

**Web-Based Thesis Archiving System
with Plagiarism Scanner and AI Detector**

A Thesis

Presented to the Faculty of

College of Computing Studies

Don Honorio Ventura State University

In Partial Fulfillment

of the Requirements for the Degree

Bachelor of Science in Information Technology

by:

Naluz, Alfredo Roi B.

Actub, John Ford C.

Sapnu, Adrian Paul M.

Guina, Edmyr Jerome B.

Manuel, King Jhusfer G.

Tique, Dhen Andrei P.

Javier, Mc Karl Kennedy C.

September 2023

TABLE OF CONTENTS

	Page
Title Page.....	i
Approval Sheet.....	ii
Acknowledgement.....	iii
Dedication.....	iv
Table of Contents.....	v
List of Tables.....	vi
List of Figures.....	vii
List of Appendices.....	viii
Abstract.....	ix
Chapter I	1
Introduction.....	2
Project Context.....	
Purpose and Description.....	
Objectives of the Study.....	
Scope and Limitation.....	
Definition of Terms.....	
Chapter II	
Related Literature.....	
System Technical Background.....	
Conceptual Framework.....	
Chapter III	
Methodology.....	
Requirements Analysis.....	
Requirements Documentation.....	
Design of Software System.....	
Product and Process.....	
Development and Testing.....	
Implementation Plan.....	
ISO/IEC 25010 or Software Quality Model.....	
Respondents of the Study.....	
Instruments of the Study.....	
Research Tools and Instruments.....	
Statistical Treatment of the Data.....	
Chapter IV	
Result and Discussion.....	
Chapter V	
Summary, Conclusion and Recommendations.....	
Appendices	



LIST OF TABLES

Table		Page
1	Hardware Specification	
2	Software Specification.....	
3	Likert Scale	
4	Overall Evaluation Rating Scale and Interpretation	
5	Alpha Testers of the Study	
6	Beta Testers of the Study	
	Assessment of the Alpha Testers based from the ISO 25010 software quality standards in terms of the Functional Suitability	
	Assessment of the Alpha Testers based from the ISO 25010 software quality standards in terms of the Performance Efficiency.....	
	Assessment of the Alpha Testers based from the ISO 25010 software quality standards in terms of the Usability	
	Assessment of the Alpha Testers based from the ISO 25010 software quality standards in terms of the Reliability	
	Assessment of the Beta Testers based from the ISO 25010 software quality standards in terms of the Functional Suitability	
	Assessment of the Beta Testers based from the ISO 25010 software quality standards in terms of the Performance Efficiency.....	
	Assessment of the Beta Testers based from the ISO 25010 software quality standards in terms of the Usability	
	Assessment of the Beta Testers based from the ISO 25010 software quality standards in terms of the Reliability	
	Summary and Frad Mean of the Proposed Study.....	

Only if applicable.

LIST OF FIGURES

FIGURE		Page
1	Conceptual Framework of the Study.....	
2	Development Methodology.....	
3	Organizational Chart.....	
4	Conceptual Framework.....	
5	Context Diagram.....	
6	Entity Relationship Diagram.....	
7	System Process Flowchart.....	
8	Data Dictionary.....	

Only if applicable.

LIST OF APPENDICES

APPENDIX		Page
A	Request Letters (Adviser).....	123
B	Request Letters (Locale).....	
C	Plagiarism Checker Certificate.....	
D	Grammar Check Certificate.....	
E	Pre-Survey Questionnaire.....	
F	Alpha Test Questionnaire.....	
G	Beta Test Questionnaire.....	
H	Alpha Test Results Summary.....	
I	Beta Test Results Summary.....	
J	System Screenshots.....	
K	User Manual.....	
L	Source Code.....	
M	Expert/Alpha Tester Profile.....	
N	Researchers Curriculum Vitae.....	
O	Research Documentation Pictures.....	

CHAPTER I



INTRODUCTION

In today's digital age, educational institutions are continuously seeking ways to improve their academic services. With the increasing number of students and the growing demand for online learning, universities and colleges are facing the challenge of organizing and managing large volumes of academic documents, including theses and dissertations. To address this challenge, a web-based thesis archiving system can be developed to provide a secure and efficient way of storing and accessing academic documents.

Technological developments in the field of education have opened up new innovations for the advancement of science and technology, making it simpler for us to carry out various academic activities. One technology that is frequently utilized are websites, they are used in data management, servers, and repositories. Websites may also be used for a variety of educational purposes, this includes, accepting new students, administration payments, e-learning, student portals, etc. Campus/School Institutions these days are required to be more creative, innovative, modern, advanced, and evolving as a result of technology's impact on the area of education. [3] Alade (2023) indicates that the electronic document management system would increase user satisfaction, raise productivity, and ensure time and data efficiency. As a result, well-known document management systems surely aid in the storage and management of a significant amount of an organization's knowledge assets, which comprise documents and other related objects. [1] Document management is now crucial to the development of an organization. It is essential to distribute the right paper to the right people in a

Commented [AN1]: Introduction is too common

Commented [AN2]: What's your basis?

