CHAPTER 1

Abstract

In today's fast-changing academic digital environment, archiving systems are essential, particularly given the exponential rise in specialized knowledge in domains like thesis and capstone projects. These initiatives preserve arduous investigations and ground-breaking ideas as the foundational archives for original research. They serve as comprehensive, cross-disciplinary knowledge stores. These perfectly organized archives serve as a tribute to the commitment and rigorous thought that went into their construction, and they stand as a symbol of a collective effort to expand knowledge and foster novel ideas and innovations.

Final year student theses are stored in a specific room, organized on book racks without clear categories, authors, courses, or titles. Softcopies are burned into Compact Discs and also stored there. This manual system makes it challenging for junior students to locate desired theses, with potential for human error leading to disorder. Insufficient information on covers and compact disks further complicates retrieval. Compact disks are also susceptible to damage or theft. Ensuring thesis integrity is crucial for future reference. Time constraints further hinder access. Students lack continuous availability to visit the thesis room (Ismail, 2013). As stated by Sengupta (2014), in spite of being a main source of scholarly communication, the print theses and dissertations is mostly not accessible to outside world. In this way, the important information remains unused and unknown to users.

According to Vieko (2014) study, he emphasized that the existing method of administration, physical files are used to keep documents. The manual process takes up physical space and the data cannot be kept current to save time. There is no editing allowed to save time. The educators' archiving system is to bring about enhancements.

It is for this reason the researcher's motivation behind conducting this study is the basic shortcomings in the current thesis archiving system are discussed. It draws attention to problems including disarray, a lack of distinct categorization, and reliance on unstable formats like Compact Discs. These issues impede accessibility and endanger the integrity of the thesis. Time restraints and students' limited availability further aggravate the issue. Vieko's (2014) study underlines the need for physical file-based systems to be improved further and calls for more simplified and flexible archiving techniques.

ND Information Technology Department (2014), as cited by del Rosario et.al (2016), the process of Electronic Document Management System has to do with capturing, storing, indexing and retrieval. It improves the accessing of information by the users of the organization wherein they can easily search and retrieve the documents needed. Also, there will be security in such a way that the only authorized users are allowed to access the particular documents.

A Knowledge-Based System (KBS) is a special computer program that uses a structured database of specific knowledge to help make decisions and solve problems in a particular field. This database holds a wide range of information collected from experts and trusted sources. The aim of a KBS is to mimic the thinking process of human experts, allowing it to analyze complex situations, draw conclusions, and offer informed advice or solutions.

Expert System is a computer system that emulates the decision-making ability of human expert in a restricted domain. It is mainly developed using artificial intelligence concepts, tools and technologies. An expert system is typically designed to provide capabilities similar to those of a human expert when performing a task. Moreover, it can be used to drive vehicles, provide financial forecasts or do things that human experts do (Dheir and Abu-Naser, 2019).

The suggested archiving system has been enhanced with the inclusion of a chatbot featuring natural language processing (NLP) capabilities. This chatbot acts as a responsive interface for researchers interacting with the system. It functions as a knowledgeable research companion, proficient in comprehending user queries and engaging in conversations. The chatbot seamlessly accesses the subject-specific knowledge stored in the Knowledge-Based System, enabling it to provide real-time assistance in various ways. This encompasses identifying pertinent research materials, generating citations, offering tailored recommendations, and addressing user queries related to their studies. This integration elevates the archiving platform from a static repository to a dynamic and adaptable tool, significantly enhancing its overall utility and effectiveness for researchers and academics alike.

The proposed development to the archiving system in JRMSU-Dapitan Campus is expected to have various positive implications, according to the study. These include thesis credibility, quicker retrieval, improved accessibility, time efficiency, and a more user-friendly interface made possible by the use of a chatbot. Furthermore, the addition of a knowledge-based system is anticipated to increase its value for academics and researchers.

