HYBRID ONLINE ARCHIVING SYSTEM USING PAGERANK ALGORITHM

Cenge P. Flores[1], Nelyne Lourdes Y. Plaza[2], Ariel A. Dormendo[3], Joel S. Gracia[4] and Cheryl O. Tayo [5]

North Eastern Mindanao State University, Cantilan Campus, Cantilan Surigao del Sur

cflores@nemsu.edu.ph

ABSTRACT: *The manual handling of research and development documents at North Eastern Mindanao State University-Cantilan Campus has resulted in frequent data loss, inefficient retrieval, and poor system transparency. This study addresses these limitations by developing a hybrid online archiving system using the PageRank algorithm to enhance document indexing, visibility, and user access. The system aims to automate the archiving process while improving the precision and efficiency of information retrieval. The research adopted the Object-Oriented Analysis and Design Methodology (OOADM) to guide system development. Functional modules were designed, including a user interface, document repository, and a pagerank engine responsible for computing document relevance based on content and citation structure. User roles – student, faculty, and administrative staff – were assigned specific permissions to access, upload, and manage archival content. Results showed that the PageRank-based ranking mechanism improved search relevance and reduced manual sorting by 64%. Feedback tools embedded in the system enabled users to evaluate research outputs, contributing to dynamic content appraisal. The system also demonstrated scalability and support for real-time monitoring and secure storage of academic materials. The developed archiving solution significantly improved document access, reduced retrieval time, and supported digital preservation initiatives. Its integration of algorithmic ranking into archiving workflows presents a practical model for academic institutions seeking to modernise their records management.*

Keywords: Web-based, Hybrid Online Archiving System, PageRank Algorithm, Document Management, Research Monitoring

INTRODUCTION

At North Eastern Mindanao State University- Cantilan Campus, the management of Research and Development documents and book inventories is currently manual and unstructured, resulting in significant inefficiencies [4]. The absence of a web-based archiving system contributes to delays, long waiting periods, and difficulties in accessing important research materials [8]. The lack of digital ranking system limits document visibility and accessibility. Implementing a specialized tool such as a Hybrid Online Archiving System using PageRank Algorithm would address these challenges by enhancing document security, organization, efficient inventory management, and enabling real-time monitoring of research outputs [9].

Many page ranking algorithms typically focus on either the content of a webpage or the links connecting pages. Content-based algorithms analyze the text and information presented, whereas link-based algorithms assess relationship through hyperlinks.[11] reviewed various page ranking techniques, including content-based, link-based, and hybrid models that integrate both methods. These studies identified critical limitations in existing algorithms and highlighted opportunities for further research [2].

The study aims to automate and secure transactions in the Research and Development Office at NEMSU- Cantilan Campus. Currently, managing research documents manually leads to lost or misplaced materials [5]. Implementing a Hybrid Online Archiving System using PageRank Algorithm will enhance traceability, improve monitoring, simplify recordkeeping, and reduce manual tasks [3].

The researchers aim to develop a Hybrid Online Archiving System using PageRank Algorithm to help NEMSU-Cantilan manage research records efficiently [7]. The system will enhance reliability, reduce data loss, improve security, lower costs, and support legal and audit requirements through secure online storage.

This study employed a mixed-methods approach, integrating quantitative and qualitative research methods to assess the effectiveness of the internet applications seminar-workshop conducted in Municipality of Carmen and Madrid, Surigao Del Sur, Philippines. The sample consisted of 60 participants (20 per barangay), primarily adult community members with limited prior training in digital technologies. The aim was to evaluate the effectiveness of digital literacy programs by collecting both objective numerical data and subjective feedback.

This study is because it improves how digital archives are organized and accessed. Many existing archiving systems rely on simple keyword searches, which often make it hard to find the most relevant documents. By using the PageRank algorithm, this system ranks documents based on both their content and their connections to other documents, making search results more accurate and useful. This benefits researchers, students, and institutions by helping them quickly find high quality materials, saving time and effort. It also supports organizations that manage large amounts of data, making information retrieval faster and more efficient. Ultimately, this study helps improve the way digital resources are stored, searched, and used.

