

# The growth and trends information technology endangered language revitalization research: insight from a bibliometric study

Leonardi Paris Hasugian<sup>1,2</sup>, Syifaul Fuada<sup>3</sup>, Triana Mugia Rahayu<sup>4</sup>, Apridio Edward Katili<sup>5</sup>,  
Feby Artwodini Muqtadiroh<sup>1</sup>, Nur Aini Rakhmawati<sup>1</sup>

<sup>1</sup>Department of Information Systems, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia

<sup>2</sup>Department of Information Systems, Universitas Komputer Indonesia, Bandung, Indonesia

<sup>3</sup>Department of Telecommunications, Universitas Pendidikan Indonesia, Purwakarta, Indonesia

<sup>4</sup>Department of Mechatronics Engineering, Parahyangan Catholic University, Bandung, Indonesia

<sup>5</sup>Badan Riset dan Inovasi Nasional, Jayapura, Indonesia

## Article Info

### Article history:

Received Jul 23, 2024

Revised Mar 29, 2025

Accepted May 24, 2025

### Keywords:

Bibliometric analysis

Endangered language

Information technology

Potential research

Revitalization

## ABSTRACT

Since United Nations Educational, Scientific and Cultural Organization (UNESCO) declared endangered languages, researchers have revitalized endangered languages in many fields. This study discusses a bibliometric analysis conducted to investigate research on the topic of revitalization of endangered languages in information technology. The study's aim is to assess research topics by identifying authors, institutions, and countries that influence research collaboration. The Scopus dataset (from 2002-2024) was obtained from journal articles (n=62) and conference papers (n=76) and visualized using VOSviewer 1.6.20. The analysis outcomes reveal a fluctuating trend with an increasing pattern. The United States, Canada, and China were identified as the top three countries in terms of publications. Meanwhile, the University of Alberta, Université du Québec à Montréal, University of Auckland, and University of Hawai'i at Mānoa are the most prolific institutions on this topic, with two authors from the Université du Québec à Montréal, Sadat and Le, being the most productive. The dominant research is related to computational linguistics. Meanwhile, topics such as phonetic posterograms, integrated frameworks, and artificial intelligence are some of the potential research areas that can be explored in the future. Its implications for exposing the extent to which the development of endangered language revitalization can be accommodated in the field of information technology.

This is an open access article under the [CC BY-SA](#) license.



## Corresponding Author:

Nur Aini Rakhmawati

Department of Information Systems, Institut Teknologi Sepuluh Nopember

Keputih Road, Sukolilo District, Surabaya 60117, Indonesia

Email: nur.aini@its.ac.id

## 1. INTRODUCTION

United Nations Educational, Scientific and Cultural Organization (UNESCO) stated that of around 7,117 languages spoken worldwide, 2,500 languages are classified as endangered [1], approximately 35.13%, and one language disappears every 14 days [2]. This condition threatens culture and heritage which contain knowledge [3], [4], traditions, and identity [5]. One of the six important indicators influencing whether a language could be threatened with extinction is the decrease of language vitality [6]. Language vitality refers to strength, health, intensity of use, and existence of a language as a communication tool [7]. High language

vitality indicates a high number of speakers, intergenerational transmission, and support or policies in the language [8]–[11]. Tsunoda put nine factors for the degradation in language vitality [12], then Bromham presented 51 global predictors of extinction grouped into ten categories [10]. These two research results intersect on how economic, social, and environment aspects can concretely degrade the vitality of language.

Those previously studied three factors are further reaffirmed through various other studies. Economic factors [13], [14], social rejection of identity [15], colonization of a nation [16], [17] and the educational environment through the hegemony of the use of English [18]–[23] in fact encourages the degradation of the vitality of the indigenous language. In response to this challenging situation, global efforts have been directed toward various research to maintain the existence of languages, with a focus on increasing language vitality [24]–[27]. One of these efforts is through various language revitalization programs, such as those carried out by Finland, Sweden, and Norway to preserve the Saami language through the Sámi Giellálgald institution [28], revitalization in the education sector [29], [30], language digitization (text, audio, and video) [6], [31], language corpus building through natural language processing [32]–[36] and the presence of various general applications for popular languages such as Babbel, Duolingo, Google Translate, and Rosetta Stone as well as special applications for endangered languages such as Save Lingo [37], IndyLan [38], Woolaroo [39], Karelian learning games [40], and HUQARIQ [41]. In those efforts, information technology plays a vital role in saving endangered languages. Mobile apps, cloud computing, social media, and artificial intelligence provide innovative solutions for capturing, sharing linguistic data, increasing the accessibility of language resources, empowering communities to preserve their heritage, and enhancing linguistic diversity around the globe [42]–[44].

Language revitalization is a research domain that is closely related to regional characteristics [45] considering that a language serves as a vital communication tool within a region, being an integral part of the culture and reflecting the spirit of a nation or state [46], [47]. Research on language revitalization carried out by certain countries tends to focus on their own language and is very dependent on their own socio-cultural characteristics. For this reason, research performance will vary depending on the country, considering that research in the field of language is complex involving many variables [48] and as a result influences research developments and trends.

Based on that, we conducted a research study using bibliometric analysis. As an introduction, so far, many methods have been implemented to identify research trends, such as meta-analysis [49], literature review [50], systematic review [51], integrative review [52], and bibliometric analysis [53]. Among these various methods, bibliometric analysis holds several advantages, including its ability to quickly provide information about multiple studies and details on potential collaboration opportunities [54]. With these various advantages, bibliometrics is a popular literature review method and can be used in various fields, such as nurse education [55], Asian linguistics [45], social life [56], augmented reality [57], biomedical [58], and others. Bibliometric analysis is a valuable tool for appraising multiple aspects of scientific literature, including the assessment of academic journals' performance and the exploration of a field's intellectual structure. By examining the existing scientific literature, this analysis can reveal patterns that emerge during the modernization process, thereby enabling researchers to comprehend the field better, identify gaps in knowledge, and discover new avenues for investigation. Ultimately, bibliometric analysis can help position expected contributions and provide a comprehensive understanding of the field [59].

Nevertheless, there is no bibliometric information technology on endangered language revival, despite the topic's growing importance. This is the gap we are trying to fill by doing an extensive bibliometric analysis and concentrating on the use of information technology for the revitalization of endangered languages. This study has implications to expose the extent to which the development of endangered language revitalization can be accommodated in the field of information technology. The main objective of this study is to give readers a comprehensive understanding of the growth of research on information technology for the revitalization of endangered languages by identifying countries that are at the forefront of this research domain, mapping revitalization of endangered languages on information technology research in various countries, its research challenges, and open topics for future research.

## 2. METHOD

Bibliometric analysis is a method for determining research trends on specific quantitative topics [60]. Bibliometric research in this study was compiled into three stages based on seven main steps explained by Passas [61], as shown in Figure 1. The data source was taken from the Scopus database in the data collection stage. Scopus was chosen because it can curate high-quality abstracts and citations covering scientific journals, conference proceedings, books, and patents [62], [63]. Apart from that, Scopus also covers many journals [64], [65] and articles [66], [67] compared to Web of Science and Dimensions. The selected dataset was relevant to the preservation of endangered languages. We explore keywords related to "Revitalization of Endangered Language" with the following parameters: *TITLE-ABS-KEY ("Preservation"*

---

*The growth and trends information technology endangered language ... (Leonardi Paris Hasugian)*

OR "Conservancy" OR "Conservation" OR "Revitalization" OR "Revival" OR "Protection" OR "Maintenance" OR "Care" AND "Endanger Language" OR "Endangered Language" OR "Endangerment Language" OR "Extinct Language" OR "Minority Language" OR "Vulnerable Language" OR "Death Language" OR "Indigenous Language") AND (LIMIT-TO (SUBJAREA, "COMP")) AND (LIMIT-TO (DOCTYPE, "ar") OR LIMIT-TO (DOCTYPE, "cp")))). These keywords were chosen to capture various word choices that are appropriate to the context of 'Endangered Languages Revitalization' contained in the title, abstract, and keywords on publication. For publications, from journal articles and conference papers. Both were chosen considering the performance evaluation and promotion of many researchers through these documents [68], [69]. Then, to exclude irrelevant search results, the search results were limited to the field of 'Computer Science,' which is relevant to information technology without limiting the inclusive year of publication. The search yielded a corpus of 138 articles, including 62 journal articles (44.9%) and 76 conference papers (55.1%). After fulfilling the data search criteria, metadata such as authors, affiliations, keywords, abstracts, and number of citations were converted into .csv format for analysis.

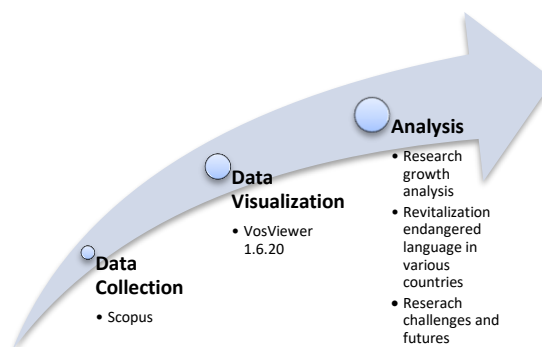


Figure 1. Research stages of bibliometric analysis

Data visualization is essential because it creates graphical representations of the analysis results to aid their interpretation and presentation [70] for comprehensive bibliometric analysis, and numerous software alternatives are accessible. In this study, data visualization was performed using VOSviewer (version 1.6.20). The platform provides a variety of data visualization approaches [71], allowing us efficiently. After that, we analyzed the data visualization. We investigated by examining research according to the number of publications and then mapped each author by institution and country. Next, we analyzed the keywords to get a research trend study. By investigating the number of publications, authors, and keywords, we could interpret the detailed and insightful report that provides recommendations and highlights significant trends and patterns. This analysis shows the growth of research on the revitalization of endangered languages in the field of information technology by identifying countries that are leading in this research domain, highlighting opportunities for collaboration in various countries, and pointing out future challenges and open topics.

