eISSN : 2287-4577
pISSN : 2287-9099
https://www.jistap.org
Vol. 12 No. 3 September 30, 2024

Journal of Information Science Theory and Practice

Journal of Information Science Theory and Practice

JISTaP

01 Exploring the Possibility of Using Chatbots as Educational Tools for School Libraries
14 Breaking the Code of Silence: A Qualitative Exploration of Cyberbullying Through the Lens of Habermas's Theory of Communicative Action
36 Incorporating Deep Median Networks for Arabic Document Retrieval Using Word Embeddings-Based Query Expansion
49 The MapDS-Onto Framework for Matching Formula Factors of KPIs and Database Schema: A Case Study of the Prince of Songkla University
63 A Digital Thesaurus of the Traditional Common Culture of the Greater Mekong Subregion
75 Assessment and Support Measures of Academic Journals in the National Open Access Platform AccessON

Vol. 12 | No. 3 | September 30, 2024

KISTI www.kisti.re.kr

Indexed/Covered by SCOPUS, LISA, DOAJ, CrossRef, and KCI

JISTaP
Journal of Information Science Theory and Practice

General Information

Aims and Scope

The Journal of Information Science Theory and Practice (*JISTaP*) is an international journal that aims at publishing original studies, review papers and brief communications on information science theory and practice. The journal provides an international forum for practical as well as theoretical research in the interdisciplinary areas of information science, such as information processing and management, knowledge organization, scholarly communication and bibliometrics. JISTaP will be published quarterly, issued on the 30th of March, June, September, and December. JISTaP is indexed in the Scopus, Korea Citation Index (KCI), Directory of Open Access Journal (DOAJ), Korea Science Citation Index (KSCI) and KoreaScience by the Korea Institute of Science and Technology Information (KISTI) as well as CrossRef. The full text of this journal is available on the website at http://www.jistap.org

Indexed/Covered by

Scopus LISA DOAJ DIRECTORY OF OPEN ACCESS JOURNALS Crossref Cited-by 한국학술지인용색인 Korea Citation Index

Publisher

Korea Institute of Science and Technology Information
66, Hoegi-ro, Dongdaemun-gu, Seoul, Republic of Korea
(T) +82-2-3299-6105
(F) +82-2-3299-6067
E-mail: jistap@kisti.re.kr
URL: http://www.jistap.org

Managing Editor: Seonhee Lee, Eungi Kim

Copy Editor: Ken Eckert

Design & Printing Company: MEDrang Inc.

2nd Floor, Hyoryeong Building, 32 Mugyo-ro, Jung-gu, Seoul 04521, Korea
(T) +82-2-325-2093
(F) +82-2-325-2095
E-mail: info@medrang.co.kr
Homepage: www.medrang.co.kr

Open Access and Creative Commons License Statement

All JISTaP content is Open Access, meaning it is accessible online to everyone, without fee and authors' permission. Open Access articles are automatically archived in the Korea Institute of Science and Technology Information (KISTI)'s Open Access repository (AccessON). All JISTaP content is published and distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/). Under this license, the authors retain full ownership of their work, while permitting anyone to use, distribute, and reproduce the content in any medium, as long as the original authors and source are cited. For any reuse, redistribution, or reproduction of a work, users must clarify the license terms under which the work was produced.

This paper meets the requirements of KS X ISO 9706, ISO 9706-1994 and ANSI/NISO Z39.48-1992 (Permanence of Paper).

2024 Copyright © Korea Institute of Science and Technology Information

JISTaP
Journal of Information Science Theory and Practice

Editorial Board

Co-Editors-in-Chief

Gary Marchionini
University of North Carolina, USA

Dong-Geun Oh
Keimyung University, Korea

Co-Associate Editors

Kiduk Yang
Youngnam Old Document Archive Center, Korea

Wan-Jong Kim
Korea Institute of Science and Technology Information, Korea

Co-Managing Editors

Seonhee Lee
Korea Institute of Science and Technology Information, Korea

Eungi Kim
Keimyung University, Korea

Editorial Board

Dan Albertson
University at Buffalo, SUNY, USA

Daniel Martínez Ávila
San Paulo State University, Brazil

Beeraka Ramesh Babu
University of Madras, India

Pia Borlund
University of Copenhagen, Denmark

France Bouthillier
McGill University, Canada

Kathleen Burnett
Florida State University, USA

Ina Fourie
University of Pretoria, South Africa

Boryung Ju
Louisiana State University, USA

Thomas Mandl
Universiät Hildesheim, Germany

Lokman I. Meho
American University of Beirut, Lebanon

Jin Cheon Na
Nanyang Technological University, Singapore

Dan O'Connor
Rutgers University, USA

M. P. Satija
Guru Nanak Dev University, India

Ou Shiyan
Nanjing University, China

Paul Solomon
University of South Carolina, USA

Kulthida Tuamsuk
Khon Kaen University, Thailand

Lihong Zhou
Wuhan University, China

Xiao-Zhi Gao
University of Eastern Finland, Finland

Wole Michael Olatokun
University of Ibadan, Nigeria

Consulting Editors

Reysa Alenzuela
Oriental Institute, ASCR, Czech Republic

Wayne Buente
University of Hawaii, USA

Sujin Butdisuwan
Mahasarakham University, Thailand

Seon Heui Choi
Korea Institute of Science and Technology Information, Korea

M. Krishnamurthy
DRTC, Indian Statistical Institute, India

S.K. Asok Kumar
The Tamil Nadu Dr Ambedkar Law University, India

P. Rajendran
SRM University, India

2024 Copyright © Korea Institute of Science and Technology Information

JISTaP
Journal of Information Science Theory and Practice

Table of Contents

Vol. 12 No. 3 September 30, 2024
Journal of Information Science Theory and Practice • https://www.jistap.org

Articles

01 Exploring the Possibility of Using Chatbots as Educational Tools for School Libraries
Seong-Kwan Lim

14 Breaking the Code of Silence: A Qualitative Exploration of Cyberbullying Through the Lens of Habermas's Theory of Communicative Action
January Febro Naga, Joshua Isaguirre, Elanie Vizconde, Raymund Sison

36 Incorporating Deep Median Networks for Arabic Document Retrieval Using Word Embeddings-Based Query Expansion
Yasir Hadi Farhan, Mohanaad Shakir, Mustafa Abd Tareq, Boumedyen Shannaq

49 The MapDS-Onto Framework for Matching Formula Factors of KPIs and Database Schema: A Case Study of the Prince of Songkla University
Kittisak Kaewninprasert, Supaporn Chai-Arayalert, Narueban Yamaqupta

63 A Digital Thesaurus of the Traditional Common Culture of the Greater Mekong Subregion
Suwannee Hoaihongthong, Kanyarat Kwiecien

75 Assessment and Support Measures of Academic Journals in the National Open Access Platform AccessON
Hyekyong Hwang, Eun Jee Lee, Wan Jong Kim, Jin Ho Park

89 **Call for Paper**

90 **Information for Authors**

2024 Copyright © Korea Institute of Science and Technology Information

JISTaP http://www.jistap.org
Journal of Information Science Theory and Practice

RESEARCH PAPER
J Inf Sci Theory Pract 12(3): 1-13, 2024

eISSN : 2287-4577 pISSN : 2287-9099

https://doi.org/10.1633/JISTaP.2024.12.3.1

Exploring the Possibility of Using Chatbots as Educational Tools for School Libraries

Seong-Kwan Lim*
Educational School Librarianship, Graduate School of Education,
Kyonggi University, Suwon, Korea
E-mail: kglimpro@kyonggi.ac.kr

ABSTRACT

The purpose of this study is to investigate the possibility of using chatbots as a school library educational tool. In order to achieve the purpose of the study, 116 librarian teachers first investigated the types and contents of education conducted in the school library setting and the perception of chatbots there. In addition, 15 librarians (five elementary, five middle, and five high school) were asked to complete a structured questionnaire after using Google's Bard, Microsoft's Bing, and OpenAI's Nova to find out if it is possible to use chatbots in school library education. As a result, user and reading education chatbots were found to be common in school libraries, and 99% of librarians knew about them in some detail. However, the average chatbot performance by area was 2.9 out of 5 (2.6 points being the lowest). Nevertheless, chatbots are being developed utilizing deep learning methodologies and have excellent performance, and are very effective for content-based library education through problem-solving activities.

Keywords: school library, school library education, educational tools, chatbots, teacher-librarians

Received: August 29, 2023 **Revised:** February 7, 2024
Accepted: April 9, 2024 **Published:** September 30, 2024

***Corresponding Author:** Seong-Kwan Lim
https://orcid.org/0000-0002-9735-1717
E-mail: kglimpro@kyonggi.ac.kr

All JISTaP content is Open Access, meaning it is accessible online to everyone, without fee and authors' permission. Open Access articles are automatically archived in the Korea Institute of Science and Technology Information (KISTI)'s Open Access repository (AccessON). All JISTaP content is published and distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/). Under this license, the authors retain full ownership of their work, while permitting anyone to use, distribute, and reproduce the content in any medium, as long as the original authors and source are cited. For any reuse, redistribution, or reproduction of a work, users must clarify the license terms under which the work was produced.

© 2024 [Seong-Kwan Lim] This article is distributed under the Creative Commons Attribution License (CC BY), allowing unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. It is published by the Korea Institute of Science and Technology Information (KISTI).

1. INTRODUCTION

Interest in chatbots is strong all over the world. A *chatbot* is a representative model of Generative artificial intelligence (AI), the world's most popular IT technology, and is based on generative pre-trained transformer (GPT) technology, which is a transformer technology that can generate answers based on self-learning and process a large amount of data and context. ChatGPT, an interactive AI chatbot developed by Open AI in the U.S., has been at the center of discussion since the GPT-3.5 version was released in November 2022. Naturally, some countries are actively discussing regulations, such as temporarily banning their use from the perspective of privacy, and there are many concerns about the side effects of active use, but this attitude is rapidly changing.

Chatbots are a representative example of the Large Language Model, which is a super-large AI model that generates sentences similar to humans. The strength of chatbots is natural language processing (NLP) technology (the language that humans use on a daily basis), and chatbots learn, understand, and generate answers as appropriate. In other words, chatbots help users perform tasks in various fields, especially regarding translation, document writing, and coding, because they give answers accordingly when users enter text in the chat window. Predictive models using existing AI have difficulty communicating with people even if they have good performance. However, chatbots naturally connect people and AI because they can create text as if they were talking, based NLP.

Generative AI technology led by chatbots is connected to various fields and leads to tangible results. Among them, because questions are standardized, AI is already replacing tasks normally performed in places such as call centers where answers are repeated, and in the field of education, where more and more attempts are being made to use chatbots as a tool to assist classes.

The use of chatbots in Korean libraries is gradually increasing, and services are being operated as a homepage-based platform or a social network service platform, depending on the type and size of the service. According to Min (2021), there were 15 university libraries that applied chatbot services as of 2021, and among public libraries, Anyang city and Yongin city in Gyeonggi province, Geoje city in Gyeongsangnam province, and the Seodaemun, Seongdong, and Eunpyeong districts in the Metropolis of Seoul were providing chatbot services by integrating public libraries in their districts. However, a study that synthesizes and analyzes the use of chatbot services in school libraries has not yet been published. Therefore, the purpose of this study is to explore the possibility of the ever-increasing general use of chatbot technology as an educational tool in school libraries, and to seek and propose appropriate measures for its evolution.

2. LITERATURE REVIEW

2.1. School Library Education

In 1999, the International Federation of Library Associations and Institutions/United Nations Educational, Scientific and Cultural Organization's School Library Manifesto defined the role of a school library as "to develop students' enjoyment of reading and learning, habits and attitudes to use the library throughout their lives, and to guide them to evaluate and utilize information containing various materials" (International Federation of Library Associations and Institutions & United Nations Educational, Scientific and Cultural Organization, 1999).

The Role of the School Library (American Association of School Librarians, 2019) released in 2019 stated that the school library is an essential component of the educational environment, because it is a dynamic learning environment that bridges the gap between access and opportunities for all students. In addition, the education and services provided through school libraries are developed through the following six essential sharing bases, listed as "Inquire," "Include," "Collaborate," "Curate," "Exploration," and "Engage."

Song (2018) stated that school library education is an educational information service performed by school libraries to achieve school education goals. As the role of school libraries in school education increases, library use and reading education, which were previously operated as informal curricula, are expanding to information literacy instruction. In addition, school library education is divided into education on school libraries and education using school libraries. Education on school libraries aims to increase ability toward self-directed production of new information, and knowledge based on the ability to utilize various learning materials owned by school libraries. In addition, school library-based education is an activity that enhances learning ability by allowing students to self-directedly solve the subject's learning topics by using various learning tools owned by school libraries.

However, there are many school libraries in Korea where librarians do not work. Even if one is working, there may not be any assistant, and so many tasks other than training must be handled alone. Therefore, if there

 https://doi.org/10.1633/JISTaP.2024.12.3.1

is a way to reduce workload and focus on education, it should be seriously investigated.

Song (2019) suggests that librarians should be in charge of school libraries, which are physical and digital learning spaces that perform teaching/learning activities based on reading, exploration, research, thinking, imagination, and creativity. To help or aid librarians with their ever-increasing workload, ChatGPT should be used. ChatGPT is expected to be an essential element as a future learning tool.

Hence, what activities can librarians at school libraries do using ChatGPT? According to Yorio (2023), who received and organized the answers to this question from a chatbot: "Providing reference assistance," "Recommending books and other materials," "Helping with writing assignments," "Providing digital resources and online tutorials," "Creating interactive learning experiences," and "Providing language assistance for non-native speakers" were possible.

2.2. Chatbot Education Use

ChatGPT is attracting attention as an educational tool that realizes 'learning initiative,' which is emphasized in the current curriculum and future educational settings direction (Shin et al., 2023).

Hetler (2024) stated that because ChatGPT is versatile, it can be used to solve math problems or find keywords optimized for search engines when people code computer programs or compose music. However, because it does not fully understand the complexity of language used by humans, the answers may be inaccurate. Various statistics can be cited and provided, but there are limitations, such as not being able to make practical comments on what it means or relevant topics. In addition, ChatGPT can be helpful for some tasks, but it is also necessary to use it with caution because it has ethical problems, such as those associated with plagiarism, privacy, and cyber security. Rouse (2024) said that some schools have completely banned the use of ChatGPT, but instead of this, teachers need to teach students how to use this type of AI ethically so that they can develop critical thinking skills.

Dilmegani (2024) cited the advantages of using ChatGPT in education as three categories: speed, availability, and personalized learning, and presented seven different cases from the perspective of teachers and students, respectively, as summarized in Table 1.

In Korea, measures to utilize ChatGPT for education are also being actively sought, and in July 2023, the Ulsan Metropolitan Office of Education developed a class resource book using ChatGPT for English teachers and distributed it to schools. The data collection includes instructional design using ChatGPT, project class cases, and evaluation methods, and includes how to organize and utilize prompts, reconstruct or produce dialogues presented in textbooks at various levels, and create role-play or scripted theater class materials.

Additionally, Jeong and Shin (2023) developed an AI chatbot that can be used in elementary school science classes. It was applied to sixth grade students and results confirmed that using it improved scientific academic achievement and science positivity.

As previously stated, research on the use of chatbots and specific measures for school library education in Korea has not yet been published. Nevertheless, cases

Table 1. Top 7 ChatGPT education use cases

Target	Case	Detail
For teachers	Content creation	Use ChatGPT to create or select educational materials to suit students' needs, such as presentations, worksheets, and quizzes
	Grammar and writing check	Leverage AI's natural language understanding capabilities to help evaluate and improve the quality of writing (spelling, grammar, punctuation, and syntax errors)
	Grading	Review and score students' essays by analyzing the content, structure, and consistency of the text
	Designing syllabus outlines	Create and structure lesson plans in a consistent and effective manner
For students	Help with homework	Ask questions about a particular topic or concept, and improve the quality of a report or writing assignment
	Research	Supports research processes on academic projects, tasks, and personal interests
	Language learning	ChatGPT helps improve language skills because it provides translation, grammar description, vocabulary practice, and conversation simulation

AI, artificial intelligence.

are being introduced for training of librarian teachers with magazines such as school library journals and Internet newspapers. Among them, according to the case of Seoul Sinlim High School, published in the Internet Newspaper Education Plus (Woo, 2023), ChatGPT was used as an auxiliary or complementary tool in cooperation with classes with chemistry subjects. At this time, the chemistry teacher selected the keywords "medicine," "new material," "biochemistry," and "cosmetics/fashion," and the librarian teacher recommended a book suitable for each topic. Afterwards, students selected a topic to read and discuss from these books, and they had the opportunity to ask and correct whether it was appropriate as a topic and whether the arguments for or against were valid.

The above literature review results show that there are several issues not only in the use of AI chatbots themselves, but also in their use in school education. If they can be used positively while diminishing any possible negative aspect associated with this technology, it raises expectations that the effectiveness of education can be increased.

3. RESEARCH OBJECTIVES

The purpose of this study is to investigate the perceptions of librarian teachers regarding the possibility of using chatbots as a school library educational tool. Recognition is knowing things accurately and understanding their meaning correctly, and in this study it means knowledge of the meaning and function of chatbots. The research question set chosen to achieve the research purpose are as follows.

RQ1. Do you know about chatbots?

RQ2. Do you think it is necessary to use chatbots for school library education?

RQ3. What would be the most effective school library education if you used a chatbot?

RQ4. What preparations are needed to use chatbots for school library education?

RQ5. How does each chatbot perform in terms of educational usability?

4. RESEARCH METHOD

4.1. Chatbot Selections

Chatbot service can play the following roles in improving the learning experience by connecting educational institutions and students. First, the chatbot can improve the curriculum and adjust the learning path according to the progress of the student. Secondly, it is possible to provide an appropriate learning experience to each student. Thirdly, students can receive answers to solve tasks through the chatbot. Fourthly, student progress data generated by the chatbot can be used to track performance and provide improvement measures or study pathways. Fifthly, since the chatbot provides a conversation-based experience, it can increase interest in learning. Sixthly, learning support can be received whenever desired.

However, since the learning effect may vary depending on which chatbot is selected, it is necessary to first explore the possible viability of each as an educational tool. The following Table 2 summarizes the characteristics of the three chatbots selected to explore possibilities for use in school libraries.

Table 2. Chatbots analyzed

Chatbot	Development company	Characteristic
Bard	Google	Bard is an artificial intelligence chatbot released in March 2023, and opened to 180 countries around the world on May 10. Bard features Google's latest large-scale language model, Palm2 (PALM), an upgraded version of Farm, introduced in April 2022, which supports more than 100 languages and enables reasoning and coding in science and mathematics, based on 530 billion parameters
Bing	Microsoft	Bing is a search service launched by Microsoft on June 1, 2009, which combines live search and power search technology acquired in 2008. On February 7, 2023, 'the new Bing' premiered, based on OpenAI's large language model, including chatbot functionality
Nova	OpenAI	Nova is a smart AI Chatbot with the latest ChatGPT technology from OpenAI GPT-4, released in March 2023 to answer unlimited questions. The industry's first cross-compatible platform, ChatGPT-based application, supports more than 140 multilingual applications and can help with anything, including writing and composing

AI, artificial intelligence; GPT, generative pre-trained transformer.

4.2. Survey

The main premise of this study is to investigate the possibility of using chatbots as educational tools for school libraries. In order to achieve the research objectives, a survey questionnaire was developed from the librarian teacher recognition study (Kim et al., 2023) on the introduction of ChatGPT in school libraries and the librarian recognition study using metaverse in university libraries (Oh & Lee, 2022). Subsequently, an online survey was conducted for five days from August 17 to August 21, 2023, composed of questions in the form provided by the Naver Office. The survey used a snow sampling method that asks some librarian teachers working in the Metropolis of Seoul and Gyeonggi province to participate by sending links, and then gradually expanding them to recruit samples. Instead of predicting the number of samples, only the results of the number of responses during the survey period were set as analysis targets. The reason the survey target area was limited to Seoul and Gyeonggi province is because Gyeonggi province has the largest number of school libraries in the country and Seoul is the capital of Korea and also the center of education. Table 3 summarizes the survey questions.

4.3. Structured Questionnaire

Next, in order to practically evaluate whether each chatbot could be used as a school library educational tool, a structured questionnaire was created to ask questions directly and check the answers. The structured questionnaire was composed of the following Table 4 materials by applying "social dialogue ability," "information provision ability in various fields," "language command ability," "creative ability," and "moral discernment," which were developed by Shin et al. (2023) to explore the possibility of using ChatGPT.

The performance missions assessment for each chatbot evaluation summarized in Table 3 includes five different areas with a total of 15 detailed questions. Among them, the "social conversation ability" area was included for priority evaluation because chatbots generally focus on topics of daily conversation. The performance mission of "ability to provide information in various fields" and "creative ability" were evaluated by chatbots and was a choice that took into account its function. Next, "moral discernment" took

Table 3. Survey items

No.	Question content	Selection
1	What school do you work for?	① Elementary school ② Middle school ③ High school
2	What education is being conducted in the school library where you are working? (multiple responses)	① User education ② Reading education ③ Information utilization education ④ Curriculum cooperation classes ⑤ Etc.
3	Do you know about chatbots?	① I know ② I don't know
4	Do you think it is necessary to use chatbots for school library education?	① They need to be used ② There is no need to use them
4-1	Why did you respond that you don't need to use one? (multiple responses)	① I don't know how to use one ② Because of the problems that an incorrect answer will cause ③ On account of ethical issues ④ Because of privacy issues ⑤ I wasn't ready ⑥ Etc.
5	What would be the most effective school library education if you used a chatbot?	① User education ② Reading education ③ Information utilization education ④ Curriculum cooperation classes ⑤ Etc.
6	What preparations are needed to use chatbots for school library education?	A subjective response

Table 4. Chatbot assessment area performance mission

Area	Performance mission
Social conversation skills	1. Say hello, ask your name, ask about your hobbies and job 2. Seeking advice on personal matters
Understanding of the library	3. Ask what the library does 4. Find the world's first library 5. Learn how to use the library 6. Learn about the library's resources
Ability to provide information in a variety of fields	7. Learn how to study well 8. Find out when humans first landed on the moon 9. Find out the world's most famous sports players 10. Find out the different types of coffee 11. Learn about the Korean War 12. Learn about Beethoven, a musician 13. Learn about the Nobel Prize in Literature
Creative ability	14. Solving learning tasks for each subject (e.g., how do I make war disappear? / What should I do to protect the Earth from environmental pollution? / How can I draw the most beautiful picture in the world?)
Moral discernment	15. Asking morally sensitive questions (e.g. what do you think of the death penalty? / What do you think about abortion? / What do you think about euthanasia?)

Table 5. Chatbot performance evaluation paper by area

Area	1. Very inadequate	2. Insufficient	3. Usually	4. Good	5. Excellent
Social conversation skills					
Understanding of the library					
Ability to provide information in a variety of fields					
Creative ability					
Moral discernment					
Educational applicability					

into account the controversy swirling over the unethicality of chatbots. Finally, "ability to understand libraries" was set as an implementation mission in terms of exploring the possibility of using it in school library education.

In addition, Table 5 is a performance evaluation paper for chatbots by area, and "educational utilization possibility" is included in the five areas presented in Table 3. The level for each area was composed of a five-point scale ranging from insufficient to excellent.

Table 6. Demographic characteristics of survey respondents

Area	Category	No. of participants (%)
Gender	Female	113 (97.4)
	Male	3 (2.6)
Age range	20s	4 (3.5)
	30s	47 (40.5)
	40s	44 (37.9)
	50s	21 (18.1)

5. ANALYSIS

5.1. Analysis of Survey Results

5.1.1. Demographic Characteristics of Survey Respondents

A total of 116 librarian teachers participated in the online survey, conducted for five days from August 17 to August 21, 2023. This number corresponds to approximately 26.4% of the 228 librarian teachers based in Seoul (1,307 school libraries) and 212 librarian teachers in Gyeonggi province (2,469 school libraries) based on 2023 education

 https://doi.org/10.1633/JISTaP.2024.12.3.1

Table 7. Preparation necessary to use chatbots for school library education

Answer	Category	Subject
Chatbots are useful for information utilization education. What students need when talking to AIs is questioning and discernment. The two abilities presented should be developed through education	Training and ethics on chatbots	Appropriate education
The librarian teacher should accurately identify the pros and cons of using chatbots, and education is needed to determine whether the information used by students during class is correct		
Correct information ethics education		
Information literacy education should be provided to students in advance to develop the ability to critically verify information, including the contents provided by chatbots		
Humanities reading education that can generate necessary questions using chatbots		
Education on how to use chatbots in class		
Need for training to fact-check information provided by chatbots		
Teacher-librarian training		
The teacher-librarian must be professionally aware of the chatbot		
Writing education and information ethics education		
Copyright education, AI ethics education, privacy education, flexible application of reading and writing programs		
Training librarian teachers for accurate information, artificial intelligence use education for students, and continuous information ethics education		
Utilization and ethics education for teachers		
Familiarization with how to use a chatbot, pay for it, and train media literacy		
Training on how to use chatbots, including pros and cons		
Parallel education on the risk of chatbots		
Ability to verify answers; appropriate comparative screening capabilities through web searches	Ability to assess and utilize accuracy of information	
Ability to grasp the accuracy of information		
Ensuring that feedback is well applied to students' levels and ethical issues		
Accuracy of information		
Fact checking of information presented in chatbot		
Accurate understanding of the purpose of use and awareness of negative consequences		
Checking for incorrect answers		
Thinking about how to use the strengths of chatbots for education		
Literacy ability to evaluate the accuracy of information calculated through chatbots and use it in the right place		
Ability to judge inaccuracies		
In elementary schools, it is necessary to establish a system that can play an auxiliary role in tasks such as answering questions about billing symbols and book location, reading education and curriculum cooperation classes, information utilization education, loans, returns, overdue loans, etc.	Building a chatbot utilization environment	Environmental construction
Information devices to be used and comparable reference book materials must be prepared		
Information ethics education and devices that enable individual student use		
Establishing a library system and producing educational materials		

Table 7. Continued

Answer	Category	Subject
Purchasing books that fit the purpose of the library		
Utilization manual		
Sufficient understanding of chatbots by librarians, and establishing an environment where chatbots can be used		
Education for teacher-librarians, utilizing available devices and spaces		
Personal media devices or beam projectors		
Building a seamless internet system		
In order to study various previous studies through chatbots and apply them to school libraries, it is necessary to share cases through professional learning communities with those in the field		

AI, artificial intelligence.

statistics (Korean Education Statistics Service, 2023). Table 6 summarizes the demographic characteristics of survey respondents.

It should be pointed out that according to the results summarized in Table 6, 97.4% of survey respondents were women, and people in their 30s and 40s accounted for 78.4% of the total.

5.1.2. Survey Participation Status

Regarding the schools where the librarian teachers who participated in the survey are working, there were a total of 60 participating librarian teachers working in elementary schools, accounting for 51.7% of the total. Next were middle schools with 32 teachers or 27.6%, and high schools contributed to 24 teachers or 20.7%.

5.1.3. Education in the School Library

"User education" (40.9%) was provided the most in school libraries. Usually, school libraries provide user education for all students at the beginning of the school year, so this can be said to be a natural result. The reason why the response rate to "curricular cooperation classes" was relatively small is because cooperation with teachers by grade is necessary for education. In other words, it can be impossible to implement – it is only up to the will of the librarian teacher.

5.1.4. Teacher-Librarians' Perception of Chatbots

When librarian teachers were asked about their awareness of chatbots, 96.6% answered that they knew about chatbots to varying degrees, so it can be surmised that almost all librarian teachers know of chatbots.

5.1.5. The Necessity of Using Chatbots for School Library Education

When asked whether chatbots need to be used in school library education, 81% of librarian teachers responded that there was a need to use chatbots in school library education.

5.1.6. Reasons for Responding That There is No Need to Use Chatbots for School Library Education

The biggest reason why librarians do not think they need to use chatbots for school library education was because they were not ready for this technology (17.6%). The "preparation" referred to here includes everything from the facility aspect of the school library to what education to use and how to use it, and even the personal readiness of librarian teachers. In addition, 16.2% of the respondents said they did not need to use chatbot answers because they were not accurate at all moments – the technology is not mature enough.

5.1.7. Type of School Library Education That Will Be Most Effective When Using Chatbots

Librarian teachers cited the "information utilization education" (70.7%) category as the most effective form of school library education when using chatbots. Since chatbots also use information and communication technology, it can be analyzed that they intend to use it to develop students' abilities to effectively find and utilize the information they need. Furthermore, this result also gave implications for this study to explore the possibility of using chatbots as an educational tool for school libraries.

The following Table 7 shows librarian teachers' free answers to the question about what preparations are

Table 8. Key points of study participants

School library at work	Teacher-librarian	Work experience (yr)
Elementary school	A	12
	B	10
	C	17
	D	8
	E	14
Middle school	F	9
	G	11
	H	12
	I	15
	J	8
High school	K	6
	L	5
	M	13
	N	9
	O	10

needed if chatbots are to be used in school library education, using open coding among the grounded theories presented by Morgan and Scannell (1998). The analysis was conducted to find overarching themes and categories. In other words, the work was done to name and categorize concepts with similar content.

According to the results summarized in Table 7, librarians responded that "environmental construction" and "appropriate education (training)" are needed first in order to use chatbots for school library education. "Environmental construction" includes an Internet environment and devices that can access spaces and chatbots in the school library, and "appropriate education" was considered necessary for librarians and students, respectively. In addition, the educational content needs to address the advantages and weaknesses of chatbots, copyright laws and information ethics, and use case manuals in school libraries.

5.2. Analysis of Structured Questionnaire

As shown in Table 4, the structured questionnaire contains missions by area of "social conversation ability," "information provision ability in various fields," "language command ability," "creative ability," and "moral discern-

Table 9. Evaluation of Bard

Teacher-librarian	Category					
	Social conversation skills	Understanding of the library	Ability to provide information in a variety of fields	Creative ability	Moral discernment	Educational applicability
A	3	3	3	2	3	3
B	4	3	4	3	3	3
C	3	3	3	3	3	3
D	3	2	3	3	3	2
E	4	3	3	3	3	2
F	4	2	3	2	2	2
G	3	3	4	3	3	2
H	3	2	3	2	3	2
I	4	3	3	3	3	2
J	3	3	3	3	3	3
K	3	2	3	3	3	3
L	3	2	3	2	2	2
M	3	2	3	3	3	2
N	3	3	3	3	2	2
O	4	3	3	3	3	3
Average	3.3	2.6	3.1	2.7	2.8	2.4

Table 10. Evaluation of Bing

Teacher-librarian	Category					
	Social conversation skills	Understanding of the library	Ability to provide information in a variety of fields	Creative ability	Moral discernment	Educational applicability
A	3	2	3	2	3	2
B	3	3	3	3	3	3
C	3	2	3	3	3	2
D	3	2	4	3	3	2
E	3	2	3	3	3	2
F	3	2	3	2	2	2
G	3	2	3	3	3	2
H	3	3	3	2	3	2
I	3	3	3	3	2	2
J	3	3	3	3	3	3
K	3	3	3	3	3	3
L	3	2	4	3	2	2
M	3	2	3	3	3	2
N	3	3	3	3	2	3
O	3	3	3	3	3	3
Average	3	2.5	3.1	2.8	2.7	2.3

ment." This questionnaire was delivered to 15 librarians working in elementary, middle, and high school libraries to carry out missions through Google's Bard, Microsoft's Bing, and OpenAI's Nova, and to reflect the performance evaluation scores of Table 4. Table 8 summarizes the mix of work experience levels for librarian teachers who participated in the structured questionnaire.

According to the results summarized in Table 8, a total of 15 librarians participated in missions by area using chatbots, including five elementary schools, five middle schools, and five high schools. Among those who accepted participation in the study were students of the researcher or those who met at conferences and made friends, and the criteria for selection included librarian teachers with at least five years of experience working in the school library. The selection criteria were librarians with at least five years of experience working in school libraries. As a result, the average work experience of the librarians who participated in answering the structured questionnaire in this study was 12.2 years in elementary school, 11 years in middle school, and 8.6 years in high school.

Tables 9-11 are summarized by analyzing the evaluation results for Bard, Bing, and Nova, respectively.

In the evaluation results of the chatbot Bard summarized in Table 9, the area with the highest average score was "social conversation ability," and the area with the lowest score was "educational usability." Since "library understanding ability" ranks 5th out of 6 areas, the possibility of using it as a school library educational tool can be said to be low.

In the evaluation results of the chatbot Bing summarized in Table 10, the area with the highest average score was "ability to provide information in various fields," and the area with the lowest score was "educational usability." In addition, since "library understanding ability" ranks 5th out of 6 areas, the possibility of using it as a school library educational tool can also be said to be low.

In the evaluation results of the chatbot Nova summarized in Table 11, the area with the highest average score was "ability to provide information in various fields," and the areas with the lowest were "ability to understand libraries" and "moral discernment." However, since "educational usability" ranks 3rd out of 6 areas, it can be said that it has the highest possibility of being used as an educational tool.

Finally, Table 12 summarizes the results from Tables

 https://doi.org/10.1633/JISTaP.2024.12.3.1

Table 11. Evaluation of Nova

Teacher-librarian	Category					
	Social conversation skills	Understanding of the library	Ability to provide information in a variety of fields	Creative ability	Moral discernment	Educational applicability
A	4	3	4	3	3	3
B	4	3	4	3	3	3
C	3	3	3	3	3	3
D	4	3	4	3	3	4
E	4	3	4	3	3	3
F	3	3	3	4	3	3
G	4	3	4	3	3	3
H	4	3	4	4	3	4
I	4	3	4	3	3	3
J	3	3	3	3	3	3
K	3	3	3	3	3	3
L	4	3	4	4	3	4
M	4	3	4	3	3	3
N	3	3	4	3	3	3
O	4	3	4	3	3	4
Average	3.6	3	3.7	3.2	3	3.3

Table 12. Results of chatbot performance evaluation by area

Area	Chatbot			
	Bard	Bing	Nova	Average
Social conversation skills	3.3	3	3.6	3.3
Understanding of the library	2.6	2.5	3	2.7
Ability to provide information in a variety of fields	3.1	3.1	3.7	3.3
Creative ability	2.7	2.8	3.2	2.9
Moral discernment	2.8	2.7	3	2.8
Educational applicability	2.4	2.3	3.3	2.6
Average	2.8	2.7	3.3	2.9

9-11.

According to the results summarized in Table 12, among the three chatbots that were subject to performance evaluation, Nova received the highest score in the entire performance for each of the six areas. However, the overall average score was usually less than 2.9 points, and in particular, the "educational use possibility" was the lowest at 2.6 points. From the above results, librarians evaluated that while chatbots are still insufficient in other areas, they are higher than average in "social conversation ability" or "information provision ability in various fields," so it is possible to use them as an educational tool in school libraries. Although it is clear that these results were evaluated by librarian teachers working in elementary schools, middle schools, and high schools, they do not take into account all differences in the context of information or

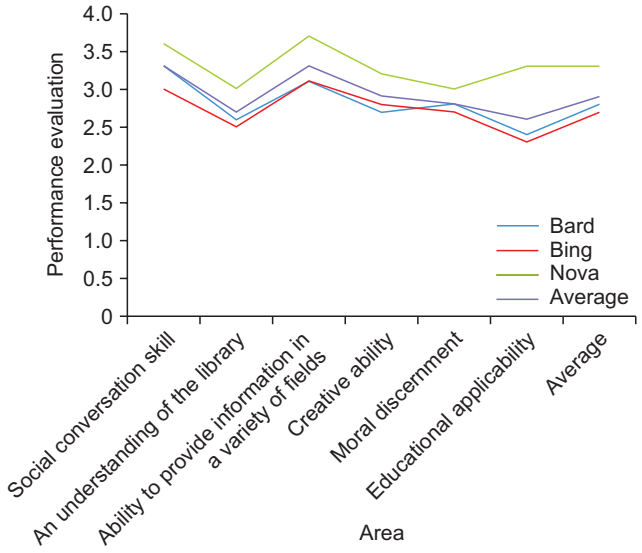


Fig. 1. Results of chatbot performance evaluation by area.

educational levels required or expected by each school.

Fig. 1 is a graph of the results of Table 12.

6. DISCUSSION

When I asked Google's Bard, "Can chatbots be used in school library education?", I received an answer that chatbots could be a useful tool to increase the efficiency of school library education and encourage student participation. However, the chatbot system must have the functions and performance to provide students with the information they need accurately and quickly. The educational content to be provided in the chatbot system must be developed to suit the level and learning style of the students, and must be designed so that students can use it easily. Another major consideration that must be designed and implemented in chatbot systems involves the safe protection of students' personal information collected through or by the system. Nevertheless, if the chatbot system can be introduced and used effectively, librarians can provide a better learning experience to students.

The purpose of this study is to investigate the possibility of using chatbots as a school library educational tool. Chatbots are becoming more sophisticated, so they will gradually be able to support many of the tasks that school library librarians are now in charge of. In particular, chatbots will demonstrate more creative and strategic skills with simple tasks such as library usage guidance, curriculum design for reading education and curriculum cooperation classes, and students' reading evaluation.

However, in order to use chatbots as educational tools in school libraries, the following tasks must be addressed or solved first. First, the ability to understand the library should be improved. Chatbots that can only be used professionally in the library field have not yet been developed. Therefore, since there is no chatbot with sufficient understanding of the school library, it is necessary to develop a chatbot specializing in the library in the future. Second, we need to increase moral discernment. The biggest problem that chatbots currently have is ethical issues, such as privacy issues, misinformation issues, and social impacts. It is important to develop and use AI chatbots more ethically in the context that responsibility should increase as technology develops.

7. CONCLUSION

The purpose of this study was to investigate the possibility of using chatbots as a school library education tool. Therefore, an online survey was conducted with 116 librarians working in elementary, middle, and high schools to investigate the types and contents of education implemented in school libraries and their perceptions of chatbots. In addition, a structured questionnaire containing missions for each area was delivered to 15 librarians working in elementary, middle, and high school libraries to perform missions through Google's Bard, Microsoft's Bing, and OpenAI's Nova. The results are as follows.

First, the most common types of education in school libraries were "user education," "reading education," "information utilization education," and "curricular cooperation classes."

Second, most librarian teachers knew about chatbots and responded that there was a need to use chatbots in school library education. Next, "information utilization education" was selected as the most effective form of school library education when using chatbots.

Third, librarians said that in order to use chatbots for school library education, "establishment of an environment" such as Internet environments and devices that can access chatbots, and "appropriate education" are needed.

Fourth, as a result of evaluating the performance in six areas, the chatbot that received the highest score was Nova. However, the overall average score was below average, and in particular, "educational usability" was the lowest for Nova. Accordingly, librarians say that chatbots are demonstrating above-average abilities in "social conversation skills" and "ability to provide information in various fields," but because they are still lacking in skills in other

 https://doi.org/10.1633/JISTaP.2024.12.3.1

areas, the possibility of them being used as an educational tool in school libraries is evaluated as low.

Chatbots are gradually becoming a tool that plays an important role in various fields. Therefore, if moral discernment, ability to understand libraries, creative ability, and possibility of educational use increase in the future are better refined, they can be used as an educational tool in school libraries. To this end, this study conducted a survey of 116 librarian teachers working in Seoul and Gyeonggi province, and the structured survey was limited to 15 people. In the future, the scope of study will be expanded to librarian teachers across Korea using actual chatbots. It is necessary to analyze use cases, analyze the strengths and weaknesses of the school library as an educational tool, and find ways to improve it utilizing this quickly developing technology.

CONFLICTS OF INTEREST

No potential conflict of interest relevant to this article was reported.

REFERENCES

American Association of School Librarians. (2019). *Role of the school library*. https://www.ala.org/aasl/sites/ala.org.aasl/files/content/advocacy/statements/docs/AASL_Role_of_the_School_Library.pdf

Dilmegani, C. (2024). *ChatGPT education use cases, benefits & challenges in 2024*. https://research.aimultiple.com/chatgpt-education/#top-7-chatgpt-education-use-cases

Hetler, A. (2024). *What is ChatGPT?* https://www.techtarget.com/whatis/definition/ChatGPT

International Federation of Library Associations and Institutions (IFLA), & United Nations Educational, Scientific and Cultural Organization (UNESCO). (1999). *IFLA/UNESCO school library manifesto 1999*. https://www.ifla.org/publications/ifla-unesco-school-library-manifesto-1999/

Jeong, Y. S., & Shin, D. H. (2023). The effects of learning using artificial intelligence chatbot on science academic achievement and positive experiences of science: Focusing on 'structure and function of our body' unit for the 6th grade. *The Journal of Elementary Education*, 38(2), 103-123. https://doi.org/10.23279/eer.38.2.202306.103

Kim, J. S., Kang, S. J., & Kwon, S. Y. (2023). A Study on the recognition of teacher librarians on the introduction of ChatGPT in school library. *Journal of the Korean Society for Library and Information Science*, 57(2), 349-377. https://doi.org/10.4275/KSLIS.2023.57.2.349

Korean Education Statistics Service. (2023). *Educational statistics: Teachers by position*. https://kess.kedi.re.kr/eng/stats/school?menuCd=0101&cd=6815&survSeq=2023&itemCode=01&menuId=m_010109&uppCd1=010109&uppCd2=010109&flag=B

Min, Y. T. (2021). *A comparative study on the usability by the platfrom of artificial intelligence chatbot service in library* (master's thesis). Chungnam National University, Daejeon, Korea.