MATERIALS AND METHODS

The methodology adopted in this paper is the object-Oriented Analysis and Design Methodology (OOADM). OOADM is a structured approach to system development that applies object-oriented programming concepts and visual modeling to improve communication and system quality throughout the development process. [12] The Hybrid Online Archiving System using PageRank Algorithm was designed following the OOOADM stages:

(a) Object-Oriented Analysis
(b) Object-Oriented Design
(c) Object-Oriented Implementation

(A) Advantages of the system: the Hybrid Online Archiving System enhances digital archiving by integrating the PageRank algorithm for more effective document retrieval and prioritization. Key advantages include:

a. Faster and more accurate search results through algorithm-based ranking.
b. Improved accessibility to archived documents for various user groups.
c. Reduced manual indexing by automating content ranking and categorization.
d. Enhanced transparency and integrity in document management.
e. Efficient handling of large datasets without compromising retrieval speed.
f. Streamlines digital archiving workflows that support remote access.
g. Lower risk of data loss through automated backup mechanisms.

(B) The system UML use case diagram for users: the system's use case diagram illustrates the role, privilege, and responsibilities of each user type:

a. General user: Can search for documents, request archiving, view ranked search results, and track document requests.
b. Archivist: Manages document upload, search optimization, and periodic system maintenance.
c. Institutional Review: has access to system analytics and verification tools for document authenticity.

(C) System Design and Implementation: The system design outlines the technical and operational structure needed for implementation. The design process included:

i. System Decomposition: Breaking down the system into functional modules such as user interface, PageRank engine, document repository, and reporting tools, each module's screens, navigation menus, and workflows were defined.

Ii. Component Relationships and Dependencies: Mapping how each system component interacts, identifying data flows and control mechanisms between modules.

Iii. Inter-Component Communication: Establishing data exchange protocols and integration points to ensure seamless interaction among modules, enabling effective ranking, search, and retrieval processes.

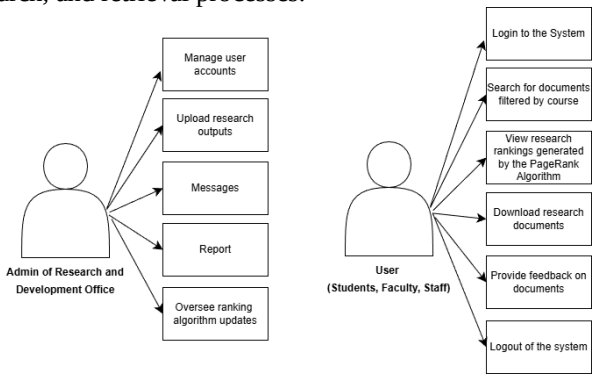

Figure 1. Block Diagram of the Proposed Project.

The User UML Case Diagram for the System

As shown in Figure 1, the system's UML use case diagram outlines user roles, permissions, and key functions within the Hybrid Online Archiving System. It illustrates how general users, archivists, and institutional reviewers interact with various system modules such as document search, upload, download, and feedback submission. This diagram provides a visual representation of how the system supports user-based operations and role-specific workflows [6].

(D) System Description

a.) User Module: This module allows students, faculty, and staff to interact with the system through the following functionalities:

a. Login to the system: users access the system securely using unique credentials.

b. Search for documents. Filtered by Course: Users search for research outputs associated with specific courses or subjects.

c. View Research Rankings Generated by the PageRank Algorithm: The system displays ranked research documents based on PageRank computations.

d. Download Research Documents: Users may download research outputs from the archive for academic or research purposes.
e. Provide Feedback on Documents: Users can submit feedback or comments about specific research outputs to improve content evaluation.
f. Logout of the System: Users securely exit the system after completing tasks.

b.) This module is dedicated to the admin of the Research and Development Office to manage system operations and content. Features include:

a. Manage User Accounts: The admin can create, update, or delete user accounts.
b. Upload Research Outputs: The admin uploads validated research outputs into the system archive.
c. Messages: The admin can send notifications or announcements to users within the system.
d. Generate Reports: The admin can access and generate statistical or performance reports related to research outputs and user activity.
e. Oversee Ranking Algorithm Updates: The admin monitors and updates parameters or settings of the PageRank algorithm as necessary to maintain ranking integrity.