Theoretical and Conceptual Framework

This study is anchored on "Naïve bayes and text classification introduction and theory" by Raschka (2014) which stated that Naive Bayes classifiers, a family of classifiers that are based abstract on the popular Bayes' probability theorem, are known for creating simple yet well performing models, especially in the fields of document classification. Additionally, Raschka used the example of spam filtering via naïve bayes classifiers in order to predict whether a new text message can be categorized as spam or not spam.

Furthermore, the study of Rahim, Zulkarnain, and Jaya (2017) entitled "Double hashing technique in closed hashing search process" featured Hash Search method is almost similar to the method of direct search that is by using a certain formula both at the time of placement and search data. Hash Search has better place space efficiency than direct search. The function used by the direct search method has poor place use efficiency, in addition to their study, Hash Search method improves the modulo operation function. Modulo operation function is often referred to as the Hash Function and the data shelter is called the Hash Table. The Hash function is not a one-to-one function such as the sharing function of the direct search method so it is possible that some data have the same function results. This results in collisions occurring at the time of placement of data into tables so that a strategy is needed to overcome this collision. Thus, the researchers come up with the double hashing technique where the data will be formed into the table with same length and then search, the results of their study indicate that the search process with double hashing technique allows faster searching than the usual search techniques, and allows to search the solution by dividing the value into the main table and overflow table so that the

search process is expected faster than the data stacked in the form of one table and collision data could avoided.

Improving the Knowledge-Based System Lifecycle by Millette (2013), where Millette used the Knowledge-based system components that are composed of several independent parts. The first component is the knowledge base in which heuristic knowledge of the domain experts as well as pertinent facts about the problem are stored [Grosan11] [Hoplin90]. The second component is the inference engine that utilizes strategies from the searching and sorting domains to test the rules contained in the knowledge base on a particular problem. The inference engine accomplishes this by querying information from the knowledge base and applying the returned results. The knowledge acquisition module, the third component, facilitates the transfer of knowledge into the knowledge base for future use [Grosan11] [Hoplin90]. The fourth component is the user interface that allows users to interact with the knowledge-based expert system by presenting the problem to the inference engine and viewing solutions. The fifth component is the working storage which the knowledge-based expert system uses to store information while a specific problem is being solved [Aniba09] and then contains the information pertaining to the solution. The researchers decided to only use this part of Millete's study as it directly aligns with their chosen field of study, which is the thesis and capstone archiving system with integrated knowledge-based referencing system.

Researchers developing an ingenious strategy that makes it simpler to find academic papers. To assist students and researchers, it makes use of specialized computer science approaches. This system will neatly arrange paperwork and provide immediate, specific suggestions. It is far superior to the previous method of putting papers on shelves.

To make the system function properly, the researchers combining various computer science concepts. Research is made simple because of it; it is like having a helpful assistant right at your fingers. The researchers anticipate that this will significantly increase learning and discovery.