Commented [AN3]: Redundant

Commented [AN4]: Redundant



DON HONORIO VENTURA STATE UNIVERSITY

secure and effective manner. Digital repositories are now more crucial for the collection and dissemination of scholarly materials because of the growing trend toward online scholarly communication and the absence of scholarly content management systems among universities. [1,2]

Commented [AN5]: Missing Reference

This study focuses on the design and implementation of an efficient web-based thesis archiving system for Don Honorio Ventura State University (DHVSU) in the Philippines. The system intends to enable students, faculty, and researchers to easily store and access electronic theses and dissertations online.

Commented [AN6]: What's your basis

This study is intended to benefit the academic community of DHVSU by creating a reliable and user-friendly resource for accessing and storing theses and dissertations. In order to achieve this goal, the system will be designed to have a simple and intuitive interface, with a focus on security and efficiency. By improving accessibility and reducing the effort required to manage academic documents, this system will offer a valuable tool for students, faculty, and researchers alike. Ultimately, the success of this study will be measured by the improved accessibility and functionality of the system.

Commented [AN7]: Wdym by academic community

Commented [AN8]: what



PROJECT CONTEXT

The proposed system aims to enhance accessibility by providing a user-friendly platform for the academic community to submit, store, search, and access thesis papers. It will feature secure document storage, advanced search capabilities, metadata management, version control for collaboration. In addition, a plagiarism scanner and AI detector functionality will be incorporated to ensure the integrity of the research outputs. The implementation of system will transform the research paper management of DHVSU. The system will have efficient access, secure preservation, and facilitating the sharing of knowledge within the academic community which aligns with the commitment of the university to academic excellence and research advancement.

PURPOSE AND DESCRIPTION

The purpose of this study is to develop and implement a comprehensive thesis management system that addresses the existing challenges in submitting, storing, searching, and accessing academic documents at DHVSU. The proposed system will provide a user-friendly platform for students, faculty, and researchers, offering secure document storage, advanced search capabilities, metadata management, version control, and collaboration tracking mechanisms. Additionally, a plagiarism detector functionality will be integrated to ensure the authenticity and originality of the research outputs.

Commented [AN9]: commas, again

Commented [AN10]: already said this in the Project Context



The thesis management system aims to streamline and automate the entire process of thesis submission and management, replacing the manual and cumbersome procedures currently in place. By implementing this system, DHVSU seeks to enhance accessibility to academic documents, promote efficient knowledge sharing, secure preservation of research outputs, and facilitate collaboration within the academic community.

Commented [AN11]: Again, what's an academic community

OBJECTIVES OF THE STUDY

The objectives of this study are as follows:

1. **Develop a user-friendly thesis management system that provides an intuitive interface for students, faculty, and researchers to submit, store, search, and access academic documents.**
 - a. Implement secure document storage mechanisms to ensure the confidentiality and integrity of academic documents, protecting them from unauthorized access or loss.
 - b. Incorporate advanced search capabilities to enable efficient and accurate retrieval of academic documents based on various criteria, such as keywords, authors, departments, and publication dates.
 - c. Integrate metadata management features to enable the categorization, organization, and easy identification of academic documents based on relevant attributes, such as discipline, topic, and research supervisor.



DON HONORIO VENTURA STATE UNIVERSITY

d. Establish version control mechanisms to track the revisions, updates, and changes made to academic documents over time, ensuring the availability of the most recent and accurate versions.

e. Develop collaboration tracking mechanisms to enable efficient and transparent collaboration among multiple authors, supervisors, and reviewers involved in the thesis writing and evaluation process.

2. Integrate credibility functionalities to detect and prevent any instances of academic dishonesty, ensuring the originality and integrity of the research outputs.

a. Integrate with reputable plagiarism detection tools or services (e.g., Turnitin, Copyscape) to the Admin App of the Website, the process is basically that the admin/professor will use the plagiarism checker and then manually decide if the document is ready for publishing.

b. Incorporate an AI detector functionality that will check for the possible usage of Generative AI such as ChatGPT for the writing of academic papers.

Commented [AN12]: Reformat this please



SCOPE AND LIMITATION

The scope of this study encompasses the development and implementation of the thesis management system specifically tailored for DHVSU. It will cover the entire lifecycle of academic documents, starting from the submission phase to storage, search, retrieval, collaboration, and plagiarism detection.

However, it is important to note that this study does not address the process of thesis evaluation or grading. The system will focus on facilitating the management and accessibility of academic documents but will not provide functionalities for evaluating or assigning grades to the submitted works.

Furthermore, the study will be limited to the specific requirements and constraints of DHVSU, considering the university's existing infrastructure, policies, and resources. The system will be designed to align with the university's needs and expectations while adhering to any technical, financial, or administrative limitations that may arise during the implementation process.

Commented [AN13]: Reformat

DEFINITION OF TERMS

Academic Community – A group of people in higher educational institutions who continuously engage in core intellectual activities such as teaching, learning and research.

Document Management - is defined as the organization and maintenance of paperwork related to specific tasks and procedures as their significance and use have grown over time.