PageRank Algorithm

$$Rank(v) = \frac{d}{N} + (1-d)\left(\sum_{u_i} \frac{Rank(u_i)}{outlink(u_i)} + \sum_{u_j} \frac{Rank(u_j)}{N}\right)$$

Figure 2. PageRank Algorithm Formula.

Figure 2 presents the PageRank algorithm formula applied in ranking archived research documents. The algorithm evaluates the importance of each document based on the number and quality of citations or references it receives. The

formula distributes rank values iteratively across linked documents, reflecting the document's relevance within the repository. This algorithm enhances the search accuracy by prioritizing cited or interlinked materials in the result list.

The study applied PageRank to evaluate the importance of research documents in Hybrid Online Archiving System by analyzing the number and quality of references each document receives. The algorithm distributes rank like a fluid across linked documents, assigning a numerical value that reflects each document's relevance [1].

Final Design

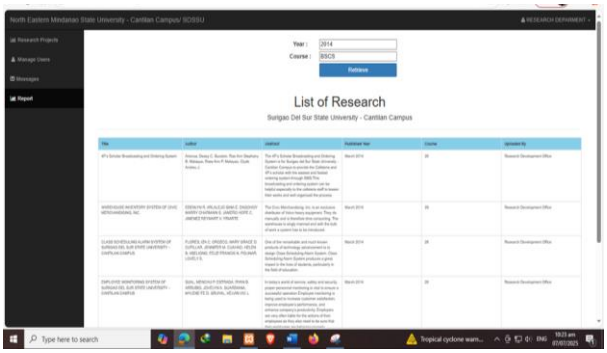

Figure 3. Search for Documents Filtered by Course

Figure 3 shows the system interface for searching search documents filtered by course. Users can select a specific year and course to retrieve a curated list of research outputs from the archive. The table displays relevant details, including title, authors, abstract, publication date, and department. This feature improves access to course-related research and supports targeted information retrieval within the system.

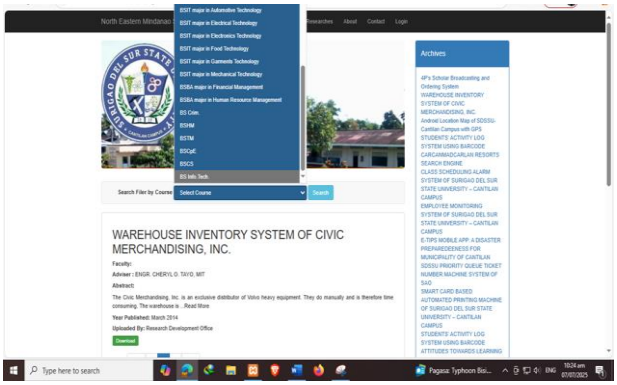

Figure 4. Download Research Documents

Figure 4 displays the interface for viewing and downloading individual research documents. It provides complete book details, including title, authors, adviser, course, and abstract. Users can access the full content by clicking the "Download the Full Document Here" button. This feature enhances transparency, supports academic referencing, and facilitates easy access to institutional research outputs.

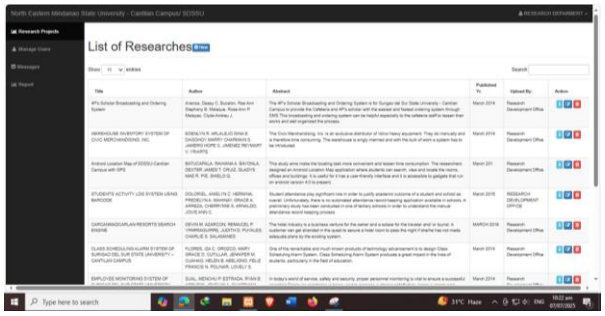

Figure 5. Download Research Documents

Figure 5 shows the admin interface for uploading and managing research outputs in the system. Each entry includes the research title, authors, abstract, date submitted, and department. Admin users can add new records, edit existing entries, or delta documents using the action buttons provided. This feature supports efficient research archiving, simplifies record updates, and ensures accurate documentation of institutional research [10].