### 3. RESULTS AND DISCUSSION

#### 3.1. Analysis of research growth

According to data collection stage from the Scopus database, as many as 138 scientific articles have been collected on the revitalization of endangered languages in the field of information technology since the start of publication findings from 2002 to February 2024, as presented in Figure 2. During those years, the quantity of publications associated with information technology for the revitalization of endangered languages varies but shows an increasing trend in the development of the domain of revitalization of endangered languages in the field of information technology. As a note, we made an exception in 2024 considering the data taken at the beginning of the year, February 29, 2024.

In the Journal Language Learning & Technology in 2002, Tracey McHenry published early research on the subject. McHenry is affiliated with Eastern Washington University. Scopus has indexed the article, and it has been referenced 12 times, while Google Scholar records 38 citations. The study highlighted the exploration of the complexities of language research for native Americans, "Indian Languages", before website technology could be used in an educational environment. The result is for educators and language specialists that the internet offers the latest context. Although traditions and stereotypes are contrary to their

culture, website technology makes their language available on website media. It can reach a wide range of native Americans in nurturing and developing the vitality of their language while remaining in the dominance of a predominantly English-speaking world [72]. Later in the same year, Daniel J. Villa, who is affiliated with New Mexico State University, presented a study on the pilot project in applying technology in the right way and overcoming the problems faced *out-group language researcher* by training a society of Navajo speakers (*in-group member*) in preserving their language. He is a pioneer in the application of interactive learning technology to seek language revitalization. As a result, native speakers are trained in methodologies and technologies to acquire, record, and preserve the Navajo language. They continue to use their language in the family. There is a fact that *in-group members* do not need to rely on *out-group language researchers* to provide materials. In addition, for this process, the technology needed is interactive language learning programs using CD-ROM, it is not too expensive and even tends to be cheap [73]. This article's recognition has been shown by its 87 citations on Google Scholar and 21 citations on the Scopus database.

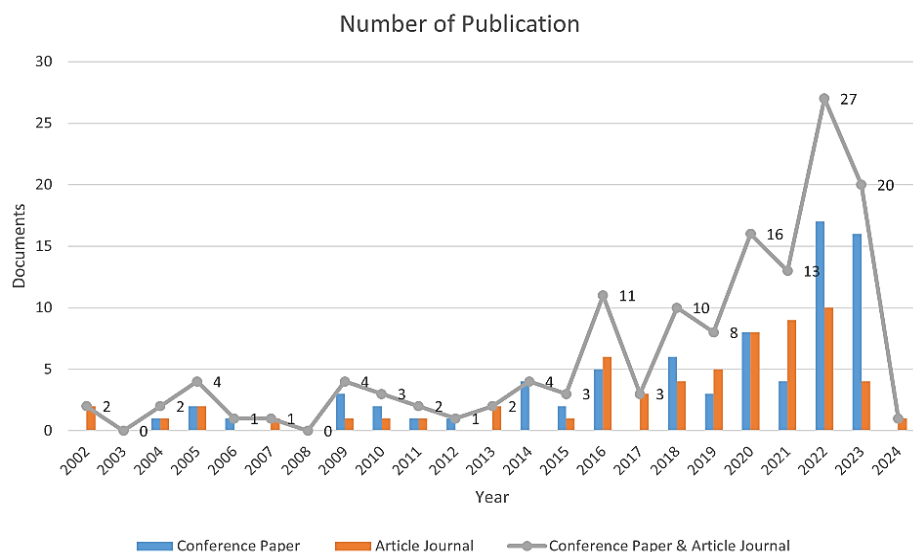


Figure 2. Growth of publications in revitalization endangered language research

The following year 2003, there was a scarcity of publications, and this also happened in 2008. Between 2004-2007, there were 4 journal articles and 4 conference papers. The emphasis of publication topics this year spans intersects with topics regarding the development of teaching materials, digital libraries (digital documentation), and bilingual systems to support minority languages, including endangered languages. Authors [74], [75] emphasized how a (computer-assisted) system can support the development of materials for endangered languages. Then the authors [75]–[77] and [78] agreed to find ways regarding how digital libraries can help maintain endangered languages through the distribution of digital information in text, audio, and video formats. Specifically, Lu put forward the development of ImageSpace, which is a tool for building ontologies and annotations that support text and images [76]. In line with Lu *et al.*, Chaisade *et al.* developed annotated verbal corpora [78]. Furthermore, authors [77], [79]–[81] focused on bilingual systems for minority languages. Cunliffe and Roberts-Young designed [79] and promoted [80] a bilingual system to society in Wales, but the trend showed an increase in the use of the English language, not the Welsh language. This shows that implementing a bilingual system for the language revitalization process for minority languages is not necessarily successful according to the objectives. This is also confirmed in Chowdurry's findings. In his research, he describes his experience introducing a multilingual electronic patient information technology for the first time in the country at a health organization in North England. The implementation failed to achieve progress even though it had great potential. This case highlights a variety of issues linked with the way of design and implementation of this project emerged from the approach taken [81] and also reinforced by the involvement of various factors [10], [12] in its implementation.

From 2009 to 2024, the quantity of publications connected to the subject has an increasing trend. For the record, in 2024, there will only be 1 journal article recorded. However, it should be noted that this data was gathered until February 29, 2024. The highest number of publications is in 2022, publishing 10 journal articles and 17 conference papers. From 2009 to 2024, there are various publications in addition to

the topic of teaching material development, digital library (digital documentation), and bilingual systems, including: educational games, gamification [82]–[86]; mobile design [87]–[89]; speech acoustics, speech recognition system [90]–[96]; crowdsourcing, rapid collection, corpus [97]–[99]; artificial intelligence: natural language processing [34], [100]–[102], machine learning [103], and neural machine [104], [105]; and social media [106], [107]. These identified topics are relevant to any topic related to the use of information technology mentioned [42]–[44].

Table 1 shows the top six authors on renewing endangered languages in information technology. Individual's Scopus H-Index, number of citations, publications, current affiliation, nationality, and Scopus author ID were noted. Rankings based on several publications and citations show the authors' impact in the research field. According to the statistics, Sadat and Le are the top-ranked authors, with 4 publications and 15 citations. They have the same association, Université du Québec à Montréal, and have research interests in natural language processing and machine learning. Then Ward, Koole, Sundaram, and Mirza each have three publications. Nevertheless, Ward has the most citations, with 19.

Table 1. Top authors with the number of publications and citations (counted on March 5, 2024)

Rank	Author	Citations	Total Publication	Scopus H-Index	Affiliation	Country	Scopus ID
1	Sadat, F. (4)	15	83	12	Université du Québec à Montréal	Canada	6507647914
1	Le, N.T. (4)	15	19	4	Université du Québec à Montréal	Canada	57202711514
2	Ward, M. (3)	19	19	5	Dublin City University	Ireland	14120491600
2	Koole, M.L. (3)	9	24	9	University of Saskatchewan	Canada	15061587000
2	Sundaram, D. (3)	5	202	18	University of Auckland	New Zealand	6602562990
2	Mirza, A. (3)	5	12	2	University of Auckland	New Zealand	57190279002

### 3.2. Mapping revitalization endangered language on information technology research in various countries

This section provides the top ten institutions that have achieved high publication in the revitalization of endangered languages, with a focus on the use of information technology, as shown in Table 2. Our findings indicate that the University of Alberta has emerged as the most productive institution, with 5 publications. The Université du Québec à Montréal, the University of Auckland, and the University of Hawai'i at Mānoa each have 4 publications. Furthermore, other universities contributed through three publications.

Table 2. Top institutions with the number of publications

Rank	Institution	Country	Number of publications
1	University of Alberta	Canada	5
2	Université du Québec à Montréal	Canada	4
2	University of Auckland	New Zealand	4
2	University of Hawai'i at Mānoa	United States	4
3	Australian National University	Australia	3
3	University of British Columbia	Canada	3
3	University of Saskatchewan	Canada	3
3	George Mason University	United States	3
3	Dublin City University	Ireland	3
3	Carnegie Mellon University	United States	3

The United States and Canada scored the most on this topic, as shown in Table 3, with one of them influenced by the efforts of academics Sadat, Le, and Koole, as shown in Table 1. In particular, the United States leads in the geographical distribution of scientific publications, with the best contributors based on the publication of  $\geq 2$  scientific articles. This has been achieved through various research partnerships with various institutions and universities at home and abroad as illustrated in Figure 3, which puts the United States at the forefront. Figure 3 shows a significant collaboration between the United States and Australia about revitalizing endangered languages in the field of information technology. Figure 3 was created using VOSviewer 1.6.20 software to analyze co-authorship (by country), with a *minimum threshold of 5 country documents*. Co-authorship analysis is applied to show collaboration in a *university-wide collaboration network* [108]. Only 9 of the 41 countries surveyed met this threshold requirement. Only 5 of the 9 countries

display interconnectedness. China, India, Mexico, and New Zealand did not participate in the collaboration with the five countries; hence they were not featured in the visualization. The United States has a *Link Strength* of 6. In addition, the country shows linkages with the United Kingdom, Canada, Japan, and Australia, with link strengths of 5, 3, 3, and 1, respectively. In total, the five interconnected countries contribute to a combined *Total Link Strength* of 18.

Table 3 shows the top three countries regarding productivity: the United States, Canada, and China, with 37, 22, and 11 publications, respectively. This means that endangered language revitalization in the information technology field is thoroughly researched and generally acknowledged in North America and East Asia. This finding aligns with Simon's findings, which mapped the percentage of languages threatened with extinction or extinct in 25 years, where the United States, Canada, and Australia are the countries experiencing the most massive degradation of language vitality [109], not including China. This strengthens the reason why there is a lot of active research to develop the revitalization of endangered languages in those countries. Overall, all the top contributors were from universities, except for the ARC Centre of Excellence for the Dynamics of Language, which was not a university.