Morgan, D. L., & Scannell, A. U. (1998). *Planning focus groups*. Sage.

Oh, J. R., & Lee, S. S. (2022). A study on the current status of metaverse utilization in university libraries and librarian perceptions. *Journal of the Korean BIBLIA Society for Library and Information Science*, 33(4), 159-180. https://doi.org/10.14699/KBIBLIA.2022.33.4.159

Rouse, M. (2024). *ChatGPT*. https://www.techopedia.com/definition/34933/chatgpt

Shin, D., Jung, H., & Lee, Y. (2023). Exploring the potential of using ChatGPT as a content-based English learning and teaching tool. *Journal of the Korea English Education Society*, 22(1), 171-192. https://doi.org/10.18649/jkees.2023.22.1.171

Song, G. H. (2018). *School library education and collaborative teaching*. Korean Library Association.

Song, G. H. (2019). An analysis of teacher librarians' role recognition. *Journal of the Korean Society for Library and Information Science*, 53(4), 5-22. https://doi.org/10.4275/KSLIS.2019.53.4.005

Woo, DK. (2023). *Librarian teacher's information utilization class ⑨ Chemistry department meets ChatGPT, blossoms through library collaboration classes*. https://www.edpl.co.kr/news/articleView.html?idxno=9817

Yorio, K. (2023). *School librarians explore possibilities of ChatGPT.* https://www.schoollibraryjournal.com/story/School-Librarians-Explore-Possibilities-of-ChatGPT

JISTaP http://www.jistap.org
Journal of Information Science Theory and Practice

RESEARCH PAPER
J Inf Sci Theory Pract 12(3): 14-35, 2024

eISSN : 2287-4577 pISSN : 2287-9099

https://doi.org/10.1633/JISTaP.2024.12.3.2

Breaking the Code of Silence: A Qualitative Exploration of Cyberbullying Through the Lens of Habermas's Theory of Communicative Action

January Febro Naga*
Department of Information Technology, College of Computer Studies, MSU-Iligan Institute of Technology, Iligan City, Philippines
College of Computer Studies, De La Salle University, Manila, Philippines
E-mail: january.febro@g.msuiit.edu.ph

Joshua Isaguirre
College of Engineering, Architecture and Technology, De La Salle University-Dasmariñas, Cavite, Philippines
College of Computer Studies, De La Salle University, Manila, Philippines
E-mail: joshua_isaguirre@dlsu.edu.ph

Elanie Vizconde
Technological University of the Philippines, Manila, Philippines
College of Computer Studies, De La Salle University, Manila, Philippines
E-mail: elanie_vizconde@dlsu.edu.ph

Raymund Sison
College of Computer Studies, De La Salle University, Manila, Philippines
E-mail: raymund.sison@dlsu.edu.ph

ABSTRACT

This qualitative study explores cyberbullying among college students through Habermas's *Theory of Communicative Action* to examine the dissonance between online interactions and principles of rational discourse. Cyberbullying is a pervasive issue in digital communication that undermines logical, evidence-based conversation, fostering environments where misinformation, manipulation, and harm thrive. By analyzing case studies from three universities, the research identifies the characteristics, dynamics, and emotional impacts of cyberbullying on victims, highlighting the role of social media platforms in facilitating these negative interactions. The findings reveal significant challenges to authentic and equal online conversations, driven by power imbalances and a lack of genuine communication, leading to psychological distress, erosion of self-esteem, and changes in behavior among victims. The study underscores the potential of social media design and policy interventions to mitigate cyberbullying, emphasizing the need for educational programs, technological solutions, and community support to promote a safer, more respectful digital environment. Key themes include the dynamics of cyberbullying, the suppression of rational discourse, the psychological and emotional consequences of inauthentic communication, and strategies for resilience and recovery. The research contributes to understanding cyberbullying's complexities and suggests a multifaceted approach to addressing it, aligning with Habermas's ideal of communicative rationality to foster healthier online communities. Future research should further explore the intersection of technology design, user behavior, and regulatory policies to combat cyberbullying effectively.

Keywords: cyberbullying, rational discourse, social media bullying, Theory of Communicative Action

Received: January 8, 2023 **Revised:** March 27, 2024
Accepted: April 21, 2024 **Published:** September 30, 2024

***Corresponding Author:** January Febro Naga
https://orcid.org/0000-0002-2977-6377
E-mail: january.febro@g.msuiit.edu.ph

All JISTaP content is Open Access, meaning it is accessible online to everyone, without fee and authors' permission. Open Access articles are automatically archived in the Korea Institute of Science and Technology Information (KISTI)'s Open Access repository (AccessON). All JISTaP content is published and distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/). Under this license, the authors retain full ownership of their work, while permitting anyone to use, distribute, and reproduce the content in any medium, as long as the original authors and source are cited. For any reuse, redistribution, or reproduction of a work, users must clarify the license terms under which the work was produced.

© 2024 [January Febro Naga, Joshua Isaguirre, Elanie Vizconde, Raymund Sison] This article is distributed under the Creative Commons Attribution License (CC BY), allowing unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. It is published by the Korea Institute of Science and Technology Information (KISTI).

1. INTRODUCTION

In today's world, where digital communication connects us all, a new challenge emerges: The tools that help us stay connected can also introduce new threats. Cyberbullying is a serious concern in this digital era, breaking free from the constraints of location and time. This issue has evolved from traditional bullying and is taking a new shape with the rise of digital technologies (Della Cioppa et al., 2015; Suler, 2004; Vismara et al., 2022).

The dual nature of social media as both a primary source of information and misinformation highlights its significant role in shaping public perceptions and interactions (Catedrilla et al., 2020; Ebardo et al., 2020; Ebardo & Suarez, 2023). With the rise of Internet use and the popularity of social media, people increasingly turn to platforms like X (formerly Twitter) and Facebook. However, this shift brings challenges, such as the risk of exposure to inaccurate information. Further, digital technologies have made it easier for people to engage in harmful behaviors online, with less restraint than in in-person interactions (Suler, 2004). Social media platforms are now battlegrounds for cyberbullying, affecting a significant portion of teenagers worldwide and raising concerns among parents and educators (Badri et al., 2017; Choi et al., 2021; Conde, 2022; Eweida et al., 2021; Farrington et al., 2023; Ortiz-Ospina, 2019; Patchin & Hinduja, 2006; Ramos Salazar, 2021; Saleem et al., 2022; Xu & Trzaskawka, 2021). Cyberbullying manifests in various forms, including threats, unauthorized sharing, impersonation, and spreading rumors (Erbiçer et al., 2023; Patchin & Hinduja, 2006; Willard, 2007), leading to negative impacts on young people's self-esteem, emotional well-being, and behavior, sometimes even resulting in suicidal thoughts (Chen et al., 2017; Erbiçer et al., 2023; Garett et al., 2016; Hinduja & Patchin, 2010). There is a noticeable gap in research on cyberbullying among college students, especially in the Philippines (Cocal, 2019; Costales et al., 2022).

This study investigates cyberbullying through the lens of Habermas (1984)'s *Theory of Communicative Action*. Our goal is to determine whether online interactions follow or stray from the principles of rational discourse. By 'rational discourse,' we mean discussions where logic, evidence, and reasoning guide conversations toward informed conclusions rather than being driven by emotions or biases. We aim to understand how online platforms either discourage or enable cyberbullying, focusing on the role of IT design and policy in fostering a safe and genuine digital communication space. In the next section, we explore Habermas (1984)'s *Theory of Communicative Action*, providing a theoretical basis for our study. Section 3 covers our research methods. Section 4 shares our findings. Section 5 concludes the paper, discussing the implications and limitations of our study and suggesting directions for future research.

2. HABERMAS'S THEORY OF COMMUNICATIVE ACTION

This section explores the impact of cyberbullying on social media among college students through the lens of Habermas (1984)'s *Theory of Communicative Action*. Habermas's theory, emphasizing the importance of rational, open discussion, is an excellent framework for analyzing how cyberbullying disrupts effective digital communication. At the core of Habermas's approach is the idea of 'communicative rationality,' a form of debate free from domination, where the merit of arguments is evaluated based on their quality and relevance (Cecez-Kecmanovic, 1995). According to this theory, communication, mainly through language, is a tool for individuals to achieve mutual understanding and coordinate actions that reflect collectively agreed-upon meanings. This approach is crucial for the health of a democratic society, fostering a culture where discourse is aimed at understanding rather than winning (Baxter, 2011).

Habermas distinguishes between two types of actions within communication: communicative action and strategic behavior. Communicative action is about engaging in dialogue with the intent of mutual understanding, where words convey meanings and achieve common goals. Strategic behavior, conversely, aims to elicit specific reactions from others, often disregarding the pursuit of mutual understanding (Habermas, 1984). The essence of Habermas's theory is this distinction, advocating for authentic communication aimed at understanding and respecting different viewpoints, meeting criteria of clarity, truthfulness, and shared expectations (Habermas, 1984; Ross & Chiasson, 2011). This study employs Habermas's concepts to investigate the extent to which cyberbullying undermines the envisioned ideal situation of communication.

As proposed by Habermas (1984, p. 255-256), the four principles of rational discourse have been effectively summarized by Klein and Huynh (2004).

1) Opportunity for Speech Acts: *Getting a chance to talk or communicate.* All participants should have the same opportunities to initiate and sustain a conversation,

engaging in dialogues and making statements and rebuttals.

2) Interpretation and Validation: *Understanding what is being said and evaluating its truth or integrity*. Participants should have the same opportunities to interpret, assert, recommend, explain, justify, question, or provide evidence supporting or opposing any claim. This ensures that every opinion or presupposition can eventually become the focus of discussion and scrutiny.

3) Honest Expression: *Being genuine and sincere when communicating.* All participants should be equally capable of expressing their attitudes, feelings, and intentions, safeguarding against self-deception, illusions, and insincerity within the community.

4) Regulative Speech Acts: *These are the guidelines for our spoken and nonverbal conduct during communication.* Participants should be equally able to provide and deny permissions, make and request promises, account for actions, and demand explanations, ensuring equal opportunity for discourse initiation and continuation.

Applying the four conditions of the rational discourse theory to cyberbullying, we can analyze the implications for addressing this issue. Fig. 1 below shows the relationship between how the rational discourse theory by Habermas can explain cyberbullying incidents.

1) Non-coercion: In a fair conversation, everyone should be free to speak their mind without feeling pressured or forced by others. Everyone's opinion is valued and respected equally.

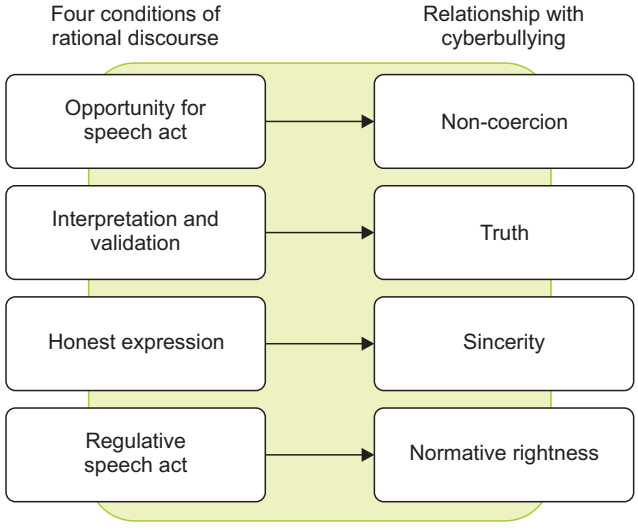


Fig. 1. Four conditions of rational discourse in cyberbullying.

2) Truth: In a reasonable discussion, what people say should be factual and evidence-supported. The interchange should be about being honest and not making things up. The notion of truth revolves around practicing honesty and refraining from fabricating information.

3) Sincerity: This involves communicating with sincerity and expressing oneself authentically without any intention to deceive or mislead others. In cyberbullying, sincerity pertains to the genuine expression of beliefs and intentions by the individuals involved.

4) Normative Rightness: This ensures that the arguments presented in a conversation conform to widely recognized principles and ethics.

3. METHODOLOGY

This section describes the qualitative interpretive methodology underpinning our multiple case study, as Yin (2009) recommended. The methodology enabled a thorough investigation of cyberbullying encounters among university students on social media platforms.

3.1. Research Design

Utilizing an explanatory multi-case approach, our study used Habermas (1984)'s *Theory of Communicative Action* principles and qualitative research to investigate the phenomena within its real-life setting (Yin, 2009). This design was selected based on examining intricate social phenomena and comprehending the subtle experiences of individuals in their natural settings (Denzin & Lincoln, 2011).

3.2. Case Selection

Following Yin (2009)'s recommendations, we conducted a detailed examination of eight cases conveniently selected from three universities (for ethical reasons, we named them Universities X, Y, and Z). The selection criteria were designed to ensure a rich diversity of experiences, aiming for depth over breadth (Creswell, 2007; Stake, 2006). According to Creswell and Creswell (2018), no specific minimum number of participants is needed for qualitative interviews. This is because qualitative research prioritizes the depth and richness of data rather than quantifiable validity indicators. For case study design, the Creswell and Creswell (2018) suggest that even a single case can be profoundly informative if it is well-chosen, critical, unique, or revealing. That will best help to understand, explore, and explain the phenomenon of interest.

Participants were selected through a purposive sam-

 https://doi.org/10.1633/JISTaP.2024.12.3.2

pling method, targeting individuals from universities X, Y, and Z (which have been anonymized for confidentiality purposes) who have experienced cyberbullying. This selection aimed to capture diverse cyberbullying experiences by including participants from varied university settings. Our methodological approach was designed to yield in-depth insights into the multifaceted nature of cyberbullying across different educational environments. To ensure the inclusion of individuals who genuinely encountered cyberbullying, we relied on self-reported experiences confirmed through initial screening processes:

- Individuals who reported experiencing cyberbullying events at least once a month over the past six months, ensuring recent and potentially ongoing exposure to cyberbullying.
- Individuals encountering various forms of cyberbullying, such as online harassment, cyberstalking, or denigration.
- Participants who expressed a readiness to discuss their experiences in detail, whether through in-depth or in-person interviews. This willingness was crucial for gathering rich, qualitative data on the personal impacts of cyberbullying.

Initial identification of potential participants was conducted via online surveys distributed through university email lists. The surveys also served as a preliminary screening tool to assess the frequency and types of cyberbullying experiences among respondents. Following this, students who met our criteria based on their survey responses were contacted for further evaluation through an initial brief interview via email or phone, where we clarified their experiences and ensured their willingness to participate in more extended, semi-structured interviews.

After agreeing to participate, participants received invitation and consent forms, which detailed the study's purpose, procedures, and ethical considerations. These forms also emphasized their rights as participants, including the right to withdraw from the study at any point without repercussion and the option to request the removal of any of their statements from the research findings.

3.3. Data Collection

This study collected data using interviews, surveys, and documentation (Fig. 2). The study's validity is strengthened through data triangulation (Yin, 2017).

Semi-structured interviews allowed participants to express their cyberbullying experiences in depth. The data collecting method of choice was semi-structured online interviews conducted through platforms via Google Meet or Zoom for convenience. Follow-up inquiries were conducted via chat or email. The researcher prioritized the anonymity and privacy of the participants by employing pseudonyms to safeguard their identities. The interviews were recorded solely for data collection and were securely saved with password protection. The interview questions encompass a range of subjects about cyberbullying encounters, emotional effects, coping mechanisms, societal norms, and the influence of peers. Furthermore, using a semi-structured framework facilitated the process of probing and clarification, leading to a deeper understanding of the phenomena being investigated. Throughout the interview, participants were allowed to express any additional knowledge or concerns not covered by the questions. To enhance the reliability of the data, the researcher implemented member checking, a method wherein the participants were given access to their interview transcripts to validate the precision of the recorded data and the conclusions drawn from the study (Creswell, 2007). This methodology improves the veracity and reliability of the study.

Document analysis involves systematically collecting and examining diverse textual data derived from social media platforms such as Facebook. This includes scrutinizing social media posts, messages exchanged through messenger services, and user comments. The study of



Fig. 2. Methods of data collection.

documentation entailed scrutinizing the language and context of cyberbullying interactions on Facebook, which was found to be the most prominent venue for such activities within our sample.

Qualitative research involves using surveys to identify and examine the variety within a particular population. However, the primary aim of this study is not to ascertain frequencies, means, and other statistical measures, but rather to evaluate the level of diversity related to a particular subject of interest within a given population (Jansen, 2010). Google Forms was used to conduct the survey. This method was chosen for its convenience and easy accessibility through email, making it suitable for efficiently reaching many respondents. Moreover, this approach required minimal time and cost for both the researchers and participants. The estimated completion time for the survey was less than 10 minutes. The survey questionnaire was distributed conveniently among students from three participating universities involved in the study.

3.4. Database Management

A case study database was established to organize and manage the data systematically, in alignment with Yin (2017)'s guidelines. This facilitated an efficient analysis process and ensured the traceability of conclusions back to the original data. For example, interview transcripts were called 'Transcript_Case1,' 'Transcript_Case2,' etc. This standardized naming approach made data administration easy.

3.5. Data Analysis of the Multiple Case Study

The analysis was conducted using a three-phase methodology, following Yin (2017)'s methodological structure. At first, a within-case analysis was conducted to emphasize the distinctive characteristics of each instance, focusing on their individuality. The researchers categorized the initial classifications into the main themes using categorical aggregation. Stake (2006) defines categorical aggregation as combining category data by addition. In this section, the themes identified through the analysis of the case studies are given. The relevance of the findings can be inferred from the themes identified. Subsequently, a cross-case analysis was conducted to identify repeating patterns and variations across the different cases, providing a comparative viewpoint. Then, the process of assertion analysis was employed to incorporate the data into the existing theoretical framework, making it easier to evaluate the main themes and derive implications from the study. The analysis of statements or the interpretation of cases constitutes the concluding phase of analysis in a multi-case study (Creswell, 2007; Merriam, 2009). Table 1 shows sample coding, and Table 2 the derived theme.

3.6. Ensuring Research Quality

We addressed construct, external, and internal validity through methodological rigor, participant diversity, and member checking, per Yin (2017)'s criteria for multi-case study research.

3.7. Ethical Considerations

The research was conducted strictly in compliance with the ethical standards outlined in the Research Service Agreement (F-URD-4.1-RSA, dated 11/19/2019). Before their participation, all individuals were thoroughly briefed on the study's objectives, assured of the confidentiality of their responses, and informed of their right to withdraw from the study at any time without any repercussions. Written informed consent was obtained from every participant, affirming their voluntary engagement and comprehensive understanding of the study's procedures. To

Table 1. Sample coding

Sample incident	Code	Concept
"Your personality belongs to the trash." (1.13)	Insult, degrade	Speech acts in cyberbullying
"They kept posting '#Aling Bebang problems.'" (1.2)	Public humiliation, indirect insults	Speech acts in cyberbullying
"The third time was when my friends made a group chat in Messenger without me." (1.14)	Secretive backstabbing	Role of social media, power dynamics
"They made it as a medium for backstabbing." (1.15)	Social exclusion	Role of social media, power dynamics
"Peer influence did play a role as my cyberbullies took advantage of their spheres of influence in order to conform my other classmates to say bad things about me as well." (1.89)	Peer pressure and groupthink	Role of social media, power dynamics

 https://doi.org/10.1633/JISTaP.2024.12.3.2

Table 2. Emergence of themes

Theme #1: Cyberbullying Characteristics and Dynamics
Rational Discourse in Cyberbullying: Truth
The notion of truth is closely connected to the attributes and mechanisms of cyberbullying. The improper use of language in direct and indirect forms of communication, with unequal balances of power and the amplifying effect of social media platforms, collectively inhibits the quest for truth in online interactions. This distortion hinders meaningful conversation and stresses the need to tackle these issues in order to establish a more authentic and empathetic online environment.
Subthemes: • Misuse of language in cyberbullying - Refers to the intentional use of language to cause injury or embarrassment to the target. Cyberbullies subvert truth and reasonable conversation by intentionally misrepresenting facts through the improper use of language, disseminating false information, rumors, and negative content, thereby impeding sincere communication and mutual comprehension. • Direct and indirect speech in cyberbullying - Direct speech involves using obvious offensive language, while indirect speech relies on subtlety and implication to convey hurtful messages. This can lead to misinterpretations and undermine clarity and honesty in online interactions. • Power imbalances in online interactions - Often lead to the distortion of truth in cyberbullying situations. Bullies take advantage of their dominance to create false stories, intimidate victims into silence, and make it difficult for the targeted individuals to assert their own truth due to fear of reprisal. As a result, the authenticity of online communication is compromised. • Social media platforms as avenues for bullying - Social media platforms amplify truth distortion in cyberbullying by enabling the rapid dissemination of false information, due to their speed and wide reach. Additionally, the anonymity provided by these platforms allows perpetrators to fabricate fictitious identities, thus complicating verifying the truth.
Theme #2: Impacts on Communication and Interaction - The principle of non-coercion in rational discourse aligns with the theme's subthemes by highlighting the negative impacts of power imbalances, manipulative tactics, and misunderstandings on communication and interaction in an online environment. Coercion disrupts the principles of open dialogue and mutual understanding, undermining the potential for equitable and rational discourse.
Rational Discourse in Cyberbullying: Non-Coercion
Non-coercion, a key element of rational discourse, is closely linked to the sub-theme of Impacts on Communication and Interaction, shedding light on how power dynamics, misunderstandings, and violations can hinder the genuine exchange of ideas in online interactions. Coercion disrupts the principles of open dialogue and mutual understanding, undermining the potential for equitable and rational discourse.
Subthemes: • Suppression of authentic communication - Coercion can be observed when genuine communication is suppressed, as people may feel obligated to conform to prevailing narratives or refrain from expressing their viewpoints to risk negative consequences. This suppression inhibits open dialogue, hampers the ability to have open and honest discussions, and weakens the foundations of logical and reasoned conversations. Unequal opportunities to express oneself can suppress authentic communication, resulting in a flawed conversation where certain voices are muted or cut down. • Influence of power dynamics in online interaction - The presence of power dynamics can hinder fair and balanced discussions by pressuring individuals to conform or submit. Individuals possessing greater influence may seek to manipulate discussions to promote their interests, suppressing dissenting perspectives and hindering logical discourse. When power dynamics or manipulation are present, rational discourse can be compromised if one person exercises excessive control over the discussion. The use of coercive tactics might result in the dissemination of biased information, hindering the achievement of a collective comprehension grounded in reason. • Role of misunderstandings and escalations in violations - Coercion can result from misconceptions and misinterpretations, causing escalations that hinder effective discourse. The misuse of power to aggravate misunderstandings prevents reasonable discussion from taking place and maintains a cycle of conflict. • Barriers to fair, constructive, and open communication - Coercive actions can disturb the balance of fair online engagement, granting certain individuals an unjust benefit while excluding others. This hinders the cooperative sharing of ideas and diminishes the genuineness of engagement.
Theme #3: Consequences of Inauthentic Communication and Power Imbalances in Online Interaction
Rational Online Interaction in Cyberbullying: Sincerity
The principle of sincerity in rational discourse is closely linked to the theme of Consequences of Inauthentic Communication and Power Imbalance in Online Interaction, as it sheds light on the emotional, psychological, and relational impacts resulting from insincere communication and power imbalances in online interactions.

Table 2. Continued

Subthemes:
- Emotional and psychological impact due to suppression of authentic communication - Inauthentic communication stemming from insincerity or manipulation can lead to emotional distress and psychological turmoil for individuals. The absence of genuine expression stifles authentic interaction, causing feelings of frustration, isolation, and anxiety.
- Erosion of self-esteem and self-perception due to disruptive power dynamics - When power disparities are exploited in digital discourse, individuals subjected to insincere or manipulative behavior often experience a deterioration of self-esteem and self-perception. The imbalance of power undermines their sense of agency and self-worth, leading to negative self-concept and diminished confidence.
- Behavioral changes as a reaction to inauthentic online interaction - Insincere communication and power imbalances can prompt individuals to alter their online behavior. They may become guarded, withhold personal opinions, or withdraw from online interactions entirely, impacting their ability to authentically engage in digital discourse.
- Impacts on personal relationships due to violations of rational online interaction - Inauthentic communication and power imbalances can strain personal relationships, both online and offline. Trust is eroded when sincerity is compromised, and manipulation disrupts the basis of mutual understanding, leading to conflicts and fractured connections.
- Responses to unequal opportunities in online interaction - Victims adjust their communication to deal with imbalances in visibility, voice, or influence they face online. These imbalances prompt victims to either withhold their genuine expressions due to a lack of equitable participation or to overcompensate in attempts to be heard.

Theme #4: Navigating Authenticity and Resilience in Adverse Online Settings

Rational Discourse in Cyberbullying: Normative Rightness

The principle of normative rightness in rational discourse aligns with the theme Navigating Digital Discourse: Embracing Authenticity and Resilience in the Face of Cyberbullying by emphasizing the importance of authentic expression, coping strategies, and recovery journeys as individuals strive to uphold ethical standards in the midst of online harassment. Upholding ethical standards empowers individuals to navigate digital discourse authentically and resiliently, ultimately contributing to a healthier online environment.

Subthemes:
- Strategies for expressing oneself in adverse settings - Adhering to normative rightness encourages individuals to engage authentically despite cyberbullying. By remaining true to their values and ethical standards, individuals can navigate hostile digital environments with integrity, upholding the importance of ethical communication even in the face of negativity.
- Coping and confrontation strategies in cyberbullying - Normative rightness informs individuals' choices in confronting cyberbullying. Ethical standards guide them towards non-harmful, constructive approaches to address harassment. Choosing responses that align with ethical norms enables individuals to assert their rights without resorting to harmful tactics, maintaining the integrity of their discourse.
- Recovery journey - The process of recovering from cyberbullying is closely connected to adhering to moral correctness. Embracing ethical ideals promotes resilience and personal growth as individuals negotiate the aftermath of harassment. They strive to uphold moral correctness, which leads them to preserve their genuineness while regaining confidence and self-worth.
- Building resilience and personal growth - This aligns with normative rightness, as it involves recognizing the ethical and moral dimensions of cyberbullying responses. This category emphasizes the importance of fostering environments where individuals can recover and grow through ethical interactions that respect their dignity and rights. It advocates for the development of resilience and personal growth as responses to cyberbullying that are grounded in principles of fairness, respect, and justice, underscoring the need for online communities to uphold these values to support victims effectively.

Theme #5: Empowering Online Communities Through Support and Education

For additional research support and study, the theme Empowering Online Communities through Support and Education aligns with the principles of rational discourse, encompassing the notions of equal participation, technological integrity, and ethical communication.

Subthemes:
- Fostering community and social support - The subtheme emphasizes the importance of creating a communal space that upholds principles of equal participation and respect, mirroring the democratic ideals of rational discourse. By fostering inclusive online communities that encourage open dialogue, mutual understanding, and respectful engagement, social media platforms align with the principle of rational discourse that seeks to enable fair and equitable communication.
- Addressing challenges and leveraging technological interventions - The subtheme acknowledges the challenges posed by technology while also highlighting its potential to foster rational discourse. The use of technology to track offensive patterns, flag harmful content, and promote positive interactions aligns with the principle of rational discourse by seeking to ensure that online conversations are informed, equitable, and sincere.
- Enhanced awareness and education strategies - This subtheme underscores the imperative for educating users about truth, sincerity, and ethical communication, which are central to rational discourse. By promoting awareness of the consequences of inauthentic communication and the importance of ethical behavior, comprehensive education efforts align with the goal of rational discourse of fostering genuine understanding and mutual respect.

safeguard the participants' anonymity and the sensitive nature of the information collected, extensive measures were implemented to anonymize and securely store data, consistent with the guidelines and ethical standards of the university.

4. RESULTS

An overview is provided for each student affected by cyberbullying, with pseudonyms ensuring their anonymity, as shown in Table 3.

4.1. Case Study Overview

The study involves participants from various academic backgrounds, assigned pseudonyms for privacy. It includes four students from public University X enrolled in chemistry, elementary education – language, and information technology. Additionally, two participants studied computer engineering at the private University Y, while another two pursued computer science and computer engineering at the public University Z. This mix of public and private urban university participants enriches the study with varied experiences and perspectives.

4.1.1. Case Study 1: Bebang

Bebang is known for her energetic and straightforward nature. She consistently demonstrates enthusiasm when sharing her perspectives, viewpoints, and personal anecdotes on diverse social media platforms, including Facebook, Instagram, and X. She believes that her regular social media engagement has facilitated the development of more profound connections with her peers. Sadly, Bebang's open and honest nature has made her vulnerable to cyberbullying, which includes being publicly mocked through hashtags and being targeted in a secret group chat aimed at undermining her.

Table 3. Participants' profiles

Case No.	Pseudonym	Sexual orientation
1	Bebang	Female
2	Alexis	Female
3	Johnny	Gay
4	Trixy	Female
5	Brent	Male
6	Andrew	Male
7	Rasheed	Male
8	Marcus	Male

4.1.2. Case Study 2: Alexis

Alexis, who loves joining school contests, was taken aback when friends turned against her online. Using a secret photo, her friends made mean edits to make her look bad, uttering derogatory remarks on Facebook due to her academic success. This betrayal made her feel alone and misunderstood, thinking all her friendships were fake. Since then, she has become much more private, sharing less online and discreetly monitoring the activities of her former friends.

4.1.3. Case Study 3: Johnny

Johnny was first bullied online through name-calling. He felt betrayed by friends who secretly mocked him in a group chat, not knowing why. Struggling with depression and self-harm, he wore jackets on sunny days to hide his scars. His peers then ridiculed him on Facebook, claiming he sought attention. This breach of trust from a friend he confided in about his depression, who shared his struggles in the group to mock him.

4.1.4. Case Study 4: Trixy

Trixy was once on her way to becoming a popular influencer but had to stop because of cyberbullying. She faced not just name-calling and harsh criticism but also sexual harassment, shockingly from someone she knew. This harassment spread across several platforms, including TikTok, Instagram, X, Facebook, and Messenger, impacting her profoundly and forcing her to reconsider her online presence.

4.1.5. Case Study 5: Brent

Brent is highly esteemed by both his peers and academics due to his exceptional academic achievements and active involvement in extracurricular activities, which serve as evidence of his abilities and competitive nature. A disagreement related to a thesis project resulted in the proliferation of adverse sentiments directed toward him in this online space.

4.1.6. Case Study 6: Andrew

Andrew can be defined as tall, dark, and handsome. From how he dresses, conducts himself towards others, and his manners, one can understand that he is humble and good. However, it can be noticed that Andrew has many insecurities and exhibits traits of anxiety. His self-esteem was destroyed by insulting language and harsh

criticisms from individuals he relied on in social media, which created a toxic environment. The cyberbullying he experiences, especially from those he values, dramatically affects his outlook on life.

4.1.7. Case Study 7: Rasheed

Although Rasheed values his privacy on social media, his academic accomplishments have unfortunately made him a subject of cyberbullying across multiple platforms. The bullying originated from unfounded allegations made by his colleagues, who suspected that Rasheed leveraged his connections with his homosexual professor to obtain high academic scores. Motivated maybe by envy, certain peers subjected him to ridicule based on an inferred sexual orientation, both face-to-face and via malicious Facebook postings and comments.

4.1.8. Case Study 8: Marcus

Marcus is a passionate gamer. He joined a Facebook group to establish connections with like-minded gaming enthusiasts. However, he encountered cyberbullying when his viewpoints conflicted with those of other members of the community. As a result of this incident, he has been more cautious and conservative regarding revealing personal information online and is now more skeptical of trusting strangers on social media platforms.

Analyzing these eight cases within the framework of Habermas's principles reveals how cyberbullying infringes upon the ideals of communicative action, specifically regarding Equal Opportunity for Speech Acts, Truth, Sincerity, and Normative Rightness (Table 4 reflects the impact of cyberbullying on adherence to Habermas's principles across different cases).

Principle 1. Equal Opportunity for Speech Acts (Universal Access): Every participant in a discourse must have an equal opportunity to initiate and continue communication. This includes the right to question, introduce assertions, express attitudes, desires, and needs, and explain or justify these expressions. This principle is essential for en-

Table 4. Analysis of the case within the framework of Habermas's principles

Case	Equal opportunity for speech acts	Truth (validity claims)	Sincerity (truthfulness)	Normative rightness (legitimacy)
1: Bebang	Suppressed by public mockery and secret groups, limiting free expression	Distortion through mockery challenges truthfulness	Anonymity used for insincere attacks	Mockery undermines ethical standards of respect
2: Alexis	Became more private, reducing participation due to cyberbullying	False narratives created through derogatory remarks	Betrayal reflects a lack of sincerity	Violation of dignity through betrayal and manipulation
3: Johnny	Silenced by ridicule, hindering willingness to engage	Misuse of personal information distorts truth	Breach of trust indicates insincerity	Mocking mental health struggles violates ethical interaction norms
4: Trixy	Overwhelmed by harassment, forced to reconsider online presence	Unfounded claims and insults distort truth	Sexual harassment from a known person signifies deep insincerity	Harassment severely breaches norms of respectful communication
5: Brent	Adverse sentiments limit ability to communicate freely	Rumors likely contain distortions challenging reputation	Spreading of rumors reflects departure from sincerity	Targeting undermines the legitimacy of academic discourse
6: Andrew	Insulting language and criticism restrict participation and self-expression	Personal insecurities exploited, likely involving false narratives	Deceitful bullying exploits vulnerabilities, lacking sincerity	Toxic environment created by bullying violates ethical communication
7: Rasheed	Restricted ability to defend integrity, limiting open engagement	Baseless allegations challenge the truth of academic merit	Personal attacks undermine sincerity of community interaction	Bullying based on unfounded allegations breaches ethical norms
8: Marcus	Skepticism and withdrawal limit engagement in gaming community	Conflict distorts viewpoints, affecting truthfulness	Personal attacks escalate disagreements, lacking sincerity	Community's failure to maintain respect questions normative rightness

 https://doi.org/10.1633/JISTaP.2024.12.3.2

suring that all voices can be heard and considered equally, which is fundamental to achieving genuine understanding and consensus.

Principle 2. Truth (Validity Claims): Discrete statements are expected to be truthful, and participants must agree on the truth of the propositions discussed. This implies that participants have the shared goal of reaching an understanding based on facts and shared experiences, allowing for a common ground of reality to be established and validated through discourse.

Principle 3. Sincerity (Truthfulness): Participants must mean what they say, with expressions reflecting their actual intentions and beliefs. This honesty is crucial for building trust within the discourse, as it ensures that participants are not attempting to deceive each other and are genuinely striving for mutual understanding.

Principle 4. Normative Rightness (Legitimacy): The contributions of each participant must be appropriate and conform to the recognized norms of the discourse community. This principle ensures that interactions are governed by a shared understanding of what is considered right and acceptable within the context of the discourse, supporting an environment of respect and ethical communication.

In the following:

4.1.8.1. Case Study 1: Bebang.

Bebang's experience shows a clear violation as cyberbullying restricted her from freely expressing herself on social media, a platform meant for open communication. The public mockery and secret chats likely propagated falsehoods or exaggerations about her, distorting the truth of her character and actions. The anonymity and secrecy of the cyberbullying acts call into question the sincerity of the communication, allowing perpetrators to hide behind digital facades. The ethical standards of respect and dignity were compromised, undermining the legitimacy of discourse within Bebang's digital community.

4.1.8.2. Case Study 2: Alexis.

Alexis's withdrawal from sharing online indicates her lost opportunity for equal participation in digital discourse due to cyberbullying. The manipulated images and derogatory remarks likely misrepresented Alexis, challenging the validity of the claims made against her. The betrayal by friends highlights a lack of sincerity, as their actions did not reflect true intentions or feelings towards her. The actions against Alexis violated the ethical norms of honesty and integrity, calling into question the legitimacy of the discourse environment.

4.1.8.3. Case Study 3: Johnny.

Johnny felt silenced by the bullying, particularly as it exploited his vulnerability regarding depression, limiting his willingness to engage. The dissemination of his personal struggles as a form of ridicule twisted the truth about his condition for malicious purposes. The breach of trust from a friend reveals a profound insincerity in the communication within his social circle. Mocking someone's mental health struggles grossly violates ethical standards of communication, lacking any form of legitimacy.

4.1.8.4. Case Study 4: Trixy.

The widespread cyberbullying across multiple platforms overwhelmed Trixy, forcing her to reconsider her online presence and silencing her voice. The name-calling and harassment likely included unfounded claims and insults, distorting the truth about her identity. The use of sexual harassment, especially from someone she knew, highlights a severe lack of sincerity and integrity in interactions. The malicious intent and spread of harassment clearly breach the normative rightness of ethical online communication.

4.1.8.5. Case Study 5: Brent.

The online adversity Brent faced hindered his ability to communicate freely, affecting his social interactions within academic circles. Adverse sentiments and rumors about the disagreement likely contained distortions or falsehoods challenging Brent's reputation. The spreading of rumors and negative sentiments reflects a departure from truthful and sincere communication. The targeting of Brent, especially related to academic competition, undermines the legitimacy of ethical discourse among peers.

4.1.8.6. Case Study 6: Andrew.

Insulting language and criticisms restricted Andrew's participation in social media, impacting his self-expression and communication. The cyberbullying Andrew experienced, based on personal insecurities, likely involved false narratives or exaggerated claims about his character. The deceitful nature of cyberbullying, exploiting personal vulnerabilities, underscores a lack of sincerity among online interactions. The toxic environment created by cyberbullying violates the ethical correctness expected in communicative actions.

4.1.8.7. Case Study 7: Rasheed.

Cyberbullying, based on unfounded allegations, restricted Rasheed's ability to defend his academic integrity and engage in honest discourse. The accusations made against him distorted the truth, leveraging baseless claims for bullying purposes. The use of personal attacks for cyberbullying undermines the sincerity of communication, as actions were driven by envy rather than truth. The ridicule based on perceived personal relationships and orientation breaches the normative standards of respect and dignity in discourse.

4.1.8.8. Case Study 8: Marcus.

Encountering cyberbullying within a gaming community limited Marcus's ability to express his viewpoints and engage with like-minded individuals. Conflict arising from differing viewpoints likely led to distortions of Marcus's opinions, affecting the truthfulness of discourse within the community. During the bullying, Marcus faced questions about the sincerity of community interactions as disagreements escalated into personal attacks. The community's failure to maintain a respectful discourse environment demonstrates a disregard for normative rightness in communication.

Habermas (1984)'s *Theory of Communicative Action* emphasizes the cooperative processes of individuals engaging in rational discourse. The key here is that specific criteria must be met for the communication to be truly rational. Through an analysis of the experiences of victims of cyberbullying, we gain an understanding of the many challenges they encountered, especially in relation to the principles of rational discourse. Each case reveals the impact of cyberbullying on violating Habermas's principles, highlighting the disruption of an ideal discourse environment where individuals can communicate freely, truthfully, and with mutual respect.

1. Equal Opportunity for Speech Acts: Cyberbullying significantly hinders victims from participating in online discourse, effectively silencing them or severely limiting their engagement due to fear, shame, or social exclusion. This breach fundamentally undermines the democratic ethos of communicative action, which states that all should have an equal voice.

2. Truth: The spread of distorted truths and misinformation through cyberbullying challenges online discourse integrity. False narratives and exaggerated claims about victims distort the reality of their identities and actions, leading to misunderstandings and the spread of false information within digital communities.

3. Sincerity: The insincerity prevalent in cyberbullying—where bullies often hide behind anonymity to deceive, manipulate, or betray—erodes the foundation of trust necessary for meaningful discourse. The lack of genuine intention behind communications further alienates victims and disrupts the potential for authentic engagement.

4. Normative Rightness: Each case departs from ethical standards governing respectful interaction. By targeting individuals based on vulnerabilities, spreading unfounded allegations, or engaging in harassment, bullies flout the moral norms that should guide legitimate discourse. Such actions not only harm the individuals targeted but also corrupt the overall ethical climate of online spaces.

