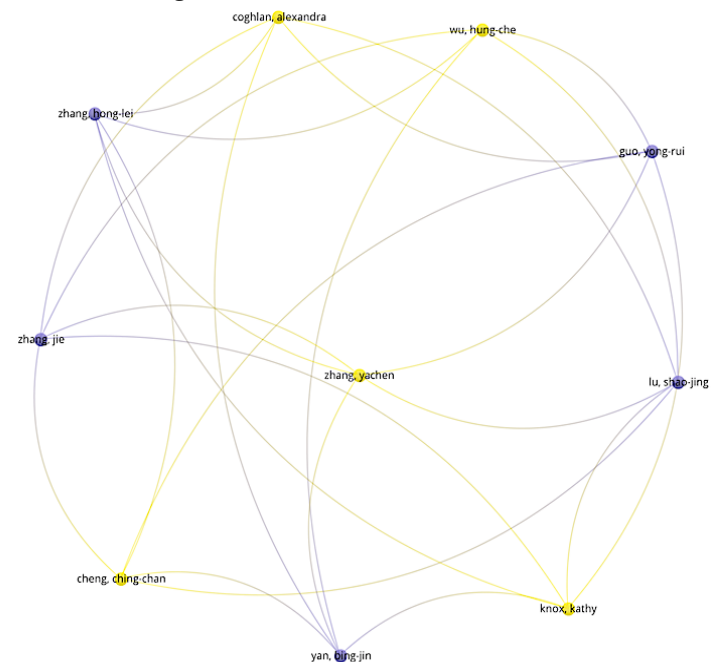
In these years (2002-2021), co-authors publish and have a strong network. Eleven Authors show strong weighted strength among themselves, as shown in figure 11. They are geographically distributed in five countries: Ghana, Spain, the People's Republic of China, Taiwan, and Slovenia. All these co-authors have 40 papers among them. Citation-wise, if we see, then Ghana and Spain co-authors have 39 citations, Taiwanese authors have 79, and Slovenia authors have 11 citations. Chinese authors are having maximum citation of 124. It shows that Chinese researchers are active in this subject and have a robust network among themselves, enhancing their total output.

Figure 11: Network of prominent researchers



Source: Analysis of Data

Figure 12: Authors and their citations



Source: Analysis of Data

RQ5: What are the co-citation and citation networks among prominent researchers?

Figure 12 represents the citation network of prominent authors on this subject of relics tourism. We can see there are 2 clusters (yellow and blue) of authors and their citation network. Yellow clustered authors have maximum strength, especially "zhang, y," "zhang, j," and "Huo, x." The blue clustered authors have more minor citations than the yellow ones. We can say yellow ones are popular in this field.

5 DISCUSSION AND CONCLUSIONS

We can see how the keywords revolve around three clusters of culture, heritage, and tourism from the analysis. Mainly the words in the title and similar and present in the abstract. Interestingly, some words like sacred and religion are also used with protection, value, and belief. Also, some words like the Internet of things (IoT) and sensors are present with protection. So, we can say there may be a slow trend developing which points towards modern technology use for relics tourism domain by researchers.

Similarly, for heritage studies, many recent papers are exploring a new trend post-2019. Words like the place and destinations are now linked with heritage directly, which were previously associated with tourism (Rezaei et al., 2021; Duy et al., 2020). So, we can say the marketing of places is now exploring relics tourism. And for tourism, the words do not have a drastic change. They are the same throughout the years of consideration (2002-2021).

Authors of prominent publications are very much strongly interrelated with each other. They have an excellent connection in the field of cultural tourism. Relics tourism is directly represented under cultural topics. But authors explore new technology applications when we consider publications and protection words in the title and abstract. Generally, authors have used the terms culture, relics, and tourism in the abstracts more. Culture has a maximum frequency in abstracts. Similarly, the term relics has a maximum frequency in titles than culture and tourism.

Here, we find that relics certainly rally for cultural tourism in all the selected articles. Researchers signify relics with culture and religion. Relics connect with tourism with words like destination, place, and religion (Van et al., 2020). The findings of this research help future scholars in terms of guidance in employing ML in text mining (Singh et al., 2020). This technique will help in enhancing their research ability. As several research publications are published in various streams, conventional research for literature review is fast becoming more impractical and complex, which requires a lot of time and effort from the researcher (Sharma & Das, 2021). Data extraction and a reduction will be smoother and more technical with text mining by ML tools. The ability to review the title, abstract and full texts of an article within the same ML tools will undoubtedly save valuable time than other contemporary methods (Das et al, 2022).