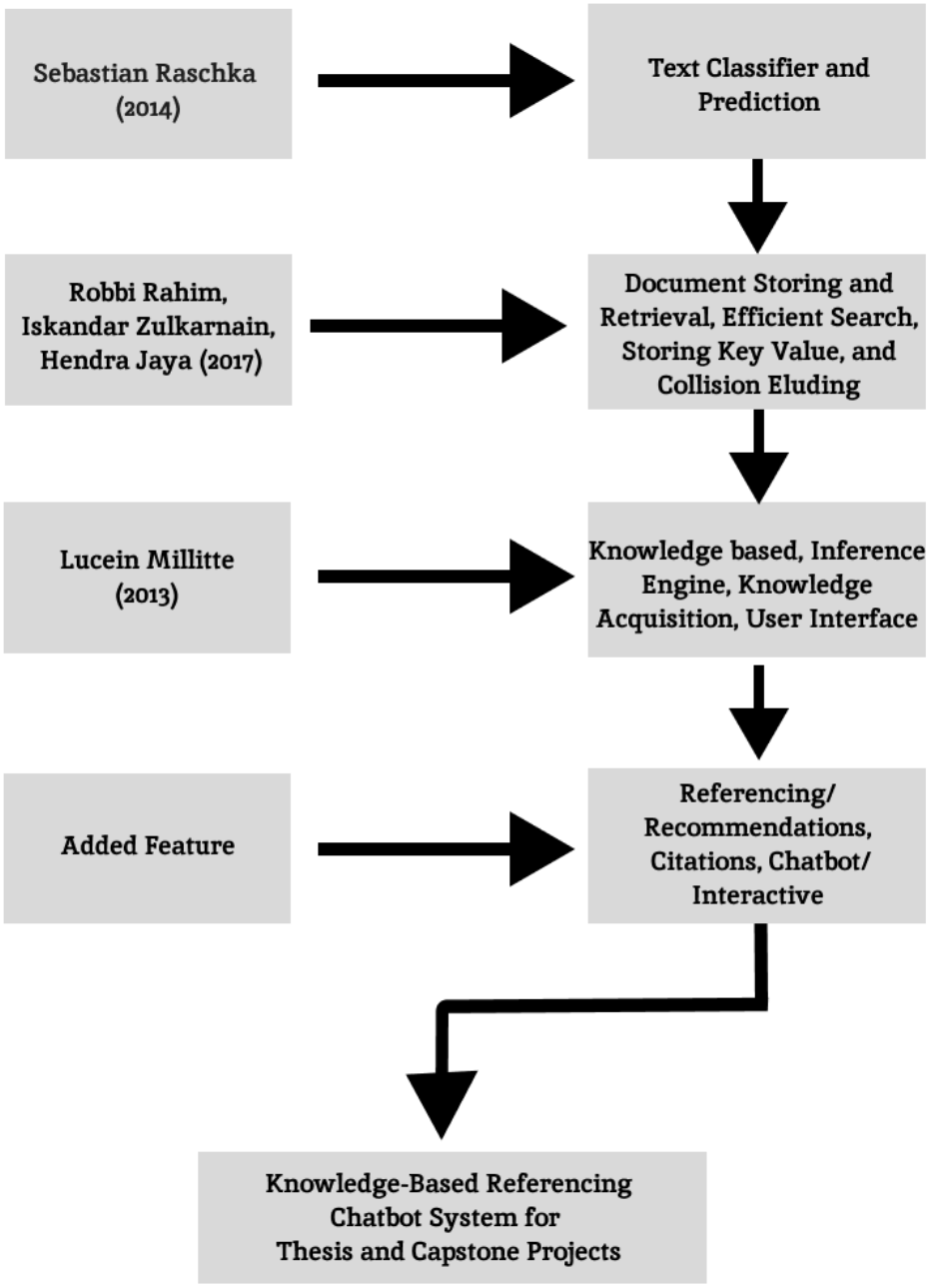

Figure 1: Conceptual Framework

Schema of the Study

In the figure 2 shows the flow of the processes of data starting from input, process, output and feedback which gives off a representation of how each box are utilize in the research study.

The first box which is input, it includes Software, HTML and CSS, JavaScript, ReactJS, NodeJS, ExpressJS and MySQL serves as the foundation for web development on creating the Archiving System that is integrated with Knowledge-Base Referencing Chatbot.

Second box labeled as process includes Data Gathering, Requirement Analysis, Designing, Coding, Testing and Evaluation, Implementation and Documentation. To further discuss, this is where the uploading of documents such as Thesis and Capstones to the system to be then scrutinize in which classification will it fall unto. The Naïve Bayes places a role in deciding which category each document belongs to and automatically tags them to which relevant base on their content and topics. To much more precise and coherent categorization, Double Hashing takes place in avoiding collision of data especially in the process of classifying the documents to their respective categories, with that being said the retrieval of metadata related to the thesis and capstones will become more efficient when it comes to labeling.