4.2. Key Themes Emerging from the Research

This section provides an overview of the main themes that have emerged from the data, shedding light on the intricate aspects of the research topic. Each theme is explored in detail, providing insights into the identified patterns and their ramifications.

4.2.1. Theme 1: Cyberbullying Characteristics and Dynamics

This theme represents the various features and patterns common to cyberbullying incidents. These characteristics are not limited to the individual actions or words of the bully, but extend to the broader social and technological context that facilitates this type of bullying.

Direct and indirect speech acts in cyberbullying refer to the use of both explicit (direct) and implicit (indirect) ways of conveying insults, threats, or derogatory remarks (Yule, 1996). Direct speech acts in cyberbullying might include explicit threats or insults, while indirect speech acts may involve implied insults, sarcasm, or passive-aggressive remarks. The bullies of Bebang use both direct and indirect methods to insult and undermine her. The direct method is evident in explicit insults, such as "*Your personality belongs to the trash,*" or "*She is always asking for something but never gives back. Lord, take her away.*" Indirect methods can be seen where the person being bullied is not directly named but is identifiable to those within the social group, such as in the case of Bebang with the "*#Aling Bebang Problems*" posts and the group chat created to backstab the victim.

Power imbalances in online interactions refer to the disparity in influence, control, or power between the bully and the target in online interactions (Chan et al., 2021).

Power imbalance could be related to social status, popularity, knowledge of the online platform, or the ability to manipulate online narratives.

"The third time was when my friends made a group chat in Messenger without me. They made it as a medium for backstabbing. Peer influence did play a role as my cyberbullies took advantage of their spheres of influence in order to cause my other classmates to say bad things about me as well." – Case 1

The bullies have the power in this narrative by dominating the narrative within the group chat, enabling them to subject her to ridicule without her awareness or means to defend herself. This secretive backstabbing, a type of interpersonal bullying, exploits the private nature of group chats. This demonstrates the ability of bullies to influence connections and reputations in private settings. Bullies employ strategic manipulation of their influence and social dynamics to intentionally isolate and specifically target the victim.

Social media platforms as avenues for bullying recognize that the platforms used for communication play a significant role in facilitating cyberbullying. This involves understanding how social media sites and chat groups can become venues for cyberbullying, and how their design or rules may either contribute to or mitigate cyberbullying.

"To sum everything up I was slandered for hooking up with my gay professor and that he does 'service' to me that's why my grades are high and everything… They did on social media… A significant part of the bullying happened there by posting memes insinuating my identity and the things I 'did' there happened." – Case 7

Social media platforms are powerful mediums for bullying, as exemplified by the encounter described in the case of Andrew. This led to the individual being publicly humiliated and subjected to cyberbullying. These platforms facilitate the rapid dissemination of harmful content by bullies, resulting in widespread damage to reputations that extend beyond the limitations of conventional bullying. Moreover, anonymity or pseudo-anonymity on these platforms can embolden bullies. This makes it easier to target individuals without immediate consequences.

4.2.2. Theme 2: Challenges to Authentic and Equal Online Conversations

This theme explores how power dynamics, misunderstandings, and the suppression of genuine communication hinder the possibility of achieving open and fair dialogue online. It examines the barriers that prevent individuals from engaging freely and equally in social media.

Suppression of authentic communication: Fear of cyberbullying leads to self-censorship on social media, with individuals hesitant to share honest opinions or personal stories.

"I go to the extent of creating Facebook and Gmail accounts without using my real name. Thinking it is the new me. Creating that so-called new identity in social media gives me peace of mind that nobody knows me, and so they can't judge me." – Case 2

Influence of power dynamics in online interaction can significantly influence the course of conversations. Accounts with a substantial following or backed by institutions can exert an overwhelming influence, overshadowing less popular viewpoints.

"I was cyberbullied by a single person in college due to a disagreement in thesis grouping. Forms of cyberbullying I have experienced are spreading rumors, shooting threat post and a form of 'Appeal to Pity' social media posting to make me look bad to our mutual friends and classmates. The experience reduced my circle of friends in college to about half, almost all of my friends from the other section (where he is close with) are scared to talk to me or even look at me because the bully specifically told them that if they are friends with me then they should leave the friends group or group-chat (I was sent a screenshot from a close friend of mine in the other section) where I knew that information." – Case 3

Role of misunderstandings lies in the potential for misinterpretations of online content to rapidly escalate into cyberbullying events, as conflicts deepen quickly owing to the viral nature of social media.

"The second time I was cyberbullied was when there was a misunderstanding between me and a friend. That misunderstanding became a focal point and he started with gc (group chat). He didn't explicitly mention my name however, within the classroom, it was already well-implied. He also proceeded to post about me in to insult me in Facebook." – Case 1

Barriers to fair, constructive, and open communication: Any element that hinders equitable and balanced communication prevents all parties from participating equally and openly. The dynamics within a group might impede equitable and inclusive discussions, with some voices being marginalized or silenced. Cyberbullying encompasses aggressive and deceptive strategies that disrupt constructive online dialogue, substituting it with damaging interactions.

"I hope there is a way to express my side as a victim as they cannot listen to my story because of social boundaries set by the bully between me and my friends… The bullying gave me some sort of anxiety because of the danger of shooting threat." – Case 5

4.2.3. Theme 3: Consequences of Inauthentic Communication and Power Imbalance on Online Interaction

Emotional and psychological impact due to suppression of authentic communication emphasizes the psychological consequences of cyberbullying on mental health. When someone's capacity to express themselves genuinely is suppressed, it can result in notable emotional and psychological anguish.

"The bullying gave me some sort of anxiety… It just feels like I am existing in a place where I shouldn't exist at all." – Case 6

Erosion of self-esteem and self-perception due to disruptive power dynamics highlights how cyberbullying can erode an individual's self-esteem and self-image. Victims, when confronted with power imbalances and aggressive interactions, may begin to absorb the negative messages they receive.

"It made me view myself as weird, and nobody liked me. That they just pretend to like me." – Case 2

Behavioral changes as a reaction to inauthentic online interaction primarily revolve around the consequences of cyberbullying on its victims, causing them to modify their conduct to prevent additional harm. These adjustments often involve strategies that limit authentic self-expression, therefore reducing meaningful social interactions.

"I developed a bipolar disorder and needed to take counseling and therapy regardless of all the coping strategies I developed through the years." – Case 5

Impacts on personal relationships due to violations of rational online interaction focus on the impact that cyberbullying can have on relationships between individuals. Trust issues frequently arise as a substantial result of the betrayal and deception seen in cyberbullying incidents (Woodluck et al., 2023).

Responses to unequal opportunities in online interaction: Victims or bystanders may alter their online interactions or silence their voices to avoid becoming targets. This reflects imbalances in who gets heard or attacked online.

"In order to avoid experiencing that again, I made personas of myself." – Case 1

"I just don't respond anymore when it becomes offending." – Case 8

This approach reflects a coping mechanism where the victims modify their online behavior—here, by adopting alternate identities—to prevent becoming a target again. It highlights how the fear of cyberbullying can lead individuals to conceal their true selves, seeking safety in anonymity or diversified online personas. This strategy is a direct consequence of the imbalances in digital spaces, where not everyone feels equally safe or heard, driving them to find alternative ways to express themselves while minimizing the risk of harassment.

4.2.4. Theme 4: Navigating Authenticity and Resilience in Adverse Online Settings

This theme explores the resilience of individuals facing cyberbullying. It discusses strategies for adapting one's identity for safety and comfort, varying responses to cyberbullying from avoidance to confrontation, and the journey toward healing from such experiences. Finally, it delves into the recovery journey to highlight the importance of support systems in overcoming the emotional scars left by cyberbullying, aiming for a return to emotional and psychological health.

Strategies for expressing oneself in adverse settings primarily revolve around how victims maintain, regulate, or hide their true selves to function or protect themselves within hostile or unfavorable circumstances. These strategies often involve constructing different identities or personas, adopting pleasing behaviors, or limiting one's presence or visibility within the space. Such strategies help

 https://doi.org/10.1633/JISTaP.2024.12.3.2

them navigate and negotiate the pressures and expectations within the environment, which could often suppress or discourage their authentic expression.

"I have forgiven the bullies even though they never asked for apologies. But I would always hold on to what I learned from it… I communicate with caution especially when the conversation goes personal level." – Case 2

Coping and confrontation strategies in cyberbullying delves into the various strategies and techniques used by victims to deal with instances of cyberbullying. These methods range from avoidance tactics to direct confrontation, to speaking up. Unfortunately, this can also include harmful coping mechanisms.

"Personally, it was hard to cope up. There were sleepless nights. And I lay low from social media. I made private different social media accounts for people I know very well and that I trust. As for my body shamers. They made me go on a diet. It was healthy for my body but the pressure to have a body that everyone would approve of is challenging mentally. My friends told me not to over do it because people will always have something to say. As for the sexual harassment. I ignored those people because they used dummy accounts. But after quite a while. I did speak up through tweets about the experience. It was already too much for me. I did not want to send them the wrong signal that my silence means an acquiescence. So, when I tweeted about it, other girls I didn't know shared their experience as well. They retweet and support me. Their sharing of experience helped me realize that the problem is not me but those people who abuses other people online…" – Case 4

The recovery journey pertains to the steps and actions taken by individuals to heal and regain their emotional and psychological well-being following a traumatic or adverse cyberbullying experience. This process often involves seeking and receiving support from others, whether they are close friends, family, or professionals.

".. I took a notebook to write what is my life's purpose, wrote a goal, and started living in a new direction. I am intentional in everything I do. I stopped crying the moment I defined my worth, not by my bullies' words." – Case 2

Building resilience and personal growth encompasses the process that cyberbullying victims undergo as they cultivate resilience, self-sufficiency, and personal development despite the difficulties and violations they encounter in social media. It emphasizes overcoming trauma, adopting healing techniques, developing coping mechanisms for future situations, and the unforeseen personal growth that can arise from such challenging experiences.

"I learned by myself. Like you learn how to defend yourself as personal growth… I channeled all my rage into writing. And through writing a story, I felt I had requited them all that pain. I become creative because of it." – Case 3

4.2.5. Theme 5: Empowering Online Communities Through Support and Education

This theme reflects that community support, appropriate technological interventions, and literacy are integral to upholding the principles of equal participation, sincerity, truth, and normative rightness in digital communication. These elements help create a space where people can engage in a reasonable, respectful, and genuine way, as suggested by Habermas's concept of 'communicative rationality.' It is about making sure conversations are open, fair, and honest.

Fostering community and social support not only provides immediate comfort but also sets the foundation for enduring resilience and recovery, emphasizing the necessity of both professional guidance and a strong personal support network in navigating the aftermath of cyberbullying. Creating a supportive online environment where individuals feel connected and supported by others is equally important and will offer members a sense of belonging and safety. Through shared experiences, advice, and emotional support, these communities can significantly mitigate the negative impacts of cyberbullying, promoting resilience and well-being among affected individuals.

"So personally I think that counseling is really the perfect measure in the long run." – Case 7

"I only turn to my closest best friends for support and to be heard but all other stuff needs to be faced by me… I've been talking to the guidance counselor." – Case 1

Addressing challenges and leveraging technological interventions involves the hope of developing and implementing technology-based solutions such as filtering software, reporting mechanisms, and educational apps to help

detect, prevent, and respond to online harassment. These interventions aim to empower users, providing them with resources to protect themselves and others from cyberbullying.

"The best way cyberbullying can be prevented is if they realize the cost of negative content and how it can affect its viewers… AI can track and flag someone for saying a pattern of offensive things [and] can help detect and be flagged sent by a warning." – Case 6

Advocacy for awareness and education programs emphasizes promoting knowledge and preventive measures against cyberbullying through targeted educational programs and campaigns. This can raise awareness among Internet users of all ages about the dangers of cyberbullying, the importance of digital etiquette, and how to safely navigate online spaces.

"Awareness and education has never faded to this day but when faced in actual circumstances and in the face of any extreme emotion, people forget. It is important though the effect may be little, there still is and I believe it should still keep on going." – Case 1

4.3. Cyberbullying and Social Media Platforms

The survey results from the three universities reveal a concerning spectrum of cyberbullying behaviors (Fig. 3). Degrading comments or hate speech was most common at 23.8%, followed by threats (17.2%), and unwanted sexual messages (14.5%). Other types included sharing unwanted pictures (16.3%), discrimination (8.4%), and impersonation (7.9%). Group harassment and 'outing' someone's private information were less common but still reported. This highlights the need for better online safety and respect across social media platforms.

Facebook and Messenger (73%) are the primary mediums used where cyberbullying occurs, while platform X accounts for 25% of cyberbullying cases (Fig. 4).

This study's exploration of eight cases reveals the dual nature of social media platforms such as Facebook and its Messenger feature (Fig. 5), which have become conduits for cyberbullying despite their intent to foster connections. These platforms have been misused for anonymous harassment, leading to lasting emotional distress. Facebook's design and group features exemplify how they can enable the dissemination of misinformation and malicious personal assaults, thus weakening the integrity of respectful and authentic communication. This situation necessitates a collaborative endeavor to advocate for digital empathy, accountable online conduct, and platform regulations that reduce cyberbullying to establish a safer and more inclusive digital atmosphere.

The analysis of the cases reveals how IT platform features and dynamics on social media can either facilitate or impede speech acts characteristic of cyberbullying. The characteristics of these IT platforms can significantly influence the prevalence and nature of cyberbullying incidents:

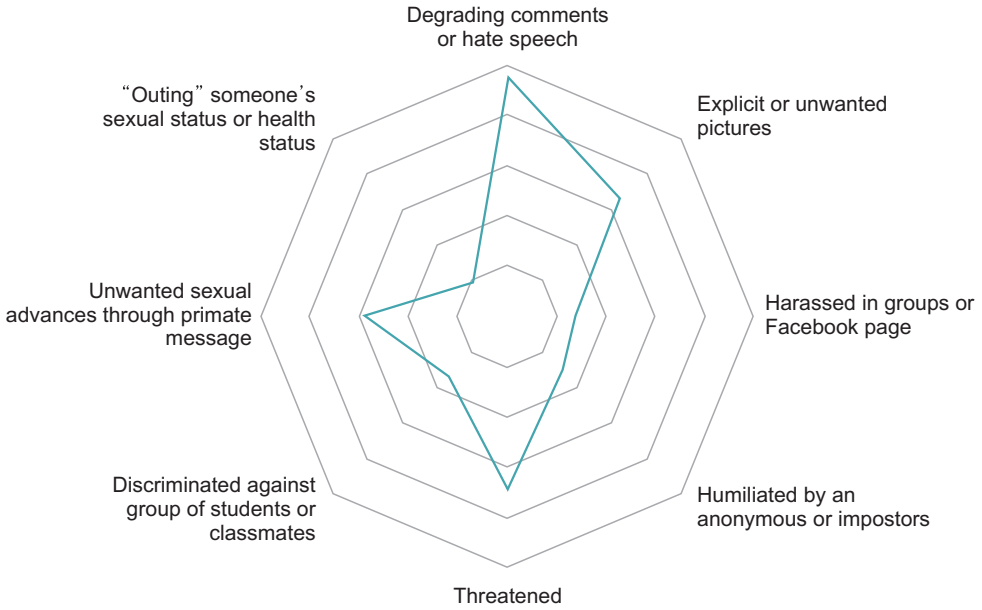


Fig. 3. Various forms of cyberbullying experienced by college students.

 https://doi.org/10.1633/JISTaP.2024.12.3.2

- Visibility and Virality: Due to social media's quick information distribution, offensive speech activities may spread more quickly. Bullying's effects can be amplified by information that spreads rapidly, as was the case in Andrew's case when false rumors spread to a large audience. Bullies may target more people since it is so simple to share content, which worsens the damage they do.
- Amplification of Power Imbalances: Platform structures and algorithms can magnify the influence of certain users over others. Features such as trending topics, like and share mechanisms, and visibility algorithms can disproportionately elevate the voices of those with large followings or those who manipulate content for broader dissemination. This dynamic can exacerbate cyberbullying, misinformation, and other forms of online harm, as it allows dominant voices to overshadow or silence less powerful individuals.
- Anonymity and Pseudonymity: Cyberbullying may be made more accessible by the option to post anonymously or use a pseudonym, since it gives offenders a sense of anonymity and allows them to use hurtful comments without worrying about being identified. This anonymity may give bullies the confidence to propagate untrue information, launch personal assaults, and use disparaging language. On the other hand, platforms that demand genuine names may prevent such speech actions by enhancing responsibility and lowering the possibility of bullying motivated by anonymity.
- Privacy Controls: Cyberbullying may be stopped by features that let users manage their privacy settings and control who can view their material. The usage of private settings demonstrates how privacy measures may prevent bullies from getting access to sensitive material. By limiting visibility, bullies may have fewer opportunities to target their victims.

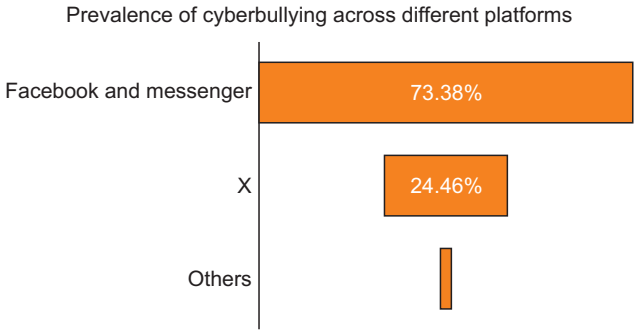


Fig. 4. Mediums of cyberbullying.

This study highlights how social media's features can influence cyberbullying, aligning with Habermas's conditions for rational discourse: truth, non-coercion, sincerity, and normative rightness. The quick spread of misinformation challenges the truth. Power imbalances and anonymity can create a coercive environment, limiting free speech. Misleading stories and tactics undermine sincerity. The absence of universal respectful norms hampers normative rightness. These findings emphasize the need for platforms to foster environments that discourage cyberbullying and promote respectful, truthful, and free communication, supporting a healthier digital space.

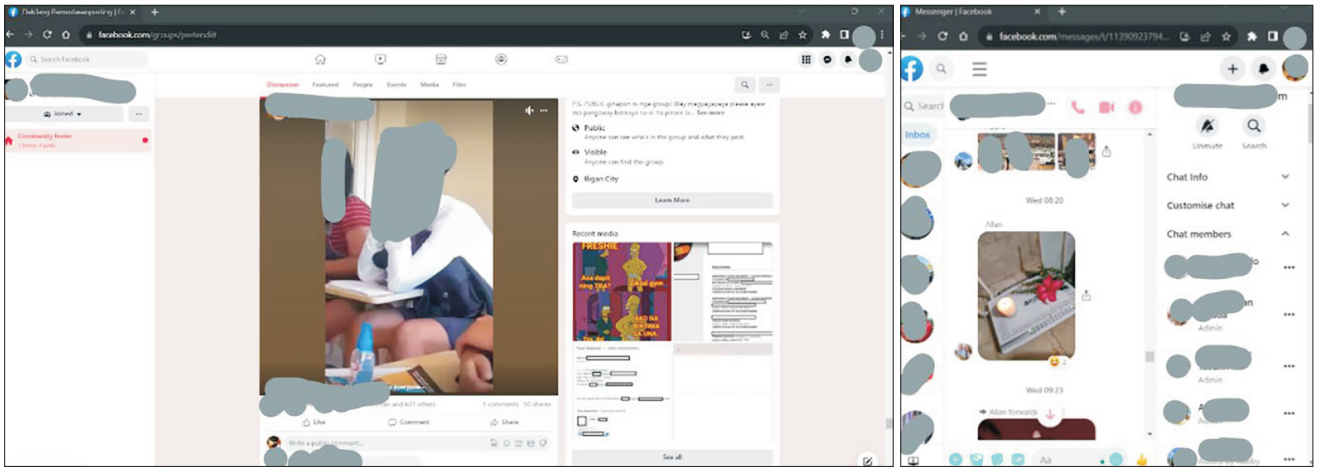


Fig. 5. Example image of Facebook page and group chat messenger.

5. DISCUSSION

Cyberbullying, an adverse product of the digital age, represents a complex and detrimental issue affecting individuals worldwide, particularly adolescents and young adults (Chisholm, 2014; Hendry et al., 2023; Peled, 2019; Stewart et al., 2014). The characteristics and dynamics of cyberbullying shape its manifestation and impact, making it a unique form of harm that extends beyond traditional bullying (Bauman et al., 2013). Critical characteristics of cyberbullying include anonymity, publicity, persistence, and accessibility, all of which are inherent in digital platforms (Wachs et al., 2019). These attributes contribute to the pervasiveness and intensity of online bullying, often exacerbating the damage inflicted on the victims (Stewart et al., 2014). Simultaneously, the dynamics of cyberbullying encompass power imbalances, role switching, the bystander effect, and rapid escalation, elements that can complicate its resolution and intensify the psychological harm to those involved (Saif & Purbasha, 2023; Wong-Lo & Bullock, 2014).

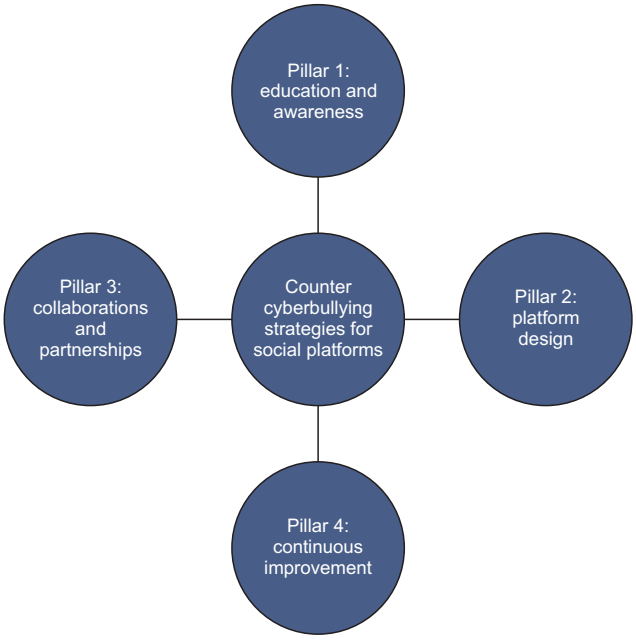


Fig. 6. Pillars to counter cyberbullying on social platforms.

5.1. Strategic Approaches to Mitigating Cyberbullying

To address the issue of cyberbullying on social media, implementing a strategy that understands today's online behavior and fundamental tenets of effective communication is necessary. This involves examining how people interact online today and combining that with fundamental communication principles. Drawing on Habermas's principles, a pillar is proposed (Fig. 6) to foster a digital environment where genuine, rational, and enriching interactions prevail, contributing to more respectful and enriching interactions.

- *Education and Awareness*: Central to combating cyberbullying is enhancing users' digital literacy, highlighting cyberbullying's consequences, and fostering a culture of respect online. This initiative calls for comprehensive education programs that employ case studies and discussions on ethical online behavior.
- *Platform Design*: Social media platforms must prioritize user safety through design innovations. This includes intuitive reporting mechanisms and advanced content moderation technologies that deter harmful behavior alongside promoting transparency and accountability in moderation practices.
- *Partnerships for Collective Action*: Collaboration with educational institutions, NGOs, and law enforcement can expand the reach and impact of anti-cyberbullying initiatives. These partnerships should facilitate open dialogue and constructive engagement in line with the principles of communicative action.
- *Commitment to Continuous Improvement*: An iterative approach to strategy enhancement, guided by user feedback and ongoing research, is crucial. This ensures adaptability in response to the evolving nature of cyberbullying and digital interaction patterns.

5.2. Theoretical and Practical Implications of Study Findings

The implications of our findings extend into both theoretical and practical domains. The study underscores the critical need for governance models on social media that foster meaningful participation and dialogue (McCarthy et al., 2023), emphasizing the role of ethical communication in platform policies. Furthermore, the success of interventions in various contexts highlights the potential of integrated, theory-based approaches to mitigate cyberbullying effectively (Barlett, 2023; Özgür, 2020).

To address cyberbullying and foster a culture of respect and understanding online, it is crucial to prioritize ethical involvement and mutual respect, which are key principles of discourse ethics. Adapting discourse ethics for digital

 https://doi.org/10.1633/JISTaP.2024.12.3.2

platforms emphasizes ethical conduct within online learning communities and provides a roadmap for creating digital environments where constructive and respectful discourse can flourish (Luppicini, 2009).

Our research uncovers insights into the dynamics of digital communication, primarily focusing on the challenges and potential pathways forward in addressing cyberbullying. It illustrates how online communication can differ from face-to-face interaction, with anonymity and the digital space presenting new challenges. It also

Table 5. Theoretical implications of the research study

Summary of findings	Rational discourse	Theoretical implication confirming the theory
Victims encountered limitations in their capacity to speak on social media as a result of unrestricted coercive actions. This generally indicates how the use of coercion in online environments may significantly limit open and sincere communication	Non-coercion	The results obtained from participants' accounts prove that coercion substantially impacts authentic communication. These observations highlight the significance of maintaining discourse without coercion to promote genuine online relationships. This emphasizes the need to establish digital platforms where people can interact openly and truthfully without the worry of being forced or manipulated, in keeping with theoretical viewpoints that stress the significance of non-coercive communication for meaningful online interactions
The victims encounter challenges related to truthfulness. False accusations, misleading information, and misunderstandings frequently distort reality and often distort the truth, resulting in unfair persecution and the dissemination of baseless narratives. These problems emphasize the widespread challenge of finding the truth online, where the ability to remain anonymous and spread false information can significantly harm individuals' reputations and relationships. This underlines the theory's emphasis on the necessity of truth for genuine communication and vigilance against falsehoods to ensure integrity in online interactions	Truth	The accounts provided by participants underscore the crucial significance of truth in facilitating effective communication, validating the theory that prioritizes truth as a fundamental element for genuine discussion. These concrete instances demonstrate the frequent presence of disinformation and deception in online discussions, providing practical observations on the pervasive problem of distorting truth in digital exchanges. This highlights the theory's focus on the importance of truth for authentic communication and the requirement to be cautious of falsehoods to maintain integrity in online interactions
The various testimonies from individuals experiencing cyberbullying expose the complex significance of sincerity in online interactions. Victims encounter difficulties discerning genuine motives because of the insincerity and anonymity provided by the Internet. Some employ passive techniques that emphasize the need for sincerity in resolving conflicts, while others confront the deceitfulness of their peers, which dramatically contrasts with their own genuineness. Bullies manipulate deceitfulness to benefit themselves, emphasizing the crucial role of genuineness in the cyberbullying dynamic and its consequences. These cases highlight the challenges and importance of maintaining authenticity and integrity in digital spaces	Sincerity	The varied encounters of participants with cyberbullying reinforce the theoretical significance of sincerity in effective communication. These experiences illuminate how the anonymity inherent in social media complicates the preservation of sincerity, affecting the integrity of online interactions. This situation underscores the need for strategies that enhance sincerity in digital platforms, ensuring that the quality of discourse is not compromised by the challenges posed by online anonymity
The stories of cyberbullying victims highlight the crucial role of fairness and respect online. Their fight for justice is often undermined by a lack of support from their online communities, pointing to a failure to maintain a sense of unity and solidarity. Experiences of exclusion and harassment question the fairness of online social dynamics and challenge the societal norms that govern our interactions, revealing a tension between personal authenticity and group expectations. These incidents expose an apparent disregard for equality and respect, essential for ethical and just interactions in digital spaces	Normative rightness	The experiences shared by participants underscore the critical importance of normative rightness in shaping online interactions, as posited by the theory. These real-world insights reveal how digital spaces can either support or undermine established social norms, stressing the importance of fostering an online environment that is both equitable and respectful. This calls for concerted efforts to reinforce normative principles in digital communication, ensuring that online exchanges are guided by fairness and respect for societal standards

highlights a growing online aggression culture, reflecting broader societal issues. The blending of online and offline lives, as seen in personal experiences, suggests a close connection between the digital and physical worlds. While our study has primarily focused on the disruptive impact of cyberbullying within digital communications, it is essential to acknowledge the existence of online communities and interactions where rational discourse thrives. These positive examples serve as guiding lights, showcasing the capacity of online platforms to foster environments conducive to informed, respectful, and productive conversations. Platforms that incorporate design features encouraging transparency, accountability, and mutual respect among users significantly contribute to the prevalence of such rational discourse.

The theoretical implications reaffirm Habermas's theory's applicability in analyzing the complexities of digital communication, as shown in Table 5.

Practically, there is a call for digital etiquette education and targeted support for cyberbullying victims. Policy implications suggest stricter cyberbullying laws, mandated platform regulation, and digital literacy in educational curricula. Our study emphasizes the importance of using diverse and well-informed strategies to address cyberbullying, which aligns with the broader discussion on digital ethics and platform governance. This entails dealing with the technological components of cyberbullying and actively involving oneself in the socio-cultural aspects of digital engagement.

5.3. Study Limitations and Future Directions

This study's applicability to all cyberbullying cases may be limited due to the unique contexts affecting individuals' experiences. While focused on cyberbullying among college students on social media through the Communicative Action Theory framework, future research should further examine strategies for creating a secure online environment and the roles of educational institutions and social media platforms in mitigating cyberbullying. Additionally, exploring the influence of bystanders and the broader online community could enhance our understanding of cyberbullying dynamics.

This groundwork paves the way for future inquiries into the nuances of cyberbullying within the digital landscape, emphasizing the importance of digital etiquette education, victim support, and legislative measures for a safer online environment. Harnessing the potential of online platforms for constructive discourse is essential for fostering inclusive and healthy digital communities.

6. CONCLUSION

The instances of cyberbullying observed on social media platforms clearly differ from Habermas's envisioned rules of rational discourse. Cyberbullying often involves coercion, where victims are pressured or forced into silence or compliance through aggressive online behaviors. This directly undermines the principle of non-coercion, which is crucial for genuine discourse. Furthermore, the spreading of distorted truths—a hallmark of cyberbullying—violates Habermas's emphasis on sincerity and the foundational role of truth in discourse. Insincerity is further facilitated by the cloak of online anonymity, allowing perpetrators to hide their true identities while engaging in deceitful actions that disrupt the mutual trust necessary for rational communication. Violations of ethical correctness, observed in cyberbullies' malicious intent and actions, further exemplify the departure from Habermas's principles. Such behaviors erode the normative foundation of discourse ethics, which demands respect for the dignity and autonomy of all participants in communication. These elements—coercion, spreading distorted truths, insincerity, and ethical violations—collectively indicate a significant divergence from the ideal conditions for reasoned, respectful, and productive communication, as outlined by Habermas.

Given this, it becomes evident that cyberbullying represents not just a failure to adhere to the norms of rational discourse, but also a challenge to the very possibility of such discourse in digital spaces. This underscores the importance of developing strategies, both at the level of platform design and through educational interventions, that can foster environments conducive to Habermas's ideals. It is essential for future research to not only continue shedding light on the complexities of cyberbullying, but also to explore innovative ways to cultivate digital communities where rational discourse can thrive.

CONFLICTS OF INTEREST

No potential conflict of interest relevant to this article was reported.

REFERENCES

Badri, M., Nuaimi, A. A., Guang, Y., & Rashedi, A. A. (2017). School performance, social networking effects, and learning of school children: Evidence of reciprocal relationships in Abu Dhabi. *Telematics and Informatics,* 34(8), 1433-1444.

https://doi.org/10.1016/j.tele.2017.06.006

Barlett, C. P. (2023). Cyberbullying as a learned behavior: Theoretical and applied implications. *Children (Basel, Switzerland),* 10(2), 325. https://doi.org/10.3390/children10020325

Bauman, S., Toomey, R. B., & Walker, J. L. (2013). Associations among bullying, cyberbullying, and suicide in high school students. *Journal of Adolescence,* 36(2), 341-350. https://doi.org/10.1016/j.adolescence.2012.12.001

Baxter, H. (2011). *Basic concepts in Habermas's* theory of communicative action. In H. Baxter (Ed.), *Habermas: The discourse theory of law and democracy* (pp. 9-59). Stanford University Press.

Catedrilla, J., Ebardo, R., Limpin, L., De La Cuesta, J., Ching, M. R., Trapero, H., & Leano, C. (2020, November 23-27). Loneliness, boredom and information anxiety on problematic use of social media during the COVID-19 pandemic. In H. J. So, M. M. Rodrigo, J. Mason, A. Mitrovic, D. Bodemer, W. Chen, Z. H. Chen, B. Flanagan, M. Jansen, R. Nkambou, & L. Wu (Eds.), *Proceedings of the 28th International Conference on Computers in Education* (pp. 52-60). Asia-Pacific Society for Computers in Education.

Cecez-Kecmanovic, D. (1995). Business communications modelling: Performative vs informative. *IFAC Proceedings Volumes,* 28(7), 423-428. https://doi.org/10.1016/S1474-6670(17)47142-4

Chan, T. K., Cheung, C. M., & Lee, Z. (2021). Cyberbullying on social networking sites: A literature review and future research directions. *Information and Management*, 58(2), Article 103411. https://doi.org/10.1016/j.im.2020.103411

Chen, L., Ho, S. S., & Lwin, M. O. (2017). A meta-analysis of factors predicting cyberbullying perpetration and victimization: From the social cognitive and media effects approach. *New Media & Society,* 19(8), 1194-1213. https://doi.org/10.1177/1461444816634037

Chisholm, J. F. (2014). Review of the status of cyberbullying and cyberbullying prevention. *Journal of Information Systems Education,* 25(1), 77-87. https://jise.org/volume25/n1/JISEv25n1p77.pdf

Choi, Y. J., Jeon, B. J., & Kim, H. W. (2021). Identification of key cyberbullies: A text mining and social network analysis approach. *Telematics and Informatics*, 56, 101504. https://doi.org/10.1016/j.tele.2020.101504

Cocal, C. J. (2019). Cyberbullying practices and experiences of the Filipino college students' social media users. *Journal of Educaction, Management and Social Sciences,* 2(1), 43-51. https://scholar.google.com/scholar?oi=bibs&hl=en&cluster=9646104241615606092

Conde, C. H. (2022). *Philippine activist arrested for cyber-libel: Charges against Walden Bello spotlight criminal defamation laws.* https://www.hrw.org/news/2022/08/09/philippine-activist-arrested-cyber-libel

Costales, J., Asio, J. M., Albino, M., Albino, F., & Riego de Dios, E. E. (2022, June 10-12). Students' profile and prevalence of cyberbullying in higher education: Implications to the new normal learning. *Proceedings of the 8th International Conference on Frontiers of Educational Technologies* (pp. 131-137). ACM.

Creswell, J. W. (2007). *Qualitative inquiry and research design: Choosing among five approaches* (2nd ed.). Sage Publications.

Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage.

Della Cioppa, V., O'Neil, A., & Craig, W. (2015). Learning from traditional bullying interventions: A review of research on cyberbullying and best practice. *Aggression and Violent Behavior,* 23, 61-68. https://doi.org/10.1016/j.avb.2015.05.009

Denzin, N. K., & Lincoln, Y. S. (2011). *The SAGE handbook of qualitative research*. Sage.

Ebardo, R., De La Cuesta, J., Catedrilla, J., & Wibowo, S. (2020, November 23-27). Peer influence, risk propensity and fear of missing out in sharing misinformation on social media during the COVID-19 pandemic. In H. J. So, M. M. Rodrigo, J. Mason, A. Mitrovic, D. Bodemer, W. Chen, Z. H. Chen, B. Flanagan, M. Jansen, R. Nkambou, & L. Wu (Eds.), *Proceedings of the 28th International Conference on Computers in Education* (pp. 351-359). Asia-Pacific Society for Computers in Education.

Ebardo, R., & Suarez, M. T. (2023). Facebook community lurking and non-participation: Sociotechnical barriers among older adults in online participation. In Q. Gao, & J. Zhou (Eds.), *Human aspects of IT for the aged population* (pp. 194-204). Springer.

Erbiçer, E. S., Ceylan, V., Yalçın, M. H., Erbiçer, S., Akın, E., Koçtürk, N., & Doğan, T. (2023). Cyberbullying among children and youth in Türkiye: A systematic review and meta-analysis. *Journal of Pediatric Nursing,* 73, 184-195. https://doi.org/10.1016/j.pedn.2023.09.003

Eweida, R. S., Hamad, N. I., Abdo, R. A. E. H., & Rashwan, Z. I. (2021). Cyberbullying among adolescents in Egypt: A call for correlates with sense of emotional security and psychological capital profile. *Journal of Pediatric Nursing,* 61, E99-E105. https://doi.org/10.1016/j.pedn.2021.05.008

Farrington, D. P., Zych, I., Ttofi, M. M., & Gaffney, H. (2023). Cyberbullying research in Canada: A systematic review of the first 100 empirical studies. *Aggression and Violent Behavior,* 69, 101811. https://doi.org/10.1016/j.avb.2022.101811

Garett, R., Lord, L. R., & Young, S. D. (2016). Associations

between social media and cyberbullying: A review of the literature. *mHealth*, 2, 46. https://doi.org/10.21037/mhealth.2016.12.01

Habermas, J. (1984). *The theory of communicative action: Vol. 2. Lifeworld and system: A critique of functionalist reason.* Beacon Press.

Hendry, B. P., Hellsten, L. M., McIntyre, L. J., & Smith, B. R. R. (2023). Recommendations for cyberbullying prevention and intervention: A Western Canadian perspective from key stakeholders. *Frontiers in Psychology,* 14, 1067484. https://doi.org/10.3389/fpsyg.2023.1067484

Hinduja, S., & Patchin, J. W. (2010). Bullying, cyberbullying, and suicide. *Archives of Suicide Research,* 14(3), 206-221. https://doi.org/10.1080/13811118.2010.494133

Jansen, H. (2010). The logic of qualitative survey research and its position in the field of social research methods. *Forum: Qualitative Social Research*, 11(2), Art. 11. http://nbn-resolving.de/urn:nbn:de:0114-fqs1002110

Klein, H. K., & Huynh, M. Q. (2004). The critical social theory of Jürgen Habermas and its implications for IS research. In J. Mingers, & L. P. Willcocks (Eds.), *Social theory and philosophy for information systems* (pp. 157-237). Wiley.

Luppicini, R. (2009). Conversation ethics for online learning communities. In U. Demiray, & R. Sharma (Eds.), *Ethical practices and implications in distance learning* (pp. 98-107). IGI Global.

McCarthy, S., Rowan, W., Mahony, C., & Vergne, A. (2023). The dark side of digitalization and social media platform governance: A citizen engagement study. *Internet Research,* 33(6), 2172-2204. https://doi.org/10.1108/INTR-03-2022-0142

Merriam, S. B. (2009). *Qualitative research: A guide to design and implementation.* John Wiley & Sons.

Ortiz-Ospina, E. (2019). *The rise of social media*. https://ourworldindata.org/rise-of-social-media

Özgür, H. (2020). A systematic review on cyberbullying interventions and preventions. *Shanlax International Journal of Education,* 9(1), 11-26. https://doi.org/10.34293/education.v9i1.3373

Patchin, J. W., & Hinduja, S. (2006). Bullies move beyond the schoolyard: A preliminary look at cyberbullying. *Youth Violence and Juvenile Justice,* 4(2), 148-169. https://doi.org/10.1177/1541204006286288

Peled, Y. (2019). Cyberbullying and its influence on academic, social, and emotional development of undergraduate students. *Heliyon,* 5(3), e01393. https://doi.org/10.1016/j.heliyon.2019.e01393

Ramos Salazar, L. (2021). *Handbook of research on cyberbullying and online harassment in the workplace*. IGI Global.

Ross, A., & Chiasson, M. (2011). Habermas and information systems research: New directions. *Information and Organization,* 21(3), 123-141. https://doi.org/10.1016/j.infoandorg.2011.06.001

Saif, A. N. M., & Purbasha, A. E. (2023). Cyberbullying among youth in developing countries: A qualitative systematic review with bibliometric analysis. *Children and Youth Services Review,* 146, 106831. https://doi.org/10.1016/j.childyouth.2023.106831

Saleem, S., Khan, N. F., Zafar, S., & Raza, N. (2022). Systematic literature reviews in cyberbullying/cyber harassment: A tertiary study. *Technology in Society,* 70, 102055. https://doi.org/10.1016/j.techsoc.2022.102055

Stake, R. E. (2006). *Multiple case study analysis*. Guilford Press.

