TECHNOLOGICAL UNIVERSITY OF THE PHILIPPINES

COLLEGE OF SCIENCE

DEVELOPMENT OF WEB-BASED INFORMATION RETRIEVAL SYSTEM OF ACADEMIC PAPERS WITH AUTOMATIC TAGGING USING PARTICLE SWARM OPTIMIZATION ALGORITHM IN THE TECHNOLOGICAL UNIVERSITY OF THE PHILIPPINES – MANILA

A Research Presented to the

Faculty of the College of Science

Technological University of the Philippines – Manila

In Partial Fulfillment

of the Requirements for the subject

Masteral in Information Systems

by:

Brillo, Alexandria Lee G.

Capistrano, Deazelle M.

Maximo, Raine Francesca

Velasquez, Erika F.

Villaruz, Maria Lourdes T.

Professor Dolores Montesines

S.Y 2024 – 2025

TABLE OF CONTENTS

CHAPTER I

INTRODUCTION

This chapter provides an overview of the context, the subject under consideration, and the significance of the study. The study covers the objectives, scope, and limitations of the research. This chapter will present a comprehensive summary of the research and a concise analysis of the concepts addressed in the study.

Introduction

This study presents the development of a web-based Information Retrieval System (IRS) for Academic Papers, also known as Research Papers. The system incorporates Automatic Tagging using the Particle Swarm Optimization (PSO) algorithm. Its goal is to improve research paper accessibility for students and faculty at the Technological University of the Philippines, Manila Campus. The system offers a convenient and efficient way for users to locate materials that are relevant to their respective courses. According to the study of Martin-Martin et al. (2020), academic research greatly depends on the ability to obtain and investigate scholarly articles, making Information Retrieval Systems (IRS) extremely important for researchers.

Background of the Study

In today's digital age, where information is readily available online, the ability to find and evaluate credible sources is essential (Long, C. et al., 2021). The internet has become the primary source of information for many individuals (Feng, S. et al., 2021). In academic research, information retrieval plays a crucial role allowing scholars to access and explore a wide range of scholarly publications (Zhang et al., 2020). Academic research heavily relies on accessing and exploring scholarly publications, making Information Retrieval Systems crucial for scholars

(Martín-Martín et al., 2021). These systems provide scholars with the ability to search for relevant papers in their specific area, helping them stay updated with the latest research and findings (Chen & Zhang, 2021). While search engines provide convenience and efficiency in information gathering (Giomelakis Dimitrios et al., 2021), students at the Technological University of the Philippines, Manila often face challenges in accessing high-quality academic research materials tailored to their specific courses.

The Technological University of the Philippines (TUP) is a state university in Manila, Philippines, established in 1901 as the Philippine School of Arts and Trades, which later became the Philippine College of Arts and Trades in 1937. TUP was officially established on July 11, 1978, through Presidential Decree No. 1518, which transformed the Philippine College of Arts and Trades into the Technological University of the Philippines. The university is mandated to provide higher and advanced vocational, technical, industrial, technological, and professional education and training in the industries, technology, and practical arts leading to certificates, diplomas, and degrees.

TUP is recognized for its excellence in engineering and technology education, offering a wide range of programs across various colleges, including the College of Engineering, College of Industrial Technology, College of Industrial Education, College of Architecture and Fine Arts, College of Science, and College of Liberal Arts. The College of Industrial Technology, in particular, has its roots in the Technical Department of the Philippine School of Arts and Trades, which was established in 1937. The university is committed to providing quality education, progressive leadership in applied research, developmental studies in technical, industrial, and technological fields, and production using indigenous materials. TUP also aims to effect

technology transfer in the countryside and assist in the development of small and medium-scale industries in identified growth centers.

To facilitate research and training, TUP has established the Integrated Research and Training Center (IRTC), which aims to provide up-to-date training and research outcomes. The TUP Library System, which consists of libraries on four campuses, subscribes to the philosophy of cooperation and partnership, aiding students, and faculty in search of diverse academic library collections. Despite the vast amount of information available, students at TUP struggle to find current and relevant research papers.