Knowledge-Base in an archiving system complements the archived thesis and capstone projects by providing context, and guidance. Users can discover related content more effectively. When searching for a specific topic, the Knowledge-Base can suggest relevant thesis and capstone projects, expanding the scope of search results. An Inference

Engine role is to draw logical conclusions, make recommendations, and provide intelligent insights based on the information stored in the system. When a user searches for specific thesis or capstone projects, the inference engine can analyze the user's query and preferences to determine the most relevant results, considering factors like keywords, and user history. The third box named Output is responsible for the user's experience and this is where the interaction happens to the extent of exchanging queries and recommending thesis and capstone projects. User Interface role is to ensure that users can interchange with the system efficiently and effectively. With the integration of Chatbot, will answer user queries, provide recommendations, and offer guidance on using the archiving system. It provides immediate and personalized assistance to the users. They can help users navigate the archiving system, answer questions, and guide them to relevant documents.

Lastly, feedback comes last as it takes the role in allowing observation of the user to be catered with the concern that they appear to encounter and so as queries that are asked in searching for related content and topics for their respective thesis, capstones or research projects. With that being said, through receiving feedbacks this will improve the system to provide much information, and data. Furthermore, it can interact with input since the said operation takes a vital role in providing information and are responsible for scrutinizing data that are uploaded in the system and categorizing each of its contents and topics of the thesis and capstones projects.