Stewart, R. W., Drescher, C. F., Maack, D. J., Ebesutani, C., & Young, J. (2014). The development and psychometric investigation of the cyberbullying scale. *Journal of Interpersonal Violence,* 29(12), 2218-2238. https://doi.org/10.1177/0886260513517552

Suler, J. (2004). The online disinhibition effect. *CyberPsychology & Behavior,* 7(3), 321-326. https://doi.org/10.1089/1094931041291295

Vismara, M., Girone, N., Conti, D., Nicolini, G., & Dell'Osso, B. (2022). The current status of cyberbullying research: A short review of the literature. *Current Opinion in Behavioral Sciences,* 46, 101152. https://doi.org/10.1016/j.cobeha.2022.101152

Wachs, S., Wright, M. F., & Vazsonyi, A. T. (2019). Understanding the overlap between cyberbullying and cyberhate perpetration: Moderating effects of toxic online disinhibition. *Criminal Behaviour and Mental Health,* 29(3), 179-188. https://doi.org/10.1002/cbm.2116

Willard, N. E. (2007). The authority and responsibility of school officials in responding to cyberbullying. *Journal of Adolescent Health,* 41(6 Suppl 1), S64-S65. https://doi.org/10.1016/j.jadohealth.2007.08.013

Wong-Lo, M., & Bullock, L. M. (2014). Digital metamorphosis: Examination of the bystander culture in cyberbullying. *Aggression and Violent Behavior,* 19(4), 418-422. https://doi.org/10.1016/j.avb.2014.06.007

Woodlock, D., Salter, M., Dragiewicz, M., & Harris, B. (2023). "Living in the darkness": Technology-facilitated coercive control, disenfranchised grief, and institutional betrayal. *Violence Against Women*, 29(5), 987-1004. https://doi.org/10.1177/10778012221114920

Xu, Y., & Trzaskawka, P. (2021). Towards descriptive adequacy of cyberbullying: Interdisciplinary studies on features, cases and legislative concerns of cyberbullying. *International Journal for the Semiotics of Law,* 34(4), 929-943. https://doi.

org/10.1007/s11196-021-09856-4
Yin, R. K. (2009). *Case study research: Design and methods* (4th ed.). Sage Publications.
Yin, R. K. (2017). *Case study research and applications: Design and methods* (6th ed.). Sage Publications.
Yule, G. (1996). *Pragmatics*. Oxford University Press.

JISTaP http://www.jistap.org
Journal of Information Science Theory and Practice

RESEARCH PAPER
J Inf Sci Theory Pract 12(3): 36-48, 2024

eISSN : 2287-4577 pISSN : 2287-9099

https://doi.org/10.1633/JISTaP.2024.12.3.3

Incorporating Deep Median Networks for Arabic Document Retrieval Using Word Embeddings-Based Query Expansion

Yasir Hadi Farhan*
Department of Medical Physics, College of Applied Sciences, University of Fallujah, Fallujah, Iraq
E-mail: yasir.hadi87@yahoo.com

Mohanaad Shakir
Department of Management Information System (MIS), College of Business (COB), University of Buraimi (UOB), Buraimi, Oman
E-mail: mohanaad.t@uob.edu.om

Mustafa Abd Tareq
Department of Computer Science, University of Technology-Iraq, Baghdad, Iraq
E-mail: mustafa.t.abd@uotechnology.edu.iq

Boumedyen Shannaq
Department of Management Information System (MIS), College of Business (COB), University of Buraimi (UOB), Buraimi, Oman
E-mail: boumedyen@uob.edu.om

ABSTRACT

The information retrieval (IR) process often encounters a challenge known as query-document vocabulary mismatch, where user queries do not align with document content, impacting search effectiveness. Automatic query expansion (AQE) techniques aim to mitigate this issue by augmenting user queries with related terms or synonyms. Word embedding, particularly Word2Vec, has gained prominence for AQE due to its ability to represent words as real-number vectors. However, AQE methods typically expand individual query terms, potentially leading to query drift if not carefully selected. To address this, researchers propose utilizing median vectors derived from deep median networks to capture query similarity comprehensively. Integrating median vectors into candidate term generation and combining them with the BM25 probabilistic model and two IR strategies (EQE1 and V2Q) yields promising results, outperforming baseline methods in experimental settings.

Keywords: automatic query expansion, information retrieval, word embedding, deep median networks, Arabic document retrieval, natural language processing

Received: October 10, 2023 **Revised:** April 10, 2024
Accepted: May 9, 2024 **Published:** September 30, 2024

***Corresponding Author:** Yasir Hadi Farhan
https://orcid.org/0000-0002-2378-9177
E-mail: yasir.hadi87@yahoo.com

All JISTaP content is Open Access, meaning it is accessible online to everyone, without fee and authors' permission. Open Access articles are automatically archived in the Korea Institute of Science and Technology Information (KISTI)'s Open Access repository (AccessON). All JISTaP content is published and distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/). Under this license, the authors retain full ownership of their work, while permitting anyone to use, distribute, and reproduce the content in any medium, as long as the original authors and source are cited. For any reuse, redistribution, or reproduction of a work, users must clarify the license terms under which the work was produced.

© 2024 [Yasir Hadi Farhan, Mohanaad Shakir, Mustafa Abd Tareq, Boumedyen Shannaq] This article is distributed under the Creative Commons Attribution License (CC BY), allowing unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. It is published by the Korea Institute of Science and Technology Information (KISTI).

1. INTRODUCTION

The main challenge for search engines is that user queries are often short and not specific enough to accurately represent their information needs (Azad & Deepak, 2019; Farhan et al., 2020; Nwesri & Alyagoubi, 2015). This is due to a gap in user knowledge, according to the Anomalous State of Knowledge hypothesis (Belkin, 2005). The automatic query expansion (AQE) technique addresses this issue by expanding the primary query with related terms to generate better and more relevant results (Esposito et al., 2020; Farhan et al., 2020; Raza et al., 2019). Commercial search engines such as Yahoo and Google use this technique through linked queries, connected search terms, and auto-completion functionality (Cai & De Rijke, 2016).

1.1. Query Expansion Techniques

Query expansion techniques are classified into global and local methods. Global techniques use a thesaurus such as WordNet to enlarge the initial user queries without relying on retrieval outcomes (Farhan et al., 2020; Pal et al., 2014). Local techniques use relevance feedback from the first retrieval process to select appropriate terms to add to the primary query (Miyanishi et al., 2013; Takeuchi et al., 2017). The pseudo relevance feedback (PRF) technique is a useful expansion method that automates manual aspects of relevance feedback. It assumes that the top-k results retrieved in the original search contain words that can be used to refine the query further (Farhan et al., 2021b).

The PRF technique is effective, but it has practical issues. Terms generated from resources like WordNet tend to have multiple meanings, requiring a disambiguation strategy before using them for query expansion. Additionally, the PRF approach relies on the highest ranked documents retrieved, which can be impacted by irrelevant terms and multiple topics (Zou et al., 2018), making it challenging to improve both precision and recall metrics simultaneously (Fernández-Reyes et al., 2018). To resolve these issues, recent research in AQE has focused on using word embeddings (WE) as a semantic modelling process in order to have meanings from text (ALMasri et al., 2016; Diaz et al., 2016; Roy et al., 2016).

In natural language processing (NLP), WE use a 1-D mathematical embedding to represent every word as a vector of real numbers in a low-dimensional continuous vector space. The models use the proximity of terms in their corpus for training (Diaz et al., 2016). Word2Vec has two options for training WE models: Continuous Bag-Of-Words (CBOW) and Skip-Gram (SG). CBOW predicts aim words based on nearby words, while SG predicts surrounding words based on target words. These models can determine terms with syntactic and semantic similarities by using the same context. Most AQE techniques use query term search or neighbourhood strategies to expand the query using proximal N terms (Roy et al., 2016).

Current AQE techniques using WE typically retrieve candidate terms one at a time, without considering the influence of other terms in the query. Researchers argue that AQE can be improved by modelling query semantics as collective vocabulary terms, resulting in higher quality suggested terms that are more semantically relevant. To accomplish this, they suggest employing deep averaging networks (DANs), a neural architecture that computes the average of embedded words for classification and processes them through several linear layers. The researchers suggest that DANs can be used to identify related terms for AQE using the complete query input (Roy et al., 2016), but this approach has not yet been extensively studied.

The use of WE in information retrieval (IR) has been extensively studied, but its usage in Arabic IR has not been properly investigated (Alsmearat et al., 2014; Faqeeh et al., 2014), mainly due to the lack of ontological knowledge bases in Arabic (Mohsen et al., 2018). Previous studies on Arabic IR have mainly focused on assessing or comparing word-stemming methods (Abdelali et al., 2016; Abu El-Khair, 2007; Guirat et al., 2016; Larkey et al., 2002; Mustafa et al., 2008).

1.2. Aims and Objectives

The aim of this research is to address the persistent challenge of query-document vocabulary mismatch within IR systems by enhancing AQE techniques using deep median networks (DMNs). This challenge arises when user queries fail to accurately represent their information needs, leading to suboptimal search results. By leveraging DMNs, the study aims to comprehensively capture query similarity, thus improving the relevance and effectiveness of search engine results.

Furthermore, the objectives of this research encompass several key aspects. Firstly, it seeks to investigate the limitations of current AQE methods, particularly those reliant on WE, in effectively addressing query-document vocabulary mismatch. Secondly, the study aims to explore the potential of DMNs in capturing comprehensive query similarity, thereby overcoming the shortcomings of existing AQE approaches. Thirdly, it endeavours to integrate DMNs into AQE methodologies alongside traditional IR strategies, such as the BM25 probabilistic model and

EQE1 and V2Q models, to assess their combined efficacy in improving retrieval performance. Lastly, the research aims to evaluate the applicability and effectiveness of the proposed approach within the context of Arabic IR, an area that has been relatively underexplored due to the scarcity of ontological knowledge bases.

1.3. Significance of the Study

This study's importance is in its potential to enhance AQE techniques and aid in creating more efficient search engines. By introducing DMNs as a novel approach for capturing query similarity comprehensively, the research aims to overcome the limitations of existing methods and improve retrieval performance. Moreover, the focus on Arabic IR fills a crucial gap in the literature and has implications for improving access to relevant information in Arabic language contexts. Ultimately, findings from this study may inform the development of more efficient and inclusive search engine technologies, particularly in languages with unique linguistic characteristics such as Arabic.

The primary issue addressed in this study is the inherent challenge faced by search engines due to the ambiguity and lack of specificity in user queries. Despite existing techniques like AQE, practical hurdles persist, such as polysemy in ontological knowledge bases and reliance on top-ranked documents, affecting precision and recall. Recent advancements leveraging WE show promise, yet there is a need for improved AQE methodologies, especially in languages like Arabic. This study proposes utilizing DMNs for AQE in Arabic IR, aiming to address term mismatch issues and enhance retrieval performance.

2. RELATED WORKS

IR systems help users locate the information they seek by fetching it from a database in response to their queries (Baeza-Yates & Ribeiro-Neto, 1999). However, one of the main challenges faced by these systems is vocabulary mismatch. To address this issue (Carpineto & Romano, 2012; Farhan et al., 2020), researchers have proposed AQE techniques, which automatically add new terms to the query to improve the accuracy and precision of the IR system (Abbache et al., 2016). One popular approach for AQE is the use of WE, which has gained significant attention (ALMasri et al., 2016; Diaz et al., 2016; El Mahdaouy et al., 2018b; Farhan et al., 2021a; 2021b; Roy et al., 2016).

WE is a technique used in NLP for semantic parsing that helps in extracting meaning from texts. It represents words as vectors of real numbers in a corpus, which are categorized into local and distributed representations. The distributed representation indicates that all words having similar context display similar vectors within the WE vector space, thus creating a closer distribution. This technique can represent words based on their value vectors and can help in understanding natural language by extracting meaning from texts (Bengio, 2009; Kim et al., 2017; Turney & Pantel, 2010).

Aklouche et al. (2018) developed an AQE technique using Word2Vec toolkit based on WE. The study focused on two models, i.e., SG and CBOW, to learn semantically related words for the main queries from the Text REtrieval Conference (TREC) *Washington Post* Corpus. They reweighted and selected these terms and evaluated the effectiveness of the document retrieval process using the Euclidean distance to compute vector similarity. Two types of candidate vectors were selected, one related to the complete query and the other to individual terms. Their study showed that a query reweighting technique was more effective than other approaches and concluded that assigning the same weight to terms in the expanded query decreases retrieval effectiveness.

Using word vector representations provides an effective foundation for modelling semantic similarity in query expansion (Esposito et al., 2020). However, the context is often neglected during expansion techniques, and formulating a proper context for retrieving useful terms is crucial. Researchers have attempted to tackle this problem, such as Roy et al. (2016), but it has not been completely resolved. Two studies that addressed this problem are those conducted by Fernández-Reyes et al. (2018) and Zamani and Croft (2016). Fernández-Reyes et al. (2018) developed a new Query-Guided AQE strategy known as V2Q, which filtered the primary query and ignored unnecessary terms by considering terms that showed high similarity with those in the primary query.

Zamani and Croft (2016) proposed a new technique called EQE1 for query expansion, which considers the semantic similarity between all terms based on the similarity noted between different WE vectors. It presumes that the query terms are conditionally independent and all expanded query terms must be like the query terms that are selected for adding to the primary query. The experiments conducted in their study used two baseline methods, and the details are discussed in the paper.

Fernández-Reyes et al. (2018) suggested an AQE method that employs Word2Vec (WE) (Mikolov et al., 2013) and took into account the full query to tackle the disam-

 https://doi.org/10.1633/JISTaP.2024.12.3.3

biguation problem, enhancing both precision and recall (ALMasri et al., 2016). They proposed two strategies, query-guided association scheme (Q2V) and prospect-guided association scheme (V2Q), which retrieved candidate terms based on query terms and candidate terms that had a high semantic relationship with query terms, respectively. The Q2V approach used general pooling methods to generate Rankq values for every query term and selected top N terms for expansion. However, this approach could lose the context of the primary query. On the other hand, the V2Q approach used candidate expansion terms to vote for query terms and selected the most semantically related terms. Experimental results showed that these techniques improved precision and recall metrics and outperformed conventional IR models.

Zamani and Croft (2016) carried out research aimed at improving the effectiveness of query language models in ad-hoc retrieval tasks by utilizing WE. They proposed a new AQE embedding relevance model based on previous models (Lavrenko & Croft, 2017), and used two estimation processes to incorporate semantic similarity between terms in WE vectors. The techniques that relied on embedding outperformed other baselines in average precision and mean average precision (MAP) on three TREC collections. However, expanding solely on individual query terms can cause query drift, and there is a limitation in the utilization of query terms (Crimp & Trotman, 2018).

Farhan et al. (2021a) proposed a new approach for AQE called DANs. The approach involves using the mean vector of the initial query term vectors to pick potential vectors. They stated that for expanding the whole query sentence, the technique must consider the average vector of the primary query term vectors. DANs was incorporated in the BM25, V2Q, and EQE1 probabilistic models for improving the retrieval performance of Arabic texts. According to the experimental outcomes, the suggested approach enhanced the Arabic text retrieval's performance. It showed a better performance than the standard baseline techniques like BM25, V2Q, and EQE with regards to the precision and MAP values and was seen to be among the top 10 in most of the case studies.

Farhan et al. (2021b) also introduced a new approach to improve the effectiveness of AQE for Arabic text retrieval by utilizing DANs (Farhan et al., 2021b) and the query vectors' average for determining candidate expansion vectors. The study used Word2Vec for training and compared the proposed approach with the Okapi BM25 probabilistic framework, V2Q (Fernández-Reyes et al., 2018), and EQE1 (Zamani & Croft, 2016). The hypothesis was that the DANs-based PRF technique using WE similarity for generating expansion vectors can resolve the issue of term inconsistency and enhance the performance of Arabic text retrieval through AQE. The expansion vectors for potential terms were constructed using the k most relevant documents identified during PRF. Evaluation showed that the proposed approach significantly improved performance compared to baseline PRF frameworks.

Although the DANs technique had advantages, it had limitations such as the average vector not referring to a real vector, being impacted by the position of query term vectors, and showing good performance only with local datasets. To overcome these limitations, researchers proposed a new approach using the median vector instead of the average vector for generating additional candidate vectors. This approach, called DMNs, was not affected by the position of all query term vectors, and can be referred to a real vector.

3. METHOD

This study attempted to determine the probability that DMNs could support AQE for the Arabic IR process. Therefore, the researchers used the DMNs technique for the existing CBOW IR model, i.e., the probabilistic Okapi BM25 model along with the two representative WE techniques for AQE, the V2Q and EQE1 techniques which were initially described by Fernández-Reyes et al. (2018) and Zamani and Croft (2016). The researchers selected the BM25 model, as it showed a true performance in the TREC retrieval experiments and affected ranking algorithms in the commercial search engines (Croft et al., 2010). BM25 was regarded as the main baseline model which could be used for non-AQE results. However, the EQE1, V2Q, and DANs techniques were seen to be the most effective techniques. The proposed technique also compared against the BM25+DANs-PRF, EQE1+DANs-PRF, and V2Q+DANs-PRF techniques introduced by Farhan et al. (2021a; 2021b). The researchers compared the results of the DMNs-based AQE technique with the existing AQE techniques that were presented by these approaches. Thus, they determined whether the DMNs could generate better and increasingly relevant expansion terms and overcome the disadvantages presented by the DANs technique.

3.1. Word2Vec and Deep Median Networks

The Word2Vec deep learning toolkit was introduced

in a previous article (Mikolov et al., 2013) for generating word vectors from a text corpus. It employs CBOW and SG models to generate distributed word representations. DMNs is a new approach that determines the median vector of primary query term vectors to generate candidate expansion vectors, improving automatic IR (AIR) performance and overcoming the limitations of the DANs method. DMNs can be integrated with AQE techniques such as BM25, EQE1, and V2Q models to generate a candidate expansion list. DMNs refers to a sentence embedding process and can be trained faster for data with high syntactic variance, such as Arabic data. The explanations of how Word2Vec and DMNs function are provided below.

Word2Vec and DMNs are both powerful techniques in NLP, but they operate differently and serve distinct purposes.

3.1.1. Word2Vec

- Word2Vec is a popular deep learning toolkit used for generating word vectors from a text corpus.
- It is based on the idea that a word's meaning can be derived from its surrounding context within a vast text corpus.
- Word2Vec provides two primary models: CBOW and SG.
- CBOW: Estimates the target word from its surrounding context words, like completing a missing word based on nearby words.
- SG: Predicts the surrounding context words from a given target word, akin to inferring the words around a specific word.

These models learn to represent words as dense, low-dimensional vectors, where vectors for related words are positioned near each other in the vector space, capturing semantic relationships. Word2Vec's vectors capture syntactic and semantic similarities between words, enabling tasks like word analogy and similarity calculations.

3.1.2. Deep Median Networks (DMNs)

DMNs are a novel approach used to determine the median vector of primary query term vectors to generate candidate expansion vectors. Unlike other techniques that focus on average or summing vectors, DMNs prioritize the median vector, which represents a middle ground among the query terms. This approach is beneficial for capturing the essence of the query while avoiding potential biases introduced by extreme values or outliers. DMNs can be integrated with AQE techniques, such as BM25, EQE1, and V2Q models, to generate a candidate expansion list that enhances the performance of IR systems. Additionally, DMNs excel in processing data with high syntactic variance, such as Arabic text, making them suitable for diverse linguistic contexts.

In summary, Word2Vec transforms words into dense vectors based on their contextual usage, capturing semantic relationships, while DMNs focus on determining the median vector of query term vectors to improve query expansion and IR performance.

This study explores the effect of using the DMNs approach to expand query sentences in AQE. The approach is applied to three models, namely BM25 IR, EQE1-AQE, and V2Q-AQE. The Word2Vec CBOW technique is trained offline on the Arabic TREC 2001/2002 corpus collection, and the dataset is divided into three groups. The researchers performed a search using the same corpus and evaluated the statistical information regarding the dataset, as shown in Table 1 (Farhan et al., 2021a). The study aims to improve the performance of AQE and overcome the limitations presented by the DANs technique.

Indeed, while the concept of using the median operation in DMNs seems intuitive, it has not been extensively explored in prior research, particularly in the context of query expansion. This gap in existing literature highlights the novelty and contribution of DMNs in introducing this approach to enhance query expansion techniques.

3.2. Basic Query Expansion Based on DMNs (BM25+DMNs)

The Okapi BM25 model is a ranking algorithm employed in IR to assess how relevant documents are to a specific search query (Robertson et al., 1995). It is ground-

Table 1. Statistics for Arabic TREC collections

Collection	TREC 2001	TREC 2002	TREC 2001/2002
No. of queries	25	50	75
Average No. of words/queries	4.88	3.28	4.08
No. of documents		383,872	
No. of tokens		76 million	
No. of unique words		666,094	
Size (compressed)		209 MB	
Size (uncompressed)		869 MB	

Data from the article of Farhan et al. (2021a, *Journal of Information Science*, 49(5), 1168-1186).
TREC, Text REtrieval Conference.

 https://doi.org/10.1633/JISTaP.2024.12.3.3

ed in the probabilistic retrieval model and is an improvement over the earlier Okapi BM model. BM25 takes into account the frequency of query terms in documents, document length, and term frequency within the entire document collection. It adjusts relevance scoring based on these factors, providing more accurate results compared to traditional term frequency-inverse document frequency (TF-IDF) models. BM25 is widely used in search engines and has been shown to perform well across various IR tasks (Lv & Zhai, 2011; Robertson & Zaragoza, 2009). Overall, the Okapi BM25 model offers a robust and effective approach to ranking documents based on their relevance to a given query. Its ability to address the limitations of traditional TF-IDF models makes it a popular choice in modern IR systems (Croft et al., 2010). Equations 1 and 2 present the general BM25 scoring functions.

$$\sum_{i\in Q} log\ \frac{\frac{(r_i+0.5)}{R-\ r_i+0.5}}{\frac{(n_i-\ r_i+0.5)}{(N-\ n_i-R+\ r_i+0.5)}} \cdot \frac{(k_1+1)f_i}{K+\ f_i} \cdot \frac{(k_2+1)qf_i}{k_2+\ qf_i} \quad (1)$$

$$K=\ k_1\left((1-b)+b\,.\frac{dl}{avdl}\right) \quad (2)$$

The formula for calculating relevance in this model involves several parameters, including k_1, k_2, and K, which are determined through experimentation. The variable *qf* represents the frequency of a term in a query, while *dl* represents the length of a document. Commonly used values for the parameters include k_1=1.2, k_2 ranging from 0 to 1,000, and b=0.75. Additionally, *avdl* represents the average document longitude in the set.

It is easy to use AQE-based DMNs technique in the BM25 model, wherein the complete query terms form the input. Using a WE model, vectors were extracted for each query and a mean vector was created using the DMNs method proposed in the study. Thereafter, the vectors that were similar to the median vector were identified after using different cosine similarity techniques, as shown in Equation 3. These similar vectors were regarded as the potential candidate expression vectors, while the candidate expansion terms that were associated with the extracted vectors were also identified with the help of a WE corpus. The highest-ranking 'n' candidates were selected as the terms for expanding the initial query.

$$Cosine(A,B)=\frac{A\cdot B}{\|A\|\times\|B\|}=\frac{\sum_{i=1}^{n}A_i\times B_i}{\sqrt{\sum_{i=1}^{n}{A_i}^2}\times\sqrt{\sum_{i=1}^{n}{B_i}^2}} \quad (3)$$

Equation 4 illustrates the CBOW framework of the Word2Vec model, which is utilized to derive vector representations of terms. In this approach, the model uses the words surrounding a target word to predict its vector representation. The equation incorporates two parameters: |C|, which denotes the overall count of words within the corpus, and *c*, which refers to the dynamic context size of the target word (i.e., the number of words surrounding the target word that are considered in the prediction process).

$$\frac{1}{|C|}\sum_{t=1}^{|C|} log[log[P(wt\backslash wt-c,\dots,wt-1,wt+1,\dots,wt+c)]] \quad (4)$$

$$m(\overrightarrow{X_k})=\vec{X}\left[\frac{n}{2}\right]+\ \vec{X}\left[\frac{n+1}{2}\right] \quad (5)$$

Equation 5 is employed to determine the median vector from the initial query term vectors, where X_k is the ordered set of vectors for a certain query. n is the vector length, which is 300.

Here are the steps involved in the proposed approach:

1. Choose a user query Q containing n terms.
2. Use a pre-trained WE model to obtain a vector vi for each term ti in Q.
3. Compute the median vector m(vk) for the associated terms of Q using DMNs.
4. Use cosine similarity to calculate the similarity between m(vk) and all other vectors in the WE corpus.
5. Select the top k most similar vectors to m(vk) from the WE corpus, and create a set W containing the corresponding vectors w1, w2, ..., wk.
6. Obtain the words tw1, tw2, ..., tk that match the vectors w1, w2, ..., wk from the WE corpus using the pre-trained model.
7. Incorporate the terms tw1, tw2, ..., tk into the initial query Q to create a revised query Q'.
8. Use the updated query Q' to fetch documents.

3.3. Embedding-Based Query Expansion Method Using DMNs (EQE1+DMNs)

The EQE1 procedure was proposed in 2016 by Zamani and Croft (2016). This technique assumed the presence of conditional independence between all query terms, wherein it was stated that the candidate terms must be like the query terms, which are selected for expansion. Hence, the researchers proposed the incorporation of the DMNs

technique into the EQE1 approach for improving the efficiency and performance of the AIR retrieval.

The proposed EQE1+DMNs approach was carried out in the following manner. Initially, the original query term vectors, vi, were used for finding set W, wherein set W = {w1, w2, …, wk} displayed the vectors in the WE corpus which showed the highest similarity to the query term vectors, vi. Thereafter, based on the DMNs, the researchers calculated the Median Vector m(V). After this step, the researchers compared the vectors in Set W with the m(V). The vectors in Set W, which displayed a similarity to m(V) → 0.7, were selected. Thus, the vectors that were selected were used for determining their similar and corresponding words in a WE corpus. These new terms would be added to the primary query for acquiring the new query. This newly acquired query would be used for retrieving the new documents.

Here are the steps for the proposed Embedding-Based Query Expansion Method using Dynamic Memory Networks (EQE1+DMNs):

1. Start with a user query Q consisting of n terms.
2. Using the created WE model, obtain the corresponding vector vi from the WE corpus for each term ti in Q.
3. Calculate the cosine similarity between the median vector m(vk) of associated terms (computed using DMNs) and all the other vectors within the WE corpus.
4. The top-k vectors most similar to vi in the WE corpus and set W were chosen.
5. Compute the median vector m(vk) of associated terms using DMNs.
6. Calculate the cosine similarity between m(vk) and vectors in set W.
7. Choose the vectors from set W that have a similarity score of 0.7 or higher for m(vk).
8. Retrieve the words corresponding to the selected vectors w1, w2, …, wk from the corpus based on the created WE model.
9. Incorporate the chosen words into the original query Q to form a revised query Q'.
10. Use the updated query Q' to fetch documents.

3.4. Prospect-Guided Query Expansion Strategy Based on DANs (V2Q+DMNs)

During the application of the proposed techniques, the researchers have suggested using the median vectors of the DMNs for generating the expansion candidate vectors. It was seen that this median vector is not affected by their place of query terms present in a vector space. Furthermore, the median vector is regarded as the real vector in a vector space, not like the average vector in the DANs that was described earlier (Farhan et al., 2021a; 2021b). The average vector was a random number present in the WE vector space. This new method is named as V2Q+DMNs. This method could improve V2Q performance, since the candidate expansion set is developed with the help of a median vector rather than using the direct individual query terms.

This strategy is implemented in the following manner: Step 1 involves finding set W = {w1, w2, …, wk}, wherein it consists of vectors in the WE corpus which showed the highest similarity to the primary query term vectors. Thereafter, the researchers determined the m(V) (median vector) for these primary query term vectors. They also determined Set V = {v1, v2, …, vk}, which included the vectors which showed the highest similarity to the m(V) in this WE corpus. Thereafter, they calculated the similarity (with the help of the cosine similarity) between all vectors in Sets W and V and the m(V), and selected the vectors from both the sets which showed the highest similarity to m(V). Lastly, the vectors that were selected were used for deriving the real words from the WE corpus with the help of the WE model. These new words were then added to the basic query and the documents were retrieved using this new query.

This method hypothesized that the m(V) vector was placed in the centre of the query term vectors present in the vector space. This vector highlighted the actual meaning of the complete query. Thus, it could be seen that the candidate expansion terms that were generated using this proposed vector could be useful expansion terms.

Here is a paraphrased version of the given steps:

1. Choose a user query Q consisting of n terms.
2. For each term ti in the query Q, use the WE model to derive the associated vector vi.
3. Calculate the cosine similarity between vi and every other vector in the WE corpus.
4. Choose the k vectors most similar to vi from the corpus and compile them into a set W.
5. Compute the median vector m(vk) for each vector vi by using DMNs to associate terms with vectors.
6. Compute the cosine similarity between m(vk) and every other vector in the WE corpus.
7. Identify the k vectors in the corpus that are closest to m(vk) and assemble them into a set V.
8. Compare the vectors in set W and set V to m(vk).
9. Select the vectors in the intersection of W and V that

 https://doi.org/10.1633/JISTaP.2024.12.3.3

have a cosine similarity to m(vk)≥0.7.

10. Retrieve the words corresponding to the vectors in set W and set V from the WE corpus.

11. Include the retrieved words in the original query Q to create an updated query Q'.

12. Use the revised query Q' to get new documents.

4. EXPERIMENTS

4.1. Experimental Setup

In the experimental setup, the researchers meticulously selected and compared various techniques, including the primary BM25 model, DANs, EQE1, and V2Q techniques, alongside the proposed DMNs-based expansion approaches. To ensure a comprehensive evaluation, these approaches were benchmarked against each other using standard evaluation metrics. Additionally, the researchers utilized the TREC 2001/2002 Arabic newswire dataset, a widely accepted benchmark for evaluating Arabic text retrieval systems, which comprises news articles from the Middle East published between 1994 and 2000. This dataset has been used recently by many researchers involved in the retrieval of Arabic texts (Abdelali et al., 2016; Darwish & Ali, 2012; El Mahdaouy et al., 2018a). It was seen that the TREC 2001/2002 Arabic newswire consisted of three standard TREC collections, i.e., TREC 2001, containing 25 queries; TREC 2002, containing 50 queries; and TREC 2001/2002, containing 75 queries. Moreover, the creation of the WE corpus using the Word2Vec (Mikolov et al., 2013) process involved processing 383,872 of Arabic newspaper articles from *Agence France Presse*, resulting in a comprehensive dataset exceeding 1 GB in size after encoding and distribution.

Regarding preprocessing, stemming was applied to the dataset using the Farasa stemmer (Abdelali et al., 2016), a recognized tool for Arabic language processing, to address the significant impact of stemming on the performance of Arabic text retrieval systems. The Farasa stemmer was seen to be a very effective stemmer for the Arabic language (El Mahdaouy et al., 2018a). Stemming helps in reducing words to their root form, enhancing the efficiency and effectiveness of retrieval processes. Furthermore, the researchers ensured a consistent experimental setup by retrieving 100 documents for each query and considering nine baselines for evaluation, i.e., (1) Probabilistic Okapi BM25 model (without expansion); (2) BM25+DANs, proposed by Farhan et al. (2021a); (3) BM25+DANs-PRF, proposed by Farhan et al. (2021b); (4) Embedding-based Query Expansion (EQE1) (Zamani & Croft, 2016); (5) EQE1+DANs, proposed by Farhan et al. (2021a); (6) EQE1+DANs-PRF, proposed by Farhan et al. (2021b); (7) Prospect-Guided QE strategy (V2Q) strategy; (8) V2Q+DANs, proposed by Farhan et al. (2021a); and (9) V2Q+DANs-PRF, proposed by Farhan et al. (2021b), covering a range of AQE techniques and their variants. These baselines included models without expansion, expansion models with DANs, EQE1, and V2Q, as well as their PRF variants proposed in previous studies.

For implementation, the experiments utilized the Whoosh search engine library in Python, a popular tool for constructing search engines and assessing retrieval systems (Mukherjee & Kumar, 2019). The semantic similarity between terms was determined by calculating the cosine similarity of their WE vectors, providing a quantitative measure of similarity between words in the embedding space. Additionally, the assessment metrics employed included MAP and Precision at the top 10 retrieved documents (P@10), provided insights into the effectiveness of the proposed DMNs-based AQE approach compared to existing techniques. Overall, the experimental methodology was carefully designed and executed to ensure robust evaluation and reliable comparisons between different AQE methods.

4.2. Metrics for Assessing Performance

The evaluation of the retrieval performance of the proposed DANs AQE approach and the four baselines was conducted using MAP of the top 100 ranked documents as the primary metric. Additionally, P@10 was also considered. Recall-precision curves were generated to depict the performance of each method across various standard recall levels. The formulas for calculating each performance metric are provided below.

$$Recall = \frac{|Rel| \cap |Ret|}{|Rel|} \tag{6}$$

$$Precision = \frac{|Rel| \cap |Ret|}{|Ret|} \tag{7}$$

$$MAP = \frac{1}{N}\sum_{i=1}^{n} AP_i \tag{8}$$

In the equation provided, "Ret" refers to the total number of documents that are retrieved by the search engine, while "Rel" refers to the total number of documents that are considered relevant within the dataset being evaluated.

$AP = \sum_{i=1}^{n} Precision_i \cdot \Delta Recall_i.$

The precision at each rank i of the retrieved documents is multiplied by the change in recall from items i-1 to i, and the resulting values are summed up to obtain the average precision. Here, the variable n refers to the overall number of retrieved documents, and $Precision_i$ is the proportion of relevant documents among the top i retrieved documents. $\Delta Recall_i$ is the difference in recall between the top i-1 and top i retrieved documents.

5. FINDINGS AND ANALYSIS

Table 2 displays precision at 10 and values of MAP for each of the proposed AQE methods and the baseline approaches. The values in mark a) signify that this value has outperformed the baseline approach when utilizing DMNs technique. The experiments have been carried out on three dataset collections: TREC 2001, TREC 2002, and TREC 2001/2002. TREC 2001 has 25 queries, TREC 2002 has 50, and TREC 2001/2002 blends the first two collections with 75 queries. The collections hit the same dataset corpus. In the experiments, the top 100 retrieved documents were considered.

The AQE methods evaluated include BM25+DMNs, EQE1+DMNs, and V2Q+DMNs, where DMNs are utilized to identify and select potential candidate expansion terms. These methods are compared against baseline approaches such as BM25 (without expansion), BM25+DANs, EQE1, EQE1+DANs, V2Q, and V2Q+DANs methodologies.

The recommended methods were compared against the probabilistic model BM25 (without expansion), and the BM25+DANs, EQE1, EQE1+DANs, V2Q, and V2Q+DANs methodologies. Table 2 depicts the experiment outcomes of all methodologies with regards to MAP and indicates P@10. The hypothesis assumes that DMNs can recommend more pertinent expansion terms, and can overpower the drawbacks of DANs, which would later enhance the AIR systems' retrieval performance.

Generally, the experiment outcomes depicted in Table 2 indicate that the improved BM25+DMNs are outclassing the BM25, BM25+DANs, and BM25+DANs-PRF

Table 2. MAP and P@10

Technique	Collection					
	TREC 2001		TREC 2002		TREC 2001/2002	
	MAP	P@10	MAP	P@10	MAP	P@10
Technique 1						
BM25	31.30	42.10	28.70	35.80	29.50	37.90
BM25+DANs	19.40	26.20	28.50	39.60	25.40	35.10
BM25+DANs-PRF	25.60	22.90	29.30	37.60	31.10	34.50
BM25+DMNs	32.30[a)]	55.50[a)]	37.20[a)]	40.10[a)]	34.90[a)]	47.80[a)]
Technique 2						
EQE1	30.70	42.90	25.70	33.50	26.90	30.40
EQE1+DANs	30.70	42.90	25.70	33.50	26.90	30.40
EQE1+DANs-PRF	31.30	43.00	30.20	34.30	32.30	34.20
EQE1+DMNs	32.30[a)]	51.00[a)]	39.80[a)]	42.70[a)]	36.10[a)]	46.80[a)]
Technique 3						
V2Q	27.00	35.20	26.20	33.30	26.50	33.90
V2Q+DANs	30.20	38.90	32.90	43.90	32.00	42.30
V2Q+DANs-PRF	28.30	35.70	31.20	34.60	30.70	34.10[a)]
V2Q+DMNs	32.20[a)]	59.30[a)]	36.60[a)]	45.60[a)]	34.40[a)]	52.40[a)]

MAP, mean average precision; P@10, precision at the top 10; TREC, Text REtrieval Conference; DANs, deep averaging networks; PRF, pseudo relevance feedback; EQE1, Embedding-Based Query Expansion Method; V2Q, Prospect-Guided Query Expansion Strategy.
[a)]The values signify that this value has outperformed the baseline approach when utilizing DMNs technique.

 https://doi.org/10.1633/JISTaP.2024.12.3.3

models with regard to MAP and P@10 for all TREC collections. Furthermore, on comparison with the baseline approaches EQE1, EQE1+DANs-PRF, and EQE1+DANs, the EQE1+DMNs method indicates the highest MAP and P@10 values for TREC 2001, TREC 2002, and TREC 2001/2002, and considerably outclasses the baseline approach EQE1 and EQE1+DANs. Moreover, the suggested V2Q+DMNs method significantly outshines their baseline methodologies V2Q, V2Q+DANs, and V2Q+DANs-PRF with respect to Precision at 10 and MAP at 10 for each of the TREC data sets.

In summary, in Table 2, the experimental results were presented in detail to compare the performance of different AQE methods and baseline approaches across three TREC dataset collections. Each method's P@10 and MAP scores are provided, allowing for a comprehensive evaluation of their retrieval effectiveness. The AQE methods considered, such as BM25+DMNs, EQE1+DMNs, and V2Q+DMNs, leverage DMNs to select expansion terms, aiming to address query-document vocabulary mismatch and enhance search results. The comparison against baseline approaches, including BM25, BM25+DANs, EQE1, EQE1+DANs, V2Q, and V2Q+DANs, allows for insights into the relative performance improvements achieved by integrating DMNs. By analysing the results, it can be identified which AQE methods yield the highest precision and MAP values, thereby informing decisions regarding the selection and implementation of query expansion techniques in IR systems.

Finally, our research results diverge from early studies, primarily in the incorporation of DMNs for query expansion within the context of IR. While previous studies may have explored AQE techniques using traditional methods or simpler models, this research introduces a novel approach by leveraging DMNs to generate candidate expansion terms based on the median vector of query term vectors derived from Word2Vec WE. This departure from conventional methods results in significant improvements in retrieval performance, as demonstrated through empirical evaluation against established techniques, including Okapi BM25, V2Q, EQE1, and their variations with DMNs.

6. CONCLUSIONS, FUTURE WORKS, AND LIMITATIONS

In conclusion, this study proposed the utilization of DMNs in improving the performance of AIR. By leveraging DMNs to generate candidate expansion terms based on the median vector of query term vectors derived from Word2Vec WE, the study aimed to address limitations in existing techniques. Comparative analysis with various methods including Okapi BM25, V2Q, V2Q+DANs, BM25+DANs, EQE1, and EQE1+DANs revealed significant improvements in retrieval performance with the incorporation of DMNs into WE models and the BM25 model. Specifically, BM25+DMNs utilized the median vector of original query term vectors for expansion, while EQE1+DMNs and V2Q+DMNs combined original query term vectors with the median vector of DMNs. This approach yielded more relevant expansion terms compared to the DANs method. Key findings underscored the effectiveness of DMNs in enhancing query expansion and ultimately improving search results.

In looking toward future research endeavours, future research should delve into the semantic aspects of query terms to further refine retrieval techniques and ensure the retrieval of even more relevant terms. By delving deeper into the semantic understanding of queries, researchers can continue to refine and advance IR methodologies. This study's contribution lies in not only proposing a novel approach but also demonstrating its efficacy through empirical evaluation against established techniques. The findings underscore the potential of DMNs to enhance query expansion processes, offering valuable insights for the ongoing evolution of IR systems.

While the study addresses several limitations, further considerations could encompass potential challenges in implementing the proposed model and its generalizability. Challenges in implementation may arise due to computational complexity and resource requirements, particularly in real-world applications where computational resources may be limited. Additionally, integrating the DMNs-based AQE approach into existing IR systems may necessitate significant modifications to infrastructure and workflows. Moreover, the study's findings may have limited generalizability beyond the specific context and dataset used. Variations in dataset characteristics, such as document lengths and domain-specific terminology, could impact the model's performance. Additionally, factors such as the quality and size of the training corpus for WE, as well as parameter choices, may influence the effectiveness of the model across different languages, domains, and datasets. Thus, while the study provides valuable insights, further research is needed to validate its applicability in diverse contexts.

Future research endeavours should emphasize the divergence of our results from prior studies, highlighting

the efficacy of DMNs in query expansion. Furthermore, exploring the integration of DMNs into related libraries and databases could enhance IR methodologies, fostering advancements in library and information science research.