Consequently, applying ML tools for text mining brings a semi-automatic alternative to the systematic literature review research repertoire. It is a fact that ML will undoubtedly decrease the human effort in the systematic literature review,

but still, human interventions are required. So, ultimately, we can hope one day, the features of ML will be upgraded, which will make the process fully automated and limit human efforts ultimately. So, the scope of performing text analysis of various databases will undoubtedly increase the possibility of a more reliable, complete systematic review. These will surely contribute to future research endeavors not only in tourism disciplines but also in others too.

5.1 Practical implications

In this novel study, the researchers found the trend of relics tourism study that different researchers can use to form their titles and abstracts more effectively. Journals can use this research to propose special issues for bringing out more concerning research in relics tourism. Researchers generally venture out for cultural and heritage studies. They are not so keen on relics tourism. However, they go for dark tourism research, a different domain. Though dark tourism is related to this scope of relics, it is other than relics. So, authors can now mainly be precise in their research of these areas. Like cultural, heritage, or dark tourism, relics tourism attracts many future researchers.

The technique of text mining and its machine learning applications can also attract systematic literature review researchers for better exploration. Here the researchers do believe that the future is for machines. So, ML will slowly venture into tourism research, and researchers will benefit from this technology intrusion. Most contemporary researchers use hybrid methods and Vos viewer software for bibliography studies which are command-based. But text mining and BERT are Programme based and more accurate. So, we think future researchers can explore different topics for educational use and practically explore more ML tools.

This paper aims to help general management and tourism scholars put ML into practice. It will provide them with the necessary characteristics to make informed decisions. We summarize these points as follows.

1. The type of literature review an ML system can perform differs depending on the software used. There are three main types of literature reviews that it can perform: searching, data extraction, and screening.
2. To choose the right tool for a particular type of review analysis, choose a set of tools with the necessary features.
3. Free and paid tools are also available. However, the choice should be based on the scope of the research.
4. Before using a free tool, make sure that it has a trial period. Doing so will allow you to try it out and see if it works well. Some free tools might not allow you to add more data to your project.

5.2 Limitations and future research

We have considered 109 articles for this study, whereas Web of Science has 143 papers on relics tourism search. So, there may be some analytical output we may miss. But that will be another scope of research where researchers can study social and physical science in relics tourism. We have considered text mining and keyword word cloud for our study with BERT. There are other tools available in ML. But we are not conversant with other devices, so researchers can explore different mechanisms that may give more in-depth analysis in the future. Also, the citation and authorship analysis have

not included the journal's prestige or popularity, which can be a limitation.