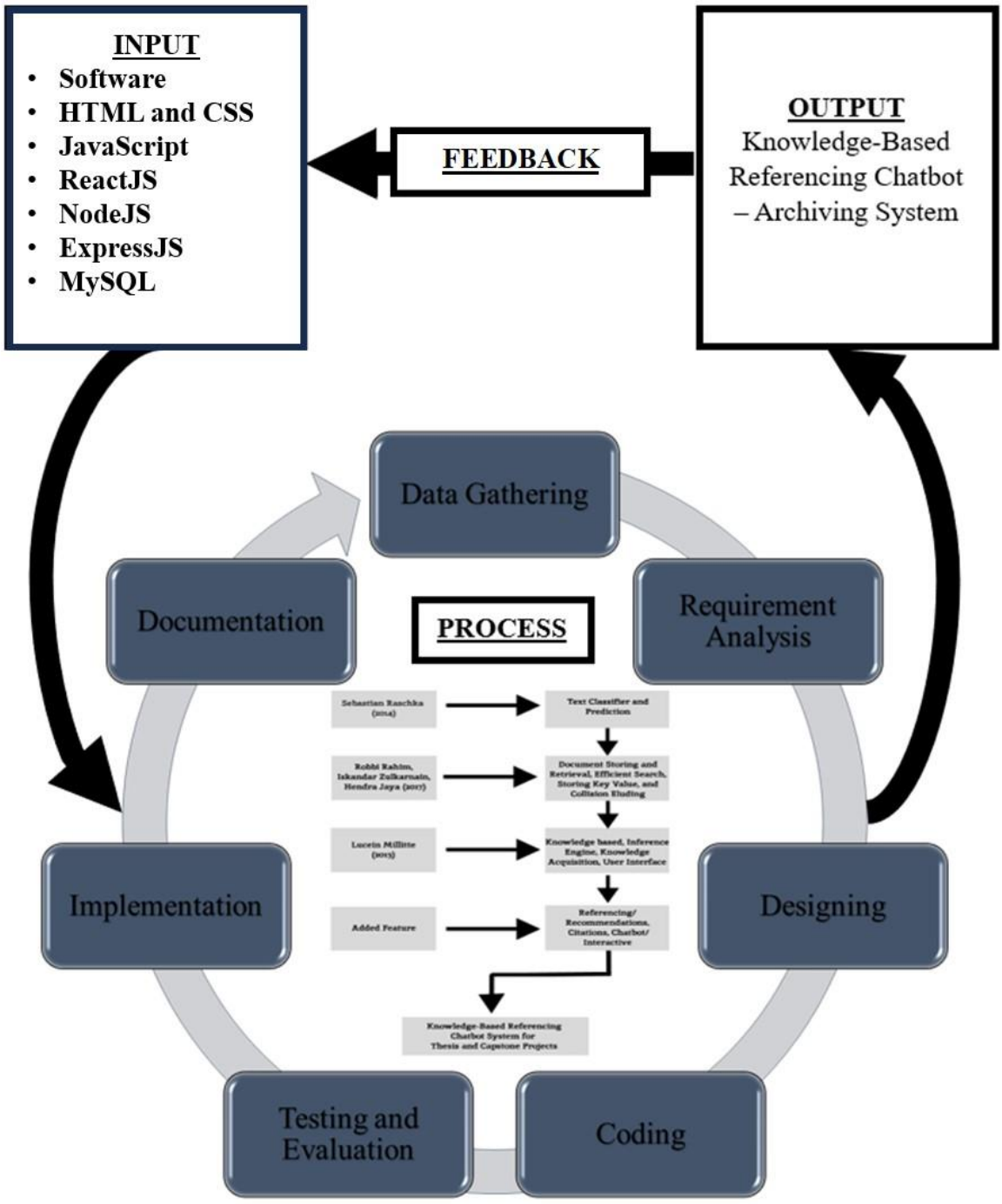

Figure 2: Schematic Diagram

Statement of the Problem

Generally, the study aims to develop and provide the efficiency of implementing the Archiving System with Integrated Knowledge-Base Reference Chatbot in Jose Rizal Memorial State University- Dapitan Campus during the school year 2023-2024.

Specifically, the study sought to answer the following questions.

1. What is the current status in archiving of researches output and studies of Jose Rizal Memorial State University Main Campus
2. What system can be developed to enhance the efficiency of research access to researchers and students?
3. What are the processes involving in the development of Archiving System with Integrated Knowledge-Base Referencing Chatbot?
4. What are the features of the existing Archiving System can be integrated into a hybrid system?
5. How does “Enhancing Research Access: A Thesis and Capstone Archiving System with Integrated Knowledge-Base Referencing Chatbot” system can be assessed in terms of the following:
    a. Functionality;
    b. Security;
    c. Maintainability;
    d. Reliability; and
    e. Usability?

Significance of the Study

Significance of the study the integration of Knowledge base system and the reference chatbots in archiving system can benefit a lot with different field of study. Emerging with the vast area of field in research we tend to cope up with the idea of a faster pace way.

Researchers: Academic researchers, faculty members, and scholars can retrieve files efficiently. This can also fasten their way of processing research to discover important data for their projects.

Students: Students, can benefit to our study. Adhering them a more convenient way in accessing research materials, thesis and capstone projects which helps them advancing their way of learning and research attempt.

Librarians: Librarians are in charge in retrieving obscure data, thus making them access the data in a feasible way.

Educational Institutions. The university can improve the academic infrastructure by implementing our proposed system. It enhances the reputation of the institution as a hub for research and knowledge access.

Library Staff. Staff members in the library can focus on providing more specialized support and services to students and researchers, contributing to a more enriching academic environment.

Web and Chatbot Developers. Our study can serve as sub reference for the researchers and developers who delves on the field of website, chatbots and knowledge base system in

different field of the industry. To conclude, our research tends to benefits with a lot of diverse possessors that involve the students, researchers, librarians, future students, educational institutions, library staff, administrators, web and Chatbots Developers

System Administrators. it can benefit from the system's data analytics capabilities, which can give more insights knowledge in accessing, researching topics and gaining informative data.