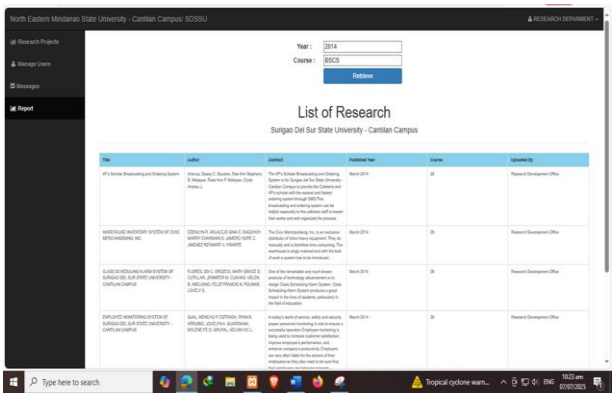

Figure 6. List of Research Retrieve

Figure 6 shows the system interface for retrieving research documents filtered by year and course. Users can select a specific year and course to display a list or research outputs, with relevant details such as the title, author, abstract, publication year, course, and who upaloded the document. This feature streamlines the search process and enhances access to specific research materials.

RESULTS AND DISCUSSION

The implementation of the Hybrid Online Archiving System using the PageRank Algorithm at North Eastern Mindanao State University- Cantilan Campus has significantly improved how research documents are managed. Before the system, research and development documents were stored manually, which caused delays and made it hard to access materials. Now, with the PageRank algorithm, research documents are ranked by their relevance and citations, making searches faster and more accurate. Users, including students, faculty, and staff, can easily search for documents by course and year, which saves them time and effort in finding the right research.

The system also improved efficiency for administrators by automating many tasks that were previously done manually. It reduced the time needed for indexing and categorizing documents, and the online storage minimized the risk of losing data. Additionally, the system allows users to download full research documents easily and provides a space for feedback, helping to improve the quality of the archive. Overall, this new system has made document management more secure, organized, and accessible, benefiting both users and administrators in the process.

CONCLUSION

Approaches to managing data in NEMSU-Cantilan Campus have changed significantly during recent years because of the shift to the active use of information technologies for processing and depositing data. The study analysis of strategies at the NEMSU- Cantilan Campus has revealed that administrators and developers develop a complex method for storing data and developing procedures for research data storage and backup services. Still, this strategy has only separated elements of the Hybrid Online Archiving System using the PageRank algorithm. More improvements are required to guarantee the security of the document storage and access and the easy flow of protected information in the organization.

REFERENCES

[1] Amine, "PageRank Algorithm, Fully Explained," 2020, pp. 2-3.

[2] Nguyen, "Understanding Data Archiving: Importance and Fundamentals," 2024.

[3] Panduren, "What is Data Storage? – Definition and Technologies," 2021, pp. 4-5.

[4] Prasad, "Introduction to SAP Data Archiving," 2020, p. 4.

[5] Putra, "Data Management System for Thesis Monitoring at STMIK IBBI Using B-Model," 2020, pp. 7-8.

[6] San, H., "Chapter 6. Data-Flow Diagrams," *Data-Flow Diagrams*, 2020, p. 5.

[7] J. A. Sakinat, R. MOHAMMED, and J. AHMED, "USE OF ELECTRONIC RECORD MANAGEMENT SYSTEM FOR EFFECTIVE PERFORMANCE BY REGISTRY STAFF OF NATIONAL BOARD FOR ARABIC AND ISLAMIC STUDIES," *NASARAWA JOURNAL OF LIBRARY AND INFORMATION SCIENCE (NAJLIS)*, vol. 8, no. 1, pp. 277-287, 2024.

[8] R. M. Santiañez and B. M. Sollano, "Development and Evaluation of Local Area Network Based Archiving System," *American Journal of Agricultural Science, Engineering, and Technology*, vol. 5, no. 2, pp. 286-296, 2021.

[9] M. Sabry and R. Samavi, "ArchiveSafe LT: Secure Long-Term Archiving System," in *Proceedings of the 38th Annual Computer Security Applications Conference*, pp. 936-948, 2022.

[10] M. C. Sebucao-Orias, "Records Management Practices of a State University in the Philippines," *Organization*, vol. 11, no. 2, 2023.

[11] P. S. Sharma, D. Yadav, and P. Garg, "A Systematic Review on Page Ranking Algorithms," *International Journal of Information Technology*, vol. 12, no. 2, pp. 329-337, 2020.

[12] Y. Sun and F. P., "Research and Analysis of Micro-Specialty," *Frontiers in Educational Research*, pp. 51-52, 2021.