Commented [AN14]: Put this on the Definition of Terms



CHAPTER II



RELATED LITERATURE

Web-Archives

Web archives enable users to access and study past versions of websites, helping to document the evolution of the web and providing valuable resources for researchers, historians, and anyone interested in understanding the development of online content and digital culture over time. highlighting the challenges it presents to researchers while also emphasizing the diverse opportunities it offers for interpretation. Brügger goes on to discuss how historians can adapt traditional research methods to effectively study the archived web, with a focus on aspects such as establishing the origin of web content, creating comprehensive overviews of archived materials, evaluating different versions, and appropriately citing digital sources. In essence, this chapter sets the stage for the development of a methodology to conduct web history in the digital age, providing valuable guidance to both contemporary media scholars and future historians. (Brügger N 2018)

A web archive, often referred to as a web archive or web collection, is a digital repository or database that systematically captures and stores snapshots of web pages and websites as they appear at different points in time.

The data archive system described in the paper offers the advantage of automatically determining the storage destination for archiving, reducing the manual work involved in the process. By utilizing control information and an archive rule table, the system efficiently decides whether to archive data to the on-premise environment or the cloud storage.



(Mori Hajime, Mimatsu Yasuyuki, Inomata Hirofumi, 2018)

Web archiving has been in existence for decades now, dating back to the 1990s and the establishment of the Internet Archive. In the late 1990s a variety of nonprofits and universities began to archive various aspects of digital content, but the formation of the Internet Archive in 1996 launched the first comprehensive Web archiving program in the world (Toyoda & Kitsuregawa, 2012).

Early Web archiving was sporadic in nature, as collections were often built from donated datasets or were constructed in an adhoc nature by sampling across web domains. In the subsequent decades the degree of Web archiving has accelerated and become far more comprehensive (Milligan, 2016; Schneider et al., 2003).

Web archives serve as invaluable time capsules, granting users access to historical iterations of websites. This access not only documents the web's evolution but also provides researchers, historians, and enthusiasts with rich resources.

A fundamental aspect of digital preservation, is indispensable in safeguarding the vast and dynamic digital landscape. In an era characterized by unprecedented internet usage, preserving web content has become a pressing concern to maintain our cultural, historical, and scholarly heritage. Libraries, archives, and cultural heritage institutions have become acutely aware of the pivotal role played by web archiving in ensuring that the wealth of information disseminated online is not lost to the passage of time (Besser, 2001).



The Internet Archive's Wayback Machine, initiated by Kahle (1996), stands as a testament to the power of web archiving, allowing users to access historical snapshots of websites dating back to the early days of the internet. Such initiatives highlight the collaborative nature of web archiving, where institutions pool resources to preserve the ever-expanding digital corpus (Koerbin & Ross, 2017).

Plagiarism

Plagiarism involves the unauthorized use of someone else's work without providing proper credit to the original creator, thereby diminishing the originality and authenticity of the content. It encompasses all types of creative works, whether published or unpublished, such as manuscripts, printed materials, and digital formats. Plagiarism can be intentional, careless, or accidental. Deliberate or negligent acts of plagiarism are strictly prohibited according to examination regulations, and disciplinary actions may be taken as a consequence. Plagiarism is morally wrong because it involves falsely claiming ownership of another person's work, essentially constituting theft. Furthermore, it violates the principles of academic integrity, which teach students how to responsibly share information and acknowledge the contributions of creators. Therefore, plagiarism is in direct violation of these principles. Plagiarism not only has ethical implications but also diminishes the value of college degrees. If companies perceive that a significant number of students have engaged in cheating to obtain their degrees, it undermines the credibility and reputation of those degrees. While academic plagiarism may not be illegal, it is a violation of academic honesty policies set by



DON HONORIO VENTURA STATE UNIVERSITY

institutions. Consequently, students who commit plagiarism can face severe consequences such as expulsion or academic probation. To address the issue of plagiarism, tutors and instructors utilize AI-based plagiarism detection software. These programs employ scanning and coding techniques to identify instances of plagiarism in students' work. By using such technology, educators can more effectively detect and address cases of plagiarism, promoting academic integrity and ensuring that students' work is original and properly credited (Zoting, P.G., & Dharmadhikari, S.C. 2023).

Plagiarism is undoubtedly a disgraceful and selfish act as it involves stealing and causes harm to others. It often occurs due to reasons such as lack of intelligence, creativity, or a lazy attitude towards learning, with individuals seeking shortcuts in their research endeavors. Cheating and theft, which plagiarism essentially entails, are considered sinful actions. It is crucial to acknowledge and address every instance of sin through appropriate rules and sanctions. In the academic context, this involves enforcing student and academic regulations, while lecturers who commit plagiarism are subject to sanctions based on their ethical code (Wijaya, H., & Gruber, K.E. 2018).