Table 3. Top leading countries

Rank	Country	Number of publications	Top contributors institution
1	United States	37	University of Hawai'i at Mānoa, George Mason University, Carnegie Mellon University, University of Illinois Urbana-Champaign, Boston College
2	Canada	22	University of Alberta, Université du Québec à Montréal, University of Saskatchewan, University of British Columbia, Simon Fraser University
3	China	11	Northwest Minzu University, Yunnan Normal University, Beijing Technology and Business University
4	Australia	9	Australian National University, University of Melbourne, Curtin University, ARC Centre of Excellence for the Dynamics of Language*
4	United Kingdom	9	University of South Wales, University of Edinburgh
5	New Zealand	6	University of Auckland

\*) Non-university-based Institution

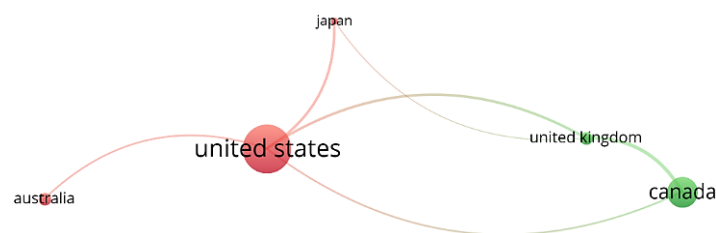


Figure 3. Network visualization among countries containing the United States, Australia, Canada, United Kingdom, and Japan

Our findings indicate that the United States and Canada have the best chances for research collaboration on endangered language revival in information technology field. This is validated by the number of articles seen in Table 3. Regarding the revitalization of endangered languages in the field of information technology, there is potential for collaboration with the Sadat and Le research groups at the Université du Québec à Montréal, as well as the Sundaram and Mirza research groups at the University of Auckland, as shown in Table 1. Both groups have great competence in various fields, as seen by the numerous citations their works have achieved. This is an opportunity to encourage collaboration in developing endangered language revitalization solutions as conveyed by Sharofova [43].

Furthermore, the data gathered in Table 3 was utilized to create Figure 4, which visually depicts the global distribution of research studies on the revitalization of endangered languages in the field of information technology in various nations. The map was colored using a 30-level gradient, with darker colors indicate more research papers and lighter shades indicate fewer articles. According to Figure 4, the United States is clearly depicted in dark red due to the large number of publications, whereas Canada is displayed in bright red. Then such as Japan, India, Argentina, Brazil, Mexico, Namibia, Republic of South Africa, and half of Europe are displayed in light red because the number of publications tends to be low. However, what is interesting is the potential to be a pioneer in conducting research and publications on the topic of revitalizing endangered languages in the field of information technology can be carried out in Russia, the West Asian region, and the African continent from the central to the north.



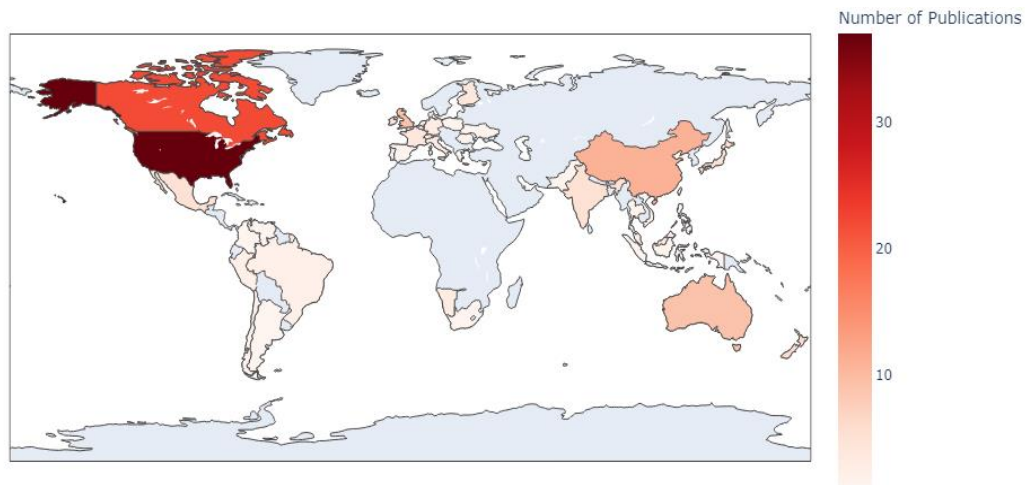


Figure 4. Map showcasing the number of research publications on the topic of revitalization of endangered languages in the information technology field across different countries

### 3.3. Mapping in revitalization endangered language on information technology research challenges and open topics

Analysis of current research trends allows for the interpretation of research challenges. These trends could be found by analyzing the quantity and relevance of keywords frequently used by academics [110], [111]. Figure 5 exhibits data on the relationship between keywords in research on the revitalization of endangered languages in information technology from 2002 until 2024. The data is presented in overlay mode, so we can analyze the progress of the research by examining color changes over the last 22 years and gain insight into current research on this topic. Keyword association mapping is possible using VOSviewer 1.6.20 software using the Scopus dataset. This mapping, also known as *keyword co-occurrence analysis*, is used to identify the most important keywords and topics in the selected publication. This work uses a threshold of one occurrence to pick a term. This indicates that only keywords appearing at least once in the given document are considered. As a result, 828 keywords met the criteria. The circle size shows the number of keywords utilized, with larger circles representing more keywords, as illustrated in Figure 5.

In addition, we discovered keywords that are uncommon in the available publications. This shows that areas need future attention and additional research to close the gap. The keywords in question are: simulation; universal design for learning; mobile design; human-computer interaction; gamification; multimedia technology; web application; social network; emotion-based learning; semantic web; text to speech; optical character recognition; phonetic posterigram; orthography development; chatbot; tokenizer; accent robot; predictive modeling; forecasting; artificial intelligent (AI); neural machine; neural networks; machine learning; natural language processing; cellular automata; analysis algorithms; collective intelligence; automatic rapid corpus collection; integrated frameworks; information system; knowledge management; agile software development; internet of things; online community; and speech community. Meanwhile, the dominant keyword is strictly in computational linguistics.

If searched for the most recent year, then the keywords in green lead to yellow label represent the most recent research domain. A phonetic posterigram is a method of performing phonetic and phonological analysis to understand how sounds are produced and recognized by listeners [112], [113]. In the process, sounds are recorded and then analyzed to determine how various acoustic features such as frequency, amplitude, and duration can change over time, visualized in graphs or diagrams, called posterigrams. The posterigram is very helpful in understanding the differences between different sounds in one or different languages and can provide insight into the sound processes involved in the process of speaking and hearing [114]–[116]. In addition, the topic of integrated frameworks has been conducted since 2005 [75], and until now only a small number of researchers have focused on this area. This indicates that the research area has the potential for further development. Integrated framework refers to a framework that is able to provide coherence and complexity to the technology and users that the study domain can map technical perspective and sociotechnical point of view through various methods and approaches [117], [118]. So that every use of technology will be easy to understand in an organization to achieve specific goals. In addition to these topics, artificial intelligence with sub-topics such as neural machines, neural networks, machine learning, and natural language processing has relatively recently been touched on the area of endangered language revitalization,

explicitly since 2020 [101]. Basically, the application of artificial intelligence in various sub-domains is in the scope of documentation in text and audio formats as the driving force behind language learning [119], [120]. Such approaches can be designed to function independently or in conjunction with communicative language teaching and potentially save languages, especially those that are endangered [35], [121]. This is in line with the revitalization needs put forward by Olaare [42] and Sharofova [42].

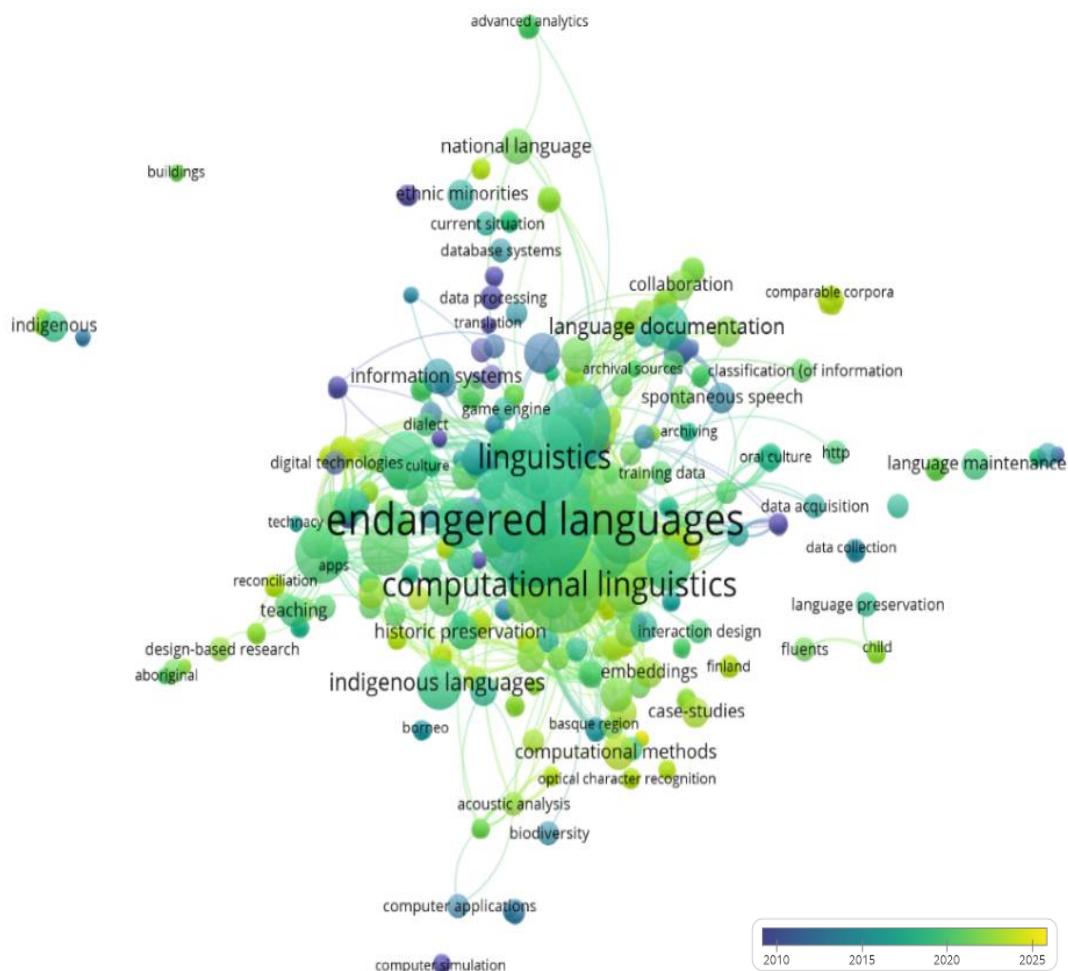


Figure 5. Keywords in the topic of endangered language on information technology field research