REFERENCES

- Albashir, T. M., Alzoubi, H., & Albatainih, M. (2020). Improving Arabic Instant Machine Translation: The Case of Arabic Triangle of Language. *Journal of Computer Science*, 16(7), 956–965. <https://doi.org/10.3844/jcssp.2020.956.965>
- Alonsopérez, M. J., Brida, J. G., & Rojas, M. L. (2022). Second homes: A bibliometric analysis and systematic literature review. *Journal of Tourism, Heritage & Services Marketing*, 8(1), 16–26. <https://doi.org/10.5281/zenodo.6581499>
- Antons, D., Grünwald, E., Cichy, P., & Salge, T. O. (2020). The application of text mining methods in innovation research: current state, evolution patterns, and development priorities. *R&D Management*, 50(3), 329–351. <https://doi.org/10.1111/radm.12408>
- Blanco-Ruiz, M., Sainz-de-Baranda, C., Gutiérrez-Martín, L., Romero-Perales, E., & López-Ongil, C. (2020). Emotion Elicitation Under Audiovisual Stimuli Reception: Should Artificial Intelligence Consider the Gender Perspective? *International Journal of Environmental Research and Public Health*, 17(22), 8534. <https://doi.org/10.3390/ijerph17228534>
- Cabitz, F., Locoro, A., & Banfi, G. (2018). Machine Learning in Orthopedics: A Literature Review. *Frontiers in Bioengineering and Biotechnology*, 6. <https://doi.org/10.3389/fbioe.2018.00075>
- Cambridge Dictionary. (2021, November 3). relic definition: 1. an object, tradition, or system from the past that continues to exist: 2. a part of the body or. . . Learn more. <https://dictionary.cambridge.org/>
- Cerquetti, M., & Ferrara, C. (2018). Marketing Research for Cultural Heritage Conservation and Sustainability: Lessons from the Field. *Sustainability*, 10(3), 774. <https://doi.org/10.3390/su10030774>
- Clifton, C., Cooley, R., & Rennie, J. (2004). TopCat: data mining for topic identification in a text corpus. *IEEE Transactions on Knowledge and Data Engineering*, 16(8), 949–964. <https://doi.org/10.1109/tkde.2004.32>
- Cripps, K. (2021). Tourism, terrorism and security: Tourism security-safety and post conflict destinations. *Journal of Tourism, Heritage & Services Marketing*, 7(2), 74–76. <https://doi.org/10.5281/zenodo.5550929>
- Das, S., Mondal, S. R., & Sandhu, K. (2022). Music logos drive digital brands: an empirical analysis of consumers' perspective. *Journal of Strategic Marketing*, 1-16. DOI: 10.1080/0965254X.2022.2098526
- Daskalaki, V. V., Voutsas, M. C., Boutsouki, C., & Hatzithomas, L. (2020). Service quality, visitor satisfaction and future behavior in the museum sector. *Journal of Tourism, Heritage & Services Marketing (JTHSM)*, 6(1), 3-8. DOI: 10.5281/zenodo.3603167
- DePaolo, C. A., & Wilkinson, K. (2014). Get Your Head into the Clouds: Using Word Clouds for Analyzing Qualitative Assessment Data. *TechTrends*, 58(3), 38–44. <https://doi.org/10.1007/s11528-014-0750-9>
- 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*.
- Duy, N. T., Mondal, S. R., Van, N. T. T., Dzong, P. T., Minh, D. X. H., & Das, S. (2020). A Study on the Role of Web 4.0 and 5.0 in the Sustainable Tourism Ecosystem of Ho Chi Minh City, Vietnam. *Sustainability*, 12(17), 7140. <https://doi.org/10.3390/su12177140>
- García, M., Maldonado, S., & Vairetti, C. (2021). Efficient n-gram construction for text categorization using feature selection techniques. *Intelligent Data Analysis*, 25(3), 509–525. <https://doi.org/10.3233/ida-205154>