Plagiarism is an increasingly significant problem in the field of academia, with the availability of abundant resources and information. Its impact on students' performance and the overall quality of education is substantial. To address this issue, numerous studies have been conducted to find ways to mitigate plagiarism. The process of plagiarism itself is a major concern within academic circles, prompting extensive research over several decades. The objective of such



research is to minimize instances of plagiarism, uphold the standards of original and creative writing, and safeguard the rights of copyright ownership. Plagiarism can be defined as the violation of an author's or authors' copyright, involving the unauthorized copying of someone's work or ideas without giving appropriate acknowledgment. This act of dishonesty significantly diminishes the quality and integrity of the work in question (Jiffriya, M., Jahan, M.A., & Ragel, R.G. 2021). Integrity plays a vital role in academic life, as it demonstrates responsible attitudes and principles in scientific work. Researchers can prevent plagiarism by gaining a proper understanding and taking correct actions. Upholding anti-plagiarism measures and ethical values, norms, and rules in higher education aligns with the principles of National Education and reflects the transmission of religious values. Honesty, integrity, and originality are essential aspects to be carefully considered when creating academic work, while plagiarism occurs when individuals fail to provide adequate source citation in their work. Ethical values that guide researchers in avoiding plagiarism include scientific truth, reasoning, honesty, justice, benefit, virtue, responsibility, diversity, and affordability. By adhering to these values, individuals can uphold the principles of ethics, maintain the integrity of their work, and avoid the detrimental consequences of plagiarism (Wijaya, H., & Gruber, K.E. 2018).

In the study of Khatri and Mohan (2016), a web-based anti-plagiarism approach to detect plagiarism at the academic level. The researchers argue that with the advancement of technology the rise of plagiarism is alarming. They described that plagiarism in academic institutions is putting a negative impact on the academic



DON HONORIO VENTURA STATE UNIVERSITY

integrity of institutions. The researchers' approach to plagiarism detection presents a way for online detection of plagiarism and attempts to put an end to the cut, copy, and paste culture that is prevalent in academic institutions. Thus, the approach's implementation will attempt to promote intellectual growth and development among students, as well as help them understand the five essential ideals of academic integrity upon which information is generated, shared, and assessed.

These plagiarism detection tools work by comparing sets of documents to identify similar passages of text, but this approach has some inherent limitations. Chief among these is the single source problem. If a student hires someone to write a term paper for them, clearly plagiarism has taken place. However, there is no second copy of this paper that can serve as evidence that the work is not the student's own. Without this second copy, these plagiarism detection tools are unable to differentiate between a plagiarized paper and legitimate work by other students. By reframing the task of "preventing plagiarism" into "ensuring authorship," a new approach has been developed that is theoretically able to detect when single source plagiarism has taken place in programming assignments involving source code. Requiring students to provide copies of their code in progress, and analyzing changes in subsequent versions can prove when a student performed their own work. An implementation of this approach was developed and tested on three courses taught by the Department of Computer Science at the Florida Institute of Technology. Results of these trials, along with additional benefits to the use of such a system are discussed (Koss, Ian Mathias, 2021).



In the study of Nwohori and Ajayi (2021), the authors implemented a system that incorporates artificial intelligence through the use of natural language processing or NPL. The researcher's findings show that a document's Similarity index is the percentage of text in that document that has been established to exist in other documents. Existing plagiarism detectors generate the similarity index; it is then up to the instructors to determine if the author properly listed his sources and whether the sources linked can safely be used. Using AI, this decision might be made by a computer that establishes a "plagiarism index" after analyzing the texts, citations, and references.

Related Studies

In the study of Mkpojiogu, Akusu, Hussain, Hashim (2020), the authors used MySQL, the world's most well-known open-source database management technology to implement the system's database to achieve an efficient and time-saving management for storing, retrieving, and moving of files to solve the difficulties and complexities in the manual system of archiving and managing of information. The research was primarily designed to handle a long term of storage and management of files which are efficient in reducing the use of hardcopy files and solving practical problems.

The development of a web-based thesis archiving system with a plagiarism scanner and AI detector addresses the limitations of existing plagiarism detection tools. These tools typically compare sets of documents to identify textual similarities but face challenges when it comes to the single source problem. In situations where a student hires someone to write a term paper, it becomes



difficult to differentiate between plagiarized work and legitimate submissions without a second copy of the paper. To overcome this limitation and ensure authorship, the proposed web-based system incorporates an innovative approach. In addition to archiving theses, it includes a plagiarism scanner and AI detector specifically designed for programming assignments involving source code. By requiring students to submit copies of their code in progress and analyzing subsequent versions, the system can determine if a student has genuinely performed their own work. The implementation of this approach has been successfully tested on three courses offered by the Department of Computer Science at the Florida Institute of Technology. The results of these trials validate the effectiveness of the system in detecting single source plagiarism in programming assignments. Furthermore, beyond plagiarism detection, the web-based thesis archiving system offers additional benefits, which can be explored and discussed in further detail. By integrating a plagiarism scanner and AI detector into a comprehensive web-based thesis archiving system, academic institutions can enhance their ability to ensure the authenticity of student work and promote academic integrity (Koss, Ian Mathias, 2021).

Information technology is currently developing rapidly so that every aspect of activity cannot be separated from the name technology, of course in the administrative aspect it also really requires fast and accurate information technology. At the Faculty of Communication and Information Technology, National University, currently still using an archive system that has not been developed again, which often causes data synchronization. With the E-Archives



application, it is hoped that archiving activities can assist administrative activities accurately so that it can become a reference in making the right decisions (Al Yarda, R.R., Fitri, I., & Benrahman. 2021).

