Republic of the PhilippinesA picture containing logo

Description automatically generatedLogo

Description automatically generated

Laguna State Polytechnic University

Province of Laguna

**COLLEGE OF COMPUTER STUDIES**

|  |
| --- |
| **Concept Paper** |
| **A. Basic Information** |
| **Project Title:**  ResearchVault: An Advanced Web-based Research Repository for Laguna State Polytechnic University - Sta. Cruz Campus with Intelligent ChatGPT Integration and Collaborative Features for Efficient Research Management |
| **Topic:** Research Management, Web Development, Collaborative Research, Artificial Intelligence, Chatbot Integration, Citation Management, PDF Viewer, Version Control, Analytics, User Interface Design |
| **Proponent**: Kristopher France A. Punio |
| **B. Technical Description** |
| The ResearchVault is a proposed advanced web-based research repository system designed to streamline research management at Laguna State Polytechnic University - Sta. Cruz Campus. The system is aimed at providing users with a comprehensive research management experience, equipped with intelligent ChatGPT integration and collaborative features to provide efficient and seamless research management.  The proposed system will be designed to accommodate different user roles and access levels, allowing for easy and secure management of research materials. Users will be able to easily upload, store, and manage their research materials, as well as share and collaborate with other users on research projects. The system will also provide users with real-time assistance in finding relevant research materials through intelligent ChatGPT integration.  The system will be built using modern web technologies, ensuring high performance and reliability. It will be designed with a user-friendly interface, providing users with easy navigation and efficient access to research materials. The proposed system will also utilize advanced encryption and authentication technologies to ensure the protection and security of sensitive research materials.  The research design for the proposed system will be qualitative in nature, utilizing case study methodology to examine the implementation and effectiveness of ResearchVault. The population for this study will consist of faculty and staff members involved in research at the university, with a sample size of 20 participants selected through purposive sampling.  Data collection methods for this study will include both primary and secondary sources. Primary data will be collected through semi-structured interviews with faculty and staff members involved in research, as well as through observations of their use of ResearchVault. Secondary data will be collected through a review of university policies and procedures related to research management.  Data collected from interviews and observations will be recorded and managed using audio and video recordings, as well as written notes.  Ethical considerations for the proposed system will be addressed by obtaining informed consent from all participants, ensuring confidentiality and anonymity of the participants, and following ethical guidelines outlined by the Institutional Review Board (IRB).  The proposed ResearchVault system aims to provide efficient and seamless research management for Laguna State Polytechnic University - Sta. Cruz Campus, allowing for easy and secure management of research materials while promoting collaboration and sharing among users. The proposed system will be designed with a focus on user experience and security, utilizing modern web technologies to ensure high performance and reliability. The case study methodology employed in the research design will provide valuable insights into the implementation and effectiveness of the proposed system, while addressing ethical considerations outlined by the IRB. |
| **Statement of the Problem:**  Laguna State Polytechnic University - Sta. Cruz Campus is a center for research and innovation, with students and faculty members conducting research across a wide range of fields. However, there is currently no comprehensive and efficient system in place for managing research activities, which can lead to a range of issues, including disorganization, duplication of efforts, and wasted time and resources. Additionally, existing solutions do not provide the necessary features for efficient research management, such as intelligent ChatGPT integration, collaborative features, citation management, and version control.  Therefore, the problem that this proposed research aims to address is the lack of a comprehensive and efficient system for managing research activities among students and faculty members at Laguna State Polytechnic University - Sta. Cruz Campus. This research will investigate the development and implementation of ResearchVault, an advanced web-based research repository that integrates intelligent ChatGPT, citation management, version control, and collaborative features to streamline research-related tasks and improve the productivity and efficiency of research projects. This research seeks to answer the following questions:   1. How can natural language processing techniques such as Generative Pre-trained Transformer be used to generate automated analytics and insights in the ResearchVault? 2. What data visualization and interactive dashboard techniques can be used to provide insights into research trends, topic areas, and user behavior within the ResearchVault? 3. How can machine learning algorithms be used to analyze research materials in the ResearchVault and make personalized recommendations to users based on their research interests and preferences? 4. What data mining and text analysis tools can be offered to allow researchers to analyze large datasets and extract insights and trends from research materials?   The proposed research aims to fill the gap in the existing research by providing a comprehensive and efficient system for managing research activities, which will lead to improved productivity, reduced duplication of efforts, and more effective use of time and resources. |