The results of this study reveal the extent to which topics on information technology fall within the scope of endangered language revitalization. Challenges ranging from resources to infrastructure reveal how information technology can support the survival of the language [122]. Despite the massive support of information technology in various fields, such as economy [123], government [124], education [125], and health [126], according to the author's findings, research on the revitalization of endangered languages in the field of information technology has not been as massive as these fields [44]. Through this study, the author wants to reveal that there are many areas that can be done to preserve endangered languages through the use of information technology considering the urgency that one language disappears every 14 days [2].

### 3.4. Strengths and limitations of the study

Following this understanding, we conducted the first bibliometric analysis of peer-reviewed literature on the revitalization of endangered languages in the field of information technology. To assess the current state of research on the topic, we employed visualizations with VOSviewer to explore data and determine various focus areas. Nonetheless, it is vital to recognize several limitations in our research. First, the data was acquired from the Scopus database, leaving out other leading databases like Web of Science, Dimensions, and Google Scholar. In addition, we only examined scientific works published in journals and conference proceedings. The use of the Scopus database that we carry out is also emphasized by the many

*The growth and trends information technology endangered language ... (Leonardi Paris Hasugian)*



uses of the database in other bibliometric analyses, as has been done by various authors in various fields [127]–[135]. The Scopus dataset was used to discover relevant research, although it is essential to note that many journals and conference proceedings are excluded from this database. This ensures that many of the missed scientific article publications come from various journals and proceedings that are not indexed by Scopus and that will distort bibliometric indicators or metrics [136]. In the search on the Scopus database, we focus only on the field of Computer Engineering in English writing with the intention of every search for scientific articles related to information technology. This is driven based on visibility, every scientific article published on the revitalization of endangered languages in the field of information technology is with the focus and scope of specific journals and conference proceedings [137], [138].

For the record, we conducted an analysis by examining the most productive authors and the number of articles cited by other authors. We also interpret co-authorship by authors and their countries. In addition, by using relevant keywords, Scopus offers precise calculations and retrieves articles that may not explicitly mention the keywords we declared. The keywords were based on the context of revitalization of endangered languages. The words in "revitalization of endangered languages" were cut into two parts, namely "revitalization" and "endangered language." In these two sections, we expand with various words that have relevant meanings, including: the word "revitalization" intersects with *preservation*, *conservancy*, *conservation*, *revival*, *protection*, *maintenance*, and *care*; while "endangered language" has a similar meaning in *endanger language*, *endangerment language*, *extinct language*, *minority language*, *vulnerable language*, *death language*, and *indigenous language*. This helps us in increasing the pool of articles available for analysis, although there may be some specific keywords missing that affect the results. This approach is quite beneficial when conducting reviews with enormous datasets, but it is impractical for manual reviews. In some circumstances, a narrow review focus with the dataset produced can be maintained via manual review, also known as a systematic review [54]. In our research, the Scopus dataset obtained was 138 articles. Although this dataset is relatively small, it is still relevant to be analyzed efficiently using the bibliometric method, as several previous studies have used datasets with a number not much different number in the range of 100–200 datasets, such as: the use of 104 datasets to analyze sustainability and smart tourism [139]; 138 datasets to analyze library services through social media [140]; 138 datasets to analyze the political connection [53]; 148 datasets to analyze research on the topic of in-body communication [141]; and 161 datasets to analyze environmental, social, and governance disclosure [142]. Visualizing the results of bibliometric analysis provides valuable information for advancing academic collaboration, especially for research endeavors involving active researchers, associations, and countries as previously Yadav stated [44]. As a result, the use of bibliometric methodologies in our current study on the revitalization of endangered languages in information technology is consistent with the established procedures in the field.

#### 4. CONCLUSION

An analysis of research on language revitalization in the information technology field from 2002 to 2024 was conducted using bibliometric techniques. This is an exploratory study that only uses data from Scopus. Our findings indicate that research on the topic of language revitalization in the information technology field has fluctuated with a steady increase from the early years until 2024, notwithstanding a pause between 2003 and 2008. The top three countries on this topic are the United States, Canada, and China. The University of Alberta, Université du Québec à Montréal, University of Auckland, and University of Hawai'i at Mānoa are notable institutions in this field with significant publications. Furthermore, Sadat and Le, who are both from the same institution, the Université du Québec à Montréal in Canada, have been identified as the most productive researchers in this field. Among all the scientific articles on this subject, computational linguistics is the dominant topic. However, the future research in making efforts to revitalize endangered languages through engineering processes in the field of information technology is still wide open, such as topics in phonetic posterograms, integrated frameworks, and artificial intelligence. These findings have implications for the extent to which the development of revitalization of endangered languages can be accommodated in the field of information technology which can revitalize endangered languages, contribute cultural preservation, promoting interdisciplinary collaboration, and setting directions for future research.

The limitation of this study is that the dataset is taken from Scopus in the field of computer engineering and published in journals and conference proceedings using the keyword "revitalization of endangered language." This is a weakness in the collection of datasets that many scientific publications can be obtained from other databases such as Web of Science, Dimensions, and Google Scholar. In addition, there may be other relevant keywords that were not included during the search and only English paper were chosen that will affect on the dataset's overall results. This is a reference for further bibliometric studies on revitalizing endangered languages in information technology.

## FUNDING INFORMATION

This research was funded by *Direktorat Penelitian dan Pengabdian kepada Masyarakat, Direktorat Jenderal Riset dan Pengembangan, Kementerian Pendidikan Tinggi, Sains, dan Teknologi Republik Indonesia*, under the master contract for the implementation of *Program Bantuan Operasional Perguruan Tinggi Negeri, Penelitian Disertasi Doktor* (Master Contract No. 017/C3/DT.05.00/PL/2025, dated May 28, 2025; Researcher Contract No. 1212/PKS/ITS/2025, dated June 3, 2025).

## AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Leonardi Paris Hasugian	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	
Syifaul Fuada		✓		✓		✓		✓		✓	✓			
Triana Mugia Rahayu	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓			
Apridio Edward Katili	✓				✓		✓						✓	
Feby Artwodini Muqtadiroh					✓					✓		✓		
Nur Aini Rakhmawati					✓	✓				✓		✓		

C : Conceptualization

M : Methodology

So : Software

Va : Validation

Fo : Formal analysis

I : Investigation

R : Resources

D : Data Curation

O : Writing - Original Draft

E : Writing - Review & Editing

Vi : Visualization

Su : Supervision

P : Project administration

Fu : Funding acquisition

## CONFLICT OF INTEREST STATEMENT

The authors declare that there is no conflict of interest regarding the publication of this paper.

## DATA AVAILABILITY

The authors confirm that the data supporting the findings of this study are available within the paper.

## REFERENCES

- [1] E. C. Moseley, *Atlas of the world's languages in danger*, 3rd ed. Paris: UNESCO Publishing, 2010.
- [2] S. Chiblow and P. J. Meighan, "Language is land, land is language: The importance of indigenous languages," *Human Geography*, vol. 15, no. 2, pp. 206–210, Jul. 2022, doi: 10.1177/19427786211022899.
- [3] R. Cámara-Leret and J. Bascompte, "Language extinction triggers the loss of unique medicinal knowledge," *Proceedings of the National Academy of Sciences*, vol. 118, no. 24, Jun. 2021, doi: 10.1073/pnas.2103683118.
- [4] K. D. Harrison, *When languages die: The extinction of the world's languages and the erosion of human knowledge*. New York: Oxford Academic, 2010.
- [5] M. Li, S. M. Croucher, and L. Shen, "Language endangerment and the linguistic vitality of Miao in China: cultural shifts and revitalisation strategies," *Journal of Multilingual and Multicultural Development*, pp. 1–16, Oct. 2024, doi: 10.1080/01434632.2024.2411006.
- [6] "SIL language technology," 2023. <https://software.sil.org/software-products/> (accessed Feb. 15, 2024).
- [7] L. R. Slavina, J. N. Mustafina, E. M. Maklakova, K. N. Gataullina, G. R. Ganieva, and E. V. Maksimova, "Assessing language vitality: Overview of existing models," *Opcion*, vol. 34, no. Special Issue 17, pp. 773–790, 2018.
- [8] C. Bowern, "Language vitality: Theorizing language loss, shift, and reclamation (Response to Mufwene)," *Language*, vol. 93, no. 4, pp. e243–e253, 2017, doi: 10.1353/lan.2017.0068.
- [9] J. E. R. Labrada, "Language vitality among the mako communities of the ventuari river," *Language Documentation and Conservation*, vol. 11, no. 2017, pp. 10–48, 2017.
- [10] L. Bromham *et al.*, "Global predictors of language endangerment and the future of linguistic diversity," *Nature Ecology & Evolution*, vol. 6, no. 2, pp. 163–173, Dec. 2021, doi: 10.1038/s41559-021-01604-y.
- [11] O. Kazakevich, "Language vitality scales and their applicability to specific language situations," *Voprosy Jazykoznanija*, no. 4, p. 7, 2022, doi: 10.31857/0373-658X.2022.4.7-47.
- [12] T. Tsunoda, *Language endangerment and Language revitalization: an introduction*. Berlin: De Gruyter Mouton, 2006.
- [13] M. C. Ewing, "Language endangerment in Indonesia," *International Journal of Education*, vol. 8, no. 1, pp. 12–22, 2014, doi: 10.1075/cilt.240.05eng.
- [14] A. Rokhman, R. E. Priyono, I. Santosa, S. Pangestuti, and M. T. Kariadi, "Existence of Banyumasan Javanese Language in digital era," *Humanities and Social Science Research*, vol. 5, no. 2, p. p1, May 2022, doi: 10.30560/hssr.v5n2p1.
- [15] S. Haidar, T. Wali, T. Tahir, and M. Parveen, "'I Am Not Punjabi, My Parents Are': Degradation of the language of dominant majority," *Acta Linguistica Asiatica*, vol. 11, no. 2, pp. 101–127, Jul. 2021, doi: 10.4312/ala.11.2.101-127.
- [16] T. Reagan, "Why language endangerment and language death matter: 'Took away our native tongue ... And taught their English to our young,'" *Linguistic Legitimacy and Social Justice*, pp. 285–314, 2019, doi: 10.1007/978-3-030-10967-7\_9.

