**HYBRID DOCUMENT MANAGEMENT SYSTEM FOR SANGGUNIANG BAYAN USING BOYER-MOORE ALGORITHM AND IMAGE TO CHARACTER RECOGNITION**

Undergraduate Thesis Outline

Submitted to the Faculty of the

Information Technology Department

Cavite State University Naic

Bucana, Naic, Cavite

In partial fulfillment

of the requirements for the degree

Bachelor of Science in Computer Science

**GABRIEL CHRISTIAN D. ARCEGA**

**STANLEY C. PANAG**

**JHONVIC B. BONIFACIO**

June 2024

**BIOGRAPHICAL DATA**

**Mr. Gabriel Christian D. Arcega** was born on December 13, 2001, in 80 Capt. C. Nazareno St. Naic, Cavite. He is the youngest child among two children of Mr. Hector L. Arcega and Mrs. Teresita D. Arcega. He presently residing at Lopez Jaena St. Kanluran, Naic, Cavite.

He took up his primary education at Naicc Elementary School located at Capt. C. Nazareno St., Naic, Cavite, and graduated in 2008. He studied his secondary education with the strand of STEM (Science, Technology, Engineering, and Mathematics) at Western Colleges Inc. located in Latoria, Naic, Cavite and he graduated in 2020. At the year of 2020, he enrolled in Cavite State University’s Naic Campus to study Computer Science at the Bachelor’s Level.

He believes that all struggles and failure happen in purpose. After he finishes his college education, he wants to focus on what path did he choose and make a better future for his family

**BIOGRAPHICAL DATA**

**Mr. Stanley C. Panag** was born on December 18, 2001 at Villa Apolonia Pediatric Hospital Naic, Cavite but he grew up at Bucana Sasahan Naic, Cavite. He is the youngest among the children of Mrs. Marissa C. Panag and Mr. Freddie C. Panag

He finished his preliminary study in Bucana Sasahan Elementary School in 2014. When he finished his study, he continued his Junior High school at Naic Coastal National High School Extension located at Ibayo Silangan Naic, Cavite. Also known as Centro De Naic National High School today. He finished his Junior High School in 2018 with a great interest in Science and Mathematics. He decided to pursue his Senior High School at Western Colleges Inc. (WCI) located at Latoria Naic, Cavite, as a Science, Technology, Engineering and Mathematics (STEM strand) student and graduated in the year 2020. After he finished his senior high school, he enrolled in Cavite State University Naic Campus and took a Bachelor of Science in Computer Science course.

**BIOGRAPHICAL DATA**

**Mr. Jhonvic B. Bonifacio** was born on September 26, 2001, at Villa Apolonia Clinic in Naic, Cavite. He is the youngest of the two children of Mrs. Juliet Bonifacio and Mr. Danilo Bonifacio. He currently resides in Timalan Concepcion, Naic, Cavite. His parents own a sari-sari store.

He completed his elementary education at Timalan Elementary School in 2014 and his Junior High School education in 2018 at Western Colleges Inc. (WCI) located in Latoria, Naic, Cavite. He finished his Senior High School at Western Colleges Inc. (WCI) in Latoria, Naic, Cavite, as a Science, Technology, Engineering, and Mathematics (STEM) strand student and graduated in 2020.

After completing his Senior High School, he enrolled at Cavite State University, Naic Campus in Bucana Malaki, Naic, Cavite, for his college education, pursuing a Bachelor of Science in Computer Science.

**ACKNOWLEDGMENTS**

The authors would like to express their sincere appreciation to the following individuals who significantly contributed to the completion of this study:

Mr. Henry R. Balanza, thesis adviser, for his invaluable guidance, knowledge, and logical thinking skills, as well as the motivation he provided to help the authors complete this study.

Mr. Jomari M. Launio, technical critic, for his insightful advice and comments, which encouraged the authors to continue improving the contents of this study.