SYSTEM TECHNICAL BACKGROUND

The thesis management system is built on a web-based architecture using various technologies and tools to ensure efficient and reliable functionality. The technical background of the system includes:

Front-End Development: The user interface of the system is developed using front-end technologies such as HTML, CSS, and JavaScript. These technologies enable the creation of an intuitive and interactive interface for users to submit, search, and access academic documents.

Back-End Development: The back-end of the system is developed using a server-side programming language such as PHP, Python, or Node.js. This allows for the implementation of the system's business logic, data processing, and integration with the database.

Database Management: The system utilizes a relational database management system (RDBMS) such as MySQL, PostgreSQL, or Oracle. The database is responsible for storing and managing thesis metadata, user information, plagiarism scan results, and other relevant data. The chosen RDBMS ensures data integrity, security, and efficient retrieval of information.



Document Storage: The system incorporates secure document storage mechanisms, which may involve storing academic documents on the server or utilizing cloud storage services such as Amazon S3, Google Cloud Storage, or Microsoft Azure Blob Storage. These storage solutions provide scalability, accessibility, and data redundancy to ensure the safe preservation of academic documents.

Plagiarism Detection: The system integrates a plagiarism detection tool or library to compare submitted theses with a database of previously published works. The tool may use techniques such as text similarity analysis, natural language processing (NLP), or machine learning algorithms to identify plagiarized content accurately.

AI Content Detection: The system incorporates AI detection methods to identify content that may have been generated by AI models. This involves training machine learning or deep learning models on AI-generated text data to recognize patterns and characteristics specific to AI-generated content.

Search and Retrieval: The system implements efficient search algorithms and indexing mechanisms to enable users to search and retrieve academic documents based on various criteria, such as keywords, authors, departments, or publication dates. Techniques like full-text search and database indexing are employed to ensure fast and accurate search results.

User Authentication and Security: The system includes user registration and authentication mechanisms to secure access to the system's functionalities. Password storage and encryption techniques are implemented to ensure the



confidentiality and integrity of user credentials. Additionally, security testing and penetration testing are performed to identify and address vulnerabilities and protect against unauthorized access and data leaks.

Performance Optimization: The system undergoes performance testing to assess its responsiveness and scalability. Techniques such as caching, query optimization, and load balancing may be employed to enhance the system's performance under various loads and stress conditions.

System Deployment: The system is deployed on a hosting platform, which may involve setting up a web server, configuring the necessary infrastructure, and ensuring proper connectivity and security measures.

Overall, the technical background of the thesis management system encompasses a combination of web development technologies, database management systems, document storage solutions, plagiarism detection tools, AI content detection techniques, search algorithms, authentication mechanisms, security measures, performance optimization strategies, and deployment processes. These technologies work together to create a robust and efficient system for managing academic documents.



CHAPTER III



METHODOLOGY

The implementation of a web-based electronic thesis paper archives system will follow this methodology. Firstly, the research objectives will be clearly defined, including the design and implementation of the system with specific functionalities and features in mind.

A comprehensive review of existing web-based electronic thesis paper archives systems will be conducted to understand their strengths, weaknesses, and limitations. This analysis will identify areas for improvement and innovation in the proposed system.

The system requirements will be identified by defining both the functional and non-functional aspects. This includes determining the necessary hardware, software, and technical infrastructure required to support the system, considering factors such as scalability, security, usability, and accessibility.

The system architecture will be designed, outlining the components and their interactions. The database structure for storing and retrieving academic documents and metadata will be defined, along with the user interface design and navigation flow to ensure intuitiveness and user-friendliness.

Appropriate development tools and technologies will be selected, including web development frameworks, programming languages, and tools suitable for building the system. A suitable database management system will be chosen for storing academic documents and metadata. Front-end technologies such as HTML, CSS, and JavaScript may be utilized for the user interface.



The system will be implemented based on the defined architecture and requirements, involving the creation of database schemas, tables, and relationships. Functionalities such as user registration, authentication, thesis submission, search, and browsing capabilities will be developed.

Comprehensive testing will be conducted to ensure the system meets the defined requirements. This includes unit testing, integration testing, and system testing to verify the system's functionality, usability, performance, and security.

Once thoroughly tested and validated, the system will be deployed on a web server or hosting platform. The server environment and database settings will be configured, considering factors such as load balancing, caching, and backup strategies.

User training sessions will be provided to familiarize users with the web-based system. Comprehensive documentation, including user manuals, system architecture, and guidelines for system administrators, will be created to assist users in effectively utilizing the system.

Continuous evaluation and improvement will be performed by gathering feedback from users and stakeholders, identifying areas for enhancement and potential improvements. The system's usage will be monitored, and any issues or bugs that arise will be addressed promptly.

By following this methodology, the implementation of the web-based electronic thesis paper archives system will be guided effectively, resulting in a well-designed, functional, and user-friendly system that meets the research



objectives.

REQUIREMENTS ANALYSIS

The web-based electronic thesis paper archives system will include the following functional requirements. Firstly, users will be able to securely register and authenticate their accounts, create personal profiles, and log in using their credentials. Secondly, users will have the ability to electronically submit their academic documents by uploading the document file and providing relevant metadata, such as the title, author, abstract, publication date, and keywords.