- [17] C. N. Ugochukwu, "From linguistic repression to revitalization: The Igbo language case," *Journal of Applied Linguistics and Language Research (JALLR)*, vol. 7, no. 4, pp. 172–183, 2020.
- [18] B. Oshodi, "An assessment of indigenous nigerian languages and factors of language endangerment: Can the indigenous languages!survive?," *Dialectologia*, no. 13, pp. 1–25, 2014.
- [19] M. D. Enaikel, "Etsako: an anthropological reflection of an endangered minority language in Nigeria," *Journal of Pan African Studies*, vol. 7, no. 4, pp. 239–255, 2014.
- [20] A. Kaharuddin, "The power of english: recognizing and utilizing the tremendous impact of the English language on the community," *English Language Teaching for EFL Learners*, vol. 1, no. 1, p. 39, Mar. 2019, doi: 10.24252/elties.v1i1.7625.
- [21] E. O. Ugwu, "English language, globalisation and the future of the Nigerian Indigenous languages," *European Scientific Journal, ESJ*, vol. 16, no. 14, p. 38, May 2020, doi: 10.19044/esj.2020.v16n14p38.
- [22] A. Y. Aikhenva, "Language contact and endangered languages," *The Oxford Handbook of Language Contact*, pp. 240–260, 2020, doi: 10.1093/oxfordhb/9780199945092.013.11.
- [23] L. Bromham, X. Hua, C. Algy, and F. Meakins, "Language endangerment: a multidimensional analysis of risk factors," *Journal of Language Evolution*, vol. 5, no. 1, pp. 75–91, Jan. 2020, doi: 10.1093/jole/lzaa002.
- [24] E. Gasser, "The right to say yes: language documentation in West Papua," *Australian Journal of Linguistics*, vol. 37, no. 4, pp. 502–526, Oct. 2017, doi: 10.1080/07268602.2017.1350131.
- [25] D. Kaufman and R. Perlin, "Language documentation in diaspora communities," *The Oxford Handbook of Endangered Languages*, pp. 398–418, 2018, doi: 10.1093/oxfordhb/9780190610029.013.20.
- [26] F. Lüpke, "Language endangerment and language documentation in Africa," *The Cambridge Handbook of African Linguistics*, pp. 468–490, 2019, doi: 10.1017/9781108283991.015.
- [27] Iskandarsyah Siregar, "Papuan tabla language preservation strategy," *LingLit Journal Scientific Journal for Linguistics and Literature*, vol. 3, no. 1, pp. 1–12, Mar. 2022, doi: 10.33258/linglit.v3i1.620.
- [28] "Sámi Giellagádu," 2023. <https://www.samediggi.fi/sami-giellagadu/> (accessed Feb. 17, 2024).
- [29] V. Ernštreits and G. Kļava, "Experiences in teaching an endangered language: finding the motivation and means to ensure the acquisition of livonian," *Teaching and Learning Resources for Endangered Languages*, pp. 66–82, 2023, doi: 10.1163/9789004544185\_006.
- [30] R.-L. Valijärvi and L. Kahn, "Teaching and learning resources for endangered languages," *Teaching and Learning Resources for Endangered Languages*, 2023, doi: 10.1163/9789004544185.
- [31] "LingTranSoft: Software," 2023. <https://lingtransoft.info/apps/results> (accessed Feb. 15, 2024).
- [32] M. Ptaszynski, M. Kazuki, and Y. Momouchi, "NLP for endangered languages: morphology analysis, translation support and shallow parsing of ainu language," in *Proceedings of the Nineteenth Annual Meeting of The Association for Natural Language Processing*, 2013, pp. 418–421.
- [33] A. Anastasopoulos, C. Cox, H. Cruz, and G. Neubig, "Endangered languages meet modern NLP," in *COLING 2020 - 28th International Conference on Computational Linguistics, Tutorial Abstracts*, 2020, pp. 39–45, doi: 10.18653/v1/2020.coling-tutorials.7.
- [34] R. Zariquiey, A. Oncevay, and J. Vera, "CLD2: Language documentation meets natural language processing for revitalising endangered languages," *COMPUTEL 2022 - 5th Workshop on the Use of Computational Methods in the Study of Endangered Languages, Proceedings of the Workshop*, pp. 20–30, 2022, doi: 10.18653/v1/2022.computel-1.4.
- [35] S. Zhang, B. Frey, and M. Bansal, "How can NLP help revitalize endangered languages? a case study and roadmap for the cherokee language," in *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2022, pp. 1529–1541, doi: 10.18653/v1/2022.acl-long.108.
- [36] G. I. Winata et al., "NusaX: multilingual parallel sentiment dataset for 10 Indonesian local languages," in *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*, 2023, pp. 815–834, doi: 10.18653/v1/2023.eacl-main.57.
- [37] A. Mirza and D. Sundaram, "Collective intelligence based endangered language revitalisation systems: design, implementation, and evaluation," *EAI Endorsed Transactions on Context-aware Systems and Applications*, vol. 4, no. 11, p. 152338, Mar. 2017, doi: 10.4108/cai.6-3-2017.152338.
- [38] K. Strani and C. Cocco, "Report on endangered Indigenous languages in partner countries and mobile learning solutions: IndyLan project," 2020.
- [39] "Woolaroo," 2021. <https://artsexperiments.withgoogle.com/woolaroo/intro/about> (accessed Feb. 17, 2024).
- [40] L. Arhippainen and P. Alaves, "Designing learning games and tools for Karelian Language - considerations based on Beginners' experiences of gameplay," in *2021 28th Conference of Open Innovations Association (FRUCT)*, Jan. 2021, pp. 3–14, doi: 10.23919/FRUCT50888.2021.9347578.
- [41] L. Camacho and R. Zavallos, "Language technology into high schools for revitalization of endangered languages," in *2020 IEEE XXVII International Conference on Electronics, Electrical Engineering and Computing (INTERCON)*, Sep. 2020, pp. 1–4, doi: 10.1109/INTERCON50315.2020.9220197.
- [42] S. Olaare, "The role of technology in language preservation," *European Journal of Linguistics*, vol. 3, no. 2, pp. 44–56, Jul. 2024, doi: 10.47941/ejl.2046.
- [43] S. Sharofova, "The impact of AI on endangered languages: Can technology save or Kill?," *Texas Journal of Philology, Culture and History*, 2023, doi: The Impact of AI on Endangered Languages: Can Technology Save or Kill?
- [44] A. K. Yadav, "Language revitalization and endangered languages: current trends and future prospects," *Journal of Advanced Research in English and Education*, vol. 10, no. 3, 2024.
- [45] D. Lee, "Bibliometric analysis of Asian 'language and linguistics' research: A case of 13 countries," *Humanities and Social Sciences Communications*, vol. 10, no. 1, p. 379, Jul. 2023, doi: 10.1057/s41599-023-01840-6.
- [46] Z. Tektigul, A. Bayadilova-Altybayeva, S. Sadykova, S. Iskindirova, A. Kushkimbayeva, and D. Zhumagul, "Language is a symbol system that carries culture," *International Journal of Society, Culture and Language*, vol. 11, no. 1, pp. 203–214, 2023, doi: 10.22034/IJSL.2022.562756.2781.
- [47] Z. N. Ghafar, "Language as a component of culture: The role of language in cultural development," *ISAR Journal of Arts, Humanities and Social Sciences*, vol. 1, no. 1, pp. 1–4, 2023, doi: Language as a Component of Culture: The Role of Language in Cultural Development.
- [48] R. Dale, J. Roche, and N. Duran, "Language is complex," *International Journal of Psychology and Psychological Therapy*, vol. 8, no. 3, pp. 351–362, 2008.
- [49] M. Fadhl, B. Brick, P. Setyosari, S. Ulfa, and D. Kuswandi, "A meta-analysis of selected studies on the effectiveness of gamification method for children," *International Journal of Instruction*, vol. 13, no. 1, pp. 845–854, 2020, doi:

- 10.29333/iji.2020.13154a.
- [50] H. Snyder, "Literature review as a research methodology: An overview and guidelines," *Journal of Business Research*, vol. 104, pp. 333–339, Nov. 2019, doi: 10.1016/j.jbusres.2019.07.039.
  - [51] A. F. Luai, A. A. Malek, E. H. Hassan, B. A. M. Sabri, and N. A. M. Radzi, "Effectiveness of immersive technologies in oral health promotion and education: A systematic review," *Journal of Dentistry*, vol. 150, p. 105324, Nov. 2024, doi: 10.1016/j.jdent.2024.105324.
  - [52] I. Bongiovanni, D. M. Herold, and S. J. Wilde, "Protecting the play: An integrative review of cybersecurity in and for sports events," *Computers & Security*, vol. 146, p. 104064, Nov. 2024, doi: 10.1016/j.cose.2024.104064.
  - [53] F. Ahmed and K. Hussainey, "A bibliometric analysis of political connections literature," *Review of Accounting and Finance*, vol. 22, no. 2, pp. 206–226, Apr. 2023, doi: 10.1108/RAF-11-2022-0306.
  - [54] N. Donthu, S. Kumar, D. Mukherjee, N. Pandey, and W. M. Lim, "How to conduct a bibliometric analysis: An overview and guidelines," *Journal of Business Research*, vol. 133, pp. 285–296, Sep. 2021, doi: 10.1016/j.jbusres.2021.04.070.
  - [55] S. O'Connor, "Over twenty years of pedagogical research from Nurse education in practice: A bibliometric analysis from 2001 to 2023," *Nurse Education in Practice*, vol. 76, p. 103912, Mar. 2024, doi: 10.1016/j.nepr.2024.103912.
  - [56] S. Ghosh, "A bibliometric analysis of social life cycle assessment (2008–2022)," in *Recent Developments in Energy and Environmental Engineering*, 2023, pp. 75–86.
  - [57] W. Min and Z. Yu, "A bibliometric analysis of augmented reality in language learning," *Sustainability*, vol. 15, no. 9, p. 7235, Apr. 2023, doi: 10.3390/su15097235.
  - [58] G. Fabiano, A. Marcellusi, and G. Favato, "Public-private contribution to biopharmaceutical discoveries: a bibliometric analysis of biomedical research in UK," *Scientometrics*, vol. 124, no. 1, pp. 153–168, Jul. 2020, doi: 10.1007/s11192-020-03429-1.
  - [59] T. Kushartadi *et al.*, "Theme mapping and bibliometric analysis of two decades of smart farming," *Information*, vol. 14, no. 7, p. 396, Jul. 2023, doi: 10.3390/info14070396.
  - [60] A. R. Mathankar, "Bibliometrics: An overview," *International Journal of Library & Information Science (IJLIS)*, vol. 7, no. 3, pp. 9–15, 2018.
  - [61] I. Passas, "Bibliometric analysis: The main steps," *Encyclopedia*, vol. 4, no. 2, pp. 1014–1025, Jun. 2024, doi: 10.3390/encyclopedia4020065.
  - [62] J. Baas, M. Schotten, A. Plume, G. Côté, and R. Karimi, "Scopus as a curated, high-quality bibliometric data source for academic research in quantitative science studies," *Quantitative Science Studies*, vol. 1, no. 1, pp. 377–386, Feb. 2020, doi: 10.1162/qss\_a\_00019.
  - [63] R. Pranckutė, "Web of science (WoS) and Scopus: The titans of bibliographic information in today's academic world," *Publications*, vol. 9, no. 1, p. 12, Mar. 2021, doi: 10.3390/publications9010012.
  - [64] P. Mongeon and A. Paul-Hus, "The journal coverage of Web of Science and Scopus: a comparative analysis," *Scientometrics*, vol. 106, no. 1, pp. 213–228, Jan. 2016, doi: 10.1007/s11192-015-1765-5.
  - [65] A. Joshi, G. Mahila, M. Kalibadi, and R. Raipur, "Comparison between Scopus & ISI web of science," *Journal Global Values*, vol. 7, no. 1, pp. 1–11, 2016, doi: Comparison Between Scopus & ISI Web of Science.
  - [66] S. Stahlschmidt and D. Stephen, *Comparison of web of science, Scopus and dimensions database*. Germany: German Centre for Higher Education Research and Science Studies, 2020.
  - [67] M. Visser, N. J. van Eck, and L. Waltman, "Large-scale comparison of bibliographic data sources: Scopus, web of science, dimensions, Crossref, and Microsoft academic," *Quantitative Science Studies*, vol. 2, no. 1, pp. 20–41, Apr. 2021, doi: 10.1162/qss\_a\_00112.
  - [68] B. González-Albo and M. Bordons, "Articles vs. proceedings papers: Do they differ in research relevance and impact? A case study in the Library and Information Science field," *Journal of Informetrics*, vol. 5, no. 3, pp. 369–381, Jul. 2011, doi: 10.1016/j.joi.2011.01.011.
  - [69] L. Zhang and W. Glänzel, "Proceeding papers in journals versus the 'regular' journal publications," *Journal of Informetrics*, vol. 6, no. 1, pp. 88–96, Jan. 2012, doi: 10.1016/j.joi.2011.06.007.
  - [70] B. Chigarev, "Visualization of bibliometric data analysis on AI & machine learning topic from the scilit abstract database for 2021–2023," Jul. 18, 2024, doi: 10.20944/preprints202407.1464.v1.
  - [71] U. A. Bakar, M. S. Sayeed, O. A. Amodu, S. F. A. Razak, S. Yogarayan, and M. Othman, "Leveraging VOSviewer approach for mapping, visualisation, and interpretation of crisis data for disaster management and decision-making," *International Journal of Information Management Data Insights*, vol. 5, no. 1, p. 100314, Jun. 2025, doi: 10.1016/j.jjimei.2024.100314.
  - [72] T. McHenry, "Words as big as the screen: Native American languages and the internet," *Language Learning and Technology*, vol. 6, no. 2, pp. 102–115, 2002, doi: 10125/25164.
  - [73] D. J. Villa, "Integrating technology into minority language preservation and teaching efforts: An inside job," *Language Learning and Technology*, vol. 6, no. 2, pp. 92–101, 2002, doi: 10125/25163.
  - [74] M. WARD, "The additional uses of CALL in the endangered language context," *ReCALL*, vol. 16, no. 2, pp. 345–359, Nov. 2004, doi: 10.1017/S0958344004000722.
  - [75] M. C. Yang and D. Victoria Rau, "An integrated framework for archiving, processing and developing learning materials for an endangered aboriginal language in Taiwan," *5th Workshop on Asian Language Resources, ALR 2005 and 1st Symposium on Asian Language Resources Network, ALRN 2005 - Proceedings*, pp. 32–39, 2005.
  - [76] S. Lu, R. Huang, and F. Fotouhi, "Annotating linguistic data with ImageSpace for the preservation of endangered languages," *19th International Conference on Computers and Their Applications 2004, CATA 2004*, pp. 193–196, 2004.
  - [77] D. M. Nichols, I. H. Witten, T. T. Keegan, D. Bainbridge, and M. Dewsnip, "Digital libraries and minority languages," *New Review of Hypermedia and Multimedia*, vol. 11, no. 2, pp. 139–155, Dec. 2005, doi: 10.1080/13614560500351071.
  - [78] A. N. Chasaide *et al.*, "Speech technology for minority languages: The case of Irish (Gaelic)," in *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, 2006, vol. 1, pp. 181–184, doi: 10.21437/interspeech.2006-46.
  - [79] D. Cunliffe and D. Roberts-Young, "Online design for bilingual civil society: A Welsh perspective," *Interacting with Computers*, vol. 17, no. 1, pp. 85–104, Jan. 2005, doi: 10.1016/j.intcom.2004.10.003.
  - [80] D. Cunliffe and R. Harries, "Promoting minority-language use in a bilingual online community," *New Review of Hypermedia and Multimedia*, vol. 11, no. 2, pp. 157–179, Dec. 2005, doi: 10.1080/13614560500350750.
  - [81] R. Chowdhury, R. E. Butler, and S. Clarke, "Healthcare IT project failure," *Journal of Cases on Information Technology*, vol. 9, no. 4, pp. 1–15, Oct. 2007, doi: 10.4018/jcit.2007100101.
  - [82] L. Levin, "Adaptable, community-controlled, language technologies for language maintenance," in *Proceedings of the 13th Annual Conference of the European Association for Machine Translation, EAMT 2009*, 2009, pp. 8–11.
  - [83] J. Byamugisha, L. Sitbon, and M. Brereton, "Cultural and linguistic localization of games to bridge the digital and cultural divide