Scope and Delimitation of the Study

The scope of the study enhancing research access: a thesis and capstone archiving system with integrated knowledge-based referencing chatbot will be conducted in Jose Rizal Memorial State University-Dapitan Campus. It only involves the inside context of the archiving system on the educational setting, focusing on the storage and retrieval of thesis and capstone projects. It fathoms to the realms of knowledge-based systems and chatbot technology to enhance how archiving system works. The main focus on the study is to help the institutions like universities and colleges where archiving academic project is a norm. Considering only the materials rather than the inclusion of the academic content of the academia. To add some, the knowledge base used in this research only compacted in the field of research purposes and the used of the chatbots responses are finite to research opportunities in the institutions. Despite on how diverse the high technological process of pieces in knowledge base systems and chatbots, referencing the complete analytical information of these medium isn’t furnished. This research project has some specific boundaries. Firstly, it's focused solely on organizing and storing academic work within Jose Rizal Memorial State University. It doesn't cover larger systems or chatbots. Additionally, how this system is put into practice might be different depending on the technological

resources available at each specific institution. Also, it's important to mention that this study won't get into the basic technical details of Artificial Intelligence and Natural Language and Processing. Instead, it will look at how these technologies can be practically used for organizing academic work. Notwithstanding, the study tend to aim the new method of action to encapsulate the advancement of retrieving data processes of academic thesis and capstone projects, within the said institutions ameliorating the accessibility and the efficiency of the archiving system.

Operational Definition of Terms

Archiving System. A platform designed to systematically collect, organize, store, and provide access to a compiled collection of thesis and capstone projects. The system incorporates features for document submission, metadata management, user-friendly search and retrieval, quality assurance, and integration with a knowledge-base and chatbot. Its objective is to enhance research accessibility and user experience by offering a centralized repository of academic works while being efficient document discovery and referencing.

Knowledge-Base System. A repository of information, explanations, references, and related concepts that enrich and contextualize the archived thesis and capstone projects. This system serves as a valuable resource to enhance the user experience by providing background knowledge, definitions, citation guidelines, cross-references, and explanations, enabling users to better understand, reference, and navigate the archived academic works. It plays a vital role in augmenting research accessibility and supporting informed exploration within the archiving system.

Expert System. An expert system is a computer program developed using artificial intelligence concepts and technologies. It mimics the decision-making capabilities of human experts within a limited domain, such as providing financial forecasts or assisting with specific tasks.

Chatbots. Conversational agents, also known as chatbots or chatterbots, are machine conversation systems designed to interact with human users using natural language. They can be employed for various applications, including customer service in e-commerce.

Natural Language Processing (NLP). Natural Language Processing is a field of artificial intelligence that focuses on enabling computers to understand, interpret, and generate human language in a valuable way. In the context of this study, NLP capabilities are utilized to enhance user interactions with the archiving system through the chatbot.

Double Hashing Technique. The double hashing technique is a method for addressing collisions and improving the efficiency of searching in data structures, particularly hash tables. It involves using two hash functions to determine the location for storing and retrieving data.

Inference Engine. The inference engine is a component of knowledge-based systems responsible for applying rules and drawing conclusions based on the information stored in the knowledge base. It plays a central role in decision-making processes.

Text Classifier. Text Classifier is an algorithm that helps categorizing and organizing the content of the archived theses and capstone projects automatically. It gives hand to the

users to easily locate and access relevant documents based on their research interests, improving their experience in archiving system.

Hash Table. Is a smart way to organize and quickly find information about the archived theses and capstone projects. It's like a well-organized index that makes searching and accessing documents faster and more efficient in your system.

User Interface. User Interface it is an interactive design and layout that users interact with when accessing the archiving system. It's like the user-friendly front door to the system, where users can search for documents, browse categories, and access the archived thesis and capstone projects with easily.

Thesis. Thesis is a research paper written by students in college or university. It's like a deep dive into a specific topic, where the author does research and shares their own insights.

Capstone. Capstone represents a significant final project or assignment that students complete at the end of their degree program. It's a practical application of their learning and often involves solving real-world challenges

Repository. A repository is like a digital library where you store and manage the thesis and capstone projects. It's a central place for archiving and organizing these academic works, making them accessible to users for research and reference purposes.

Metadata. Metadata refers to the information that provides details about the theses and capstone projects. It contains information such as the title, author, publication date, and keywords. Metadata helps users find and understand the content within the archiving system more easily.