The system will also provide search and browse functionality, enabling users to find academic documents based on criteria such as title, author, keywords, or department. Users will be able to enter search queries and receive relevant results. Additionally, a document preview feature will allow users to view a summary or excerpt of a thesis paper before downloading it.

To ensure appropriate access control, the system will implement mechanisms to manage user permissions and restrict access to academic documents. Only authorized users will have the ability to access and download the documents.

In terms of non-functional requirements, security will be a priority. The system will employ robust measures, including secure user authentication, data encryption, and access control mechanisms, to ensure the confidentiality, integrity, and availability of academic documents and user data.

Scalability is another crucial non-functional requirement. The system will be designed to handle a large number of academic documents and



concurrent users, with the ability to scale up resources as the archive grows to accommodate increased demand.

Usability will be a key consideration in the system's design, aiming for an intuitive, user-friendly, and accessible user interface. Users should be able to navigate the system easily and perform tasks without confusion.

Performance optimization will be crucial to the system, ensuring fast search results and document retrieval to provide a smooth user experience.

The system will be compatible with various web browsers and operating systems to ensure broad accessibility. Thorough testing will be conducted on popular browsers and platforms to ensure consistent functionality across devices.

To safeguard against data loss or system failures, the system will incorporate backup and recovery mechanisms, including regular data backups and established procedures for data recovery.

Maintenance and support will be prioritized, with provisions for regular updates, bug fixes, and technical support to ensure the system remains functional and up to date.

From a technical perspective, the system will be developed using a web development framework such as Django or Ruby on Rails. A reliable database management system such as MySQL or PostgreSQL will be chosen to store academic documents and metadata. The system will be hosted on an appropriate environment with web servers, storage, and bandwidth that ensure availability and performance. Additionally, document conversion mechanisms will be



DON HONORIO VENTURA STATE UNIVERSITY

implemented to convert academic documents to a standard format, such as PDF, for consistent display and easy downloading.

To address security concerns, encryption protocols, and SSL certificates will be implemented to protect against common web vulnerabilities. The system will also comply with accessibility standards, such as WCAG 2.0, to ensure accessibility for users with disabilities.

This requirement analysis will serve as a guide for the development and implementation of the web-based electronic thesis paper archives system, ensuring it effectively meets user needs while operating securely and reliably.

REQUIREMENTS DOCUMENTATION

The objective of the web-based thesis archiving system with an AI detector is to offer a productive and user-friendly platform for the storage, search, and retrieval of theses while including an AI detector to identify instances of plagiarism. The technology will speed up the archiving procedure, improve the effectiveness of plagiarism detection, and guarantee the objectivity and originality of the study effort.

User Interface Requirements:

Requirement ID	Requirement Description	Priority
UI_REQ001	User-friendly thesis submission interface	High
UI_REQ002	Intuitive search and	High



	retrieval interface	
UI_REQ003	Responsive design for various devices	Medium
UI_REQ004	Easy navigation and user guidance	Medium
UI_REQ005	Customizable user preferences and settings	Low

Plagiarism Detection Requirements:

Requirement ID	Requirement Description	Priority
PD_REQ001	Integration of AI detection algorithms	High
PD_REQ002	Comparison of submitted theses for plagiarism	High
PD_REQ003	Generation of plagiarism reports	Medium
PD_REQ004	Setting a threshold for acceptable similarity	Medium
PD_REQ005	Integration of institutional policies	Medium

Administration and Management Requirements:

Requirement ID	Requirement Description	Priority
AM_REQ001	Role-based user management	Medium
AM_REQ002	Approval workflow for thesis submissions	Medium
AM_REQ003	Statistical reporting on archiving and detection	Medium



AM_REQ004	Access control and permission settings	Medium
AM_REQ005	Audit logs for system activities	Low

DESIGN OF SOFTWARE SYSTEM

The design of the thesis archive system focuses on creating a software system that efficiently manages and archives academic documents through a web interface. It encompasses the following key aspects:

For **System Architecture and Component Design**, the system includes User Management, Thesis Management, Search Functionality, Document Preview, Plagiarism Scanner and AI Detector, and Access Control.

For **Database Design**, the database schema includes tables for users, theses, and associated metadata. Relationships are established using primary and foreign keys for data integrity and retrieval efficiency.

For **User Interface Design**, the user interface features an intuitive and responsive design with forms for thesis submission, search filters, and user profile management. It adheres to accessibility standards and compatibility with different devices.

For **Error Handling and Exception Management**, robust error handling mechanisms catch and handle exceptions, displaying user-friendly error messages. Exceptions are logged for troubleshooting and system improvement.

For **Security and Privacy**, user authentication, access control, and encryption

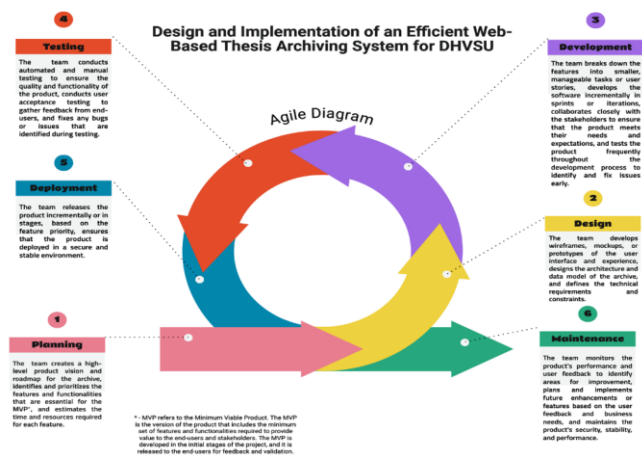


techniques ensure secure access and protect sensitive information. Privacy measures comply with data protection regulations.