- Gensim: topic modelling for humans. (n.d.). <https://radimrehurek.com/Gensim/apiref.html>
- Hamel, C., Kelly, S. E., Thavorn, K., Rice, D. B., Wells, G. A., & Hutton, B. (2020). An evaluation of DistillerSR's machine learning-based prioritization tool for title/abstract screening – impact on reviewer-relevant outcomes. *BMC Medical Research Methodology*, 20(1). <https://doi.org/10.1186/s12874-020-01129-1>
- Haq, M. I. U., Li, Q., & Hassan, S. (2019). Text Mining Techniques to Capture Facts for Cloud Computing Adoption and Big Data Processing. *IEEE Access*, 7, 162254–162267. <https://doi.org/10.1109/access.2019.2950045>
- Ho, Q. T., Nguyen, T. T. D., Khanh Le, N. Q., & Ou, Y. Y. (2021). FAD-BERT: Improved prediction of FAD binding sites using pre-training of deep bidirectional transformers. *Computers in Biology and Medicine*, 131, 104258. <https://doi.org/10.1016/j.compbiomed.2021.104258>
- Indurkha, N. (2015). Emerging directions in predictive text mining. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 5(4), 155–164. <https://doi.org/10.1002/widm.1154>
- Information Extraction from Text using Text Mining. (2020). *International Journal of Innovative Technology and Exploring Engineering*, 9(2S5), 22–24. <https://doi.org/10.35940/ijitee.b1006.1292s519>
- Jatain, A. (2020). Data Mining Applications for Predicting the COVID-19 Pattern. *International Journal of Data Mining And Emerging Technologies*, 10(2), 55–62. <https://doi.org/10.5958/2249-3220.2020.00008.7>
- Jayashankar, S., & Sridaran, R. (2016). Superlative model using word cloud for short answers evaluation in eLearning. *Education and Information Technologies*, 22(5), 2383–2402. <https://doi.org/10.1007/s10639-016-9547-0>
- Kabir, A. I., Karim, R., Newaz, S., & Hossain, M. I. (2018). The Power of Social Media Analytics: Text Analytics Based on Sentiment Analysis and Word Clouds on R. *Informatica Economica*, 22(1/2018), 25–38. <https://doi.org/10.12948/issn14531305/22.1.2018.03>
- Korstanje, M., & Séraphin, H. (Eds.). (2020). *Tourism, Terrorism and Security*. Emerald Publishing Limited.
- Krótkiewicz, M. (2018). A novel inheritance mechanism for modeling knowledge representation systems. *Computer Science and Information Systems*, 15(1), 51–78. <https://doi.org/10.2298/csis170630046k>
- LaGrandeur, K. (2020). How safe is our reliance on AI, and should we regulate it? *AI and Ethics*, 1(2), 93–99. <https://doi.org/10.1007/s43681-020-00010-7>
- Lakshmi, R., & Baskar, S. (2021). Efficient text document clustering with new similarity measures. *International Journal of Business Intelligence and Data Mining*, 18(1), 49. <https://doi.org/10.1504/ijbidm.2021.111741>
- Lehr, D. (2019). Guest Editorial: Completion-System Reliability: Where Have We Been and Where Are We Going? *Journal of Petroleum Technology*, 71(03), 14–15. <https://doi.org/10.2118/0319-0014-jpt>
- Linn, M. C. (1996). Key to the information highway. *Communications of the ACM*, 39(4), 34–35. <https://doi.org/10.1145/227210.227219>
- Manosso, F. C., & Domareski Ruiz, T. C. (2021). Using sentiment analysis in tourism research: A systematic, bibliometric, and integrative review. *Journal of Tourism, Heritage & Services Marketing*, 7(2), 16–27. <https://doi.org/10.5281/zenodo.5548426>
- McKercher, B. (2018). What is the state of hospitality and tourism research – 2018? *International Journal of Contemporary*