- in indigenous populations,” in *Proceedings of the 26th Australian Computer-Human Interaction Conference on Designing Futures: the Future of Design*, Dec. 2014, pp. 484–487, doi: 10.1145/2686612.2686689.
- [84] T. Zaman, H. Winschiers-Theophilus, A. W. Yeo, L. C. Ting, and G. Jengan, “Reviving an indigenous rainforest sign language,” in *Proceedings of the Seventh International Conference on Information and Communication Technologies and Development*, May 2015, pp. 1–4, doi: 10.1145/2737856.2737885.
- [85] Triando and L. Arhippainen, “Development and user experiences of the learn Viena Karelian mobile web game,” in *2019 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*, Oct. 2019, pp. 465–470, doi: 10.1109/ICACSIS47736.2019.8979925.
- [86] L. Xu, E. U. Dhonnchadha, and M. Ward, “User experience study of ‘Cipher: Faoi Gheasa’, a digital educational game for language learning and student engagement,” in *Proceedings of the 2nd Workshop on Games Systems*, Jun. 2022, pp. 5–8, doi: 10.1145/3534085.3534339.
- [87] D. Lackaff and W. J. Moner, “Local languages, global networks,” in *Proceedings of the 34th ACM International Conference on the Design of Communication*, Sep. 2016, pp. 1–9, doi: 10.1145/2987592.2987612.
- [88] M. Koole and K. wásakâyásiw Lewis, “Mobile learning as a tool for indigenous language revitalization and sustainability in Canada,” *International Journal of Mobile and Blended Learning*, vol. 10, no. 4, pp. 1–12, Sep. 2022, doi: 10.4018/IJMBL.2018100101.
- [89] S. Vatilifa, V. Hasheela-Mufeti, and L. Julius, “In pursuit of preserving Namibian Languages: The development of the Ndongika App, an Oshiwambo Children’s Android application,” in *Proceedings of the 19th International Conference on Smart Business Technologies*, 2022, pp. 130–136, doi: 10.5220/0011302900003280.
- [90] X. Zhou, Y. Zheng, and H. Hu, “Introduction to China minority speech acoustic parameter database Platform,” in *2009 Oriental COCOSA International Conference on Speech Database and Assessments*, Aug. 2009, pp. 132–136, doi: 10.1109/ICSDA.2009.5278361.
- [91] S. Sakti and S. Nakamura, “Recent progress in developing grapheme-based speech recognition for Indonesia ethnic languages: Javanese, Sundanese, Balinese and Batakese,” *4th Workshop on Spoken Language Technologies for Under-Resourced Languages, SLTU 2014*, pp. 46–52, 2014.
- [92] R. Jimerson, K. Simha, R. Ptucha, and E. Prudhommeaux, “Improving ASR output for endangered language documentation,” in *6th Workshop on Spoken Language Technologies for Under-Resourced Languages (SLTU 2018)*, Aug. 2018, pp. 187–191, doi: 10.21437/SLTU.2018-39.
- [93] J. Ma, H. Yu, D. Wang, C. Pu, and A. K. Singh, “RETRACTED: Design and creation of salar acoustical phonetics parameter database,” *Journal of Intelligent & Fuzzy Systems*, vol. 37, no. 3, pp. 3469–3476, Oct. 2019, doi: 10.3233/JIFS-179151.
- [94] Z. Chen and H. Yang, “Yi language speech recognition using deep learning methods,” in *2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)*, Jun. 2020, pp. 1064–1068, doi: 10.1109/ITNEC48623.2020.9084771.
- [95] N. San *et al.*, “Automated speech tools for helping communities process restricted-access corpora for language revival efforts,” in *Proceedings of the Fifth Workshop on the Use of Computational Methods in the Study of Endangered Languages*, 2022, pp. 41–51, doi: 10.18653/v1/2022.computel-1.6.
- [96] C. Yu, J. Yu, Z. Qian, and Y. Tan, “Endangered Tujia language speech recognition research based on audio-visual fusion,” in *Proceedings of the 2022 5th Artificial Intelligence and Cloud Computing Conference*, Dec. 2022, pp. 190–195, doi: 10.1145/3582099.3582128.
- [97] A. Mirza and D. Sundaram, “Architecting crowd-sourced language revitalisation systems: Generalisation and evaluation to te Reo Māori and Vietnamese,” in *5th International Conference on Computational Social Networks, CSoNet 2016*, 2016, pp. 333–344, doi: 10.1007/978-3-319-42345-6\_29.
- [98] A. A. Raza *et al.*, “Rapid collection of spontaneous speech corpora using telephonic community forums,” in *Interspeech 2018*, Sep. 2018, pp. 1021–1025, doi: 10.21437/Interspeech.2018-1139.
- [99] D. Teodorescu, J. Matalski, D. Lothian, D. Barbosa, and C. Demmans Epp, “Cree corpus: A collection of nēhiyawēwin resources,” in *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2022, pp. 6354–6364, doi: 10.18653/v1/2022.acl-long.440.
- [100] J. M. R. Corbett, “Indigenizing computer programming for cultural maintenance,” in *Conference Companion of the 2nd International Conference on Art, Science, and Engineering of Programming*, Apr. 2018, pp. 243–244, doi: 10.1145/3191697.3213802.
- [101] T. Ngoc Le and F. Sadat, “Revitalization of Indigenous languages through pre-processing and neural machine translation: The case of Inuktitut,” in *Proceedings of the 28th International Conference on Computational Linguistics*, 2020, pp. 4661–4666, doi: 10.18653/v1/2020.coling-main.410.
- [102] S. Miyagawa, K. Kato, M. Zlazli, S. Carlino, and S. Machida, “Building Okinawan Lexicon resource for language reclamation/revitalization and natural language processing tasks such as universal dependencies treebanking,” in *RESOURCEFUL 2023 - Workshop on Resources and Representations for Under-Resourced Languages and Domains, Proceedings of the 2nd*, 2023, pp. 86–91.
- [103] P. Dwivedi, C. Shraddha, S. Mathews, S. Majumder, R. Madhumathi, and M. R. Vasundhara, “Predicting language endangerment: A machine learning approach,” in *2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, Jul. 2020, pp. 1–7, doi: 10.1109/ICCCNT49239.2020.9225576.
- [104] P. Orynycz, “Say it right: AI neural machine translation empowers new speakers to revitalize Lemko,” in *3rd International Conference on Artificial Intelligence in HCI, AI-HCI 2022 Held as Part of the 24th HCI International Conference, HCII 2022*, 2022, pp. 567–580, doi: 10.1007/978-3-031-05643-7\_37.
- [105] P. Orynycz, “BLEU skies for endangered language revitalization: Lemko Rusyn and Ukrainian neural AI translation accuracy soars,” in *4th International Conference on Artificial Intelligence in HCI, AI-HCI 2023, held as part of the 25th International Conference on Human-Computer Interaction, HCII 2023*, 2023, pp. 135–149, doi: 10.1007/978-3-031-35894-4\_10.
- [106] K. A. B. Chew, “#KeepOurLanguagesStrong: Indigenous language revitalization on social media during the early COVID-19 pandemic,” *Language Documentation and Conservation*, vol. 2021, no. 15, pp. 239–266, 2021.
- [107] L. Kotut and D. S. McCrickard, “Winds of change: Seeking, preserving, and retelling Indigenous knowledge through self-organized online communities,” in *CHI Conference on Human Factors in Computing Systems*, Apr. 2022, pp. 1–15, doi: 10.1145/3491102.3502094.
- [108] H. Cho and Y. Yu, “Link prediction for interdisciplinary collaboration via co-authorship network,” *Social Network Analysis and Mining*, vol. 8, no. 1, p. 25, Dec. 2018, doi: 10.1007/s13278-018-0501-6.
- [109] G. F. Simons, “Two centuries of spreading language loss,” *Proceedings of the Linguistic Society of America*, vol. 4, no. 1, p. 27,