For **Performance and Scalability**, the system is designed to handle a large volume of academic documents with efficient indexing and query optimization. Load balancing and scalability techniques can be implemented.

For **Testing and Quality Assurance**, thorough testing includes unit testing, integration testing, and user acceptance testing. Comprehensive documentation covers system architecture, database schema, API documentation, and user manuals.

For **Maintenance**, version control systems manage updates and maintenance tasks. Regular maintenance addresses bugs, security vulnerabilities, and system enhancements.



PRODUCT AND PROCESS

Users to upload and store thesis documents securely on the cloud-based platform. It supports various file formats, including PDF, Word documents. The system maintains a centralized repository of all archived theses for easy retrieval and reference. Users can search for specific theses based on keywords, authors, titles, or other relevant parameters. DHVSU Archives utilizes advanced search algorithms to quickly retrieve the desired documents, saving time and effort for researchers and educators. The Plagiarism Scanner feature uses powerful algorithms to compare the uploaded theses against an extensive database of scholarly resources, including research articles, journals, and online repositories. It highlights potential instances of plagiarism, providing users with a detailed report for further investigation.

The web-based thesis archive system is a comprehensive platform designed to facilitate the management, storage, and access of academic documents. It provides the following features:

1. User Registration and Authentication:
 - Users can create accounts and authenticate themselves to access the system.
 - User roles and permissions are defined to control access to different system functionalities.
2. Thesis Submission and Management:
 - Users can upload their academic documents in various formats, such as PDF or Word documents.



- Metadata associated with each thesis, including title, author name, supervisor, abstract, keywords, and year of publication, are collected during the submission process.

- The system organizes and stores the uploaded academic documents, ensuring secure and reliable storage.

3. Advanced Search Functionality:

- Users can search for academic documents based on criteria like title, author, keywords, supervisor name, and year of publication.

- Advanced search options allow for more precise and tailored searches.

4. Document Preview:

- The system provides a document preview functionality, allowing users to view the content of academic documents without the need to download the entire file.

5. Plagiarism Scanner and AI Detector:

- The system integrates a plagiarism scanner and AI detector to analyze uploaded academic documents.

- These tools identify potential instances of plagiarism, ensuring academic integrity.

6. Access Control and Security:

- Role-based access control is implemented, allowing different levels of access for students, faculty, and administrators.

- The system ensures data privacy and protection by employing encryption techniques for sensitive information.

Process: The development and implementation of the web-based thesis archive system involve the following steps:

1. Requirement Analysis:

- Gather requirements from stakeholders, including users, administrators, and faculty members.

- Identify the core features, functionalities, and system constraints.

2. System Design:

- Define the system architecture, including the client-server model and three-tier structure.
- Design the database schema, considering tables for users, theses, and associated metadata.
- Create wireframes and prototypes for the user interface, ensuring a user-friendly design.

3. Development:

- Develop the system components, including User Management, Thesis Management, Search Functionality, Document Preview, Plagiarism Scanner and AI Detector, and Access Control.
- Implement the database and establish necessary relationships.
- Apply appropriate software design patterns, such as the Model-View-Controller (MVC) pattern, for code organization and maintainability.

4. Testing and Quality Assurance:

- Perform unit testing to verify the functionality of individual components.
- Conduct integration testing to ensure the seamless interaction between different system modules.
- Perform system testing to validate the overall system behavior, including user workflows and data integrity.
- Conduct user acceptance testing to gather feedback from stakeholders and refine the system.

5. Deployment and Maintenance:

- Deploy the web-based thesis archive system on a suitable server environment.
- Ensure the system's availability, performance, and security through regular monitoring and maintenance.
- Address any bug fixes, security vulnerabilities, and system enhancements identified during the maintenance phase.



- Continuously gather user feedback and incorporate new features and improvements to enhance the system.

DEVELOPMENT AND TESTING

A. System Architecture and Design:

- Create the overall system architecture, which will include the web server, database, and user interface.
- Develop the database schema to store thesis metadata, user information, and scan results.
- Develop the user interface for submitting, searching, and viewing the thesis.

B. User Registration and Authentication:

- Implement a user registration and login system to authenticate users.
- Develop password storage and encryption mechanisms that are secure.

C. Thesis Submission and Storage:

- Create a form for users to submit their theses, along with the necessary metadata (e.g., title, author, abstract).
- Implement a file upload feature to securely save the academic documents on the server or in cloud storage.

D. Plagiarism Scanner Integration:

- Integrate a plagiarism detection tool or library that can compare the submitted thesis to a database of previously published works.
- Develop an algorithm to compare similarities between the submitted thesis and the current database.
- Create a plagiarism report that includes probable plagiarized elements as well as their sources.