One common challenge they face is the difficulty of locating current and pertinent research papers tailored to their specific academic courses. Postgraduate students and early career academics face challenges in choosing a practical approach, designing an efficient search strategy, locating relevant literature, determining the appropriate scope, and effectively synthesizing and critiquing the literature (Daniel, 2022). This struggle often arises due to the constant evolution of research in their field, making it a challenge to pinpoint the most up-to-date and relevant sources. In addition to this issue, students may also have limited access to past research conducted within the school or institution (Miller, 2023). To address these challenges, students often need to employ effective search strategies, leverage academic databases, and seek guidance from professors or librarians to ensure they access the most appropriate materials for their studies.

Recognizing the difficulty of accessing research at TUP Manila Campus, this study aims to develop an Information Retrieval System (IRS) for academic papers that can provide TUP students with a more efficient and effective way to search and access relevant information. By integrating advanced search algorithms, accessibility, and interoperability with other IRS and academic platforms, the proposed IRS aims to provide a more efficient and effective way for TUP

students to search and access research papers within the school. This system will enhance students' research capabilities, enabling them to navigate the academic literature landscape with ease and precision, thereby fostering a more productive and engaging research experience.

Objective of the Study

The primary objective of this project is to develop a web-based Information Retrieval System (IRS) for the Technological University of the Philippines, Manila Campus, aimed at efficiently locating research papers conducted by students. This project specifically aims to:

- **Improve Accessibility**: Enhance the accessibility of research papers for both students and faculty members by implementing a user-friendly system that allows easy retrieval of relevant information.
- **Automatic Tagging with PSO Algorithm**: Implement the Particle Swarm Optimization (PSO) algorithm for Automatic Tagging of uploaded PDFs, facilitating effective categorization and organization of research papers based on content.
- **Search Functionality**: Develop a robust search functionality within the system to enable users to quickly and accurately locate specific research papers based on keywords, authors, topics, or other relevant criteria.
- **Systematic Organization**: Implement a systematic organization and categorization system for research papers, making it easier for users to navigate and find papers within specific subject areas or academic disciplines.
- **User Authentication**: Establish a reliable user authentication system with tiered permissions to control access to specific functionalities, ensuring that only authorized users can upload, edit, or delete research papers for enhanced security.

Significance of the Study

College students are required to complete a project known as a *thesis*, which often involves both written documentation and a practical component that varies depending on the specific course of study. Students build up and retain their skills and knowledge acquired during their four to five years stay in college in order to comply, demonstrate, and validate their abilities. This study will redound to benefit the following:

To Students, it provides easy access to past research conducted by TUP-Manila students, making it easier to find articles that are relevant to their courses. This resource reduces the process of finding appropriate materials, promoting a cooperative and enhanced academic atmosphere within the TUP-Manila community.

In the Field of Technology, is to provide an understanding of the creation of advanced technologies that are designed to make educational processes more efficient. The ultimate objective is to utilize innovation to greatly improve and elevate the overall standard of education by incorporating state-of-the-art technological solutions.

For Future Researchers, use this as a reference for further development and a guide for future studies, offering a blueprint to navigate the complexity of their research endeavors.

For Future Research Projects, it can function as a point of reference to improve and perfect the current systems. By utilizing the knowledge and techniques established in this research, researchers can expand upon a solid foundation, facilitating the development of innovative progress and the establishment of more precise systems. This reference point serves the purpose of not only duplicating successful models but also creating opportunities for customizing solutions to specific domains or enhancing discovered limitations during the research.

For the Higher Education Institutions, it serves as a method to centralize and exhibit all students' research papers. This study aims to showcase the combined scholarly work and assist students in finding relevant research for their courses. It promotes innovation and academic achievement within the university.