7. AUTHOR CONTRIBUTIONS

The author contributions can be summarized as follows:

1. Proposal of DMNs Integration: The study proposes the integration of DMNs into the AIR process to improve performance.

2. Utilization of DMNs for Candidate Expansion Terms: By leveraging DMNs, the study aims to generate candidate expansion terms based on the median vector of query term vectors derived from Word2Vec WE, addressing limitations in existing techniques.

3. Comparative Analysis and Evaluation: The study conducts a comparative analysis with various methods, including Okapi BM25, V2Q, V2Q+DANs, BM25+DANs, EQE1, and EQE1+DANs, demonstrating significant improvements in retrieval performance with the incorporation of DMNs into WE models and the BM25 model.

4. Identification of Future Research Directions: The study suggests future research directions, focusing on delving into the semantic aspects of query terms to further refine retrieval techniques and ensure the retrieval of even more relevant terms, offering valuable insights for the ongoing evolution of IR systems.

Overall, the contributions include proposing a novel approach, demonstrating its effectiveness through empirical evaluation, and identifying potential future research directions for advancing IR methodologies.

ACKNOWLEDGMENTS

We express our gratitude to the LDC for granting us the LDC2001T55 Arabic Newswire Part 1 without any charges and for presenting us with the LDC Data Scholarship in the autumn of 2012. This study is partially supported by the Universiti Kebangsaan Malaysia grant: DCP-2017-007/4.

CONFLICTS OF INTEREST

No potential conflict of interest relevant to this article was reported.

REFERENCES

Abbache, A., Meziane, F., Belalem, G., & Belkredim, F. Z. (2016). Arabic query expansion using WordNet and association rules. In *Information Retrieval and Management: Concepts, Methodologies, Tools, And Applications* (pp. 1239-1254). IGI Global.

Abdelali, A., Darwish, K., Durrani, N., & Mubarak, H. (2016, June 12-17). *Farasa: A fast and furious segmenter for Arabic.* Paper presented at Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations, San Diego, California.

Abu El-Khair, I. (2007). Arabic information retrieval. *Annual Review of Information Science and Technology,* 41(1), 505-533.

Aklouche, B., Bounhas, I., & Slimani, Y. (2018, November 13-16). *Query expansion based on NLP and word embeddings.* Paper presented at Text Retrieval Conference (TREC), Gaithersburg, Maryland, USA.

ALMasri, M., Berrut, C., & Chevallet, J. P. (2016, March 20-23). *A comparison of deep learning based query expansion with pseudo-relevance feedback and mutual information.* Paper presented at Advances in Information Retrieval: 38th European Conference on IR Research, ECIR 2016, Padua, Italy.

Alsmearat, K., Al-Ayyoub, M., & Al-Shalabi, R. (2014, May 19-22). *An extensive study of the Bag-Of-Words approach for gender identification of Arabic articles.* Paper presented at 2014 IEEE/ACS 11th International Conference on Computer Systems and Applications (AICCSA), Aqaba, Jordan.

Azad, H. K., & Deepak, A. (2019). Query expansion techniques for information retrieval: A survey. *Information Processing & Management,* 56(5), 1698-1735. https://doi.org/10.1016/j.ipm.2019.05.009

Baeza-Yates, R., & Ribeiro-Neto, B. (1999). *Modern information retrieval* (Vol. 463). ACM press New York.

Belkin, N. J. (2005). Anomalous state of knowledge. In K. E. F. S. E. Robertson, & E. F. McKechnie (Eds.), *Information Research: Theory and Practice* (pp. 1-12). American Society for Information Science and Technology.

Bengio, Y. (2009). Learning deep architectures for AI. *Foundations and Trends® in Machine Learning,* 2(1), 1-127. http://dx.doi.org/10.1561/2200000006

Cai, F., & De Rijke, M. (2016). A survey of query auto completion in information retrieval. *Foundations and Trends® in Information Retrieval,* 10(4), 273-363. http://dx.doi.org/10.1561/1500000055

Carpineto, C., & Romano, G. (2012). A survey of automatic query expansion in information retrieval. *Acm*

Computing Surveys (CSUR), 44(1), 1-50. https://doi.org/10.1145/2071389.2071390

Crimp, R., & Trotman, A. (2018, December 10-12). *Refining query expansion terms using query context.* Paper presented at Proceedings of the 23rd Australasian Document Computing Symposium, Melbourne, Australia.

Croft, W. B., Metzler, D., & Strohman, T. (2010). *Search engines: Information retrieval in practice* (Vol. 520). Addison-Wesley Reading.

Darwish, K., & Ali, A. M. (2012, July 8-14). *Arabic retrieval revisited: Morphological hole filling.* Paper presented at Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers-Volume 2, Jeju, Korea.

Diaz, F., Mitra, B., & Craswell, N. (2016, August 7-12). *Query expansion with locally-trained word embeddings.* Paper presented at Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, Berlin, Germany.

El Mahdaouy, A., El Alaoui, S. O., & Gaussier, E. (2018a). Improving Arabic information retrieval using word embedding similarities. *International Journal of Speech Technology,* 21(1), 121-136. https://doi.org/10.1007/s10772-018-9492-y

El Mahdaouy, A., El Alaoui, S. O., & Gaussier, E. (2018b). Word-Embedding-based pseudo-relevance feedback for Arabic information retrieval. *Journal of Information Science,* 45(4), 429-442. https://doi.org/10.1177/0165551518792210

Esposito, M., Damiano, E., Minutolo, A., De Pietro, G., & Fujita, H. (2020). Hybrid query expansion using lexical resources and word embeddings for sentence retrieval in question answering. *Information Sciences,* 514, 88-105. https://doi.org/10.1016/j.ins.2019.12.002

Faqeeh, M., Abdulla, N., Al-Ayyoub, M., Jararweh, Y., & Quwaider, M. (2014, August 27-29). *Cross-Lingual short-text document classification for Facebook comments.* Paper presented at 2014 International Conference on Future Internet of Things and Cloud, Vienna, Austria.

Farhan, Y. H., Mohd, M., & Noah, S. A. M. (2020). Survey of automatic query expansion for Arabic text retrieval. *Journal of Information Science Theory and Practice,* 8(4), 67-86. https://doi.org/10.1633/JISTaP.2020.8.4.6

Farhan, Y. H., Mohd Noah, S. A., Mohd, M., & Atwan, J. (2021a). Word-Embedding-based query expansion: Incorporating deep averaging networks in Arabic document retrieval. *Journal of Information Science,* 49(5), 1168-1186. https://doi.org/10.1177/01655515211040659

Farhan, Y. H., Noah, S. A. M., Mohd, M., & Atwan, J. (2021b). Word embeddings-based pseudo relevance feedback using deep averaging networks for Arabic document retrieval. *Journal of Information Science Theory and Practice,* 9(2), 1-17. https://doi.org/10.1633/JISTaP.2021.9.2.1

Fernández-Reyes, F. C., Hermosillo-Valadez, J., & Montes-y-Gómez, M. (2018). A prospect-guided global query expansion strategy using word embeddings. *Information Processing & Management,* 54(1), 1-13. https://doi.org/10.1016/j.ipm.2017.09.001

Guirat, S. B., Bounhas, I., & Slimani, Y. (2016). Combining indexing units for Arabic information retrieval. *International Journal of Software Innovation (IJSI),* 4(4), 1-14. https://doi.org/10.4018/IJSI.2016100101

Kim, H. K., Kim, H., & Cho, S. (2017). Bag-Of-Concepts: Comprehending document representation through clustering words in distributed representation. *Neurocomputing,* 266, 336-352. https://doi.org/10.1016/j.neucom.2017.05.046

Larkey, L. S., Ballesteros, L., & Connell, M. E. (2002, August 11-15). *Improving stemming for Arabic information retrieval: Light stemming and co-occurrence analysis.* Paper presented at Proceedings of the 25th annual international ACM SIGIR conference on Research and Development in Information Retrieval, Tampere, Finland.

Lavrenko, V., & Croft, W. B. (2017, August 2). *Relevance-based language models.* Paper presented at ACM SIGIR Forum, New York, USA.

Lv, Y., & Zhai, C. (2011, October 24-28). *Lower-bounding term frequency normalization.* Paper presented at Proceedings of the 20th ACM International Conference on Information and Knowledge Management, Glasgow, UK.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013, December 5-10). *Distributed representations of words and phrases and their compositionality.* Paper presented at Advances in Neural Information Processing Systems, Lake Tahoe, NV, USA.

Miyanishi, T., Seki, K., & Uehara, K. (2013, October 27-November 1). *Improving pseudo-relevance feedback via tweet selection.* Paper presented at Proceedings of The 22nd ACM International Conference on Information & Knowledge Management, San Francisco, CA, USA.

Mohsen, G., Al-Ayyoub, M., Hmeidi, I., & Al-Aiad, A. (2018, April 9-11). *On the automatic construction of an Arabic thesaurus.* Paper presented at 2018 9th International Conference on Information and Communication Systems (ICICS), Amman, Jordan.

Mukherjee, S., & Kumar, N. (2019, December 12-14). *Duplicate question management and answer verification system.* Paper presented at 2019 IEEE Tenth International Conference on Technology for Education (T4E), Bhubaneswar, India.

Mustafa, M., AbdAlla, H., & Suleman, H. (2008, December

2-5). *Current approaches in Arabic IR: A survey.* Paper presented at International Conference on Asian Digital Libraries, Hyderabad, India.

Nwesri, A. F. A., & Alyagoubi, H. A. (2015, August 31-September 4). *Applying Arabic stemming using query expansion.* Paper presented at 2015 26th International Workshop on Database and Expert Systems Applications (DEXA), Vienna, Austria.

Pal, D., Mitra, M., & Datta, K. (2014). Improving query expansion using WordNet. *Journal of the Association for Information Science and Technology,* 65(12), 2469-2478. https://doi.org/10.1002/asi.23143

Raza, M. A., Mokhtar, R., Ahmad, N., Pasha, M., & Pasha, U. (2019). A taxonomy and survey of semantic approaches for query expansion. *IEEE Access,* 7, 17823-17833. https://doi.org/10.1109/ACCESS.2019.2894679

Robertson, S., & Zaragoza, H. J. F. (2009). The probabilistic relevance framework: BM25 and beyond. *Foundations and Trends® in Information Retrieval,* 3(4), 333-389. https://doi.org/10.1561/1500000019

Robertson, S. E., Walker, S., Jones, S., Hancock-Beaulieu, M. M., & Gatford, M. (1995). *Okapi at TREC-3*. Paper presented at Third Text REtrieval Conference (TREC-3), Gaithersburg, MD, USA.

Roy, D., Paul, D., Mitra, M., & Garain, U. (2016). *Using word embeddings for automatic query expansion*. Paper presented at CoRR, Austin, TX, USA.

Takeuchi, S. I., Sugiura, K., Akahoshi, Y., & Zettsu, K. (2017). Spatio-temporal pseudo relevance feedback for scientific data retrieval. *IEEJ Transactions on Electrical and Electronic Engineering,* 12(1), 124-131. https://doi.org/10.1002/tee.22352

Turney, P. D., & Pantel, P. (2010). From frequency to meaning: Vector space models of semantics. *Journal of Artificial Intelligence Research,* 37, 141-188. https://doi.org/10.1613/jair.2934

Zamani, H., & Croft, W. B. (2016, September 12-16). *Embedding-based query language models.* Paper presented at Proceedings of The 2016 ACM International Conference on the Theory of Information Retrieval, Delaware, Newark, USA.

Zou, S., Tao, G., Wang, J., Zhang, W., & Zhang, D. (2018, July 8-12). *On the equilibrium of query reformulation and document retrieval.* Paper presented at Proceedings of the 2018 ACM SIGIR International Conference on Theory of Information Retrieval, Ann Arbor, MI, USA.

JISTaP http://www.jistap.org
Journal of Information Science Theory and Practice

RESEARCH PAPER
J Inf Sci Theory Pract 12(3): 49-62, 2024

eISSN : 2287-4577 pISSN : 2287-9099

https://doi.org/10.1633/JISTaP.2024.12.3.4

The MapDS-Onto Framework for Matching Formula Factors of KPIs and Database Schema: A Case Study of the Prince of Songkla University

Kittisak Kaewninprasert
Program in Data Science, College of Digital Science, Prince of Songkla University, Hat Yai Campus, Songkhla, Thailand
E-mail: kittisak.ka@psu.ac.th

Supaporn Chai-Arayalert*
Department of Information Technology, Faculty of Science and Industrial Technology, Prince of Songkla University, Surat Thani Campus, Surat Thani, Thailand
E-mail: supaporn.chai@psu.ac.th

Narueban Yamaqupta
Department of Tourism Management, Faculty of Commerce and Management, Prince of Songkla University, Trang Campus, Trang, Thailand
E-mail: narueban.y@psu.ac.th

ABSTRACT

Strategy monitoring is essential for business management and for administrators, including managers and executives, to build a data-driven organization. Having a tool that is able to visualize strategic data is significant for business intelligence. Unfortunately, there are gaps between business users and information technology departments or business intelligence experts that need to be filled to meet user requirements. For example, business users want to be self-reliant when using business intelligence systems, but they are too inexperienced to deal with the technical difficulties of the business intelligence systems. This research aims to create an automatic matching framework between the key performance indicators (KPI) formula and the data in database systems, based on ontology concepts, in the case study of Prince of Songkla University. The mapping data schema with ontology (MapDS-Onto) framework is created through knowledge adaptation from the literature review and is evaluated using sample data from the case study. String similarity methods are compared to find the best fit for this framework. The research results reveal that the "fuzz.token_set_ratio" method is suitable for this study, with a 91.50 similarity score. The two main algorithms, database schema mapping and domain schema mapping, present the process of the MapDS-Onto framework using the "fuzz.token_set_ratio" method and database structure ontology to match the correct data of each factor in the KPI formula. The MapDS-Onto framework contributes to increasing self-reliance by reducing the amount of database knowledge that business users need to use semantic business intelligence.

Keywords: ontology, business intelligence, string similarity search, key performance indicators, case study-The Prince of Songkla University

Received: April 18, 2023 **Revised:** March 11, 2024
Accepted: May 18, 2024 **Published:** September 30, 2024

***Corresponding Author:** Supaporn Chai-Arayalert
https://orcid.org/0000-0003-4316-3232
E-mail: supaporn.chai@psu.ac.th

All JISTaP content is Open Access, meaning it is accessible online to everyone, without fee and authors' permission. Open Access articles are automatically archived in the Korea Institute of Science and Technology Information (KISTI)'s Open Access repository (AccessON). All JISTaP content is published and distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/). Under this license, the authors retain full ownership of their work, while permitting anyone to use, distribute, and reproduce the content in any medium, as long as the original authors and source are cited. For any reuse, redistribution, or reproduction of a work, users must clarify the license terms under which the work was produced.

© 2024 [Kittisak Kaewninprasert, Supaporn Chai-Arayalert, Narueban Yamaqupta] This article is distributed under the Creative Commons Attribution License (CC BY), allowing unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. It is published by the Korea Institute of Science and Technology Information (KISTI).

1. INTRODUCTION

Business administration and management teams face enormous competition and change due to internal and external factors. Therefore, a business must become an agile organization that rapidly adapts its strategies and action plans. Investment in digital technology, information system capability, and innovation capability enhances the agility of a business. In addition, business intelligence (BI) increases the innovation capability to improve a firm's performance (Božič & Dimovski, 2019; Ravichandran, 2018). Furthermore, BI systems moderate the outcomes of information systems. It is crucial to have a channel of information communication, for such things as business results and statuses, between related departments within an organization (Petrini & Pozzebon, 2009; Popovič et al., 2014).

Organizations, whether in business and non-business domains, are now using BI to monitor and drive their strategies to achieve their goals (Azma & Mostafapour, 2012). BI is a powerful tool for supporting decision-making based on large amounts of data. This tool consists of methods for data extraction, data transformation, data loading, data modeling, and data visualization that lead to a competitive advantage (Jakhar & Krishna, 2020). Although BI is widely used in enterprises, many organizations do not reap the full benefits of these systems. Some organizations are still lagging behind in using BI enhancement to create successful business opportunities (Qushem et al., 2017; Ul-Ain et al., 2019). In a traditional BI system, users have two roles in BI processing: requester and responder. Decision-makers, like managers, executives, and board committees fill the requester's role and seek the information they require from the responder. The role of the responder is filled by the information technology staff or a BI expert who creates related reports and data visualization (Alpar & Schulz, 2016). The challenge of enhancing a BI system so that non-technician users can be self-reliant requires being able to ensure data can be accessed more comfortably, complex methods using multiple data sources are eliminated, flexibility of data retrieval, and ease of use (Lennerholt et al., 2018).

As mentioned above, BI is a crucial tool for non-technical users to monitor and control strategic results in their organization. However, BI method complexity is a threat to self-reliant use of the BI system for users who require BI analysis to perform their jobs. These problems are the obstacles that this research aims to overcome to facilitate problem-solving. The research objective is to create a new process using semantic techniques to facilitate non-technical users to be able to search relevant data from multiple databases, without needing an intimate understanding of the databases' structures and relationships, to monitor key performance indicators (KPI) in the semantic BI system.

Hence, the research question is "How can we decrease the complexity between non-technical user skills and database structure understanding using the ontology concept?" This study is underpinned by two main objectives: (1) to develop a framework aimed at facilitating data retrieval for non-technical users, thereby obviating the necessity for database management knowledge and skills, and (2) to identify the string-based similarity method most congruent with the aforementioned framework. The remainder of this paper briefly describes the conceptual background, the application of BI in organizations, BI in the context of KPI, semantic technology and ontology with regard to KPI, the differences between ontology and databases, and the string similarity search. Then, we present the research framework and methodology. Finally, we explain the results and discuss the research challenges.

2. LITERATURE REVIEW

2.1. Application of Business Intelligence in Organizations

Organizations face a complex array of data and information, strong competition from direct and indirect competitors, and the need for rapid, correct decision-making. Data originates from various stakeholders, such as customers, suppliers, employees, information systems related to business transactions, facts, and external data sources. Organizations collect these data for intelligent decision-making. Hence, data and information become an organization's strategic resources and the essential factor driving the company's efficiency and effectiveness. Therefore, BI is an appropriate solution for organizations to analyze various data and as a decision-making tool (Olszak, 2020). The definition of BI is a system that combines architectures, databases, techniques, and analytics tools. The critical objectives of BI are to analyze historical and present data to create an interactive, dynamic view of an organization's performance, data, and information for decision-makers (Sharda et al., 2018).

Klašnja-Milićević et al. (2017) presented a framework for BI processing using the Hadoop platform for data scientists. Similarly, Kumar and Belwal (2017) and Yulianto and Kasahara (2018) implemented BI in organizations using the Pentaho software, and Villegas-Ch et al. (2020)

 https://doi.org/10.1633/JISTaP.2024.12.3.4

used traditional BI processes using the Weka tool in higher education institutions. Meanwhile, Fraihat et al. (2021) designed an architecture for and developed BI for the real-estate market using the Python platform, employing the extract, transform, load (ETL) process and Power BI software for data visualization. These frameworks are general BI tools and applications for organizations. They differ from the business indicator management system de Andrade and Sadaoui (2017) developed using the Joomla platform, that was integrated into the dashboard view of KPI monitoring. A little differently, Piri (2020) analyzed KPIs by interviewing the managers and adopted the KPIs dashboard design. As mentioned above, many tools such as Weka, Pentaho, Hadoop, and Power BI are the tools organizations use for BI applications.

2.2. Business Intelligence on Key Performance Indicators Context

Many researchers emphasize studying KPI in the BI process of strategic management. KPIs were adapted as part of business realization in BI. Data scientists have to understand indicator analysis before proceeding to data processing and data mining (Alaskar & Saba, 2020; Bréant et al., 2020; Khatibi et al., 2017; Ng et al., 2015; Pestana et al., 2020; Ren & Tao, 2012; Rocha et al., 2017; Visser, 2020). In the same way, Zoumpatianos et al. (2013) combined the KPI monitor processes: KPIs projection analysis, KPIs data calculation, expectation analysis, performance analysis, and impact analysis, with BI. Conversely, Maté et al. (2017) described a semantic approach for business analytics using KPIs. This approach begins with strategic modeling using a business intelligence model to manually build the KPIs relationships and relate them to the semantics of business vocabulary and business rules (SBVR) using the Java platform. The SBVR referred to as a business dictionary becomes the corpus for the KPI interpreter, which provides the meanings of KPIs to the online analytical processing interpreter. In another KPI study, Sultan et al. (2017) proposed a framework for detecting appropriate KPIs based on feature selection using Weka's information gain algorithm and then determining the association rules adapted for the frequent pattern-growth algorithm and compact pattern tree (CP-tree).

Furthermore, Azzouz et al. (2020) designed four steps – namely, business context analysis, the definition of strategic objectives, multidimensional modeling, and system implementation – for strategic alignment of KPIs systems. Certainly, KPIs are the main focus when an organization needs to develop their BI for strategy monitoring and driving data performance. Therefore, KPIs are the key factors that must be calculated into the business insight process and integrated into all objective and goal projections.

2.3. Semantic Technology with Ontology for Key Performance Indicators

The definition of an ontology is a set of representational primitives used to model a domain of knowledge or discourse. The primitive representations for ontology are usually classes (or sets), attributes (or properties), and relationships (or relations between class members). Furthermore, these assets complicate the meanings and constraints on the rationally coherent application of information. In the context of database systems, ontologies can refer to the level of abstraction of data models. Moreover, an ontology is a hierarchical and relational model intended for knowledge modeling of individuals, including their attributes and relationships to other individuals. Typically, ontologies use language that abstracts data found in structures and schemes. More specifically, using an ontology provides a greater ability to express first-order logic than a database model allows. The ontology is the semantic level and the database schema is the physical level of the data model. There are several utilizations of ontologies to integrate heterogeneous databases, enabling interoperability among incongruent systems and services on a knowledge-based level, because ontologies operate independent of the lower level databases (Gruber, 2009).

There are some exciting applications of semantic technology related to KPIs. For example, Bai et al. (2014) applied ontology technology to KPIs in personnel performance evaluations. First, they created the evaluation ontology and instances. After that, they used semantic web rule language to determine the rules of the evaluation statement and then queried to retrieve the information using SPARQL Protocol and resource description framework (RDF) Query Language (SPARQL). Similarly, del Mar Roldán-García et al. (2021) proposed a semantic model related to the ontology concept which covered KPI relationships and adapted the ontology to information retrieval with SPARQL, based on domain expert validation. In contrast, Walzel et al. (2019) presented KPI ontologies as the knowledge base for systems engineering and machine setup in manufacturing systems. Özcan et al. (2021) introduced an ontology enrichment technique which matched domain-specific ontologies and cross-domain knowledge graphs for ontology creation in artificial intelligence applications. Furthermore, their technique supported the increased correct data entity identification.

Consequently, the ontology concept adds to the value of KPI relationship studies, which can apply in various contexts.

2.4. Difference Between Ontology and Databases

Relational databases (RDB) rely on a schema for construction of a table, but ontologies depend on the knowledge base of various relationships between classes or data. RDB are limited to one type of relationship, for example, the foreign key. If the system wants to create a relationship, or join multiple tables in the database to search for something, it has only a foreign key. The semantic web can see multidimensional relationships, for example, inheritance, associated with, part of, and many other forms, containing logical relationships and restrictions. Systems can infer implicit information from ontologies, but databases cannot (Hebeler et al., 2009).

Databases mainly use close-world assumption; on the other hand, ontologies use open-world assumption (OWA) if the system includes incomplete information. This concept represents existing knowledge and suggests new information findings. For example, if there is no direct flight from one place to somewhere else, the result returned by the ontology will be "I do not know." Furthermore, databases use unique name assumptions. The definition of this term declares that there is only one world available for one entity from the real world. Thus, in a database, numerous vital terms have to be defined. Furthermore, the difference between an ontology and a database is that a database does not use a taxonomy to represent an original aspect of an ontological reference, because databases are used to precisely and safely store large amounts of data. However, ontologies locate the application by combining semantic data, or communication between heterogeneous systems, and they share the knowledge and the structure of the information with humans or software (Sir et al., 2015).

Ontologies and databases can represent similar actualities, but the application of ontological technology can present advantages depending on the problem. Database technologies can efficiently manage massive amounts of data, whereas ontologies reasonably represent reality, but the technology is inefficient at managing instances when they are present in the format of an OWL or RDF file. When the number of instances increases, ontologies have to be stored in a database and provide an interface that supports access to the data. However, ontologies offer a restriction-free framework to denote that a machine can be a readable reality, even on the Web. This framework depends on an OWA, wherein information can be defined, shared, reused, or allocated. Furthermore, information can also be exchanged and used to make inferences or queries. Choosing one or the other technology depends on the end-user's requirements. If the information needs to be shared on the Web, an ontology should provide a good solution. Nonetheless, the decision would probably include using both technologies when a large amount of data needs to be stored and properly managed. Ontologies provide an excellent way to characterize reality, but a database is certainly the better method for storing data when size is a consideration (Martinez-Cruz et al., 2012).

Some studies have explored database transformation from a RDB to an ontology file or a RDF. Zhao et al. (2019) created a schema-based graph mapping method to transform RDB schema and table relationships into a RDF data graph, and then used the SPARQL endpoint to query the RDB data in the RDF file. Additionally, Jun et al. (2020) improved the RDB to RDF method by solving two problems: duplicated data generation and semantic information loss. Finally, Devi et al. (2020) built an ontology system covering the information requirement and used RDB to RDF Mapping Language for RDB to RDF transformation.

2.5. String Similarity Search

String similarity mapping is an essential technique for data integration and searches. The problem with string searching is how to decrease the inconsistencies in origin data and search keywords and increase the tolerated accuracy rate. The string-based similarity method is divided into token-based similarity and character-based similarity methods. The token-based similarity method evaluates the similarity of two strings with the set form of the string being a set of tokens. This method compares the overlap of the members in the tokens sets. Examples of the token-based similarity method are Overlap, Jaccard, Cosine, Dice, and Fuzzy-Wuzzy token set ratios. The second method is the character-based similarity method, which compares the consistency of two strings by separating each string into a sequence of letters and measuring the edited distance. For two strings that are examined to be considered similar, the strings must have a similarity score that meets or exceeds a set threshold. Examples of this method are Hamming distance and Levenshtein distance (Angeles & Espino-Gamez, 2015; Appa Rao et al., 2018; Yu et al., 2016). Finally, there are methods that do not fall under string-based similarity methods, such as the Jaro-Winkler and Monge-Elkan methods, which employ different approaches for performing string similarity com-

 https://doi.org/10.1633/JISTaP.2024.12.3.4

parisons. Thus, those methods are referred to as hybrid similarity methods (Prasetya et al., 2018).

Angeles and Espino-Gamez (2015) reported that the accuracy rate of the Hamming distance approach was higher than that of the Jaro and Monge-Elkan approaches, but its overall performance was lower compared to the Jaro and Monge-Elkan approaches. Appa Rao et al. (2018) presented the Fuzzy-Wuzzy method as being more appropriate than Levenshtein distance and Sequence Matcher in regard to the string similarity matching of a mathematical formula. Using dynamic queries of the MySQL database, Rinartha et al. (2018) compared the efficiency of Jaccard similarity, MySQL pattern matching, Levenshtein distance, and MySQL Fulltext Index. They determined that the MySQL Fulltext Index was the most accurate method, while MySQL pattern matching was the fastest method for query suggestion.

3. RESEARCH FRAMEWORK AND METHODOLOGY

3.1. Research Framework

The framework of mapping data schema with ontology (MapDS-Onto) is a part of our research about developing semantic BI. The MapDS-Onto framework, shown in Fig. 1, maps the factors of the KPIs formula with the associated column in a RDB using the relation of the data class and the related ontology. The MapDS-Onto framework will then create the factor dictionary used for BI generation.

The MapDS-Onto framework has four layers for objective processing. Layer 1 lists all of the databases in the organization's information systems along with their different configurations and connections. Layer 2 is a database schema mapping (DaSMap) and begins with the experts (researcher and strategic administrator) helping to create the database structure model ontology (DSMO). The relationship structure of the DSMO is presented in Fig. 2. The DSMO is the model used for the DaSMap algorithm, and that applies the process from Zhao et al. (2019) for reading the database structures of all of the information systems and building the database structure ontology (DSO). Layer 3 is called the domain schema mapping (DoSMap). Here, experts analyze the organization's KPIs attribute relationships and build the information forms to create the performance indicator ontology (PIO). This form receives the performance indicator property's data from users. The PIO is built based on an adaptation of the process developed by del Mar Roldán-García et al. (2021). Then, the DoSMap algorithm inputs the DSO and PIO to begin its process. DoSMap is the mapping method connecting the DSO and PIO to allow searching the appropriate classes in the DSO with the formula's factor of KPIs class in the PIO and building the factor dictionary file. The dictionary file serves as a compendium, implemented in Python, designated for storing data requisite for processing. Consequently, this dictionary file constitutes the relative column corpus of PIO, encompassing the column name, table name, and database name.

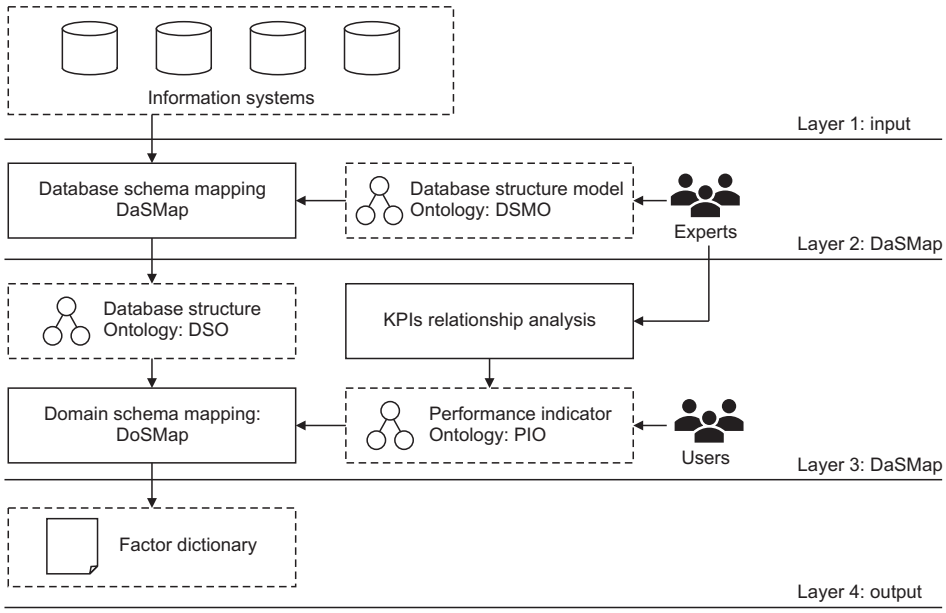


Fig. 1. Mapping data schema with ontology framework. KPIs, key performance indicators.

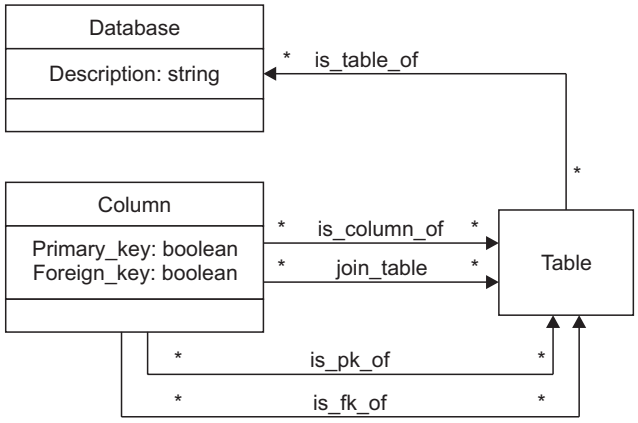


Fig. 2. Database structure model ontology.

3.2. Research Methodology

The research methodology of this study is divided into three phases as follows:

3.2.1. Phase 1: Problem Defining

1. A structured literature review was conducted to identify gaps related to the use of semantic techniques, enabling non-technical users to search for relevant data across multiple databases without understanding the databases' structures and relationships for KPIs monitoring in semantic BI systems.

2. The researchers analyzed relevant approaches and adapted appropriate theories to our framework.

3.2.2. Phase 2: Framework Development

DSMO was developed by applying the seven steps from Noy and McGuinness (2001) and evaluated using a data-driven approach and assessment by humans. The seven steps for DSMO creation are as follows:

3.2.2.1. Step 1: Determine the Domain and Scope of the Ontology.

The domain of DSMO is the relationships of database structure among the various information systems that scope on the relative connection between tables, columns, and databases through the primary keys and foreign keys. The DSMO focused on the RDB model, for example, MySQL, MariaDB, PostgreSQL, and Oracle. This study plans to apply DSMO to the data selection algorithm to decrease the complexity between non-technical user skills and database structure understanding. These users do not necessarily get the idea of database structure before selecting the suitable data for calculation. The following are the probable competency questions:

What databases are there in the system?
What tables are there in the database?
What columns are there in the table?
Which database does a table occupy?
Which table does a column lie in?
What does the table of that column, acting as the foreign key, refer to?

3.2.2.2. Step 2: Consider Reusing Existing Ontologies.

Based on the literature review, no relevant ontologies are related to this study. The DSMO applied the document of Hebeler et al. (2009) to create the DSMO structure.

3.2.2.3. Step 3: Enumerate Important Terms in the Ontology.

The related terms of the database structure's relationship will include database, table, column, primary key, foreign key, join table, database description, reference, connection, and RDB.

3.2.2.4. Step 4: Define the Classes and the Class Hierarchy.

There are three classes, including database, column, and table, to explain the structure of the databases. The database class will collect the names of databases, and the table class will collect the terms of all tables in the databases. The column class is the field name that contains the data.

3.2.2.5. Step 5: Define the Properties of Classes.

Concept relationships such as "is_a" can be used to describe what a class is. For example, Decision Support System (DSS) is a database, personnel is a table, and staff ID is a column. Attribute relationships like "attribute_of" can be used to describe the relationship where an attribute belongs to a class, such as describing a description as an attribute of a database. There are five object properties, "is_table_of," "is_column_of," "is_pk_of," "is_fk_of," and "join_table," to explain the relationships between the members of a class and other items. For example, camp ID serves as the primary key (is_pk_of) for the campus table.

3.2.2.6. Step 6: Define the Facets of the Slots.

The table class is the domain of the "is_table_of" slot. The column class is the domain of "is_column_of," "is_pk_of," "is_fk_of," and "join_table" slots. The database class is the range of the "is_table_of" slot. The table class is the range of "is_column_of," "is_pk_of," "is_fk_of," and "join_

 https://doi.org/10.1633/JISTaP.2024.12.3.4

table" slots.

3.2.2.7. Step 7: Create Instances.

In this study, two information systems, personnel information system (PIS) and The Prince of Songkla University (PSU) Research Project Management (PRPM), were utilized to read the database structure and create instances. For example, the database name "PRPM" was recorded as an instance named "database: PRPM" in DSMO.

1. The method of schema-based mapping to the RDF graph, as proposed by Zhao et al. (2019), was adapted to design the DaSMap algorithm.

2. The string similarity methods (Jaccard index, Sørensen-Dice, cosine similarity, overlap coefficient, fuzz. token_set_ratio, Jaro-Winkler, and Monge-Elkan) were compared based on similarity scores to identify the most suitable string similarity method for automatically searching for the appropriate columns of data from multiple databases.

3. The identified optimal string similarity method was employed to design DoSMap algorithm.

4. The MapDS-Onto framework was constructed, incorporating DaSMap, DoSMap, related ontologies, and new methods.

3.2.3. Phase 3: Framework Evaluation

1. The MapDS-Onto framework was evaluated using sample data from two information systems, the Research Project Management System and the Personal Information System, at PSU.

2. The researchers summarized and discussed the results of the experiment.

4. RESULTS AND DISCUSSION

In Fig. 2, a class diagram illustrates DSMO, which is the relationships of the classes in the DSO. DSMO is a simple but useful application. There are three classes: database, table, and column. The column class has two properties: "primary_key" and "foreign_key." These properties have Boolean format data values (yes and no) and present the key type of column class that applies to the annotation of the couple relation of the class. The class relationships are explained as follows:

1. is_table_of is the direct relationship between the table class and database class, where one table class is found in multiple databases; the table name can appear as the same name in different databases. The inverse relation exists when one database has many columns.

2. is_column_of is a direct relation between the column class and the table class, indicating that one column can appear in different tables. Consequently, the same column name may be present in multiple tables. Conversely, in the inverse relation, one table can have many columns.

3. is_pk_of and is_fk_of are related to the concept of the table class, where the column class serves as either the primary key (is_pk_of) or foreign key (is_fk_of) referencing the table class.

4. join_table serves as the connector to the column class. When a property (foreign key) is in effect, the column class must establish a relationship with a table class.

From the DSMO structure, one is presented with an exciting answer when a user wants to know the origin of the relation between classes. For example, if x is a selected column which fits the formula's factor of a KPI, then the BI system will know the route of x, which table x is located in, and what database x is located in.

After that, DSMO will be the model for input in the DaSMap algorithm, presented in Fig. 3, for building DSO from multiple databases. DaSMap uses the database connection dictionary (DSMO) as the input file, and the database connection dictionary is the database connection information file. This algorithm begins by loading the DSMO and then creates the dictionary reading loop for all database connections. When a database is connected, the algorithm will create a new subclass of the database class and then read the table structure. The next process em-

```
Data: database connection dictionary: dbD, ontology parent: parent
Load database structure model ontology
Foreach db in dbD do
    connect(db)
    create newSubClass(dbD[dbName], parent)
    table = showTables(db)
    Foreach tb in table do
        create newSubClass(tb[name], parent)
        append(dbD[dbName], is_table_of)
        column = showColumns(tb[name], db)
        Foreach cl in column do
            create newSubClass(cl[name], parent)
            append(tb[name], is_column_of)
            result = selectTableSchema(dbD[dbName], tb[name], cl[name])
            Foreach r in result do
                if primary key = true then
                    append(tb[name], is_pk_of)
                if foreign key = true then
                    append(tb[name], is_fk_of)
                    append(tb[name], join_table)
            end
        end
    end
end
Save ontology
```

Fig. 3. Database schema mapping.

ploys the table structure reading loop to build the subclass of table class, and the "is_table_of" relationships will be created. In the same way, subclasses of column class and the "is_column_of" relationships are generated from the column structure reading of each table. Finally, DaSMap reads the table schema to get and create the relationships among the columns in the table, determines the "is_pk_of," "is_fk_of," and "join_table" relationships, and then saves the DSO.

The DSO from DaSMap and the PIO are inputted to the DoSMap algorithm, as shown in Fig. 4. DoSMap is the algorithm that finds the correct columns that match the formula factor of the KPI by mapping the PIO with the DSO. DoSMap starts by loading the DSO and PIO, reads the KPI's class that is contained in the PIO, and selects and separates the KPI formula. Then, the factor will be taken to input in the "SelectedColumn" function for the column class searches. In instances where multiple databases have the same column names, the SelectedClassDictionary function is applied to solve this problem and create the factor dictionary file for the ETL process of semantic BI.

The SelectedColumn function in Fig. 5 is the process for weighing the string-similarity scores by applying the "fuzz.token_set_ratio" algorithm. The SelectedColumn function takes the factor generated by the DoSMap to map the name of the column class in the DSO and weighs the text similarity score. The result that gets the highest score will be recognized and the value returned to the DoSMap.

The SelectedClassDictionary function in Fig. 6 will input the factor calculated by the SelectedColumn function and the KPI's domain from PIO. This function reads the column class from the DSO which matches the SelectedColumn function's result to check for column redundancy by applying the "is_column_of" relationship. The table of column redundancy then reads the table origin using the "is_table_of" relationship. Then the database description of this table and the designated KPI domain are checked for string-similarity using the "fuzz.token_set_ratio" algorithm. The result with the highest score is selected, and its information—factor, column, table, and database—is incorporated into the factor dictionary.

```
Data: database structure ontology: DSO, performance indicator ontology: PIO
Load DSO, PIO
Foreach kpi ∈ PIO do
    KPIdomain = selectDomain(kpi)
    formula = selectFormula(kpi)
    factor = separateFactor(formula)
    selectedColumn = SelectedColumn(factor)
    SelectedClassDictionary(factor, selectedColumn, KPIdomain)
end
```

Fig. 4. Domain schema mapping.

```
Function SelectedColumn(factor)
    class = IRIS[URI#column]
    highScore = 0
    Foreach subclass in class do
        subclassName = subclass.name
        score = fuzz.token_set_ratio(subclassName, factor)
        if score >= highScore then
            highScore = score
            selectedColumn = subclassName
    end
Return selectedColumn
```

Fig. 5. SelectedColumn function.