E. AI Detection:

- Develop an AI detection method to discover content that may have been created by AI models.
- Implement a machine learning or deep learning model trained on AI-generated text data to identify suspicious content.
- Integrate the AI detection module into the system and create a report that highlights potential AI-generated portions.

F. Search and Retrieval:

- Implement a search feature to allow users to discover theses based on criteria such as author, keywords, or topic.
- Create algorithms to retrieve and display search results efficiently.

G. User Management and Administration:

- Develop an administration panel for system administrators to manage users, view reports, and perform other administrative tasks.
- Implement user roles and permissions to control access levels and privileges.



Testing:

A. Unit Testing:

- Conduct unit testing for individual components, such as user registration, file upload, search functionality, plagiarism scanner integration, and AI detection.
- Validate that each component functions properly and handles both valid and incorrect inputs.

B. Integration Testing:

- Perform integration testing to ensure that different modules of the system work together seamlessly.
- Test the interaction between user registration, thesis submission, plagiarism scanner, AI detector, and search functionalities.
- Validate that data is passed correctly between different components and that the system behaves as expected as a whole.

C. System Testing:

- Perform system-level testing to confirm the overall operation of the web-based thesis archiving system.
- Test various user scenarios, such as user registration, thesis submission, search, plagiarism scanning, AI detection, and report generation.
- Verify that the system functions as intended, satisfies all criteria, and handles every error and exceptions gracefully.

D. Security Testing:

- Perform security testing to identify and address vulnerabilities.
- Conduct penetration testing to verify the system is safe against unauthorized access and data leaks.
- Test the system's capacity to handle malicious inputs and guard against common security risks.



E. Performance Testing:

- Evaluate the system's performance under various loads and stress conditions.
- Test the response time of different functionalities, such as thesis submission, search, and report generation, to ensure acceptable performance levels.
- Identify and resolve any performance or scalability concerns.

F. User Acceptance Testing (UAT):

- Engage users to perform UAT to gather feedback on the system's usability, functionality, and overall satisfaction.
- Incorporate user feedback and make necessary improvements based on their recommendations.

IMPLEMENTATION PLAN

Testing

- Develop test cases to verify the functionality of system components.
- Conduct usability testing to ensure that the UI is intuitive and user-friendly.
- Test performance to assess the system's responsiveness and scalability.
- Find and fix any bugs and issues found during testing.

Deployment

- Set up a hosting platform to deploy the system.
- Ensure the system is accessible and secure.



Evaluation and Feedback

- Gather feedback from users through surveys to assess their satisfaction and check for areas that need improvement.
- Evaluate the system's effectiveness in achieving its objectives.

Future Enhancement and Maintenance

- Establish a report section where the user can report any issues or bugs that may appear.
- Take future enhancements into consideration such as updating or enhancing the UI, Plag Scanner, and AI Detector.
- Incorporate additional features based on user feedback and emerging technologies.



References

- [1] Al Yarda, R.R., Fitri, I., & Benrahman. (2021). Implementation of E-Archives at the Faculty of Communication and Information Technology Using the Web-Based Waterfall Method. Jurnal Mantik, 4(4), 2532-2544. E-ISSN 2685-4236. Retrieved from <https://iocscience.org/ejournal/index.php/mantik>
- [2] Wijaya, H., & Gruber, K.E. (2018). Ethics perspective and regulation of plagiarism in Higher Education. International Journal of Humanities and Innovation (IJHI), 1(1), 17-25.
- [3] Jiffriya, M., Jahan, M.A., & Ragel, R.G. (2021). Plagiarism detection tools and techniques: A comprehensive survey. Journal of Science-FAS-SEUSL, 02(02), 47-64. ISSN: 2738-2184.
- [4] Zoting, P.G., & Dharmadhikari, S.C. (2023). PLAGIARISM DETECTION USING MACHINE LEARNING. International Research Journal of Modernization in Engineering Technology and Science, 5(5), 4546. doi: 10.56726/IRJMETS39442. Retrieved from www.irjmets.com
- [5] Abbasova, V.S. (2020). Main Concepts of the Document Management System Required for Its Implementation in Enterprises. ScienceRise, 28(1), 32-37. doi: 10.21303/2313-8416.2020.001149



[6] Nwohiri, A. M., Joda, O., & Ajayi, O. (2021). AI-POWERED PLAGIARISM DETECTION: LEVERAGING FORENSIC LINGUISTICS AND NATURAL LANGUAGE PROCESSING. *Fudma Journal of Sciences*, 5(3), 207–218.
<https://doi.org/10.33003/fjs-2021-0503-700>

[7] Mkpojiogu, E., Akusu, G. E., Hussain, A., & Hashim, W. (2020). Implementation of a Web-based Data Archival Management System. ResearchGate.
https://www.researchgate.net/publication/341680207_Implementation_of_a_Web-based_Data_Archival_Management_System

[8] Khatri, J., & Mohan, V. (2016). An Approach for Implementing Web-Based Tool for Plagiarism Detection. ResearchGate.
https://www.researchgate.net/publication/304169398_An_Approach_for_Implementing_Web-Based_Tool_for_Plagiarism_Detection