| **Objectives: General and Specific General Objective:**  The general objective of this research is to develop an advanced web-based research repository for Laguna State Polytechnic University - Sta. Cruz Campus that will provide efficient research management, collaborative research, and intelligent ChatGPT integration.  Specifically, it aims to:   1. To create a web-based research repository that incorporates key features such as Key Insights, Compare & Contrast, Critique, Reminders, Sort & Filter, Notes, Split view, Citation Management, Full-text Search, PDF Viewer, Annotations, Version Control, Integration with Reference Management Software, Analytics, and Customizable Dashboard. 2. To integrate an intelligent ChatGPT feature into the research repository that will provide assistance and answer queries related to research topics and literature. 3. To develop collaborative features within the research repository that will allow researchers and team members to share literature lists, notes, and critiques and work together efficiently on research projects. 4. To provide a citation management system that will allow users to keep track of all sources used in their research papers and generate citations in various formats. 5. To create a user-friendly interface design that will ensure easy navigation and maximize the usability of the research repository. 6. To evaluate the effectiveness of the developed web-based research repository through user acceptance testing and gather feedback for further improvement. 7. To contribute to the academic community by providing an efficient and reliable platform for research management and collaborative research. |
| **How did others solve the problem?**   1. Artificial intelligence (AI) has helped to obtain accurate, fast, robust results without any human errors. Hence, it has been used in various applications in our daily lives. The Turing test has been a fundamental problem that AI systems aim to overcome. Recently developed various natural language problem (NLP) models have shown significant performances. AI language models, used in translation, digital assistant, and sentiment analysis, have improved the quality of our lives. It can perform scans on thousands of documents in seconds and report them by establishing appropriate sentence structures. Generative pre-trained transformer (GPT)-3 is a popular model developed recently has been used for many applications. Users of this model have obtained surprising results on various applications and shared them on various social media platforms. This study aims to evaluate the performance of the GPT-3 model in writing an academic article. Hence, we chose the subject of the article as tools based on artificial intelligence in academic article writing. The organized queries on GPT-3 created the flow of this article. In this article, we have made an effort to highlight the advantages and limitations of using GPT-3 for research paper writing. Authors feel that it can be used as an adjunct tool while writing research papers. (Evaluation of GPT-3 AI language model in research paper writing, Oğuzhan et. al., 2022) 2. Quality measures help identify gaps and disparities in care delivery and prioritize opportunities to improve health. Calls to enhance Systems-Based Practice and Practice-Based Learning and Improvement competencies for residency training cite the need for quality measures for trainees as central to this effort. The authors sought to demonstrate the feasibility of creating a residency program data visualization dashboard to examine individual and program quality measures for an internal medicine residency program within Kaiser Permanente Northern California. An interactive display was developed to allow for easy visualization of quality and operational measures through an iterative design process. The dashboard displays data for individual residents, residency classes (PGY1-3), and the entire program, including quality measures, systems measures, and patient diagnoses. An iterative process continues to improve the functionality and usefulness of the dashboard. It is feasible to create a dashboard to visualize individual and program quality measures and health equity measures for a residency program using a learner-centered approach and alignment with institutional goals through collaboration between education and operational teams. Future studies will examine the audit and feedback process, resident perceptions, and changes to patient outcomes. Use of dashboards in graduate medical education is feasible and can be used to help residents and residency programs identify gaps in quality of care. (Quality and Health Equity Dashboards for Internal Medicine Residents: Interactive Displays to Promote Systems-Based Practice and Practice-Based Learning and Improvement, Nardine & Lindsay, 2023) 3. With the rapid increase in Social Web applications, a significant amount of research has been dedicated to the analysis and development of personalized recommendations using artificial intelligence. Different machine learning and neural networks are applied widely to recommend relevant research books or research articles to online learners. However, conventional machine learning algorithms require additional training models to train recommendation models for processing