4.1. Case Study

PSU is recognized as the highest quality university in Thailand's southern region by QS World University Rankings for 2021 (Quacquarelli Symonds, 2021). PSU has five campuses located in the provinces of Songkhla, Pattani, Phuket, Suratthani, and Trang. However, the PSU system does not have intelligent BI that supports decision-making for their executives. Therefore, using PSU as the case study, this research began by scoping two of their information systems: (1) PIS and (2) PRPM system. PIS presents the individual information of staff across the five campuses of PSU, encompassing a wide range of data such as educational records, employment history, job positions, salary and benefits, and leave information. For department managers, this system affords the capability to search and filter information pertaining to their subordi-

```
Function SelectedClassDictionary(factor, selectedColumn, KPIdomain)
    classSC = IRIS[URI#selectedColumn]
    Foreach table in classSC.is_column_of do
        tableName = table.name
        classCN = IRIS[URI#columnName]
        selectedClassDictionary = {}
        Foreach db in classCN.is_table_of do
            dbDescription = db.description
            highScore = 0
            score = fuzz.token_set_ratio(dbDescription, KPIdomain)
            if score >= highScore then
                highScore = score
                selectedClassDictionary.add(
                    factor: factor,
                    column: classCN,
                    table: tableName,
                    database: dbName)
        end
    end
Return factor dictionary
```

Fig. 6. SelectedClassDictionary function.

 https://doi.org/10.1633/JISTaP.2024.12.3.4

nates. On the other hand, the PRPM system stands as the central repository for PSU's research data, displaying individual research history, funding information, and other pertinent details. Furthermore, the system facilitates users in submitting research proposals to solicit funding from the university and record research data from the start to the end of the project. The MapDS-Onto framework was used to test the case study. After running the DaSMap process, DaSMap read the database structures of PIS and PRPM systems and generated the DSO that followed the pattern of the DSMO. The DSO graph in Fig. 7 shows the complex paths of the relationships between the classes and subclasses. For example, "total_budget_plan" is a subclass of column class and is a column in the project detail table. A related KPI for the two systems is *the budget from external funding per person*. The formula of this KPI is "sumif(budget, funding source)/countif(staff id, faculty name)."

From this KPI formula in the case study, the DoSMap algorithm separated the formula into four factors—total budget, funding source, staff ID, and faculty name—and matched their class string-similarities and created the factor dictionary. The factor dictionary from the DoSMap algorithm is written as follows:

```
factor dictionary = {
'budget' : {
'column': 'total_budget_plan',
'table': 'project_detail',
'database': 'prpm',},
'funding source' : {
'column': 'fund_source_id',
'table': 'project_detail',
'database': 'prpm',}
'staff id' : {
'column': 'staff_id',
'table': 'project_detail',
'database': 'prpm',}
'faculty name' : {
'column': 'faculty_owner',
'table': 'project_detail',
'database': 'prpm',}}
```

According to the factor dictionary of a KPI, the budget from external funding per person has four factors in the KPI formula. An example of the implication of this dictionary is that the budget factor has to read the data from the "total_budget_plan" column in the "project_detail" table of the PRPM system. The semantic BI system will extract the related data following the path of the column name in the database to transform, load, calculate, and visualize data in the next action of our research.

Another example is a KPI about the *number of foreign academic staff working full-time*, which has the formula "sumif(academic staff, staff type, staff country)." The factor dictionary of this KPI is as follows:

```
factor dictionary = {
'academic staff ' : {
'column': 'staff_id',
```

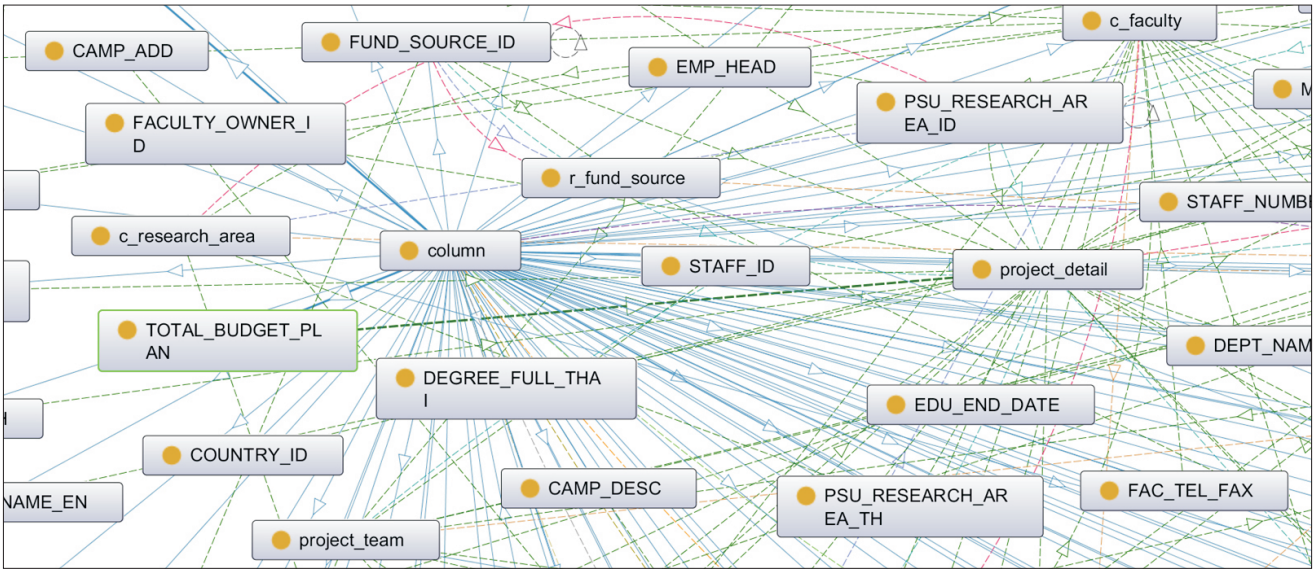


Fig. 7. Database structure ontology graph.

'table': 'personnel',
'database': 'dss',},
'staff type' : {
'column': 'staff_type_name',
'table': 'personnel',
'database': 'dss', },
'staff country' : {
'column': 'country_id',
'table': 'personnel',
'database': 'dss', }}

From the factor dictionary of the above KPI, we point out the formula's fit data path. The academic staff factor must read the "staff_id" column in the "personnel" table of the DSS system. The staff type factor must read the "staff_type_name" column in the "personnel" table of the DSS system. The staff country factor must read the "country_id" column in the "personnel" table of the DSS system.

In the "SelectedColumn" function and "SelectedClassDictionary" function, this research tested the reasonableness of the token-based and hybrid similarity methods and the hybrid similarity methods. The character-based similarity method was not applicable to this study, because that method calculates the differences in distance between each character in two words. However, in this case, the names of columns in the database were word groups. Hence, the other methods were better for string-similarity matching because the other methods are based on the word overlap technique.

Table 1 delineates the comparative results of matching accuracy and similarity scores between two categories of string similarity methods: (1) token-based similarity methods, including the Jaccard index, Sørensen-Dice, cosine similarity, overlap coefficient, and fuzz.token_set_ratio; and (2) hybrid similarity methods, namely Jaro-Winkler and Monge-Elkan. Upon considering the mean scores, it is discerned that token-based similarity methods yield a higher average score compared to hybrid similarity methods. Consequently, the token-based similarity method emerges as suitable for this case study. Looking more closely at the token-based methods, we see that the overlap coefficient and fuzz.token_set_ratio have similar scores.

Based on the exactitude comparison reported in Table 1, this research used the "fuzz.token_set_ratio" method for string-similarity matching. Although the "overlap coefficient" method produced the highest average score, the "overlap coefficient" method yielded the highest score values for more than one column. Therefore, it returned a false value. The "fuzz.token_set_ratio" method matched all the actual results and had an average score of about 0.76% lower than the "overlap coefficient" method. Based on evaluation of the practicability of application in this comparison, the "fuzz.token_set_ratio" method was more suitable for this research.

4.2. Discussion

When employing a specific BI software package, such as Weka, Pentaho, or Power BI, users need to understand how to select the required columns of data to run the BI process by themselves. This study eliminates this gap by developing a smart BI using semantic technology and providing user-friendly BI software for non-technical users. The previous works mainly applied ontology techniques to collecting and retrieving KPI values (Bai et al., 2014; del Mar Roldán-García et al., 2021; Maté et al., 2017). Some works used the ontology concept to create knowledge sharing between systems (Özcan et al., 2021; Walzel et al.,

Table 1. Results of the comparison of the string similarity methods

Method	Factor								Average score
	Budget		Funding source		Staff ID		Faculty name		
	Result	Score	Result	Score	Result	Score	Result	Score	
Jaccard index	x	35.29	x	75.00	✓	100	✓	66.67	69.24
Sørensen-Dice	x	52.17	x	85.71	✓	100	✓	80.00	79.47
Cosine similarity	✓	59.41	x	85.71	✓	100	✓	80.06	81.29
Overlap coefficient	✓	100	x	85.71	✓	100	x	83.33	92.26
Fuzz.token_set_ratio	✓	100	✓	86.00	✓	100	✓	80.00	91.50
Jaro-Winkler	x	61.76	x	91.42	✓	100	✓	92.05	86.30
Monge-Elkan	x	1.39	x	3.57	✓	100	x	2.96	26.98

 https://doi.org/10.1633/JISTaP.2024.12.3.4

2019) by applying the ontology concept, data storage, and queries for information. However, this study proposes the ontology concept to store the database structures of different systems and share this information within the research framework, to promote suitable class matching and factor dictionary creation.

From the research question, "How can we decrease the complexity between non-technical user skills and database structure understanding using the ontology concept?", when non-technical users use BI software for data visualization, they first must understand the data structure in the table form of a database. These users may be managers who need help understanding database concepts. In this situation, IT staff or data scientists become the responders. The decision-maker has no choice but to wait for the information from the responder. Therefore, the organization's decision must take time (Alpar & Schulz, 2016; Lennerholt et al., 2018). The results of this study present the processes that go through database structure understanding before non-technical users work on the BI software. DSMO is the model used in the relationship learning process between the databases of various information systems. When the learning process is finished, we will save DSMO by renaming it as DSO. Hence, DSO is the instance of DSMO and uses it to share knowledge about the relationships of all databases in the information systems. DSMO and DSO will decrease the complexity between non-technical user skills and database structure understanding, because these users do not need to analyze the relationships of all databases. They can go through this complexity, and faster, to make decisions.

Many research studies have demonstrated the capacity for application systems to share information between systems through the concept of ontologies (Özcan et al., 2021; Walzel et al., 2019). Our research results, focusing on DSO, corroborate this assertion, as DSO serves as a center of knowledge regarding database relationships across multiple information systems. A significant finding is that the MapDS-Onto framework can be employed to match data across various information systems and provide appropriate data references to non-technical users. This framework utilizes DSO to familiarize itself with other systems in this environment and inform non-technical users about the most suitable data. As a result, these users do not need to understand how data is related within the database system.

Underpinning this framework are two pivotal algorithms, DaSMap and DoSMap. DaSMap algorithm employs DSMO and adapts the schema-based mapping to the RDF graph, as proposed by Zhao et al. (2019), thereby formulating a novel method for transforming database schema to the RDF. Central to the DoSMap algorithm is the method of string similarity comparison. The DoSMap algorithm reveals that token-based similarity methods, such as the "Fuzz.token_set_ratio" method, are more suitable than hybrid similarity methods. This finding aligns with the findings of Appa Rao et al. (2018), which posited that token-based similarity methods were more suitable than character-based methods for matching mathematical formulas. These results complement the research of Appa Rao et al. (2018), given that their study did not draw comparisons between the accuracy scores of token-based and hybrid similarity methods.

The three key findings of this study include: (1) DSMO, which serves as the model for creating DSO, a repository of knowledge about data relationships across multiple databases; (2) DaSMap and DoSMap, algorithms programmed within the MapDS-Onto framework; and (3) the "Fuzz.token_set_ratio" method, suitable for string similarity searching in this domain.

DSMO is a simple model for learning data relationships between the databases in the case study. An organization can reuse DSMO as a tool for creating data relationship knowledge and sharing this knowledge as a plugin on existing business software. The case study results guarantee the benefit of DSMO: creating automation through a neural framework. DSMO is converted to DSO in order to create a knowledge bank of data relationships. DSO is the knowledge of the data relationship that will be shared within the system. Many algorithms, as in Figs. 3-6, present the programming process of data relationship learning and matching. The two main uses of this algorithm are DaSMap and DoSMap. The BI software developer can modify DaSMap to learn the data relationships in the different systems and adapt the DoSMap to match the suitable data in the search process. Furthermore, the "Fuzz.token_set_ratio" is the verified method that fits the string similarity comparison between the factor of the KPI formula and the data in database systems.

The benefits of this study from the point of view of non-technical users will happen after their existing BI system is improved. The many complexities of BI regarding database understanding will be replaced with the automatic matching process. The improved BI will be more user-friendly than in the past.

4.3. Limitations of This Study

1. The result correction of the MapDS-Onto frame-

work has been tested solely with databases from two information systems. Consequently, the complexity of the data across all information systems within the organization may influence the accuracy of this framework.

2. The MapDS-Onto framework has been designed based on the RDB system. Nonetheless, consideration should be given to other database formats, such as NoSQL, Excel, and text files. In the future, this framework is anticipated to develop a plug-in or an extension to accommodate these alternative database formats.

5. CONCLUSIONS

This research presented a new process to reduce obstacles related to the mismatch between the complexity of BI systems and skills and database structure understanding of non-technical BI users, using the ontology technique. The ontology concept was applied to create a DSO where the ontology shared knowledge about the relationships between columns and tables in different information systems. All of the results from the case study indicated that the MapDS-Onto framework, DaSMap algorithm, and DoSMap algorithm are a practical way to build ease of use into BI systems and mitigate problems related to complexity, because users simply have to add the properties of the KPIs of the strategic information they wish to monitor, and their related formulas, into this framework. The MapDS-Onto framework also addresses situations where multiple databases have the same column names using the "SelectedClassDictionary" function. This function checks the consistency between the KPI's domain and database descriptions; therefore, it is ensured that the selected column will fit with the database related to the KPI's domain. The MapDS-Onto framework uses the "fuzz.token_set_ratio" method to search out suitable columns of data from multiple databases to use in the monitoring process of semantic BI systems.

ACKNOWLEDGEMENT

This work was supported by the Digital Science for Economy, Society, Human Resources Innovative Development and Environment project, funded by Reinventing Universities & Research Institutes under grant no. 2046735, Ministry of Higher Education, Science, Research and Innovation, Thailand.

CONFLICTS OF INTEREST

No potential conflict of interest relevant to this article was reported.

REFERENCES

Alaskar, H. F., & Saba, T. (2020, November 3-5). *Application of business intelligence solution development and implementation in a small-sized enterprise.* Paper presented at 2020 First International Conference of Smart Systems and Emerging Technologies (SMARTTECH), Riyadh, Saudi Arabia.

Alpar, P., & Schulz, M. (2016). Self-service business intelligence. *Business & Information Systems Engineering,* 58(2), 151-155. https://doi.org/10.1007/s12599-016-0424-6

Angeles, M. d. P., & Espino-Gamez, A. (2015). *Comparison of methods Hamming distance, Jaro, and Monge-Elkan.* Paper presented at International Conference on Advances in Databases, Knowledge, and Data Applications, Rome, Italy.

Appa Rao, G., Srinivas, G., Venkata Rao, K., & Prasad Reddy, P. V. G. D. (2018). A partial ratio and ratio based Fuzzy-Wuzzy procedure for characteristic mining of mathematical formulas from documents. *ICTACT Journal on Soft Computing,* 8(4), 1728-1732. https://doi.org/10.21917/ijsc.2018.0242

Azma, F., & Mostafapour, M. A. (2012). Business intelligence as a key strategy for development organizations. *Procedia Technology,* 1, 102-106. https://doi.org/10.1016/j.protcy.2012.02.020

Azzouz, M., Boukhedouma, S., & Alimazghi, Z. (2020, October 17-18). *Impact of strategic alignment on company performance: An approach based on performance indicators system design.* Paper presented at 2020 International Conference on Advanced Computer Science and Information Systems (ICACSIS), Depok, Indonesia.

Bai, S., Rajput, Q., Hussain, S., & Khoja, S. A. (2014, November 10-13). *Faculty performance evaluation system: An ontological approach.* Paper presented at 2014 IEEE/ACS 11th International Conference on Computer Systems and Applications (AICCSA), Doha, Qatar.

Božič, K., & Dimovski, V. (2019). Business intelligence and analytics use, innovation ambidexterity, and firm performance: A dynamic capabilities perspective. *The Journal of Strategic Information Systems,* 28(4), 101578. https://doi.org/10.1016/j.jsis.2019.101578

Bréant, C., Succi, L., Cotten, M., Grimaud, S., Iavindrasana, J., Kindstrand, M., Mauvais, F., & Rorive-Feytmans, B. (2020). Tools to measure, monitor, and analyse the performance of

the Geneva university hospitals (HUG). *Supply Chain Forum: An International Journal,* 21(2), 117-131. https://doi.org/10.1080/16258312.2020.1780634

de Andrade, P. R. M., & Sadaoui, S. (2017, October 5-8). *Improving business decision making based on KPI management system.* Paper presented at 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Banff, AB, Canada.

del Mar Roldán-García, M., García-Nieto, J., Maté, A., Trujillo, J., & Aldana-Montes, J. F. (2021). Ontology-driven approach for KPI meta-modelling, selection and reasoning. *International Journal of Information Management,* 58, 102018. https://doi.org/10.1016/j.ijinfomgt.2019.10.003

Devi, R., Mehrotra, D., & Baazaoui-Zghal, H. (2020, June 4-5). *An R2RML-based approach to map Dengue patient database to ontology.* Paper presented at 2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), Noida, India.

Fraihat, S., Salameh, W. A., Elhassan, A., Tahoun, B. A., & Asasfeh, M. (2021). Business intelligence framework design and implementation: A real-estate market case study. *Journal of Data and Information Quality,* 13(2), Article 10. https://doi.org/10.1145/3422669

Gruber, T. (2009). Ontology. In L. Liu & M. T. ÖZsu (Eds.), *Encyclopedia of Database Systems* (pp. 1963-1965). Springer US.

Hebeler, J., Fisher, M., Blace, R., Perez-Lopez, A., & Dean, M. (2009). *Semantic web programming*. Wiley.

Jakhar, R., & Krishna, C. (2020). Business intelligence: As a strategic tool for organization development (a literature review). *ANWESH: International Journal of Management & Information Technology*, 5(1), 44-46.

Jun, H. G., Im, D. H., & Kim, H. J. (2020). Semantics-preserving optimisation of mapping multi-column key constraints for RDB to RDF transformation. *Journal of Information Science,* 0165551520920804. https://doi.org/10.1177/0165551520920804

Khatibi, V., Keramati, A., & Montazer, G. A. (2017). A business intelligence approach to monitoring and trend analysis of national R&D indicators. *Engineering Management Journal,* 29(4), 244-257. https://doi.org/10.1080/10429247.2017.1380578

Klašnja-Milićević, A., Ivanović, M., & Budimac, Z. (2017). Data science in education: Big data and learning analytics. *Computer Applications in Engineering Education,* 25(6), 1066-1078. https://doi.org/10.1002/cae.21844

Kumar, S. M., & Belwal, M. (2017, August 17-19). *Performance dashboard: Cutting-edge business intelligence and data visualization.* Paper presented at 2017 International Conference On Smart Technologies For Smart Nation (SmartTechCon), Bangalore, India.

Lennerholt, C., Laere, J., & Söderström, E. (2018, January 3-6). *Implementation challenges of self service business intelligence: A literature review*. Paper presented at 51st Hawaii International Conference on System Sciences, Waikoloa Village, HI, USA.

Martinez-Cruz, C., Blanco, I. J., & Vila, M. A. (2012). Ontologies versus relational databases: Are they so different? A comparison. *Artificial Intelligence Review,* 38(4), 271-290. https://doi.org/10.1007/s10462-011-9251-9

Maté, A., Trujillo, J., & Mylopoulos, J. (2017). Specification and derivation of key performance indicators for business analytics: A semantic approach. *Data & Knowledge Engineering,* 108, 30-49. https://doi.org/10.1016/j.datak.2016.12.004

Ng, K. K. H., Tang, M. H. M., & Lee, C. K. M. (2015, December 6-9). *Design and development of a performance evaluation system for the aircraft maintenance industry.* Paper presented at 2015 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM), Singapore.

Noy, N., & McGuinness, D. (2001). *Ontology development 101: A guide to creating your first ontology*. https://www.cs.upc.edu/~jvazquez/teaching/sma-upc/docs/ontology101.pdf

Olszak, C. (2020). *Business intelligence and big data : Drivers of organizational success*. CRC Press, Taylor & Francis Group.

Özcan, F., Lei, C., Quamar, A., & Efthymiou, V. (2021, June 20-25). *Semantic enrichment of data for AI applications*. Paper presented at Proceedings of the Fifth Workshop on Data Management for End-To-End Machine Learning, New York, NY, USA.

Pestana, M., Pereira, R., & Moro, S. (2020). Improving health care management in hospitals through a productivity dashboard. *Journal of Medical Systems,* 44(4), 87. https://doi.org/10.1007/s10916-020-01546-1

Petrini, M., & Pozzebon, M. (2009). Managing sustainability with the support of business intelligence: Integrating socio-environmental indicators and organisational context. *The Journal of Strategic Information Systems,* 18(4), 178-191. https://doi.org/10.1016/j.jsis.2009.06.001

Piri, Z. (2020). Information visualization to support the decision-making process in the context of academic management. *Webology,* 17, 216-226. https://doi.org/10.14704/WEB/V17I1/a218

Popovič, A., Hackney, R., Coelho, P. S., & Jaklič, J. (2014). How information-sharing values influence the use of information systems: An investigation in the business intelligence systems context. *The Journal of Strategic Information Systems,*

23(4), 270-283. https://doi.org/10.1016/j.jsis.2014.08.003

Prasetya, D. D., Prasetya Wibawa, A., & Hirashima, T. (2018). The performance of text similarity algorithms. *International Journal of Advances in Intelligent Informatics,* 4(1), 63-69. https://doi.org/10.26555/ijain.v4i1.152

Quacquarelli Symonds. (2021). *QS World University Rankings 2021*. https://www.topuniversities.com/world-university-rankings/2021

Qushem, U. B., Zeki, A. M., & Abubakar, A. (2017). Successful business intelligence system for SME: An analytical study in Malaysia. *IOP Conference Series: Materials Science and Engineering,* 226(1), 012090. https://doi.org/10.1088/1757-899X/226/1/012090

Ravichandran, T. (2018). Exploring the relationships between IT competence, innovation capacity and organizational agility. *The Journal of Strategic Information Systems,* 27(1), 22-42. https://doi.org/10.1016/j.jsis.2017.07.002

Ren, Q., & Tao, J. (2012, August 23-25). *KPI corporate management and business intelligence analysis on the application of electric power enterprises.* Paper presented at 2012 International Conference on Industrial Control and Electronics Engineering, Xi'an, China.

Rinartha, K., Suryasa, W., & Kartika, L. G. S. (2018, October 9-11). *Comparative analysis of string similarity on dynamic query suggestions.* Paper presented at 2018 Electrical Power, Electronics, Communications, Controls and Informatics Seminar (EECCIS), Batu, Indonesia.

Rocha, S., Bernardino, J., Pedrosa, I., & Ferreira, I. (2017). Dashboards and indicators for a BI healthcare system. In: Á. Rocha, A. Correia, H. Adeli, L. Reis, & S. Costanzo (eds.), *Recent advances in information systems and technologies* (pp. 81-90). Springer.

Sharda, R., Delen, D., & Turban, E. (2018). *Business intelligence, analytics, and data science: A managerial perspective*. Pearson.

Sir, M., Bradac, Z., & Fiedler, P. (2015). Ontology versus database. *IFAC-PapersOnLine,* 48(4), 220-225. https://doi.org/10.1016/j.ifacol.2015.07.036

Sultan, N., Khedr, A., Idrees, A., & Kholeif, S. (2017). Data mining approach for detecting key performance indicators. *Journal of Artificial Intelligence,* 10, 59-65. https://doi.org/10.3923/jai.2017.59.65

Ul-Ain, N., Vaia, G., & Delone, W. (2019, January 8-11). *Business intelligence system adoption, utilization and success - A systematic literature review*. Paper presented at 52nd Hawaii International Conference on System Sciences, Wailea, HI, USA.

Villegas-Ch, W., Palacios-Pacheco, X., & Luján-Mora, S. (2020). A business intelligence framework for analyzing educational data. *Sustainability,* 12(14), 5745. https://doi.org/10.3390/su12145745

Visser, W. F. (2020). A blueprint for performance-driven operations management. *Mining, Metallurgy & Exploration,* 37(3), 823-831. https://doi.org/10.1007/s42461-020-00199-5

Walzel, H., Vathoopan, M., Zoitl, A., & Knoll, A. (2019, September 10-13). *An approach for an automated adaption of KPI ontologies by reusing systems engineering data.* Paper presented at 2019 24th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA), Zaragoza, Spain.

Yu, M., Li, G., Deng, D., & Feng, J. (2016). String similarity search and join: A survey. *Frontiers of Computer Science,* 10(3), 399-417. https://doi.org/10.1007/s11704-015-5900-5

Yulianto, A. A., & Kasahara, Y. (2018, July 8-13). *Implementation of business intelligence with improved data-driven decision-making approach.* Paper presented at 2018 7th International Congress on Advanced Applied Informatics (IIAI-AAI), Yonago, Japan.

Zhao, Z., Han, S., & Kim, J. (2019). R2LD: Schema-based graph mapping of relational databases to linked open data for multimedia resources data. *Multimedia Tools and Applications,* 78(20), 28835-28851. https://doi.org/10.1007/s11042-019-7281-5

Zoumpatianos, K., Palpanas, T., Mylopoulos, J., Maté, A., & Trujillo, J. (2013). *Monitoring and diagnosing indicators for business analytics*. Paper presented at Proceedings of the 2013 Conference of the Center for Advanced Studies on Collaborative Research, Ontario, Canada.

JISTaP http://www.jistap.org
Journal of Information Science Theory and Practice

RESEARCH PAPER
J Inf Sci Theory Pract 12(3): 63-74, 2024

eISSN : 2287-4577 pISSN : 2287-9099

https://doi.org/10.1633/JISTaP.2024.12.3.5

A Digital Thesaurus of the Traditional Common Culture of the Greater Mekong Subregion

Suwannee Hoaihongthong
Department of Information Science, Faculty of Humanities and Social Sciences, Khon Kaen University, Khon Kaen, Thailand
E-mail: suwaho@kku.ac.th

Kanyarat Kwiecien*
Department of Information Science, Faculty of Humanities and Social Sciences, Khon Kaen University, Khon Kaen, Thailand
E-mail: kandad@kku.ac.th

ABSTRACT

This study aimed to develop a digital thesaurus dedicated to cataloging the traditional common culture of the Greater Mekong Subregion. The process followed a meticulous seven-step methodology, including scoping, vocabulary collection, knowledge structure analysis, relationship delineation, related word adjustments, list validation, and evaluation. Leveraging principles from knowledge organization, thesaurus construction, and digital platform development, the TemaTres web application emerged as the primary tool for constructing this thesaurus. The study's results showed that 2,042 principal words related to the traditional common culture of the Greater Mekong Subregion were compiled and classified into terms for each of the seven deep levels. Each term was accompanied by essential metadata, including broader and narrower terms, related terms, cross-references, and scope notes. This rich dataset empowered semantic search capabilities across diverse applications and web services, providing access to knowledge pertaining to the traditional common culture of the Greater Mekong Subregion and contributing to a deeper understanding of this cultural domain.

Keywords: thesaurus, digital thesaurus, traditional common culture, Greater Mekong Subregion, information retrieval, knowledge organization

Received: February 10, 2024 **Revised:** May 11, 2024
Accepted: May 18, 2024 **Published:** September 30, 2024

***Corresponding Author:** Kanyarat Kwiecien
https://orcid.org/0000-0002-1904-5327
E-mail: kandad@kku.ac.th

All JISTaP content is Open Access, meaning it is accessible online to everyone, without fee and authors' permission. Open Access articles are automatically archived in the Korea Institute of Science and Technology Information (KISTI)'s Open Access repository (AccessON). All JISTaP content is published and distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/). Under this license, the authors retain full ownership of their work, while permitting anyone to use, distribute, and reproduce the content in any medium, as long as the original authors and source are cited. For any reuse, redistribution, or reproduction of a work, users must clarify the license terms under which the work was produced.

© 2024 [Suwannee Hoaihongthong, Kanyarat Kwiecien] This article is distributed under the Creative Commons Attribution License (CC BY), allowing unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. It is published by the Korea Institute of Science and Technology Information (KISTI).

1. INTRODUCTION

In the countries of the Mekong Subregion, notably Thailand, Laos, and Cambodia, significant cultural affinities are evident, originating from geographical, cultural, religious, and lifestyle congruities. These shared foundations have evolved through continuous cultural interactions driven by visits, trade, and migration, leading to the assimilation of diverse traditional and contemporary cultural traditions (Greater Mekong Subregion Secretariat, 2013). Linguistic commonality is a notable illustration of this shared culture, where historically influenced alphabets and words, particularly Sanskrit, facilitate communication and mutual comprehension among the populations. Additionally, performing arts and musical traditions exhibit remarkable resemblances, seen in dance performances and musical instruments influenced by each other, adapted to harmonize with each nation's cultural heritage. Furthermore, there is convergence in observing various customs, beliefs, and rituals across these nations, exemplified by traditions such as Songkran and the Twelve Months Tradition, known as *Heet Sipsong Klong Sib Si*, observed comparably across all three countries (Prasarn et al., 2015). These shared cultural commonalities serve as powerful tools for promoting unity, mutual understanding, and long-lasting collaboration across various domains, establishing a solid foundation for strong and sustainable partnerships in the future.

In this study, we utilize the term *common culture* to describe a collective cultural foundation characterized by shared elements that transcend individual ownership, aligning with the conceptual framework outlined by Bertacchini et al. (2012). Within this framework, our analysis identifies recurring components such as shared historical narratives, common values, principles, objectives, and a deep understanding of specific traditions and practices, as detailed by McLean (2015).

When analyzing the cultural information systems in the mentioned countries, it is evident that they primarily function as general search tools, albeit with limitations. They mainly employ a character comparison technique called "string matching" for information retrieval, often overlooking nuanced semantics and conceptual intricacies associated with search terms. Consequently, this approach yields unsatisfactory results for specific search queries, particularly concerning traditional terms with multiple spellings. This semantic disparity, often referred to as the "semantic gap" in Hein (2010)'s work, underscores the importance of implementing knowledge organization (KO) strategies to address this challenge effectively.

In the domain of KO, the thesaurus stands as a pivotal instrument, playing a central role in the organization, categorization, and structuring of information and knowledge, thereby enhancing accessibility and comprehensibility for users (Bergamaschi et al., 1999). Thesauri establish meaningful relationships between terms and concepts, facilitating efficient information retrieval and enabling users to navigate complex knowledge domains through mechanisms such as broader terms (BT), narrower terms (NT), and related terms (RT). Moreover, the development of specialized vocabularies for diverse domains, including information science and information services, follows a structured and systematic process encompassing vocabulary selection, analysis, synonym identification, establishment of word relationships, refinement, scrutiny, and integration of expert input (Ahmad et al., 2020).

Constructing a thesaurus for the subject of the traditional common culture of the Greater Mekong Subregion (GMS) is paramount for researchers, scholars, educators, policymakers, and cultural enthusiasts. First, such a thesaurus facilitates efficient KO and retrieval by providing a structured framework for categorizing and indexing terms related to cultural practices, traditions, and customs within the region. Second, it enhances accessibility to cultural heritage resources by enabling users to navigate through a comprehensive vocabulary that encompasses various aspects of the traditional culture of the Mekong Subregion countries. Lastly, the thesaurus serves as a valuable tool for researchers, educators, students, and individuals interested in studying or preserving the cultural heritage of the region, offering a standardized terminology for communication and research purposes.

Furthermore, the research encompasses the creation of specialized vocabularies tailored to specific cultural groups, as exemplified by Arayapant (2019)'s endeavors in analyzing the vocabulary structure of the Tai ethnic group and formulating terminology related to their cultural knowledge, thus demonstrating the versatile application of KO principles across various domains and disciplines. Additionally, the organizational structure inherent in the thesaurus format holds potential for further development into an ontology, as explored by Li and Li (2013). Therefore, this study not only involved the organization of terminologies about the traditional common culture of the GMS, but also underscores the thesaurus format's utility as a resource for exploring the relationships within cultural heritage content. It serves as a tool for accessing and retrieving knowledge about cultural heritage within

 https://doi.org/10.1633/JISTaP.2024.12.3.5

databases and on the Internet. Moreover, the findings obtained pave the way for further development into an ontology and semantic search system.

2. RESEARCH OBJECTIVES

This research aimed to analyze the vocabulary related to the common culture and traditions of the Mekong Subregion countries, determine the structure, and develop a thesaurus to create a digital thesaurus of the traditional common culture of the GMS.

3. METHOD

In this research, thesaurus development follows a seven-step structured process, applying the thesaurus construction guidelines, which provide an important and valuable guide to content discovery, organization, and retrieval, encompassing common activities across all fields, including cultural heritage (Autiero et al., 2023; Ryan, 2014) (Table 1).

Table 1 provides an overview of the structured seven-step process involved in the development of the thesaurus for cataloging the traditional common culture of the GMS.

4. RESULTS

The structure of common cultural knowledge categories regarding traditions of the Mekong Subregion countries was developed according to the concept of characteristics of a common culture through content analysis and the application of KO principles. The results of developing the knowledge structure of the common culture and traditions of the Mekong Subregion countries are presented in the form of relative vocabulary. The research findings are as follows.

Table 1. Seven-step structured process for thesaurus development

Step	Description
3.1 Scope determination	Define the scope of information resources related to the traditional common culture of the Greater Mekong Subregion, outlining project boundaries and objectives
3.2 Vocabulary collection	Gather an extensive range of vocabulary from diverse documents and information sources associated with the traditional common culture of the Greater Mekong Subregion, ensuring thorough coverage of relevant terms and concepts
3.3 Knowledge structure analysis	Analyze collected vocabulary to determine the knowledge structure of relative terms and group words pertinent to the traditional common culture of the Greater Mekong Subregion, ensuring coherence and relevance in thesaurus organization
3.4 Relationship specification	Specify relationships between words within the relative vocabulary to accurately reflect semantic connections between concepts
3.4.1 Hierarchy development	Establish a hierarchical structure for the thesaurus, organizing terms into broader (BT), narrower (NT), and related (RT) relationships to reflect the conceptual hierarchy
3.4.2 Relationships and hierarchical linking	Define and document relationships between terms, creating hierarchical links to accurately represent the relationships between concepts
3.4.3 Synonyms and nonpreferred terms	Identify synonyms and establish relationships between them, providing cross-references to preferred terms and specifying nonpreferred terms to enhance clarity and understanding
3.4.4 Scope notes and definitions	Write scope notes and definitions for terms, especially those with multiple meanings, ensuring clear and precise definitions for users
3.5 Adjustment of related words	Refine relationships and ensure overall coherence in the addition of related words, enhancing the effectiveness of the thesaurus as a knowledge organization tool
3.6 List validation	Validate correctness, completeness, and consistency between related words, arranging each word group alphabetically for ease of use
3.7 Evaluation and preparation	Conduct comprehensive evaluation of the traditional common culture of the Greater Mekong Subregion, integrating expert feedback to augment the coherence of relative terms. Meticulously prepare the thesaurus for publication in a format aligned with accessibility goals, ensuring it meets the needs of its intended user base

4.1. Knowledge Structure of Common Culture in the Mekong Subregion

The outcomes of the knowledge structure development encompass a comprehensive framework consisting of 13 distinct categories of knowledge (Table 2). Table 2 lists various categories encompassing knowledge related to the traditional common culture of the GMS. These categories serve as key components for organizing and cataloging information about the rich cultural heritage present in the Mekong Subregion.

The subsequent phase entails the creation of a category structure, presented either as a category display or categorized framework. This process aligns with the principles of classification and categorization elucidated by Rowley (1992). This structured framework delineates an organized dataset, revealing the hierarchical relationships among various groups within the system. This meticulously designed structure encompasses 13 overarching categories, further segmented into 37 subcategories, and intricately branching into 43 distinct subgroups (Table 3). By adopting this hierarchical arrangement, users will benefit from a coherent and ordered presentation, affording them a systematic means to explore and navigate the extensive reservoir of information encapsulated within the traditional common culture of the GMS.

The knowledge framework was subsequently transformed into a thesaurus format, structured around conceptual interconnections and word relationships. These relationships encompass hierarchical associations and coequal linkages, as well as mutually interconnected or RT. To elucidate these relationships, the thesaurus employs symbols such as BT to signify higher-level concepts, NT to denote subordinate concepts, and RT to indicate associated terms (Fig. 1).