- Mar. 2019, doi: 10.3765/plsa.v4i1.4532.
- [110] W. Lu, S. Huang, J. Yang, Y. Bu, Q. Cheng, and Y. Huang, "Detecting research topic trends by author-defined keyword frequency," *Information Processing & Management*, vol. 58, no. 4, p. 102594, Jul. 2021, doi: 10.1016/j.ipm.2021.102594.
  - [111] L. Corrin, K. Thompson, G.-J. Hwang, and J. M. Lodge, "The importance of choosing the right keywords for educational technology publications," *Australasian Journal of Educational Technology*, vol. 38, no. 2, pp. 1–8, Jun. 2022, doi: 10.14742/ajet.8087.
  - [112] G. Zhao, S. Samsat, J. Levis, E. Chukharev-Hudilainen, and R. Gutierrez-Osuna, "Accent conversion using phonetic posteriorgrams," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Apr. 2018, pp. 5314–5318, doi: 10.1109/ICASSP.2018.8462258.
  - [113] B. Sisman, J. Yamagishi, S. King, and H. Li, "An overview of voice conversion and its challenges: From statistical modeling to deep learning," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 29, pp. 132–157, 2021, doi: 10.1109/TASLP.2020.3038524.
  - [114] L. Sun, K. Li, H. Wang, S. Kang, and H. Meng, "Phonetic posteriorgrams for many-to-one voice conversion without parallel data training," in *2016 IEEE International Conference on Multimedia and Expo (ICME)*, Jul. 2016, pp. 1–6, doi: 10.1109/ICME.2016.7552917.
  - [115] G. Zhao and R. Gutierrez-Osuna, "Using phonetic posteriorgram based frame pairing for segmental accent conversion," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 27, no. 10, pp. 1649–1660, Oct. 2019, doi: 10.1109/TASLP.2019.2926754.
  - [116] W.-Z. Zheng, J.-Y. Han, C.-K. Lee, Y.-Y. Lin, S.-H. Chang, and Y.-H. Lai, "Phonetic posteriorgram-based voice conversion system to improve speech intelligibility of dysarthric patients," *Computer Methods and Programs in Biomedicine*, vol. 215, p. 106602, Mar. 2022, doi: 10.1016/j.cmpb.2021.106602.
  - [117] G.-G. Lee and T. Gough, "An integrated framework for information systems planning and its initial application," *Journal of Information Technology*, vol. 8, no. 4, pp. 227–240, 1993, doi: 10.1177/026839629300800404.
  - [118] S. X. Duan, S. Wibowo, and H. Deng, "An integrated framework for understanding digital work in organizations," in *ACIS 2020 Proceedings - 31st Australasian Conference on Information Systems*, 2020, pp. 1–11.
  - [119] Gazali, I. Patekai, Ulfah, A. Bismawati, A. N. Izzati, and A. A. Yotolembah, "Bridging the gap between tradition and innovation: a survey on study of revitalization of Kailinese language as one of the precious cultural heritages of central Sulawesi (preliminary research)," in *Proceedings of the International Joint Conference on Arts and Humanities 2023 (IJCAH 2023)*, 2023, pp. 303–313, doi: 10.2991/978-2-38476-152-4\_28.
  - [120] C. S. Pinhanez, P. Cavalin, M. Vasconcelos, and J. Nogima, "Balancing social impact, opportunities, and ethical constraints of using AI in the documentation and vitalization of Indigenous languages," *IJCAI International Joint Conference on Artificial Intelligence*, vol. 2023-Augus, pp. 6174–6182, 2023, doi: 10.24963/ijcai.2023/685.
  - [121] T. Haokip, "Artificial intelligence and endangered languages," *SSRN Electronic Journal*, 2022, doi: 10.2139/ssrn.4212504.
  - [122] J. Rangel, "Challenges for language technologies in critically endangered languages," *UNESCO International Conference Language Technologies for All (LT4All)*, 2019.
  - [123] C. S. Saba, N. Ngepah, and N. M. Odhiambo, "Information and communication technology (ICT), growth and development in developing regions: evidence from a comparative analysis and a new approach," *Journal of the Knowledge Economy*, vol. 15, no. 3, pp. 14700–14748, Dec. 2023, doi: 10.1007/s13132-023-01571-8.
  - [124] K. V. Chitechi and D. A. Otanga, "Current trends in information technology: Which way for modern it experts," *International Journal of Computer Trends and Technology*, vol. 68, no. 7, pp. 31–36, Jul. 2020, doi: 10.14445/22312803/IJCTT-V68I7P105.
  - [125] A. Tatnall and A. Fluck, "Twenty-five years of the education and the information technologies journal: Past and future," *Education and Information Technologies*, vol. 27, no. 2, pp. 1359–1378, Mar. 2022, doi: 10.1007/s10639-022-10917-9.
  - [126] S. Wahyuni, S. Pranata, F. B. Setiaprabhawa, and L. Maftuchah, "Exploring the trend of technology use and innovation in health care service in hospitals through a bibliometric analysis," *Babcock University Medical Journal*, vol. 7, no. 1, pp. 51–63, Jun. 2024, doi: 10.38029/babcockuniv.med.j.v7i1.379.
  - [127] A. Alvi-Meza, M. H. Vásquez-Coronado, J. G. Delgado-Caramutti, and D. J. Blanco-Victorio, "Bibliometric analysis of fourth industrial revolution applied to heritage studies based on web of science and Scopus databases from 2016 to 2021," *Heritage Science*, vol. 10, no. 1, p. 189, Nov. 2022, doi: 10.1186/s40944-022-00821-3.
  - [128] M. Rodriguez-Marin, J. M. Saiz-Alvarez, and L. Huez-Ponce, "A bibliometric analysis on pay-per-click as an instrument for digital entrepreneurship management using VOSviewer and SCOPUS data analysis tools," *Sustainability*, vol. 14, no. 24, p. 16956, Dec. 2022, doi: 10.3390/su142416956.
  - [129] N. Z. M. Afandi, R. Umar, N. H. Sabri, S. Safei, and C. C. Lau, "Revealing trends: A bibliometric analysis of 28 years of space weather event research publications using the Scopus databases (1994–2022)," *Advances in Space Research*, vol. 72, no. 12, pp. 5753–5766, Dec. 2023, doi: 10.1016/j.asr.2023.09.029.
  - [130] Ž. Korde, S. Šuriņa, and K. Mārtinsons, "Research trends in drama therapy: a bibliometric analysis based on Scopus," *Frontiers in Psychology*, vol. 14, Dec. 2023, doi: 10.3389/fpsyg.2023.1327656.
  - [131] E. Kocyigit and T. Kocak, "Rheumatic diseases and gut microbiota publications in 2002–2022: a Scopus-based bibliometric analysis," *Rheumatology International*, vol. 43, no. 11, pp. 2115–2124, Aug. 2023, doi: 10.1007/s00296-023-05408-0.
  - [132] S. K. Patel and P. Jhalani, "Formulation of variables of environmental taxation: a bibliometric analysis of Scopus database (2001–2022)," *Environment, Development and Sustainability*, vol. 26, no. 3, pp. 7687–7714, Feb. 2023, doi: 10.1007/s10668-023-03027-0.
  - [133] B. Pham-Duc, T. Tran, D. Huu Hoang, and C. Bao Do, "Global scientific literature on human resource development: a bibliometric analysis using Scopus database," *European Journal of Training and Development*, vol. 47, no. 7/8, pp. 846–861, Aug. 2023, doi: 10.1108/EJTD-01-2022-0004.
  - [134] M. Turmuzi, I. G. P. Suharta, I. W. P. Astawa, and I. N. Suparta, "Mapping of mobile learning research directions and trends in Scopus-indexed journals: a bibliometric analysis," *International Journal of Interactive Mobile Technologies (IJIM)*, vol. 17, no. 03, pp. 39–69, Feb. 2023, doi: 10.3991/ijim.v17i03.36461.
  - [135] Z. Kannazarova, M. Juliev, A. Muratov, and J. Abuduwaili, "Groundwater in the commonwealth of independent states: A bibliometric analysis of scopus-based papers from 1972 to 2023, emphasizing the significance of drainage," *Groundwater for Sustainable Development*, vol. 25, p. 101083, May 2024, doi: 10.1016/j.gsd.2024.101083.
  - [136] F. Franceschini, D. Maisano, and L. Mastrogiovanni, "The museum of errors/horrors in Scopus," *Journal of Informetrics*, vol. 10, no. 1, pp. 174–182, Feb. 2016, doi: 10.1016/j.joi.2015.11.006.
  - [137] M. Tullu and S. Karande, "Success in publishing: Selecting an appropriate journal and braving the peer-review process," *Journal of Postgraduate Medicine*, vol. 64, no. 1, pp. 1–4, 2018, doi: 10.4103/jpgm.JPGM\_618\_17.
  - [138] P. DATTA and M. T. JONES, "Being published in reputable academic and scientific journals: Key criteria for acceptance,"






- International Journal of Higher Education Management*, vol. 05, no. 01, 2018, doi: 10.24052/ijhem/v05n01/art05.
- [139] C. Madeira, P. Rodrigues, and M. Gomez-Suarez, "A bibliometric and content Analysis of sustainability and smart tourism," *Urban Science*, vol. 7, no. 2, p. 33, Mar. 2023, doi: 10.3390/urbansci7020033.
- [140] M. Islam, M. Hossain, and M. Ali, "Bibliometric analysis of the scientific literature on library service through social media," *Library Philosophy & Practice*, 2022.
- [141] S. Fuada, G. Ma, and M. Katz, "Global growth and trends of in-body communication research—insight from bibliometric analysis," *International Journal of Online and Biomedical Engineering (IJOE)*, vol. 20, no. 01, pp. 128–149, Jan. 2024, doi: 10.3991/ijoe.v20i01.44967.
- [142] N. O. D. Ellili, "Bibliometric analysis and systematic review of environmental, social, and governance disclosure papers: current topics and recommendations for future research," *Environmental Research Communications*, vol. 4, no. 9, p. 092001, Sep. 2022, doi: 10.1088/2515-7620/ac8b67.

## BIOGRAPHIES OF AUTHORS






**Leonardi Paris Hasugian**    received a B.Eng. degree in informatics engineering from Universitas Komputer Indonesia, Indonesia, in 2011 and a double degree in the master's degree in information systems from Universitas Komputer Indonesia and computer engineering from Youngsian University in 2014. Currently, he is an assistant professor at the Department of Information Systems, Universitas Komputer Indonesia, and since 2025, he has been pursuing to PhD study on the Department of Information Systems, Institut Teknologi Sepuluh Nopember. His research interests include analysis and assessment of information systems, knowledge management, website development, and UI/UX design for the worker, education, government, social life, and culture fields. He can be contacted at 7026242002@student.its.ac.id or leonardi@email.unikom.ac.id.






**Syifaul Fuada**    received an M.Sc. in electrical engineering option microelectronics from the School of Electrical Engineering and Informatics, Institut Teknologi Bandung (ITB), Indonesia. He was with the University Center of Excellence at Microelectronics (PUI-PT Mikroelektronika) ITB from 2016-2018 as a main researcher. He is a university lecturer at the Department of Telecommunication, Universitas Pendidikan Indonesia (UPI). He is pursuing a doctoral degree at the Faculty of Information Technology and Electrical Engineering (FITEE), Centre for Wireless Communications (CWC), University of Oulu, Oulu, Finland. He can be contacted at email syifaulfuada@upi.edu.






**Triana Mugia Rahayu**    is a lecturer at the Department of Mechatronics Engineering, Parahyangan Catholic University, Indonesia. She is currently pursuing a doctoral degree at the Faculty of Medicine, University of Oulu, Finland. Her research interests are Acquistio optic sensing for brain monitoring. She can be contacted at email: triana.mr@unpar.ac.id.






**Apridio Edward Katili**    is a staff and researcher at Badan Riset dan Inovasi Nasional, Indonesia. He received a bachelor's degree from Universitas Komputer Indonesia in 2015. He is active in journal management and research in the design field. He can be contacted at email apri018@brin.go.id.



**Feby Artwodini Muqtadiroh**    received a bachelor's degree in informatics engineering from Institut Teknologi Sepuluh Nopember, Indonesia, in 2002, a master's degree in electrical engineering and informatics from Institut Teknologi Bandung in 2007, and a doctoral degree in electrical engineering from Institut Teknologi Sepuluh Nopember in 2024. She is a university lecturer at the Department of Information Systems, Institut Teknologi Sepuluh Nopember. Her research interests on educational technology, software engineering, IS/IT evaluation, decision making, and behavior analysis. She can be contacted at feby@is.its.ac.id.



**Nur Aini Rakhmawati**    received a bachelor's degree in informatics engineering from Institut Teknologi Sepuluh Nopember, Indonesia, in 2004, a master's degree in electronic engineering from National Taiwan University of Science and Technology in 2008, and a doctoral degree in information engineering from National University of Ireland, Galway in 2016. She is a Professor at the Department of Information Systems, Institut Teknologi Sepuluh Nopember. Her research interests include big data, linked data, knowledge graphs, computer ethics, and computational social science. She can be contacted at nur.aini@its.ac.id.