large-scale data. Hence, they are considered to be inappropriate for generating personalized recommendations in learning systems. This research presents an LSTM-based approach for generating appropriate learning recommendations. The proposed approach analyses the behavior of online learners and predicts their interests. The performance of the proposed approach will be validated by determining different perform (LSTM-Based Top N Recommendation System using Cognitive Data, Avick et. Al., 2022) 4. Text mining has been shown to be an auxiliary but key driver for modeling, data harmonization, and interpretation in bio-medicine. Scientific literature holds a wealth of information and embodies cumulative knowledge and remains the core basis on which mechanistic pathways, molecular databases, and models are built and refined. Text mining provides the necessary tools to automatically harness the potential of text. In this study, we show the potential of large-scale text mining for deriving novel insights, with a focus on the growing field of microbiome. We first collected the complete set of abstracts relevant to the microbiome from PubMed and used our text mining and intelligence platform Taxila for analysis. We drive the usefulness of text mining using two case studies. First, we analyze the geographical distribution of research and study locations for the field of microbiome by extracting geo mentions from text. Using this analysis, we were able to draw useful insights on the state of research in microbiome w. r.t geographical distributions and economic drivers. Next, to understand the relationships between diseases, microbiome, and food which are central to the field, we construct semantic relationship networks between these different concepts central to the field of microbiome. We show how such networks can be useful to derive useful insight with no prior knowledge encoded. (Large scale text mining for deriving useful insights: A case study focused on microbiome, Syed et. al., 2022)   **Other Related Study (data gathering and machine learning methods)**   1. Possibilities for deep learning, machine learning and natural language processing present fascinating new library service areas. These areas are being integrated into traditional academic library information, digital literacy and university research environments. Much of university faculty, graduate students and library staff work outside of Computer Science disciplines and require help to enable their data. This research overviews methodologies, infrastructures and human resources needed for building new AI services within the ‘third interdisciplinary space’ of the academic library. This work suggests pragmatic steps that may be taken to set up good foundations. Data-centered steps for setting up digital scholarly research ecosystems are reviewed. Setting needed groundwork for library AI services enables research, data and media towards global online possibilities. Library AI external scholarly communications services are discussed as well as educational methodologies. Pathways are clarified and focused steps are forwarded to move library staff researchers and graduate students towards these new AI possibilities. Data-centered ecosystems, retooling and building on existing human resource expertise are reviewed. Data research repositories and programmatic algorithmic literacy are recommended. Library AI working groups and R&D prototype methodologies for scaling future library services and human resource infrastructures are considered. Prescriptive pathways create best in class library infrastructures for a currently occurring global AI paradigm shift. (Building Library AI Infrastructures: Research Data Repositories, Scaffolding & Raymond, 2023) 2. The study seeks to illustrate the most current trends in digital library research via the use of scientometrics. The study of scientific networks is important in many scientific domains. A social network with many nodes and connections serves as the foundation for scientific network research. Nodes include authors, publications, and journals, while linkages include citations, cocitations, and coauthorship. Data was collected from the Scopus abstracting and citation database for the period of ten years from 2012 to 2021. The most relevant 1957 documents were chosen from the collection, and selected documents were analyzed using Biblioshny and VOSviewer. The research showed that digital library productivity is rising annually, the United States of America dominates the production of scholarly production on digital libraries, and research is increasingly focused on digital resource and digital collection development. However, artificial intelligence, deep learning, machine learning, big data, and other related areas of study have emerged as the most recent research trends in digital library research. The outcomes of this study will aid digital library research by providing up-to-date and reliable research information. (Recent Trends in Digital Library Publications: A Scientometric Analysis, Rajashekhar & Pavithrabai, 2022) 3. The context of digital library has changed from insufficient information to information overload, and its corresponding service mode should also change from “people looking for information” to “information looking for people.” Using a grounded theory approach, this paper extracts 78 initial concepts, 24 basic categories, and 6 main categories by coding and analyzing the raw data obtained from the interviews. On this basis, the relationship path and action mechanism between categories are discovered; based on which a theoretical model of the influence mechanism of digital library intelligent information recommendation