- Hospitality Management, 30(3), 1234–1244. <https://doi.org/10.1108/ijchm-12-2017-0809>
- Mugobi, T., & Mlozi, S. (2021). The impact of external factors on ICT usage practices at UNESCO World Heritage Sites. *Journal of Tourism, Heritage & Services Marketing*, 7(1), 3–12. <https://doi.org/10.5281/zenodo.4514800>
- Park, D., Kim, S., Lee, J., Choo, J., Diakopoulos, N., & Elmqvist, N. (2018). ConceptVector: Text Visual Analytics via Interactive Lexicon Building Using Word Embedding. *IEEE Transactions on Visualization and Computer Graphics*, 24(1), 361–370. <https://doi.org/10.1109/tvcg.2017.2744478>
- Pi, J., & Fan, Y. (2021). The impact of robots on equilibrium unemployment of unionized workers. *International Review of Economics & Finance*, 71, 663–675. <https://doi.org/10.1016/j.iref.2020.10.003>
- Porciello, J., Ivanina, M., Islam, M., Einarson, S., & Hirsh, H. (2020). Accelerating evidence-informed decision-making for the Sustainable Development Goals using machine learning. *Nature Machine Intelligence*, 2(10), 559–565. <https://doi.org/10.1038/s42256-020-00235-5>
- Porter, A. L., Chiavetta, D., & Newman, N. C. (2020). Measuring tech emergence: A contest. *Technological Forecasting and Social Change*, 159, 120176. <https://doi.org/10.1016/j.techfore.2020.120176>
- Pratt, M. G., Kaplan, S., & Whittington, R. (2019). Editorial Essay: The Tumult over Transparency: Decoupling Transparency from Replication in Establishing Trustworthy Qualitative Research. *Administrative Science Quarterly*, 65(1), 1–19. <https://doi.org/10.1177/0001839219887663>
- Ren, X., & Han, J. (2018). Mining Structures of Factual Knowledge from Text: An Effort-Light Approach. *Synthesis Lectures on Data Mining and Knowledge Discovery*, 10(1), 1–199. <https://doi.org/10.2200/s00860ed1v01y201806dmk015>
- Rezaei, N., Ghaderi, Z., & Ghanipour, M. (2021). Heritage tourism and place making: investigating the users' perspectives towards Sa'd al-Saltaneh Caravanserai in Qazvin, Iran. *Journal of Heritage Tourism*, 1–18. <https://doi.org/10.1080/1743873x.2021.1998076>
- Rüdiger, M., Antons, D., & Salge, T. O. (2017). From Text to Data: On The Role and Effect of Text Pre-Processing in Text Mining Research. *Academy of Management Proceedings*, 2017(1), 16353. <https://doi.org/10.5465/ambpp.2017.16353abstract>
- Sabahi, S., & Parast, M. M. (2020). The impact of entrepreneurship orientation on project performance: A machine learning approach. *International Journal of Production Economics*, 226, 107621. <https://doi.org/10.1016/j.ijpe.2020.107621>
- Salloum, S. A., Al-Emran, M., Monem, A. A., & Shaalan, K. (2017). A Survey of Text Mining in Social Media: Facebook and Twitter Perspectives. *Advances in Science, Technology and Engineering Systems Journal*, 2(1), 127–133. <https://doi.org/10.25046/aj020115>
- Sarica, S., & Luo, J. (2021). Stopwords in technical language processing. *PLOS ONE*, 16(8), e0254937. <https://doi.org/10.1371/journal.pone.0254937>
- Scholtz, M., & De Ridder, K. (2021). Brace for impact! COVID-19, lockdown and the initial reaction and adaptability of Flemish travel consumers. *Journal of Tourism, Heritage & Services Marketing*, 7(2), 28–37. <https://doi.org/10.5281/zenodo.5548442>
- Sharma, E., & Das, S. (2021). Integrated model for women empowerment in rural India. *Journal of International Development*, 33(3), 594–611. <https://doi.org/10.1002/jid.3539>
- Shivakumara, P., Alaei, A., & Pal, U. (2021). Mining text from natural scene and video images: A survey. *WIREs Data Mining and Knowledge Discovery*, 11(6). <https://doi.org/10.1002/widm.1428>
- Sinclair, J., & Cardew-Hall, M. (2007). The folksonomy tag cloud: when is it useful? *Journal of Information Science*, 34(1), 15–29. <https://doi.org/10.1177/0165551506078083>
- Sinclair-Maragh, G., & Simpson, S. B. (2021). Heritage tourism and ethnic identity: A deductive thematic analysis of Jamaican Maroons. *Journal of Tourism, Heritage & Services Marketing*, 7(1), 64–75. <https://doi.org/10.5281/zenodo.4521331>
- Singh, S., Mondal, S., Singh, L. B., Sahoo, K. K., & Das, S. (2020). An Empirical Evidence Study of Consumer Perception and Socioeconomic Profiles for Digital Stores in Vietnam. *Sustainability*, 12(5), 1716. <https://doi.org/10.3390/su12051716>
- Sohrabi, B., & Khalilijafarabad, A. (2018). Systematic method for finding emergence research areas as data quality. *Technological Forecasting and Social Change*, 137, 280–287. <https://doi.org/10.1016/j.techfore.2018.08.003>
- Text Mining Technique for Driving Potentially Valuable Information from Text. (2020). *Information and Knowledge Management*. Published. <https://doi.org/10.7176/ikm/10-1-01>
- Van, N. T. T., Vrana, V., Duy, N. T., Minh, D. X. H., Dzung, P. T., Mondal, S. R., & Das, S. (2020). The Role of Human–Machine Interactive Devices for Post-COVID-19 Innovative Sustainable Tourism in Ho Chi Minh City, Vietnam. *Sustainability*, 12(22), 9523. <https://doi.org/10.3390/su12229523>
- Waqas, M., Iram, M., Shahzad, S., Arshad, S., & Nawaz, T. (2018). Knowledge Extraction Using Web Usage Mining. *ICST Transactions on Scalable Information Systems*, 5(16), 154551. <https://doi.org/10.4108/eai.13-4-2018.154551>
- Watanabe, W. M., Felizardo, K. R., Candido, A., de Souza, R. F., Neto, J. E. D. C., & Vijaykumar, N. L. (2020). Reducing efforts of software engineering systematic literature reviews updates using text classification. *Information and Software Technology*, 128, 106395. <https://doi.org/10.1016/j.infsof.2020.106395>
- Yusuf, S. E., Hong, J. B., Ge, M., & Kim, D. S. (2017). Composite Metrics for Network Security Analysis. *Convergence Security*, 2017(1), 59–82. <https://doi.org/10.13052/jcs2445-9992.2017.003>
- Zanzotto, F. M. (2019). Viewpoint: Human-in-the-loop Artificial Intelligence. *Journal of Artificial Intelligence Research*, 64, 243–252. <https://doi.org/10.1613/jair.1.11345>
- Zuluaga, M., Robledo, S., Osorio Zuluaga, G. A., Yathe, L., Gonzalez, D., & Taborda, G. (2016). *Metabolómica y Pesticidas: Revisión sistemática de literatura usando teoría de grafos para el análisis de referencias*. *Nova*, 14(25), 121. <https://doi.org/10.22490/24629448>

SUBMITTED: FEB 2022

REVISION SUBMITTED: APR 2022

2nd REVISION SUBMITTED: MAY 2022

ACCEPTED: JUN 2022

REFEREED ANONYMOUSLY

PUBLISHED ONLINE: 14 NOV 2022