Scope and Delimitation

This project aims to develop a web-based IRS for research papers, specifically for TUP, Manila Campus. The primary focus is to enhance the accessibility of research papers for students and faculty by providing a convenient and efficient means of locating materials relevant to their respective courses.

An innovative aspect of this system is the implementation of the Particle Swarm Optimization (PSO) algorithm, which will be responsible for the Automatic Tagging of research papers based on their content. This will significantly improve the accuracy and speed of the retrieval process. However, this project is limited to the scope of the university's research papers and does not extend to external publications or resources. Additionally, Automatic Tagging will be constrained by the accuracy of keyword extraction and the predefined course categories established within the system.

CHAPTER II
CONCEPTUAL FRAMEWORK

This chapter presents the review of related literature, related studies, a conceptual model, and the definition of terms that are viewed by the researchers and have a significant bearing on the present study.

REVIEW RELATED LITERATIRE

Information plays a crucial role in our everyday lives. Over time, the significance of storing and retrieving information has become widely recognized. With the swift advancement of technology, it has become simpler for individuals to store vast amounts of data and extract valuable information from it. Information retrieval involves enabling users to find relevant information within unstructured documents (Wable R. et al.,2021).

Information retrieval involves the science of locating information within documents, finding the documents themselves, and searching for metadata, as well as databases containing texts, images, or sounds (Wable R. et al.,2021). The purpose of an IRS is to provide accurate information to the user. To accomplish this, information is meticulously stored, collected, and organized across different subjects, making it easily accessible when required. In modern libraries and archives, information retrieval includes searching full-text databases, finding items from bibliographic databases, and document delivery through a network (Dr. Manjunatha S et al., 2022). In information retrieval, searches can utilize full-text or other forms of content-based indexing. As stated by Agbele (2018), users frequently input ad-hoc keywords in their search queries, which are personalized terms not pre-defined within the system. It becomes the responsibility of information retrieval systems (IRS) to precisely grasp the user's information requirements and contextual

nuances. Keywords play a pivotal role in information retrieval, as they dictate document relevance, assist in document classification, and streamline the indexing process (Pin Ni et al., 2020).

According to Sahal Manasia et al. (2023), the importance of archiving and retrieving information has been recognized for several years. With the widespread use of computers, the ability to extract valuable information from large collections has become essential. As a result, information retrieval has emerged as a significant research area within computer science and has gained prominence across various fields such as business, healthcare, agriculture, medicine, law, and others. Information retrieval involves locating material, often in unstructured document form, that contains the necessary information.

Moreover, the challenges identified in information retrieval among engineering students underscore the pressing need for advancements in search algorithms and user interfaces tailored to the specific requirements of diverse fields. This highlights the ongoing evolution and refinement of information retrieval techniques to address the intricacies of accessing relevant data in specialized domains.

In the study by Navitas et al. (2022), which analyzed the types of online database websites favored by students as their primary information sources, investigated the underlying reasons for their preference of these websites, and identified the challenges encountered during information seeking through online databases, it was found that engineering students primarily rely on search engines as their main information source due to their richness and flexibility. However, despite this reliance, these students often face challenges in effectively retrieving information using different keywords, primarily because the broad scope and complexity of engineering topics require precise and nuanced search queries, making it difficult to find relevant information amidst the vast amount of available data.

Moreover, in the study of Wu et al. (2022), it is emphasized that a computer-based legal information retrieval system serves as a pivotal tool in navigating the complexities of legal data. Through its comprehensive search functionality, users can efficiently access a plethora of legal resources, ranging from statutes to case law. Moreover, the system offers curated legal information content, ensuring users have access to accurate and up-to-date data essential for decision-making processes. Additionally, by providing robust management services, the system facilitates seamless organization and retrieval of information, contributing to enhanced productivity and efficiency. As highlighted by Wu et al. (2022), the adaptability of such systems to changing demands underscores their significance in driving technological advancements and fostering growth in our economic society.