service satisfaction is constructed. The research results have shown that, under the moderating effect of user preference, the quality of data mining system, recommendation information quality, recommendation service quality, and recommendation form together have an impact on the satisfaction of digital library intelligent information recommendation service. The results of our work can provide useful reference for the optimization and healthy development of digital library services. Meanwhile, it has some theoretical and practical contributions. How to quickly obtain the information people required from a large amount of information is particularly important. Personalized construction is an inevitable service trend for the development of digital libraries in the new era. We in this paper study the current situation of digital libraries and the development of personalized services in digital libraries. We focus on data mining-related technologies, the development of digital library data exploration technologies, and the provision of Internet application services. The problems in this area were summarized, and the countermeasures were put forward based on this. It can be concluded that the concept of a digital library is not just a collection of data with information management tools, it is an environment that brings together collections, services, and people to support the entire data flow. It converts information into domain knowledge, from creation to dissemination. It guides the process from use to save. The trial registration number is ChiCTR2200055403. (Massive-Scale Data Mining to Enhance Digital Library with Applications in College Education, Xu & Bai, 2022) 4. In recent times, natural language processing (NLP) technique has received significant attention in content retrieval (CR) domain. The emergence of digital libraries, in recent years, enables people from across the globe to access and store books, documents, and literature of multiple kinds. The development of NLP models has considerably improved the performance in terms of digital library management. In this scenario, artificial intelligence-based expert systems are required to handle massive quantities of data that exist in digital libraries and achieve effective CR performance. In this background, the current study designs NLP with deep learning enabled hybrid content retrieval (NLPDL-HCR) model for digital library management. The aim of the presented NLPDL-HCR is to effectually retrieve the images as well as textual data from digital libraries based on a user's query. The proposed NLPDL-HCR model encompasses two major stages namely, text retrieval and image retrieval (IR). During text retrieval process, the proposed NLPDL-HCR model includes term frequency inverse document frequency vectorizer with optimal gated recurrent unit (GRU) model. The hyperparameters of the GRU model are optimally adjusted with the help of RMSProp approach. Besides, the IR process involves three sub-processes namely, densely connected networks-based feature extraction, butterfly optimization algorithm-based hyperparameter tuning, and Euclidean distance-based similarity measurement. The experimental analysis results, accomplished by the proposed NLPDL-HCR model using benchmark datasets, highlighted its superior performance over recent state-of-the-art approaches. (Natural language processing with deep learning enabled hybrid content retrieval model for digital library management, Almuhammadi et. al., 2022)   **Patents**   1. Disclosed herein is computer technology that applies natural language processing (NLP) techniques to training data to generate information used to train a natural language generation (NLG) system to produce output that stylistically resembles the training data. In this fashion, the NLG system can be readily trained with training data supplied by a user so that the NLG system is adapted to produce output that stylistically resembles such training data. In an example, an NLP system detects a plurality of linguistic features in the training data. These detected linguistic features are then aggregated into a specification data structure that is arranged for training the NLG system to produce natural language output that stylistically resembles the training data. Parameters in the specification data structure can be linked to objects in an ontology used by the NLG system to facilitate the training of the NLG system based on the detected linguistic features. (Applied artificial intelligence technology for using natural language processing to train a natural language generation system. (Applied artificial intelligence technology for using natural language processing to train a natural language generation system US11042713B1, United States, Platt et. al., 2021) 2. 10. A virtual agent can implement a “chatbot” to provide output based on predictive/prescriptive models for incidents. The virtual agent can integrate with natural language processor for text analysis and summary report generation. The virtual agent can integrate with cognitive search to enable processing of search requests and retrieval of search results. The virtual agent uses computing processes with self-learning systems that use data mining, pattern recognition and natural language processing to mimic the way the human brain works. The virtual agent provides an automated IT system that is capable of resolving incidents without requiring human assistance. The virtual agent can display condensed summaries of a large amount of data and can link the summaries to predictive models and operational risk models to identify risk events and provide summaries of those events. (Machine natural language processing for summarization and sentiment analysis US11563699B2, United States, CAI et. al., 2023) |