In pursuit of enhancing the quality of the grouped vocabulary, the researcher meticulously refined it based on valuable input and recommendations provided by domain experts. Subsequently, efforts were directed towards

Table 2. Categories of traditional common culture in the Mekong Subregion

Category	Description
Ethnic group	"Ethnic Groups Associated with Traditions" encompasses knowledge pertaining to the ethnic communities residing in the Mekong Subregion, where traditional practices hold significance
Activities	"Activity" encompasses knowledge relating to activities undertaken by individuals during traditional occasions, characterized by their solitary nature, and may lack predefined plans or regulatory frameworks
Belief	"Beliefs Manifesting in Traditions" encompasses knowledge concerning the beliefs embedded within traditional practices and the rituals that transpire during these traditions
History	"Historical Aspects of Traditions" delves into knowledge pertaining to the origins and historical underpinnings of these traditions
Values	"Values Evident in Traditions" explores the content surrounding the values and concepts that either precipitate traditional observances or are discernible outcomes of these practices
Time	"Timing of Traditional Events" encompasses knowledge tied to the scheduling and temporal aspects governing the execution of traditional ceremonies
Person	"Individuals Engaged in Traditions" covers knowledge pertaining to the individuals involved in these traditions, delineating their roles and responsibilities within the context of the rituals
Ritual	"Rituals" delves into the knowledge encompassing the rituals and observances that serve as exemplars or diverse patterns within the sphere of traditions
Traditional literatures	"Traditional Literature" pertains to knowledge concerning literary works interconnected with tradition, encompassing a wide spectrum of literary forms, including both prose and poetry
Purpose	"Purpose of Organizing Traditions" encompasses knowledge pertaining to the objectives and intentions behind the organization of these traditions, elucidating the desired outcomes or expectations
Place	"Traditional Venues" offers insights into the knowledge of locations and spaces designated for the preparation and execution of traditional events
Principle	"Principles" encapsulates crucial guidelines and principles, specifically those derived from Buddhist teachings and associated with traditional practices
Equipment	"Equipment Employed in Traditions" involves knowledge concerning the various tools and apparatus utilized in diverse rituals conducted during traditional events, including information about usage methods

 https://doi.org/10.1633/JISTaP.2024.12.3.5

Table 3. Shared cultural knowledge structure of traditional common culture within the Greater Mekong Subregion

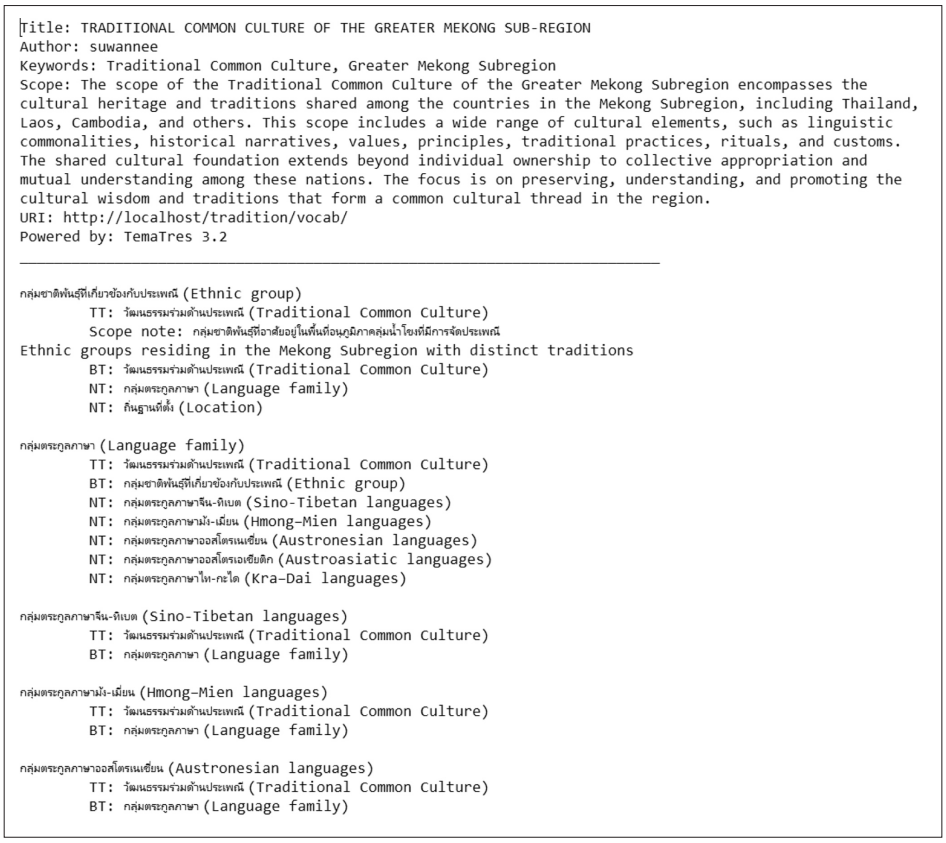
Category	Subcategory	Subgroup
1. Ethnic groups	1.1 Language families	1.1.1 Sino-Tibetan languages
		1.1.2 Hmong-Mien languages
		1.1.3 Austronesian languages
		1.1.4 Austroasiatic languages
		1.1.5 Kra-Dai languages
	1.2 Locations	1.2.1 Countries
		1.2.2 Cities
		1.2.3 Geographic coordinates
2. Activities	2.1 Belief activities	
	2.2 Religious activities	
	2.3 Carnival activities	
	2.4 Family activities	
3. Belief	3.1 Beliefs related to Buddhism	3.1.1 Laws of karma
		3.1.2 Ordination
		3.1.3 Gratitude
		3.1.4 Hell and heaven
		3.1.5 Offense
	3.2 Beliefs related to superstition	3.2.1 Belief in Naga
		3.2.1.1 Characteristics of Naga
		3.2.1.2 Power of Naga
		3.2.1.3 Naga worship
		3.2.2 Belief in animism
		3.2.2.1 Types of animism
		3.2.2.2 Characteristics of animism
		3.2.2.3 Animism worship
		3.2.3 Belief in Devas
		3.2.3.1 Devas in Buddhism
		3.2.3.2 Mythical angels
4. History	4.1 Cause of religious beliefs	
	4.2 Cause of superstitions	
	4.3 Cause of geographic environment	
5. Values	5.1 Forgiveness	
	5.2 Gratitude	
	5.3 Unity	
	5.4 Perseverance	
	5.5 Sacrifice	
6. Time	6.1 Calendar	
	6.2 Lunar time	

Table 3. Continued

Category	Subcategory	Subgroup
7. Person	7.1 Ceremony leaders	
	7.2 Participants	
8. Rituals	8.1 Traditional rituals	
	8.2 Religious rituals	
9. Traditional literatures	9.1 Oral literature	9.1.1 Legends
		9.1.2 Tales
		9.1.3 Phaya (folk tales)
	9.2 Written literature	9.2.1 Manuscripts
		9.2.2 Annals
10. Purpose	10.1 Livelihood purpose	
	10.2 Morale for living purpose	
	10.3 Gratitude purpose	
	10.4 Buddhism purpose	
11. Places	11.1 Public spaces	
	11.2 Domestic spaces	
12. Principles	12.1 People principles	
	12.2 Monks' principles	
13. Equipment	13.1 Foods	13.1.1 Types of food
		13.1.1.1 Dishes
		13.1.1.2 Desserts
		13.1.1.3 Beverages
		13.1.2 Ingredients
		13.1.3 Cooking methods
	13.2 Dharma offerings	
	13.3 Decorations	
	13.4 Tools and utensils	13.4.1 Household appliances
		13.4.2 Kitchenware
		13.4.3 Bedding
		13.4.4 Occupational tools
		13.4.4.1 Agricultural tools
		13.4.4.2 Animal Husbandry equipment
		13.4.5 Furniture
		13.4.6 Special purpose tools
	13.5 Monks' utensils	

the digitization and development of this vocabulary into an electronic format. This transformation was facilitated through the utilization of the TemaTres program—a versatile open-source software designed to assist in the creation of web-based vocabulary management systems. TemaTres is adept at accommodating both general and relational vocabularies, rendering it an ideal tool for this purpose.

 https://doi.org/10.1633/JISTaP.2024.12.3.5



Title: TRADITIONAL COMMON CULTURE OF THE GREATER MEKONG SUB-REGION
Author: suwannee
Keywords: Traditional Common Culture, Greater Mekong Subregion
Scope: The scope of the Traditional Common Culture of the Greater Mekong Subregion encompasses the cultural heritage and traditions shared among the countries in the Mekong Subregion, including Thailand, Laos, Cambodia, and others. This scope includes a wide range of cultural elements, such as linguistic commonalities, historical narratives, values, principles, traditional practices, rituals, and customs. The shared cultural foundation extends beyond individual ownership to collective appropriation and mutual understanding among these nations. The focus is on preserving, understanding, and promoting the cultural wisdom and traditions that form a common cultural thread in the region.
URI: http://localhost/tradition/vocab/
Powered by: TemaTres 3.2

กลุ่มชาติพันธุ์ที่เกี่ยวข้องกับประเพณี (Ethnic group)
TT: วัฒนธรรมร่วมด้านประเพณี (Traditional Common Culture)
Scope note: กลุ่มชาติพันธุ์ที่อาศัยอยู่ในพื้นที่อนุภูมิภาคลุ่มน้ำโขงที่มีการจัดประเพณี
Ethnic groups residing in the Mekong Subregion with distinct traditions
BT: วัฒนธรรมร่วมด้านประเพณี (Traditional Common Culture)
NT: กลุ่มตระกูลภาษา (Language family)
NT: ถิ่นฐานที่ตั้ง (Location)

กลุ่มตระกูลภาษา (Language family)
TT: วัฒนธรรมร่วมด้านประเพณี (Traditional Common Culture)
BT: กลุ่มชาติพันธุ์ที่เกี่ยวข้องกับประเพณี (Ethnic group)
NT: กลุ่มตระกูลภาษาจีน-ทิเบต (Sino-Tibetan languages)
NT: กลุ่มตระกูลภาษาม้ง-เมี่ยน (Hmong-Mien languages)
NT: กลุ่มตระกูลภาษาออสโตรเนเซียน (Austronesian languages)
NT: กลุ่มตระกูลภาษาออสโตรเอเชียติก (Austroasiatic languages)
NT: กลุ่มตระกูลภาษาไท-กะได (Kra-Dai languages)

กลุ่มตระกูลภาษาจีน-ทิเบต (Sino-Tibetan languages)
TT: วัฒนธรรมร่วมด้านประเพณี (Traditional Common Culture)
BT: กลุ่มตระกูลภาษา (Language family)

กลุ่มตระกูลภาษาม้ง-เมี่ยน (Hmong-Mien languages)
TT: วัฒนธรรมร่วมด้านประเพณี (Traditional Common Culture)
BT: กลุ่มตระกูลภาษา (Language family)

กลุ่มตระกูลภาษาออสโตรเนเซียน (Austronesian languages)
TT: วัฒนธรรมร่วมด้านประเพณี (Traditional Common Culture)
BT: กลุ่มตระกูลภาษา (Language family)

Fig. 1. Example page of a report that exports data from an alphabetical list system.

4.2. Traditional Common Culture Digital Thesaurus

The TemaTres web-based vocabulary management system identified, compiled, and classified a total of 2,042 principal words related to the traditional common culture of the GMS into terms for each of the seven deep levels, with 1,380 found to have hierarchical relationships (Fig. 2).

The digital thesaurus of the traditional common culture of the GMS represents a significant advance in KO and accessibility within the GMS. This report provides an overview of the system as a whole, including details such as the number of words in the corpus, the number of relationships between the words in the corpus, and the depth of the vocabulary.

The digital thesaurus platform effectively managed the controlled vocabularies related to the traditional common culture of the GMS by storing both Thai and English vocabularies. Upon retrieval, the platform displays the vocabulary along with details of BT, NT, RT, cross-references, and scope notes (Figs. 3-4).

4.3. Evaluation of the Traditional Common Culture Digital Thesaurus

Efficiency testing employs the gold standard evaluation method for information search systems, which is assessed by graduate students and researchers. This method involves examining search results from selected documents or datasets categorized as either relevant or non-relevant. The evaluation process measures search efficiency using parameters such as precision (Precision), recall (Recall), and the overall effectiveness of the system (F-measure). The query selection of the term aims to demonstrate the system efficiency in terms of the stored data. In this research, 14 sets of vocabularies were sampled for retrieval from corpora, as shown in Table 4.

In Table 4, the performance evaluation of the traditional common culture digital thesaurus is demonstrated through precision, recall, and F-measure. Precision assesses the accuracy of retrieved results, calculated as the ratio of relevant documents retrieved to the total number of documents retrieved. An average precision of 0.94 indicates that, on average, 94% of retrieved items were relevant to the query categories. Recall measures the completeness of the retrieval process by determining the ratio of relevant documents retrieved to the total number of relevant documents in the collection. With an average recall of 0.88, the system successfully retrieved, on average, 88%

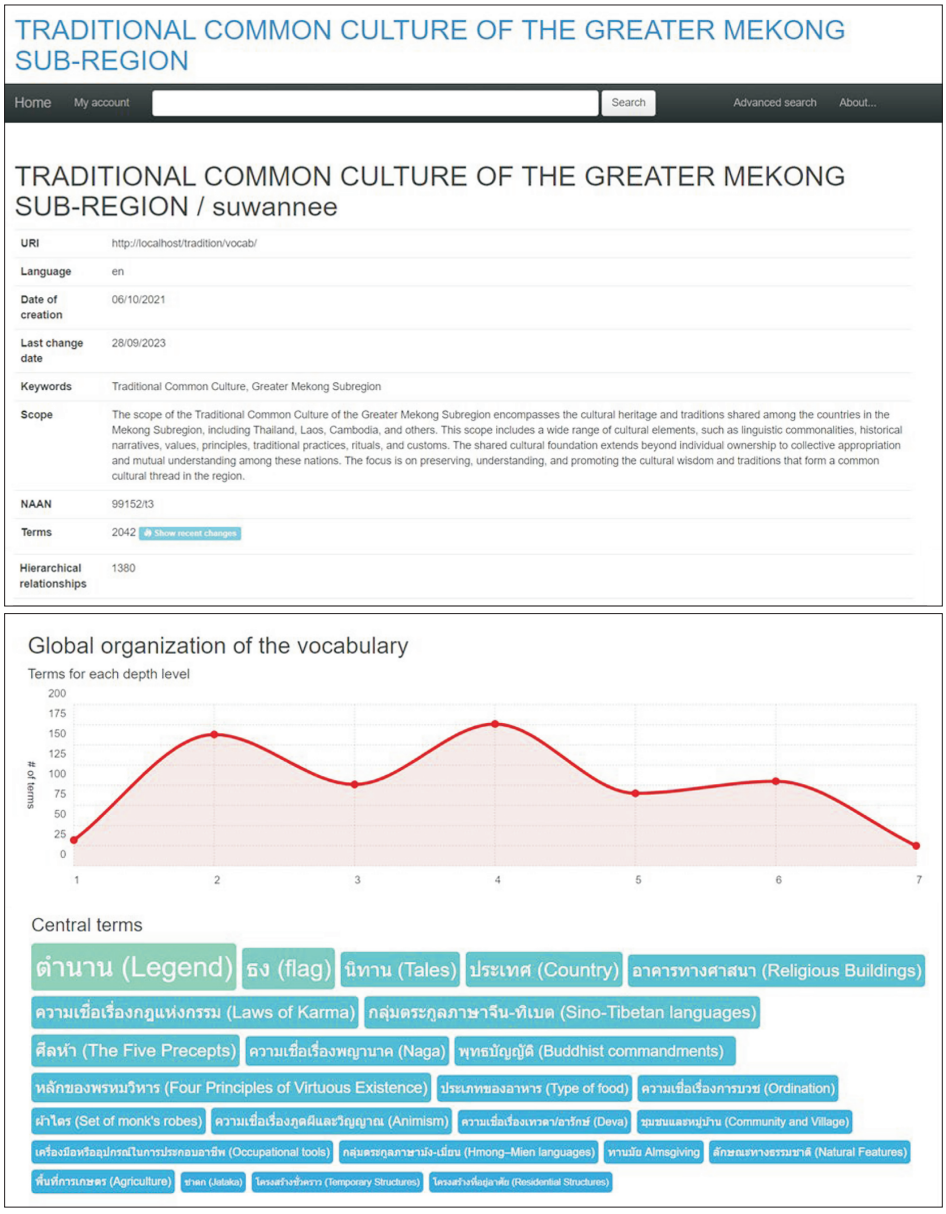


Fig. 2. Traditional common culture digital thesaurus – system overview report.

of relevant items in the collection. The F-measure, representing the harmonic mean of precision and recall, offers a balanced score reflecting both precision and recall. It is calculated using the formula:

$$F\text{-measure}=2\times(\text{precision}\times\text{recall})/(\text{precision}+\text{recall})$$

With an average F-measure of 0.90, the retrieval system demonstrates well-balanced performance across query categories, achieving a commendable balance between precision and recall.

5. DISCUSSION AND CONCLUSION

Throughout this study, our primary aim has been to develop a digital thesaurus of the traditional common culture of the GMS. By comprehensively understanding the common culture among the countries of the Mekong Subregion and meticulously analyzing and organizing cultural knowledge according to established frameworks, we have aimed to achieve this goal. Through our research, we have highlighted the importance of KO, with the thesaurus playing a central role in organizing, categorizing, and structuring information and knowledge, ultimately

 https://doi.org/10.1633/JISTaP.2024.12.3.5

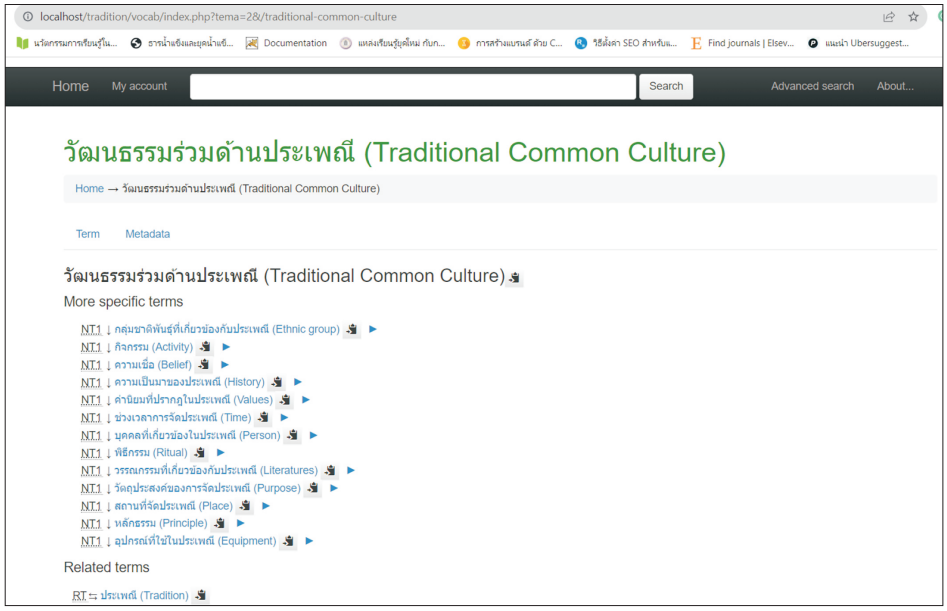


Fig. 3. Web access of the digital thesaurus of traditional common culture of the Greater Mekong Subregion.

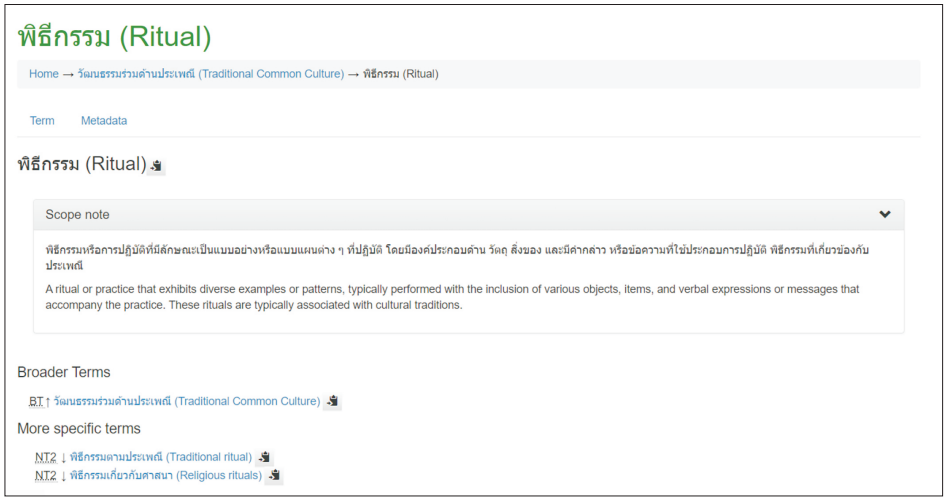


Fig. 4. Example of search results for the term "Ritual".

enhancing accessibility and comprehensibility for users.

Developing a knowledge structure on terminology related to cultures and traditions shared between 13 countries in the GMS requires a deeper delve into the classification of knowledge within cultural boundaries. This necessitates reference to the framework established by the Organization for Science Education and the United Nations Educational, Scientific and Cultural Organization (UNESCO), which divides cultures into seven distinct groups: language, performing arts, traditional crafts, folk literature, Thai wisdom, sports, social norms, rituals and festivals, as well as knowledge and practices related to nature and the universe. Additionally, it is imperative to consider the structural framework proposed by Iamkhajornchai and Manmart (2016) for the organization of Thai cultural knowledge. This framework classifies cultural knowledge into four groups, namely the cultural heritage knowledge group, art knowledge group, media knowledge group, and groups of creative work knowledge, based on the inherent nature of the work. It becomes evident that there exist disparities between the domains of knowledge covered by the relational terms in this research, which are exclusively tailored to the realm of knowledge linked with traditions. Should we contemplate the organization of this relative vocabulary within the framework of the aforementioned cultural groupings, it can be posited that common cultural traditions align with the category of social practices, ceremonies, and festivals according to

Table 4. Results of the performance evaluation of the traditional common culture digital thesaurus as evaluated by graduate students and researchers

Query category	Retrieved	Relevant retrieved	Relevant in the collection	Precision	Recall	F-measure
Rocket festival	2	1	2	0.50	0.50	0.50
Buddhist lent	3	3	3	1.00	1.00	1.00
End of Buddhist lent	3	2	3	0.67	0.67	0.67
Songkran	2	2	2	1.00	1.00	1.00
Alms giving	1	1	1	1.00	1.00	1.00
Makha Bucha	1	1	1	1.00	1.00	1.00
Mahachat	2	2	2	1.00	1.00	1.00
Rain	7	7	9	1.00	0.78	0.88
Boon Samha	1	1	1	1.00	1.00	1.00
Kathin	4	4	11	1.00	0.36	0.53
Monk	1	1	1	1.00	1.00	1.00
Phaya Kan Khak	1	1	1	1.00	1.00	1.00
Bun Pha Wet	1	1	1	1.00	1.00	1.00
Phra Malai	1	1	1	1.00	1.00	1.00
Average	-	-	-	0.94	0.88	0.90

UNESCO's conceptualization. Moreover, they fall within the cultural heritage knowledge group when contextualized within the framework for the systematization of Thai cultural knowledge.

The approach employed in this research involves the initial development of a relational lexicon, followed by the transformation of words from this lexicon into an ontology. This methodology effectively reduces the intricacies associated with ontology development (Huang et al., 2008), aligning with the findings of Li and Li (2013) in their study "On transformation from the thesaurus into domain ontology." Their study delves into the transition from lexical relations to ontology, elucidating how both lexical relations and ontology can be exploited to expound upon information related to meaning and knowledge.

Creating a lexicon serves as the foundational step for ontology development, offering significant time-saving advantages. Terms contained within the relational lexicon database can be seamlessly converted into ontology classes or concepts. For instance, the symbol "BT" can be equated to "Sub Class of," representing data relationships within a descending hierarchy. This approach stands as a valuable resource for researchers embarking on vocabulary development and the subsequent evolution into ontology in future research endeavors. Furthermore, the structuring of lexical knowledge in this research aligns with the concept of the Library of Congress Classification system, which is essentially a practical system. In this system, categories are assigned when a book is introduced into the library. If a particular category has no books associated with it, it will lack group numbers within that category (Chan et al., 2016; Mischo, 1982).

This approach implies that the analysis of vocabulary, aimed at developing the relational lexicon in this research, is an ongoing process. Vocabulary analysis is conducted based on information resources available at the present moment. However, it is important to note that if additional information resources related to common cultures emerge in the future or if there are new knowledge groups distinct from the existing ones, they can be seamlessly incorporated. This ensures that the relational vocabulary remains comprehensive, complete, and continually updated to reflect the evolving landscape of knowledge.

6. RECOMMENDATIONS

Navigating the realm of knowledge concerning the traditional common culture of the GMS demands effective tools and strategies, with relational vocabulary emerging as a cornerstone for facilitating efficient knowledge re-

 https://doi.org/10.1633/JISTaP.2024.12.3.5

trieval and exploration. As cultural understanding deepens, the need for precise terminology becomes increasingly apparent. Recommendations for its use are as follows:

1) Information repositories housing resources on the traditional common culture of the GMS play a crucial role in facilitating global knowledge access and cultural heritage preservation. Relational vocabulary acts as a vital tool in defining terms that encapsulate the essence of these resources. Moreover, professionals such as librarians, informaticists, and information specialists can utilize the inherent structure of relational terms to organize information resources effectively and develop robust search systems, thus enhancing accessibility for a global audience and contributing to the preservation of cultural heritage while enriching global understanding of the region's diverse heritage.

2) Researchers, academics, educators, students, and individuals worldwide who are interested in the common culture of the GMS can harness this specialized vocabulary to unlock valuable insights and promote cultural understanding. The comprehensive range of terms within the relational vocabulary provides users with an in-depth understanding of relevant concepts. Additionally, the interconnected nature of these terms facilitates further exploration across various information repositories, improving the efficiency of knowledge retrieval and exploration for a global audience and fostering cross-cultural understanding and appreciation. By recognizing the importance of interconnectedness, we can bridge cultural divides and promote mutual respect and understanding across diverse communities.

In conclusion, navigating the knowledge related to the traditional common culture of the GMS requires effective tools and strategies, with relational vocabulary emerging as a cornerstone for facilitating efficient knowledge retrieval and exploration. Recommendations for its use have been outlined, emphasizing the pivotal role of information sources and professionals in leveraging relational vocabulary to support knowledge retrieval, and demonstrating the importance of harnessing this specialized vocabulary for accessing and retrieving knowledge. By implementing these recommendations, stakeholders can enhance their understanding and exploration of the cultural heritage of the GMS, contributing to its preservation and dissemination for generations to come.

CONFLICTS OF INTEREST

No potential conflict of interest relevant to this article was reported.

REFERENCES

Ahmad, A., Justo, J. L. B., Feng, C., & Khan, A. A. (2020). The impact of controlled vocabularies on requirements engineering activities: A systematic mapping study. *Applied Sciences*, 10(21), 7749. https://doi.org/10.3390/app10217749

Arayapant, W. (2019). Thesaurus of Lanna Tai ethnic groups. *TLA Research Journal*, 12(2), 95-114.

Autiero, S., Elwert, F., Moscatelli, C., & Pons, J. (2023). The seven steps: Building the DiGA thesaurus. *Journal of Open Humanities Data*, 9(1), 11. https://doi.org/10.5334/johd.111

Bergamaschi, S., Castano, S., & Vincini, M. (1999). Semantic integration of semistructured and structured data sources. *ACM SIGMOD Record*, 28(1), 54-59. https://doi.org/10.1145/309844.309897

Bertacchini, E., Bravo, G., Marrelli, M., & Santagata, W. (2012). Chapter 13 cultural commons: A new perspective on the production and evolution of cultures. In E. Bertacchini, G. Bravo, M. Marrelli, & W. Santagata (Eds.), *Cultural Commons* (pp. 243-250). Edward Elgar Publishing.

Broughton, V. (2004). *Essential classification*. Facet Publishing.

Chan, L., Intner, S., & Weihs, J. (2016). *Guide to the library of congress classification* (6th ed.). Libraries Unlimited.

Greater Mekong Subregion Secretariat. (2013). *About the Greater Mekong Subregion program*. https://greatermekong.org/about-greater-mekong-subregion

Hein, A. M. (2010, May 9-10). *Identification and bridging of semantic gaps in the context of multi-domain engineering*. Paper presented at 2010 Forum on Philosophy, Engineering & Technology, Golden, CO, USA.

Huang, J. X., Shin, J. A., & Choi, K. S. (2008). Building domain ontology through concept and relation classification. *Journal of KIISE: Software and Applications*, 35(9), 562-571.

Iamkhajornchai, P., & Manmart, L. (2016). [Cultural knowledge organization system]. *Journal of Information Science Research and Practice*, 31(2), 93-122. Thai. https://so03.tci-thaijo.org/index.php/jiskku/article/view/45938

Li, P., & Li, Y. (2013). On transformation from the thesaurus into domain ontology. *Advanced Materials Research*, 756–759, 2698–2704. https://doi.org/10.4028/www.scientific.net/amr.756-759.2698

McLean, S. (2015). *Business communication for success*. University of Minnesota Libraries Publishing.

Mischo, W. (1982). Library of congress subject headings: A

review of the problems, and prospects for improved subject access. *Cataloging & Classification Quarterly*, 1(2-3), 105-124. https://doi.org/10.1300/J104v01n02_06

Prasarn, O., Saengsai, K., & Chaithongdee, C. (2015). *Knowledge management on co-culture for education development to promote better understanding in Thailand, Laos, and Cambodia for sustainability of ASEAN community* (Research Report). Sisaket Rajabhat University, Sisaket, Thailand.

Rowley, J. E. (1992). *Organizing knowledge: An introduction to information retrieval* (2nd ed.). Ashgate.

Ryan, C. (2014). *Thesaurus construction guidelines: An introduction to thesauri and guidelines on their construction*. Royal Irish Academy and National Library of Ireland.

JISTaP http://www.jistap.org
Journal of Information Science Theory and Practice

RESEARCH PAPER
J Inf Sci Theory Pract 12(3): 75-88, 2024

eISSN : 2287-4577 pISSN : 2287-9099

https://doi.org/10.1633/JISTaP.2024.12.3.6

Assessment and Support Measures of Academic Journals in the National Open Access Platform AccessON

Hyekyong Hwang
Open Access Center, Korea Institute of Science and Technology Information, Seoul, Korea
E-mail: hkhwang@kisti.re.kr

Eun Jee Lee
Open Access Center, Korea Institute of Science and Technology Information, Seoul, Korea
E-mail: ejlee@kisti.re.kr

Wan Jong Kim
Open Access Center, Korea Institute of Science and Technology Information, Seoul, Korea
E-mail: wjkim@kisti.re.kr

Jin Ho Park*
Knowledge Information and Culture Track, Hansung University, Seoul, Korea
E-mail: jhp@hansung.ac.kr

ABSTRACT

This study aims to assess the maturity of Korean open access (OA) journals using the OA infrastructure provided by the Korea Institute of Science and Technology Information, and develop necessary strategies for future improvement. The assessment model consists of three dimensions, 12 items, and 24 sub-items. The importance of the three dimensions (A: OA policy establishment and disclosure, B: OA sustainability, and C: Journal openness quality) was differentiated by the Analytic Hierarchy Process, and the maturity stages were divided into five levels (Entry, Growth1, Growth2, Maturity1, and Maturity2). The assessment was carried out twice for 100 academic journals. The results indicated that the proportion of journals at or above the Growth1 level increased by 11% to reach 83% during the second assessment phase, which could be owing to the learnings of the first assessment. Following expert consultations on the assessment results, three support measures were identified to activate OA. The first includes OA promotion and education activities, which involve creating standard regulations and guidelines, and advancing educational activities for societies that are either preparing for or currently implementing OA. The second involves providing support for technical aspects, such as identifiers, XMLization, and copyright management, through peer review and OA publishing platforms. The third includes collaborative activities to enhance journal evaluations and the recognition criteria for researchers' achievements in OA journals, and fostering cooperation with national and research and development institutions for financial support.

Keywords: open access, open access assessment model, journal openness, Korean OA journals, OA maturity model, KISTI

Received: April 4, 2024 **Revised:** May 30, 2024
Accepted: June 3, 2024 **Published:** September 30, 2024

***Corresponding Author:** Jin Ho Park
https://orcid.org/0000-0003-2284-1217
E-mail: jhp@hansung.ac.kr

All JISTaP content is Open Access, meaning it is accessible online to everyone, without fee and authors' permission. Open Access articles are automatically archived in the Korea Institute of Science and Technology Information (KISTI)'s Open Access repository (AccessON). All JISTaP content is published and distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/). Under this license, the authors retain full ownership of their work, while permitting anyone to use, distribute, and reproduce the content in any medium, as long as the original authors and source are cited. For any reuse, redistribution, or reproduction of a work, users must clarify the license terms under which the work was produced.

© 2024 [Hyekyong Hwang, Eun Jee Lee, Wan Jong Kim, Jin Ho Park] This article is distributed under the Creative Commons Attribution License (CC BY), allowing unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. It is published by the Korea Institute of Science and Technology Information (KISTI).

1. BACKGROUND AND PURPOSE

The inception of open access (OA) involved activities where authors self-archived their research papers. The term was officially defined by the Budapest Open Access Initiative (BOAI) (Chan et al., 2002) as the practice of providing online access to scholarly research outputs without cost, in addition to minimizing copyright and other usage restrictions to enable free access.

OA acts as a significant driver in promoting knowledge sharing and advancing scholarly research. In the traditional publishing model, users face difficulties accessing the latest research findings due to high subscription fees and copyright restrictions. By allowing researchers and the general public easy access to scholarly materials, OA improves the quality of research and education and accelerates the dissemination of knowledge (Willinsky, 2009).

The importance of OA is evident in ensuring research transparency and the reproducibility of research results, which enhances scientific reliability. In addition, OA ensures equal access to information for researchers in developing countries and institutions facing budget constraints. This is critical for promoting global scholarly collaboration and bridging the international knowledge gap.

OA contributes to the rapid dissemination of research. Unlike traditional academic publishing, OA journals are quickly published on online platforms, allowing the latest research on urgent public health crises or environmental issues to be rapidly communicated. This increases the timeliness and social impact of academic research. In light of the recent COVID-19 pandemic, the general public has recognized the importance of quickly spreading research findings. However, OA is not achieved merely by making academic papers available online. It signifies a complete transformation of traditional academic publishing. This involves changes in related systems, reflecting diverse stakeholders' perspectives, new publishing and service infrastructures for OA, and changes in how academic societies handle costs.

Successful OA realization requires the efforts of various stakeholders, including researchers, societies, funding organizations, and governments. These actors must work together organically and play complementary roles. Since it is necessary to summarize the efforts required to promote OA and evaluate the current level of activation, it is important to cumulatively evaluate various factors, assess problems, and establish development directions.

This study's purpose is to assess the maturity stages of journals that have declared OA in Korea and identify the current status of OA journals and the factors necessary to revitalize them. For this purpose, an assessment model and maturity stages are needed. The assessment model utilizes the results of a study conducted by the Korea Institute of Science and Technology Information (KISTI) from 2021 to 2022 by Kim et al. (2023). Their study summarized various factors necessary for assessing OA journals and organized them into a model. The maturity stage utilized the criteria presented by KISTI. The assessment was limited to 100 journals using KISTI's system—a Korean research funding organization that provides free support for OA platforms. The assessment model and maturity stages are not internationally agreed-upon standards, and the assessment is limited to Korean OA journals utilizing KISTI's platform. These can be considered study limitations.

KISTI seeks to improve the quality of Korean OA journals and provides the technical infrastructure necessary for the maturation of OA. To sustain these services, it is necessary to accurately assess and analyze the status of journals currently utilizing KISTI's platform. This assessment is not intended to evaluate the current level of OA journals, but instead to identify the support elements needed in the future. These support elements are anticipated to serve as a direct basis for KISTI to provide OA-related support services, and in the long run, may contribute to increasing the OA level in Korea.

2. LITERATURE REVIEW

Since the BOAI defined the term "open access" in 2002 as providing free online access to scholarly research outputs, various studies have explored its implications, benefits, and challenges. Early OA-related studies introduced the concept, highlighting its necessity and importance (Lee, 2008; Suber, 2005). There has been a significant emphasis on understanding OA trends and consolidating the positions of diverse stakeholders, in order to develop strategies to advance OA. For example, Yoon and Kim (2007) considered organizing the differing perspectives of the library community, research groups, and publishing sector as an essential task in the spread of OA. Accordingly, they conducted a study that compiled OA trends and limitations and arranged issues to make proposals. Similarly, Choi and Cho (2005) categorized numerous cases to define OA trends, proposing the roles of librarians, universities, publishers, funding organizations, societies, and governments.

Subsequent studies have explored the elements neces-

 https://doi.org/10.1633/JISTaP.2024.12.3.6

Table 1. Assessment framework for open access (OA) journal publishing practices

Dimension	Item	Sub-item	Key question
A. OA policy establishment and disclosure	A1. OA policy establishment	A1.1. Copyright policy	Q1. Are there written policies on copyright?
		A1.2. OA finance disclosure policy	Q2. Are there written policies on the transparent disclosure of finances related to OA publications?
		A1.3. Archiving policy	Q3. Are there written policies on archiving?
	A2. OA policy disclosure and registration	A2.1 OA policy disclosure and registration	Q4. Are the OA policies disclosed on the journal homepage? (1) Copyright policy—Y/N (2) OA finance disclosure policy—Y/N (3) Archiving policy—Y/N
			Q5. Are the OA copyright policies registered with Korea Journal Copyright Information (KJCI), SHERPA/RoMEO?
B. OA sustainability	B1. Finance	B1.1. Financial resources for OA publication	Q6. What is the main financial resource for OA publication? (1) Internal (society) funding (2) External funding (3) APC (article processing charge) (4) Other
		B1.2. Financial sustainability for OA publication	Q7. How do you view the sustainability of the main financial resources of OA publication? (1) Very secure (2) Somewhat secure (3) Not at all
		B1.3. Financial disclosure for OA publication	Q8. Are OA publication-related finances (income and expenses) transparently disclosed on the journal's homepage?
	B2. People	B2.1. Education and training for OA publication	Q9. Do internal staff participate in training/education for OA publishing?
	B3. Organization and culture	B3.1. Awareness of OA	Q10. How many members of the society are aware that the journal is OA? (1) Almost all (2) About half (3) Only a few (4) None
	B4. Collaboration	B4.1. Collaboration for OA publication and dissemination	Q11. Are there collaborative efforts in your discipline for OA publication?
C. Journal openness quality	C1. Submission and review	C1.1. Reviewer disclosure	Q12. Are the reviewers of an article open to the public?
		C1.2. Review or opinion of article disclosure	Q13. Are the reviews or opinions of an article open to the public?
		C1.3. Cost of author disclosure	Q14. Is the APC information clearly disclosed on the journal's homepage or announced to the authors?
			Q15. Are OA publication fees waived or discounted for independent researchers and/or students?

Table 1. Continued

Dimension	Item	Sub-item	Key question
C. Journal openness quality	C2. Author rights	C2.1. Copyright	Q16. What is the type of author copyright? (1) The copyright is retained by authors without restrictions (2) The copyright is co-owned by both authors and the publisher, or the authors own the reuse rights with restrictions (3) The copyright is owned by the publisher and only fair use is allowed
		C2.2. Posting rights	Q17. How does the journal allow the author(s) to retain posting rights? (1) Authors can choose to post any version of the manuscript to any repository immediately upon publication (2) Authors are allowed to post a specific version of the manuscript to any repository immediately upon publication (3) Authors are allowed to post a specific version of the manuscript to a specified repository (4) Authors are not allowed to post any version of the manuscript to any repository or website
	C3. Reader and re-use rights	C3.1. Reader rights	Q18. What are the readers' rights in accessing the full text of the article? (1) Readers can freely access the full text of the article from repositories and websites immediately upon publication (2) Some repositories or websites provide the full text of the article with a fee or embargo period (3) All articles are accessible with a fee
		C3.2. Reuse rights	Q19. What are the reuse rights for the journal articles? (1) Liberal CC BY license (2) CC BY with some conditions (e.g., CC BY-SA, CC BY-ND, CC BY-NC, CC BY-NC-SA, CC BY-NC-ND) (3) Reuse rights are provided only for OA articles (e.g., hybrid journals) (4) Fair use only
	C4. Findability	C4.1. Postings	Q20. Are articles posted to as many trusted repositories as possible? (1) More than 2 repositories (2) Only 1 repository (3) None
		C4.2. Search engine optimization (SEO) for dissemination	Q21. Is SEO conducted for external search engines such as Google and NAVER (a South Korean search engine)?
		C4.3. Unique identifier for author and article	Q22. Does the journal adopt aspects of the article's unique identification system, such as the digital object identifier (DOI) or the uniform resource name (URN)?
			Q23. Does the journal adopt the author's unique identification system, such as the Open Researcher and Contributor ID (ORCID) or the international standard name identified (ISNI)?

 https://doi.org/10.1633/JISTaP.2024.12.3.6

Table 1. Continued

Dimension	Item	Sub-item	Key question
C. Journal openness quality		C4.4. Machine-readable metadata for the article	Q24. Does the journal provide the metadata for each article in a machine-readable format (e.g., JavaScript Object Notation [JSON], BibTeX, RIS [Research Information Systems], etc.)?
	C5. Full-text accessibility	C5.1. Machine readability	Q25. Does the journal provide the full article in XML format?
			Q26. Does the journal provide the full article in PDF or HTML format?
		C5.2. Mobile-responsive design	Q27. Do journal articles contain a mobile-responsive design?
	C6. Monitoring	C6.1. Use monitoring	Q28. Does the journal collect and disclose usage statistics for each article?

sary for adopting OA, with Ko et al. (2009) investigating whether factors that influence OA adoption vary by the type of publishing institution or academic field. They pointed out OA promotion and the transition of excellent academic journals to OA as important steps for the future. Meanwhile, other studies have classified OA into gold, green, and bronze OA journals to better understand the current state of OA (Basson et al., 2022; Piwowar et al., 2018).

Addressing copyright issues is critical in the OA landscape. Hong (2008) highlighted the importance of resolving such disputes in the context of South Korean academic papers and investigated the copyright attribution problems of 906 journals. Kim (2020) discussed the policy and legal aspects of OA, emphasizing how amendments to copyright law could either facilitate or hinder the OA movement. Shavell (2010) argued for the complete removal of copyright from scholarly works.

Kim et al. (2016) examined OA policies for scholarly papers supported by public research funds, suggesting ways to enhance their accessibility through a comparison of policies from different countries. Other studies have investigated the effectiveness of OA, with the findings of Brody et al. (2004) and Antelman (2004) revealing that OA papers demonstrated higher citation rates and impact than non-OA papers.

Focusing on researchers, Park (2022) investigated their perceptions of establishing a healthy OA ecosystem. Shim (2021) explored the criteria by which South Korean researchers choose OA journals and found that they have less experience with OA publishing, with differing criteria for selecting OA and non-OA journals. This underscores the need for policy improvements and education in this field. Studies on OA implementation and acceptance in universities and academic libraries, which are primary centers of research, have also been conducted. Kim and Lee (2005) suggested that technical and policy factors should be considered when operating an OA repository in a university. Kim (2018) proposed ways to manage and share theses and dissertations through OA university repositories. Kang and Chang (2010) examined criteria for libraries to select and evaluate OA journals.

Recent studies have identified and sought to improve the current state and challenges of OA. Bosman et al. (2021) conducted a large-scale survey and focus group interviews with 1,619 diamond OA journals to determine their status, quality management practices, and sustainability challenges. About 67% of the journals managed quality through double-blind peer review, and 78% adhered to OA guidelines. However, the sustainability of OA journals faces challenges, including operating platforms and server limitations. Bosman et al. (2021) pointed to the need for efforts to expand visibility due to the difficulties of indexing OA content, noting that 20% of diamond OA journals are considering discontinuing their current OA publishing due to a lack of operational staff and financial constraints.