Furthermore, according to Lima et al. (2022), in their paper titled "Information Storage and Retrieval System: An Analysis of the Impact of Variables and Measures Aimed at the Organization and Retrieval of Information Centered on the User," the authors assert that the efficacy of an information retrieval system relies on several key components. Chief among these is the quality of organization within the system, which directly influences its ability to manage and categorize data effectively. Additionally, Lima et al. (2022) emphasize the importance of timely and precise retrieval of the most relevant information, ensuring that users can access the data they need efficiently. Furthermore, the authors highlight the significance of adopting a systemic perspective, with the user positioned as the central focus of the information retrieval process. This user-centric approach acknowledges the diverse needs and preferences of users, thereby enhancing the usability and overall effectiveness of the system. In summary, Lima et al. (2022) suggest that meticulous organization, targeted retrieval strategies, and a user-centric approach are critical for optimizing the performance of information retrieval systems.

According to MacFarlane, A. (2007), common challenges encountered in teaching mathematics for information retrieval to postgraduate students encompass overcoming misconceptions and devising practical strategies to enhance both delivery and learning outcomes. This assertion underscores the complexities inherent in conveying mathematical concepts relevant to information retrieval effectively. Moreover, MacFarlane highlights the necessity of addressing misconceptions that may impede students' comprehension and mastery of these concepts. Additionally, the author emphasizes the importance of implementing practical approaches aimed at fostering improved delivery methods and enhancing learning outcomes. By acknowledging and addressing these challenges, educators can better equip postgraduate students with the requisite mathematical skills essential for proficient information retrieval practices.

Manasia, S., Rai, S., Kalwar, M., Yadav, V., & Shinde, V. (2023). Information Retrieval System for College Search. International Journal of Scientific Research in Computer Science, Engineering and Information Technology. https://doi.org/10.32628/cseit2390224.

S, D. (2022). DIGITALISATION OF LIBRARIES (DIGITAL LIBRARIES) AND ITS SCOPE. EPRA International Journal of Research & Development (IJRD). https://doi.org/10.36713/epra10724.

Wable, R. (2021). Information Retrieval in Business. https://doi.org/10.36227/TECHRXIV.14709456.V1.

Ni, P., Li, Y., & Chang, V. (2020). Research on Text Classification Based on Automatically Extracted Keywords. Int. J. Enterp. Inf. Syst., 16, 1-16. https://doi.org/10.4018/ijeis.2020100101.

Agbele, K., Ayetiran, E. and Babalola, O. (2018). A Context-Adaptive Ranking Model for Effective Information Retrieval System. International Journal of Information Science. 8(1), 1-12

Kumar, S., Garg, A., & Haider, M. (2023). PSO based Web Documents Prioritization for Adaptive Websites using multi-Criteria. 2023 6th International Conference on Information Systems and Computer Networks (ISCON), 1-6. https://doi.org/10.1109/ISCON57294.2023.10112161.

Navitas, P., & Febrianti, Y. (2022). Analysing engineering students' information retrieval behaviour using online databases as information sources. Jurnal Kajian Informasi & Perpustakaan. https://doi.org/10.24198/jkip.v10i2.35264.

Wu, F., Ji, Y., & Shi, W. (2022). Design of a Computer-Based Legal Information Retrieval System. Computational Intelligence and Neuroscience, 2022. https://doi.org/10.1155/2022/6942773.

Lima, G., & Campos, M. (2022). Information Storage and Retrieval System: an analysis of the impact of variables and measures aimed at the organization and retrieval of information centered on the user. RDBCI: Revista Digital de Biblioteconomia e Ciência da Informação. https://doi.org/10.20396/rdbci.v20i00.8667925/28663.

MacFarlane, A. (2007). Pedagogic challenges in information retrieval - teaching mathematics to postgraduate information science students. , 6-6. https://doi.org/10.14236/EWIC/TLIR2007.6.