| **How do you intend to solve the problem?**  The proponent of the study plans to solve the problem by developing An Advanced Web-based Research Repository for Laguna State Polytechnic University - Sta. Cruz Campus with Intelligent ChatGPT Integration and Collaborative Features for Efficient Research Management.  Natural language processing techniques such as Generative Pre-trained Transformer will be utilize to generate automated analytics and insights in the research repository system. This involves using machine learning algorithms to analyze the vast amount of data available in the research repository and generating insights and trends that are relevant and useful to users.  Data visualization and interactive dashboard techniques will be used to provide insights into research trends, topic areas, and user behavior within the repository. This involves creating visually appealing and interactive dashboards that enable users to easily explore the research data and identify trends and patterns.  Machine learning algorithms will be use to analyze research materials in the repository and make personalized recommendations to users based on their research interests and preferences. This involves developing recommendation algorithms that take into account a user's research history, preferences, and behavior to provide relevant and personalized recommendations.  Data mining and text analysis tools will be use to allow researchers to analyze large datasets and extract insights and trends from research materials. This involves developing data mining and text analysis tools that enable users to perform advanced analysis on the research data, including identifying patterns, trends, and relationships in the data.  To implement these features, I plan to use a combination of programming languages such as Python and tools such as Jupyter Notebooks, Flask, and PyTorch. I will also use data visualization tools such as Tableau and D3.js to create interactive and visually appealing dashboards. Additionally, I will leverage machine learning libraries such as scikit-learn and TensorFlow to develop recommendation algorithms and perform data analysis.  To ensure the effectiveness of the research repository system, I plan to conduct usability testing and collect feedback from users to identify areas of improvement. I will also evaluate the system's performance using metrics such as accuracy, precision, and recall.    **Figure 1. Conceptual Framework**  The conceptual framework presented in Figure 1 outlines the process involved in developing the ResearchVault. It has three main components: research repository, intelligent ChatGPT integration, and collaborative features. The research repository serves as the primary source of data where researchers can store, organize, and manage their research outputs. The intelligent ChatGPT integration provides an automated system that can assist users in their research-related inquiries, including citation management, PDF viewing, version control, and analytics. The collaborative features, on the other hand, allow multiple users to work together on a research project in real-time, enhancing efficiency and productivity. The scope of this study is limited to the development and testing of ResearchVault for Laguna State Polytechnic University - Sta. Cruz Campus. |
| **Target users / Beneficiaries:(Describe each Beneficiary)**  Faculty Members: The ResearchVault system will benefit faculty members at Laguna State Polytechnic University - Sta. Cruz Campus by providing them with a comprehensive and efficient research management tool. The system will allow them to easily upload, store, and manage their research materials, collaborate with colleagues on research projects, and find relevant research materials using intelligent ChatGPT integration.  Staff Members: Staff members involved in research management and administration will also benefit from the ResearchVault system. The system will streamline the process of managing research materials, making it easier and more efficient to track and organize research projects and materials.  University Management: The ResearchVault system will be a valuable tool for university management, as it will provide them with a centralized platform for managing research activities and materials. This will allow them to easily monitor and track research projects and ensure compliance with university policies and procedures.  Students: Students conducting research at Laguna State Polytechnic University - Sta. Cruz Campus will also benefit from the ResearchVault system. The system will provide them with a centralized platform for managing their research materials and collaborating with peers and faculty members on research projects. |
| **Significance of study:**  The development of the ResearchVault system has the potential to significantly improve research management at Laguna State Polytechnic University - Sta. Cruz Campus. The system's user-friendly interface and intelligent ChatGPT integration could provide a more efficient and seamless research management experience for faculty and staff members involved in research. The collaborative features of the system could also facilitate better teamwork and information sharing among researchers.  By implementing the ResearchVault system, the university could improve the organization, storage, and accessibility of research materials, which could have a positive impact on the quality and quantity of research produced at the university. The system's advanced security features could also ensure that sensitive research materials are protected from unauthorized access. |