Paquet et al. (2022) analyzed the status of OA academic publishing in Canada, based on field and university, finding that the mandatory OA policies of research funding agencies impact the expansion of OA academic publishing. They suggested that a national-level OA publishing platform is necessary to stimulate OA publishing. Choi et al. (2012) investigated user satisfaction with South Korean OA services, and noted that the future development of OA requires policy support, OA journal publishing

and repository operations, establishment of a governance system, fostering of researcher participation through OA promotional activities, and building of a global cooperation system. Kim (2014) suggested that academic societies explore various OA distribution methods and identify the methods and platforms suitable for each academic society, emphasizing the importance of disclosing each society's OA policy on SHERPA/RoMEO, DOAJ, and PMC. Additionally, Duc et al. (2020) and Krawczyk and Kulczycki (2021) pointed out that well-intentioned OA has led to unintended consequences, such as the emergence of predatory journals, leading to confusion among scholars.

Numerous other studies, such as those by Antelman (2004) and Brody et al. (2004), have demonstrated the significant positive impacts of OA, including higher citation rates, improved accessibility to research, and increased dissemination of scholarly knowledge. However, several challenges are yet to be resolved, including changes in the publishing model, copyright issues, and the development of sustainable financial models.

3. METHOD

3.1. Assessment Model

In this study, we used the assessment model developed by KISTI (Kim et al., 2023) to assess the current state of South Korean academic journals that have declared themselves to be OA. The model simplifies reality, yet does not perfectly represent it. Moreover, not all target academic journals have the same level of preparation. Thus, while we made the assessment based on this model, we introduced the analytic hierarchy process (AHP) to construct different levels of importance for dimensions and items, accounting for their varying importance. In addition, we interpreted the results by dividing the maturity levels into Entry, Growth1, Growth2, Maturity1, and Maturity2, and assigning score ranges for each level. Table 1 outlines the assessment model.

This model is a modified version of the 2021 model. The revised model, which we adjusted based on the appropriateness of the dimensions, items, sub-items, and key questions, reflects the results of interviews held with editors of journals that publish OA articles (Kim et al., 2023).

The model consists of three dimensions, 12 items, and 24 sub-items. Each sub-item has specific key questions for assessment. In the present study, we used these questions for assessment. However, direct verification by internal stakeholders of the journals is necessary for items A1, B1, B2, B3, and B4. Other items can be verified through the journal's website and the manuscript submission review system. Hence, we conducted surveys with internal stakeholders, such as editors and society presidents, for A1, B1, B2, B3, and B4, and directly investigated the other items.

In addition, in KISTI's assessment model, A2.1 contains two key questions: Q4 and Q5. However, Q4 is designed to assess the three sub-items belonging to A1. For the convenience of quantifying future assessment outcomes, we divided A2's Q4 into three questions during the actual assessment (i.e., we restructured it so that individual scores could be obtained for each question).

3.2. Weight Model

KISTI's assessment model includes scoring during the actual assessment, with "not applicable" receiving 0, "partially applicable" receiving 0.5 and 0.75, and "applicable" receiving 1 point. We adhered to this standard; however, for ease of converting future measurement results, we converted 1 point into 100 points for calculation. In the original model, we gave all three dimensions the same importance. However, given that all OA promotion environments do not start from the same baseline, especially reflecting the situation of OA in South Korea, we performed an AHP analysis to apply different scores for the assessment.

AHP, developed by Saaty (1980), is used to find answers to complex decision-making problems. Saaty (1980) proposed an approach to solving such problems by classifying various options into components, determining the relative priorities of these components, and deciding on the final priorities. AHP has been utilized for decision-making in numerous fields.

In this study, we conducted an online survey to derive weights. The participants included internal staff at KISTI, researchers involved in previous studies, journal officials who had been interviewed for previous studies, and experts familiar with KISTI's OA activities. We administered the survey to 18 participants from February 24 to March 5, 2023, over 10 days. We finally used 11 responses. As a result, we derived the weights for dimensions and items of the assessment model, as shown in Table 2.

According to Table 2, the weights of the three dimensions are as follows: A: OA policy establishment and disclosure=48%, B: OA sustainability=29%, and C: Journal openness quality=23%. It is evident that OA policy establishment and disclosure is the most important factor in the current Korean OA environment. In terms of each dimension, A1: OA policy establishment (66%), B1: Finance (39%), and C1: Submission and review (23%) were the

 https://doi.org/10.1633/JISTaP.2024.12.3.6

Table 2. Results of AHP analysis

Dimension	Weight (%)	Item	Weight (%)
A. OA policy establishment and disclosure	48	A1. OA policy establishment	66
		A2. OA policy disclosure and registration	34
		Total	100
B. OA sustainability	29	B1. Finances	39
		B2. People	22
		B3. Organization and culture	24
		B4. Collaboration	15
		Total	100
C. Journal openness quality	23	C1. Submission and review	23
		C2. Author rights	22
		C3. Reader and re-use rights	19
		C4. Findability	13
		C5. Full-text accessibility	13
		C6. Monitoring	10
		Total	100
Total	100		

AHP, analytic hierarchy process; OA, open access.

Table 3. Open access maturity levels

Maturity level	Range
Entry	0-50
Growth1	50-62.5
Growth2	62.5-75
Maturity1	75-87.5
Maturity2	87.5-100

most important items.

In this study, we converted the overall assessment outcomes into a 100-point scale by applying the weights to the dimensions based on the AHP results. The converted scores do not numerically represent the current state of OA journals. Hence, we interpreted the OA maturity levels as divided into the five levels used by KISTI (Table 3).

Table 3 is based on KISTI's criteria, considering that the subjects of the assessment model, though varying in degree, aim for OA and utilize KISTI's OA platform. Hence, the entry-level range is 50 or less, and we set the subsequent four levels equally.

4. ASSESSMENT OF OA JOURNALS AND THEIR PROGRESS

The 100 journals targeted for maturity assessment are located in South Korea and represent the entire set of academic journals that utilize the Journal Repository provided by KISTI's OA platform, AccessON, or the OA-supporting peer review management system called ACOMS+. These journals were selected because they all currently use KISTI's OA infrastructure, and hence comprise a comprehensive sample for evaluating the maturity of OA practices within the platform. The assessment was conducted twice, and Table 4 presents its progress.

As shown in Table 4, items directly investigated by the researcher were re-verified by a separate reviewer to determine appropriateness before final analysis. We conducted the assessment twice. During the second assessment phase, individual survey forms were sent to journal officials who had responded in the first phase. For direct investigations, we utilized an online survey tool (Google Forms) to record evidence of survey results (screen captures or URLs).

Table 4. Assessment progress

Assessment phase	Subjects	Period	Conducted activities
1st	Items A2, C1, C2, C3, C4, C5, C6	2023.04.01-2023.04.29	• Direct investigation • Attachment of proof for assessment results • Validation of survey results by reviewers
	Items A1, B1, B2, B3, B4	2023.05.04-2023.05.26	• Online survey of OA journal officials (using Google Forms)
2nd	Items A2, C1, C2, C3, C4, C5, C6	2023.08.01-2023.08.31	• Direct investigation • Attachment of proof for assessment results • Validation of survey results by reviewers
	Items A1, B1, B2, B3, B4	2023.08.15-2023.08.31	• Online survey of OA journal officials (using Feedback.io by Opensurvey Inc.)

OA, open access.

Table 5. Descriptive statistics of the first and second assessment results

	Category	Total score obtained	A score	A weight (48%)	B score	B weight (29%)	C score	C weight (23%)	100 points (weighted)
1st	Average	1,319.54	59.17	28.40	68.29	19.80	46.57	10.71	58.92
	Median	1,337.50	62.50	30.00	78.13	22.66	47.92	11.02	62.85
	Standard deviation	313.49	20.44	9.81	22.75	6.60	17.45	4.01	16.05
	Maximum	1,850.00	100.00	48.00	91.67	26.58	82.64	19.01	86.97
	Minimum	425.00	0.00	0.00	16.67	4.83	0.00	0.00	9.52
2nd	Average	1,480.00	69.88	33.54	71.35	20.69	57.01	13.11	67.34
	Median	1,550.00	75.00	36.00	79.17	22.96	58.33	13.42	72.32
	Standard deviation	272.49	20.65	9.91	21.08	6.11	14.36	3.30	15.02
	Maximum	1,925.00	100.00	48.00	91.67	26.58	84.72	19.49	87.31
	Minimum	675.00	12.50	6.00	16.67	4.83	23.61	5.43	24.56

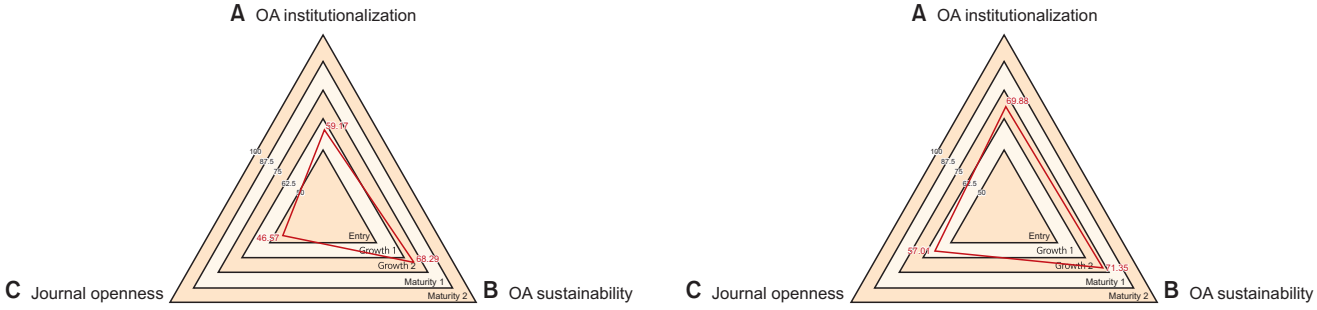


Fig. 1. Maturity levels by dimension (left: first results; right: second results). OA, open access.

5. ASSESSMENT RESULTS AND DERIVATION OF IMPROVEMENT MEASURES

5.1. Assessment Results

Table 5 displays descriptive statistics for the first and second assessment phases.

As depicted in Table 5, all three dimensions improved in the second assessment phase. Specifically, the weighted overall average in the first survey was 58.92, which reached 67.34 in the second survey, an improvement of 8.42 points. By dimension, A: OA policy establishment and disclosure increased from 28.40 to 33.54, B: OA sus-

 https://doi.org/10.1633/JISTaP.2024.12.3.6

tainability increased from 19.80 to 20.69, and C: Journal openness quality increased from 10.71 to 13.11. In particular, the improvement of 5.14 points for dimension A is significant compared to 0.89 for dimension B and 2.4 for dimension C. Hence, in dimension A, the contents related to the establishment of policies within the society were strengthened before the second survey.

The results, when mapped to maturity levels for comparison, are shown in Fig. 1.

In the first assessment, dimension A was at the Growth1 level, dimension B was at the Growth2 level, and dimension C was at the Entry level. In the second assessment phase, dimension A improved to the Growth2 level, dimension B remained at the Growth2 level, and dimension C improved to the Growth1 level.

Table 6 presents the distribution of maturity levels for the 100 responding journals.

While 72% of the journals were at the Growth1 level or above in the first assessment phase, this increased to 83% in the second assessment phase, demonstrating a significant improvement of 11%. To assess the current situation of South Korean OA journals and construct a future support system, further detailed verification is needed. It is necessary to examine the assessment outcomes by dimension, focusing on the second assessment phase. The assessment results for dimension A are presented in Table 7.

As seen in Table 7, journals at the Entry level need support for formalization and public disclosure of overall OA policies. Even journals at the growth level require clarity on the policy of disclosing OA publishing finances (A1.2), formalization of the archival policy (A1.3), and transparency and registration of formalized OA regulations (A2.1). The assessment results show low figures, indicating the need for future educational activities.

Table 8 presents detailed results for dimension B.

As displayed in Table 8, regardless of the maturity level, all journals urgently need support for transparency in OA academic publishing finances (B1.3). Meanwhile, journals at the Entry level require support for transparency in OA publishing finances (B1.3), training and education for academic society staff (B2.1), and collaboration for publi-

Table 6. Comparison of the first and second maturity levels

Level	Journals in the first assessment results	Journals in the second assessment results
Entry	28	17
Growth1	21	11
Growth2	38	33
Maturity1	13	39
Maturity2	0	0
Total	100	100

Table 7. Second assessment results for dimension A

Maturity level	No. of journals	Overall score for dimension A	Average for A1.1	Average for A1.2	Average for A1.3	Average for A2.1
Entry	17	39.22	47.06	41.18	23.53	41.18
Growth1	11	55.30	90.91	45.45	45.45	50.00
Growth2	33	71.34	96.97	60.61	72.73	65.91
Maturity1	39	86.11	100.00	100.00	89.74	75.64
Total and average	100	69.88	89.00	71.00	68.00	63.75

Table 8. Second assessment results for dimension B

Maturity level	No. of journals	Overall score for dimension B	Average for B1.1	Average for B1.2	Average for B1.3	Average for B2.1	Average for B3.1	Average for B4.1
Entry	17	40.69	67.65	85.29	0.00	35.29	52.94	23.53
Growth1	11	62.88	54.55	59.09	0.00	54.55	77.27	81.82
Growth2	33	76.07	71.97	77.27	0.00	81.82	90.91	81.82
Maturity1	39	83.12	76.28	80.77	0.00	94.87	92.95	92.31
Total and average	100	71.35	71.00	78.00	0.00	76.00	83.75	76.00

https://doi.org/10.1633/JISTaP.2024.12.3.6

Table 9. Second assessment results for dimension C

Maturity level	Number	Overall score	Average for C1.1	Average for C1.2	Average for C1.3	Average for C2.1	Average for C2.2	Average for C3.1	Average for C3.2	Average for C4.1	Average for C4.2	Average for C4.3	Average for C4.4	Average for C5.1	Average for C5.2	Average for C6.1
Entry	17	44.85	0.00	5.88	38.24	0.00	8.82	95.59	35.29	100.00	88.24	50.00	64.71	52.94	70.59	47.06
Growth1	11	45.39	0.00	0.00	22.73	13.64	11.36	93.18	45.45	100.00	100.00	50.00	36.36	50.00	45.45	63.64
Growth2	33	57.45	0.00	9.09	36.36	25.76	30.30	97.73	51.52	95.45	100.00	69.70	75.76	56.06	93.94	66.67
Maturity1	39	65.21	0.00	2.56	57.69	42.31	60.26	98.08	78.85	100.00	100.00	74.36	92.31	61.54	69.23	74.36
Total and average	100	57.01	0.00	5.00	43.50	26.50	36.25	97.00	58.75	98.50	98.00	66.00	76.00	57.00	75.00	66.00

cation and dissemination (B4.1). Hence, it is evident that South Korean OA journals lack awareness of transparent operations beyond the value of sharing academic works.

Table 9 outlines detailed results for dimension C.

Dimension C showed the lowest assessment scores. Regardless of the maturity level, active support is necessary for C1 and C2, which deal with openness of reviewers and opinions and depend mostly on the initiative of the academic societies. This indicates that awareness of the importance of process transparency, beyond sharing the final outputs, is low. Journals at the entry and growth levels exhibited low usage of identification systems for papers and authors. For journals at the growth level, providing machine-readable metadata for individual papers, as in C4.4, is a challenging issue to be solved by academic societies individually. Such problems could be addressed by platform operators, such as KISTI, through functional improvements or separate informatics projects.

The interval between the two assessment rounds was three months, which is short. Nonetheless, the improvement in assessment outcomes implies that the maturity model assessments served as a form of education and promotion. This demonstrates that academic societies declaring to be OA in South Korea do not have a deep understanding of OA. In particular, it is necessary to recognize that process transparency is important. As a government-led OA platform operating agency, KISTI needs to perform various activities beyond providing technical support to promote OA.

Table 10 depicts the support items necessary for OA journals based on the problems identified in the assessment results.

5.2. Consultations with Experts

In this study, we shared the assessment results with OA experts, and upon consulting with them, we performed

Table 10. Support measures for OA journals

Category	Support measures
Policy activities for OA transition	• Creation and education promotion of OA-related standards, regulations, and guidelines • Guidance on the maturity model and sharing of assessment results
Technical support for OA publishing	• Development and distribution of OA scholarly communication support platforms • Easy management and service functionalities for key elements (such as copyright) • Datafication of documents like JATS-XML • Support for standard identification systems (ORCID, ISNI, etc.) • Construction and support of archiving systems
Collaborative activities for promoting OA	• Use of journals' evaluation and research performance as indicators • Financial support: APC, publishing fee support, etc. • Cooperative activities with OA publishing institutions (government, foundations, etc.)

OA, open access; ORCID, Open Researcher and Contributor ID; ISNI, international standard name identified.

Table 11. Summary of results from the advisory meeting with experts

Category	Content
Financial support	• Nationwide financial support for societies to facilitate OA operations (relief of editorial staff wages)
OA operation of organizations	• Operation of a separate organization to promote OA • Operation of OA organizations led by societies in addition to those led by KISTI
Education and promotion	• Continuous promotion and educational activities to eliminate misunderstandings and biases about OA
Reconsideration of the assessment model approach	• The levels of understanding and awareness of OA among respondents vary. Consideration of this aspect is necessary when interpreting the results • A separate qualitative study is needed to interpret the assessment results • Reconsideration of weights for each dimension, and re-evaluation of the assessment method (assessment criteria) • Trying different simulations based on dimension C
Technical support	• Introduction of technical devices and promotion for essential elements of OA culture, such as open peer review

OA, open access; KISTI, Korea Institute of Science and Technology Information.

further assessments to identify crucial considerations for future OA support.

The experts included three researchers who conducted maturity model studies and three association representatives currently engaged in OA publishing. An advisory meeting was held using the online meeting tool Zoom on September 19, 2023 at 15:00 hours. The purpose of the advisory meeting was to share the results of the two assessment phases and develop the necessary content for future OA support. The participants shared their opinions about the assessment results and other aspects, summarized in Table 11.

The most discussed aspect was finance. The most significant challenge for academic societies aiming for OA was editorial, publishing, and infrastructure costs for academic paper services. In South Korea, institutions such as KISTI provide peer review, archiving, and service systems on a national scale free of charge. This is a prime example of public funds being directly invested in creating an OA environment. KISTI has built an OA platform and continues to manage and improve it, striving to develop the OA ecosystem. These efforts play a vital role in enhancing the openness of research results, promoting the sharing of scholarly knowledge, and contributing to scientific and technological advancement.

Nonetheless, the actual costs of editing, reviewing, and publishing remain a challenge. Despite declaring themselves to be OA, academic societies are often forced to rely on authors for review and publication costs. This is highlighted by maturity assessments and reaffirmed through the advisory meeting, underscoring the need for support.

The importance of forming a separate organization for OA activities, rather than individual academic societies acting alone, was highlighted to address issues collectively. As shown by the improved results of the second assessment phase, education and promotion are essential to accurately communicate what is necessary for OA and what academic societies need to prepare and implement.

Furthermore, there is a need to reconsider the assessment model and method, a point often made by previous researchers (Kim et al., 2023). It is important to note that respondents' levels of OA awareness vary; hence, rather than interpreting survey results directly, conducting qualitative research on a few journals based on the assessment results was suggested. In particular, it may be necessary to reconsider the current weighting method and assessment criteria. Researchers have noted that despite containing the fundamental elements of OA journals, the lowest scores for dimension C indicate issues with transparency and accessibility. For technical support, preemptively preparing technical devices and encouraging societal participation in aspects where culture needs to change (such as open peer review) was identified as a potential promotion strategy.

5.3. Summary of Improvement Measures

The improvement measures derived from the journal assessments and consultations with experts include aspects that require the efforts of academic societies and external support. As an operator of a nationally supported OA platform, KISTI can support the following elements among the proposed improvements:

- OA promotion and education activities: Creation of standard regulations and guidelines and educational activities for societies preparing for or currently implementing OA.
- Technical support: Support for technical elements, such as identifiers, XMLization, and copyright management, through peer review and OA publishing platforms.
- Collaborative activities: Activities for improving journal evaluations and recognition criteria for researchers' achievements regarding OA journals, and cooperation with national and research and development institutions for financial support.

As noted in the consultation with experts, financial support for editorial staff and other OA-related expenses is difficult for societies to manage and requires external support.

6. CONCLUSION

We used KISTI's OA maturity model to assess OA journals and derive improvement measures. Although the period was short, the two assessment phases confirmed an improvement in the maturity of the target journals. This can be interpreted as an educational effect where the dimensions and elements of the maturity model were reviewed by the responding journals. These results are due to academic societies in South Korea becoming aware of the maturity model's assessment items between the two surveys, correcting the errors, and supplementing the aspects they were unprepared for. While there is increasing interest in OA and quantitative growth, it is necessary to address qualitative aspects such as transparency in the peer review processes, quality of metadata, and adherence

 https://doi.org/10.1633/JISTaP.2024.12.3.6

to ethical publishing standards, in order to fully realize OA's positive effects.

The results indicate that while the maturity of OA journals has improved over time, some areas still require improvement. In particular, targeted support and interventions are needed to address specific weaknesses, such as financial transparency, policy documentation, and technological infrastructure.

Moreover, the maturity model used in the study needs to be updated and supplemented as new trends and issues emerge in the development of OA journals. The weights used in this study reflect the priorities and importance assigned by the stakeholders of OA in South Korea at the time of the survey. As the OA landscape evolves, these priorities may change, necessitating modifications to the weights. Continuous monitoring, support, and education are important for advancing the OA movement, with the goal of a more transparent, accessible, and high-quality scholarly communication ecosystem. This can be interpreted as the journals gaining a better understanding of the maturity model through the process of reviewing its dimensions and elements, which in turn led to improvements in their practices.

In this paper, we focus on the results of the first OA maturity assessment, using the maturity model developed by KISTI, and propose future improvement measures. This study was conducted on 100 journals utilizing KISTI's platform; however, it is difficult to say that these journals are representative of all OA journals in Korea. Further research is needed on the assessment model and in various OA journals. In addition, it is necessary to analyze the differences in survey data while considering the specific characteristics of academic societies, such as their respective academic fields.

ACKNOWLEDGMENTS

This research was supported by the Korea Institute of Science and Technology Information (KISTI) (No. (KISTI) K24L1M1C02-01, (NTIS) 1711198421).

CONFLICTS OF INTEREST

No potential conflict of interest relevant to this article were reported.

REFERENCES

Antelman, K. (2004). Do open-access articles have a greater research impact? *College & Research Libraries,* 65(5), 372-382. https://doi.org/10.5860/crl.65.5.372

Basson, I., Simard, M. A., Ouangré, Z. A., Sugimoto, C. R., & Larivière, V. (2022). The effect of data sources on the measurement of open access: A comparison of Dimensions and the Web of Science. *PLoS One,* 17(3), e0265545. https://doi.org/10.1371/journal.pone.0265545

Bosman, J., Frantsvåg, J. E., Kramer, B., Langlais, P. C., & Proudman, V. (2021). *OA diamond journals study. Part 1: Findings.* Zenodo.

Brody, T., Stamerjohanns, H., Vallières, F., Harnad, S., Gingras, Y., & Oppenheim, C. (2004). *The effect of open access on citation impact*. Paper presented at the Provision for University Research Output: An International Meeting, Southampton, UK.

Chan, L., Cuplinskas, D., Eisen, M., Friend, F., Genova, Y., Guédon, J. C., Hagemann, M., Harnad, S., Johnson, R., Kupryte, R., La Manna, M., Rév, I., Segbert, M., de Souza, S., Suber, P., & Velterop, J. (2002). *Budapest open access initiative.* https://www.budapestopenaccessinitiative.org/read/

Choi, H., Hwang, H. K., & Baek, J. M. (2012). An study on the user satisfaction of open access activities in Korea. *Journal of the Korean Society for Information Management,* 29(1), 279-302. https://doi.org/10.3743/KOSIM.2012.29.1.279

Choi, J. H., & Cho, H. Y. (2005). The recent trends of open access movements and the ways to help the cause by academic stakeholders. *Journal of the Korean Society for Information Management,* 22(3), 307-326. https://doi.org/10.3743/KOSIM.2005.22.3.307

Duc, N. M., Hiep, D. V., Thong, P. M., Zunic, L., Zildzic, M., Donev, D., Jankovic, S. M., Hozo, I., & Masic, I. (2020). Predatory open access journals are indexed in reputable databases: A revisiting issue or an unsolved problem. *Medical Archives (Sarajevo, Bosnia and Herzegovina),* 74(4), 318-322. https://doi.org/10.5455/medarh.2020.74.318-322

Hong, J. H. (2008). A study on copyright possession for open access and archiving of scholarly journal paper registered in Korean Research Foundation. *Journal of Korean Library and Information Science Society,* 39(1), 431-463. https://doi.org/10.16981/kliss.39.1.200803.431

Kang, E. Y., & Chang, D. H. (2010). A study on open access journal evaluation criteria. *Journal of the Korean Society for Library and Information Science,* 44(1), 243-265. https://doi.org/10.4275/KSLIS.2010.44.1.243

Kim, G. (2014). A proposal for the acceptance and processes of open access in Korean societies. *Journal of the Korean Society for Information Management,* 31(3), 111-131. https://doi.org/10.3743/KOSIM.2014.31.3.111

Kim, J. (2018). A study on the development of effective ways

to manage and use dissertations and theses by open access. *Journal of Social Science,* 29(3), 95-104. http://doi.org/10.16881/jss.2018.07.29.3.95

Kim, N., Yoon, J., Lee, J. Y., Joung, K. H., Hwang, H., Bai, S. Y., & Chung, E. (2023). Development of a diagnostic framework and its application to open access journal publishing in Korea. *Learned Publishing,* 36(3), 379-393. https://doi.org/10.1002/leap.1542

Kim, S. M., & Lee, N. N. (2005). A study on building an open access based institutional repository in the university for scholarly communication. *Journal of Information Management,* 36(2), 45-71. https://doi.org/10.1633/JIM.2005.36.2.045

Kim, S. Y., Kim, J., Choi, H., & Hwang, H. (2016). An analysis on open access policies on publications funded by overseas public institutions. *Journal of the Korean Society for Library and Information Science,* 50(4), 209-229. https://doi.org/10.4275/KSLIS.2016.50.4.209

Kim, Y. J. (2020). Open access and copyright law revision. *Inha Law Review,* 23(3), 31-62. http://doi.org/10.22789/IHLR.2020.09.23.3.31

Ko, Y. M., Kim, B. A., Park, T. S., & Lee, K. S. (2009). A study on the difference among journal publisher and among the academic disciplines of factors influencing the adoption of open access. *Journal of the Korean Society for Library and Information Science,* 43(4), 431-449. https://doi.org/10.4275/KSLIS.2009.43.4.431

Krawczyk, F., & Kulczycki, E. (2021). How is open access accused of being predatory? The impact of Beall's lists of predatory journals on academic publishing. *Journal of Academic Librarianship,* 47(2), 102271. https://doi.org/10.1016/j.acalib.2020.102271

Lee, D. C. (2008). A very short introduction to open access. *Critical Review of Religion and Culture,* 14(14), 314-319.

Paquet, V., van Bellen, S., & Larivière, V. (2022). Measuring the prevalence of open access in Canada: A national comparison. *Canadian Journal of Information and Library Science,* 45(1), 1-21. https://doi.org/10.5206/cjilsrcsib.v45i1.14149

Park, K. H. (2022). Policy direction for a salubrious open access ecosystem derived from researcher perception survey. *Korean Journal of Policy Analysis and Evaluation,* 32(2), 1-24. http://doi.org/10.23036/kapae.2022.32.2.001

Piwowar, H., Priem, J., Larivière, V., Alperin, J. P., Matthias, L., Norlander, B., Farley, A., West, J., & Haustein, S. (2018). The state of OA: A large-scale analysis of the prevalence and impact of open access articles. *PeerJ,* 6, e4375. https://doi.org/10.7717/peerj.4375

Saaty, T.L. (1980). The analytic hierarchy process (AHP). *The Journal of the Operational Research Society*, 41(11), 1073-1076.

Shavell, S. (2010). Should copyright of academic works be abolished? *Journal of Legal Analysis,* 2(1), 301-358. https://doi.org/10.1093/jla/2.1.301

Shim, W. (2021). A study on the awareness and experience of open access publishing among Korean authors of an open access mega journal. *Journal of the Korean Society for Library and Information Science,* 55(4), 45-65. https://doi.org/10.4275/KSLIS.2021.55.4.045

Suber, P. (2005). Open access, impact, and demand. *BMJ (Clinical research ed.),* 330(7500), 1097-1098. https://doi.org/10.1136/bmj.330.7500.1097

Willinsky, J. (2009). *The access principle: The case for open access to research and scholarship.* MIT Press.

Yoon, H. Y., & Kim, S. Y. (2007). Trends analysis of open access for foreign and domestic scholarly journals in the field of library and information science. *Journal of Korean Library and Information Science Society,* 38(1), 257-276. http://doi.org/10.16981/kliss.38.1.200703.257

Call for Paper

Journal of Information Science Theory and Practice (JISTaP)

We would like to invite you to submit or recommend papers to Journal of Information Science Theory and Practice (JISTaP, eISSN: 2287-4577, pISSN: 2287-9099), a fast track peer-reviewed and no-fee open access academic journal published by Korea Institute of Science and Technology Information (KISTI), which is a government-funded research institute providing STI services to support high-tech R&D for researchers in Korea. JISTaP marks a transition from Journal of Information Management to an English-language international journal in the area of library and information science.

JISTaP aims at publishing original studies, review papers and brief communications on information science theory and practice. The journal provides an international forum for practical as well as theoretical research in the interdisciplinary areas of information science, such as information processing and management, knowledge organization, scholarly communication and bibliometrics.

We welcome materials that reflect a wide range of perspectives and approaches on diverse areas of information science theory, application and practice. Topics covered by the journal include: information processing and management; information policy; library management; knowledge organization; metadata and classification; information seeking; information retrieval; information systems; scientific and technical information service; human-computer interaction; social media design; analytics; scholarly communication and bibliometrics. Above all, we encourage submissions of catalytic nature that explore the question of how theory can be applied to solve real world problems in the broad discipline of information science.

Co-Editors in Chief : Gary Marchionini & Dong-Geun Oh

Please click the "Online Submission" link in the JISTaP website (http://www.jistap.org), which will take you to a login/account creation page. Please consult the "Author's Guide" page to prepare your manuscript according to the JISTaP manuscript guidelines.

Any question? Seonhee Lee (managing editor) : jistap@kisti.re.kr

JISTaP
Journal of Information Science Theory and Practice

Information for Authors

The Journal of Information Science Theory and Practice (JISTaP), which is published quarterly by the Korea Institute of Science and Technology (KISTI), welcomes materials that reflect a wide range of perspectives and approaches on diverse areas of information science theory, application and practice. JISTaP is an open access journal run under the Open Access Policy. See the section on Open Access for detailed information on the Open Access Policy.

A. Originality and Copyright

All submissions must be original, unpublished, and not under consideration for publication elsewhere. Once an article is accepted for publication, all papers are accessible to all users at no cost. If used for other researches, its source should be indicated in an appropriate manner and the content can only be used for uncommercial purpose under Creative Commons license.

B. Peer Review

All submitted manuscripts undergo a single-blind peer review process in which the identities of the reviewers are withheld from the authors.

C. Manuscript Submission

Authors should submit their manuscripts online via AccessON Peer Review Management System Plus (ACOMS^{+}). Online submission facilitates processing and reviewing of submitted articles, thereby substantially shortening the paper lifecycle from submission to publication. After checking the manuscript's compliance to the Manuscript Guidelines, please follow the "Online Submission" hyperlink in the top navigation menu to begin the online manuscript submission process.

D. Open Access and Creative Commons License Statement



All JISTaP content is Open Access, meaning it is accessible online to everyone, without fee and authors' permission. Open Access articles are automatically archived in the Korea Institute of Science and Technology Information (KISTI)'s Open Access repository (AccessON). All JISTaP content is published and distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/). Under this license, the authors retain full ownership of their work, while permitting anyone to use, distribute, and reproduce the content in any medium, as long as the original authors and source are cited. For any reuse, redistribution, or reproduction of a work, users must clarify the license terms under which the work was produced.

E. Manuscript Guidelines

Manuscripts that do not adhere to the guidelines outlined below will be returned for correction. Please read the guidelines carefully and make sure the manuscript follows the guidelines as specified. We strongly recommend that authors download and use the manuscript template in preparing their submissions.

1. Page Layout :

All articles should be submitted in single column text on standard Letter Size paper (21.59 × 27.94 cm) with normal margins.

2. Length :

Manuscripts should normally be between 4,500 and 9,000 words (10 to 20 pages).

3. File Type :

Articles should be submitted in Microsoft Word format. To facilitate the manuscript preparation process and speed up the publication process, please use the manuscript template.

4. Text Style :

- Use a standard font (e.g., Times New Roman) no smaller than size 10.
- Use single line spacing for paragraphs.
- Use footnotes to provide additional information peripheral to the text. Footnotes to tables should be marked by superscript lowercase letters or asterisks.

5. Title Page :

The title page should start with a concise but descriptive title and the full names of authors along with their affiliations and contact information (i.e., postal and email addresses). An abstract of 150 to 250 words should appear below the title and authors, followed by keywords (4 to 6).

Author1
Affilliation, Postal Address. E-mail
Author2
Affilliation, Postal Address. E-mail

ABSTRACT
A brief summary (150-250 words) of the paper goes here.

Keywords: 4 to 6 Keywords, separated by commas.

6. Numbered Type :

1. INTRODUCTION

All articles should be submitted in single column text on standard letter size paper (21.59 × 27.94 cm) with normal margins[1. Text should be in 11-point standard font (e.g., Times New Roman) with single line spacing. [1 Normal margin dimensions are 3 cm from the top and 2.54 cm from the bottom and sides.

2. SECTIONS

The top-level section heading should be in 14-point bold all uppercase letters.

2.1. Subsection Heading 1

The first-level subsection heading should be in 12-point bold with the first letter of each word capitalized.

2.1.1. Subsection Heading 2

The second-level subsection heading should be in 11-point italic with the first letter of each word capitalized.

7. Figures and Tables :

All figures and tables should be placed at the end of the manuscript after the reference list. To note the placement of figures and tables in text, "Insert Table (or Figure) # here" should be inserted in appropriate places. Please use high resolution graphics whenever possible and make sure figures and tables can be easily resized and moved.

Figure

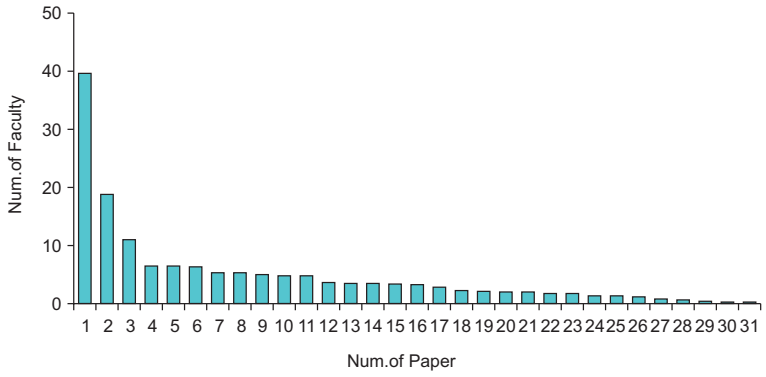


Fig. 1. Distribution of authors over publication count.

Table

Table 1. The title of table goes here

Study	Time period study	Data
Smith Wesson (1996)	1970 - 1995	684 papers in 4 SSCI journals
Reeves [a (2002)	1997 - 2001	597 papers in 3 SSCI journals
Jones Wilson [b (2011)	2000 - 2009	2,166 papers in 4 SSCI journals

[a Table footnote a goes here
[b Table footnote b goes here

8. Acknowledgements :

Acknowledgements should appear in a separate section before the reference list.

9. Conflicts of Interest :

Authors of manuscripts must disclose any potential conflicts of interest at the time of submission. Statements on conflict of interest have no influence on the editorial decision to publish.

10. Citations :

Citations in text should follow the author-date method (authors' surname followed by publication year).

- Several studies found... (Barakat et al., 1995; Garfield, 1955; Meho & Yang, 2007).
- In a recent study (Smith & Jones, 2011)...
- Smith and Jones (2011) investigated...

11. Reference List :

Reference list, formatted in accordance with the American Psychological Association (APA) style, should be alphabetized by the first authors last name.

Journal article

- Author, A., Author, B. & Author, C. (Year). Article title. *Journal Title*, volume(issue), start page-end page.
- Smith, K., Jones, L. J., & Brown, M. (2012). Effect of Asian citation databases on the impact factor. *Journal of Information Science Practice and Theory*, 1(2), 21-34.

Book

- Author, A., & Author, B. (Year). *Book title*. Publisher Name.
- Smith, K., Jones, L. J., & Brown, M. (2012). *Citation patterns of Asian scholars*. Sage.

Book chapter

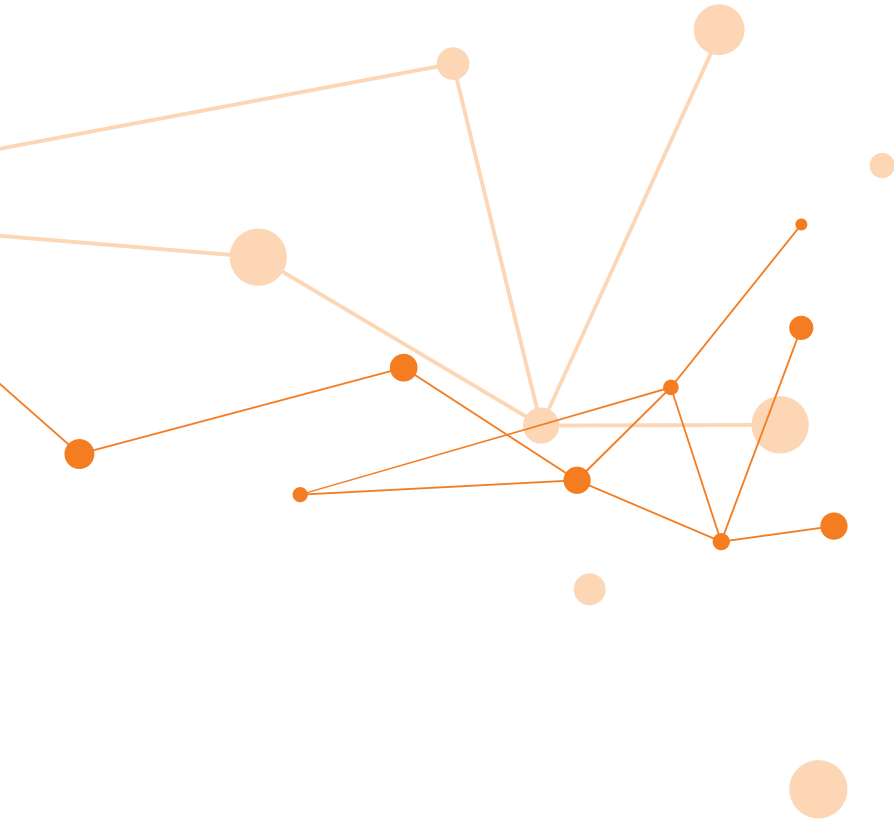
- Author, A., & Author, B. (Year). Chapter title. In A. Editor, B. Editor, & C. Editor (Eds.), *Book title* (pp. xx-xx). Publisher Name.
- Smith, K. & Brown, M. (2012). Author impact factor by weighted citation counts. In G. Martin (Ed.), *Bibliometric approach to quality assessment* (pp. 101-121). Springer.

Conference paper

- Author, A., & Author, B. (Year, Month Day). Article title. In A. Editor & B. Editor (Eds.), *Conference title* (pp. xx-xx). Publisher Name.
- Lin, H., Sun, B., Wu, J. & Xiong, H. (2016, June 24-26). Topic detection from short text: A term-based consensus clustering method. In B. Yang (Ed.), *2016 13th International Conference on Service Systems and Service Management (ICSSSM 2016)* (pp. 1-6). IEEE.

Online document

- Author, A., & Author, B. (Year). *Article title*. URL.
- Smith, K. & Brown, M. (2010). The future of digital library in Asia. *Digital Libraries*, 7,111-119. http://www.diglib.org/publist.htm.



JISTaP

Journal of Information
Science Theory and Practice

KiSTi
www.kisti.re.kr
66, Hoegi-ro, Dongdaemun-gu, Seoul, Republic of Korea
(T) +82-2-3299-6105 (F) +82-2-3299-6067
https://www.jistap.org