Mr. Ricky Jay A. Riel, Information Technology Department Chairperson, for his guidance regarding all the requirements needed to pass this project, and for motivating us to successfully complete the semester.

Ms. Armina B. Pugay, Research, Development and Extension Services (RDES) Director, for her continuous guidance and support.

Mrs. Leah Navarro, our beloved Campus Administrator, for her unwavering support in every group project.

Above all, we humbly dedicate these words to our Almighty God, grateful for the life and strength He has given us, for His guidance, and for keeping everything in His hands.

**ABSTRACT**

**ARCEGA, GABRIEL CHRISTIAN D., PANAG, STANLEY C., BONIFACIO, JHONVIC B. Hybrid Document Management System for Sangguniang Bayan Using Boyer-Moore Algorithm and Image to Character Recognition.** A Project Design. Bachelor of Science in Computer Science. Cavite State University – Naic, Cavite. June 2024. Thesis Adviser: Mr. Henry R. Balanza.

This study aimed to develop a hybrid document management system for Sangguniang Bayan, integrating the Boyer-Moore algorithm and image-to-character recognition. The researchers used React Native Android Platform and Supabase to develop the system, which enhances document management by incorporating efficient text search and character recognition from images. The study's goal is to create an approach to implement this unique document management and system, which will use effective algorithms to achieve the best possible results. The "Hybrid Document Management System for Sangguniang Bayan, integrating the Boyer-Moore Algorithm and Image to Character Recognition," introduces a new way of managing documents for local government units. By using smart technology like advanced algorithms and Optical Character Recognition (OCR), this system makes it easier to handle documents and data. It helps with tasks like finding specific information in documents and turning printed words into digital text. With careful planning and building, the system now lets users gather data faster, make fewer mistakes when managing documents, and access files more easily, even online. This abstract gives a brief overview of how the system was created and why it could be useful for other local government units facing similar challenges in managing documents effectively.

**TABLE OF CONTENTS**

|  |  |
| --- | --- |
|  | **Page** |
| **APPROVAL SHEET**…………………………………………………………………. | ii |
| **BIOGRAPHICAL DATA**..................................................................................... | iii |
| **ACKNOWLEDGEMENT**..................................................................................... | vi |
| **ABSTRACT**........................................................................................................ | vii |
| **LIST OF TABLES**............................................................................................... | x |
| **LIST OF FIGURES**.............................................................................................. | xi |
| **LIST OF APPENDIX TABLES**……………………………………………………… | xiii |
| **LIST OF APPENDIX FIGURES**…………………………………………………….. | xiv |
| **LIST OF APPENDICES**...................................................................................... | xv |
| **INTRODUCTION**................................................................................................. | 1 |
| Statement of the Problem.......................................................................... | 3 |
| Objectives of the Study............................................................................. | 4 |
| Significance of the Study........................................................................... | 5 |
| Time and Place of the Study..................................................................... | 6 |
| Scope and Limitation of the Study............................................................. | 6 |
| Definition of Terms.................................................................................... | 7 |
| Conceptual Framework of the Study......................................................... | 8 |
| **REVIEW OF RELATED LITERATURE**.............................................................. | 10 |
| Synthesis of the Literature………….......................................................... | 19 |
| **METHODOLOGY**................................................................................................ | 20 |
| Research Design ……………………………………………………………... | 20 |
| Participants of the Study …………………………………………………….. | 27 |
| Data Gathering Procedure …………………………………………………... | 27 |
| Sampling Technique ……..…………………………………………………... | 28 |
| Data Analysis …………….…………………………………………………… | 28 |
| Statistical Treatment of Data …….………………………………………….. | 29 |
| Research Instrument ……………….………………………………………… | 30 |
| **RESULTS AND DISCUSSION**........................................................................... | 31 |
| Project Description...………………………………………………………..... | 31 |
| Project Structure………...…………………………………………………….. | 31 |
| Implementation Guidelines……………….………………………………….. | 47 |
| Evaluation Results…………..………………………………………………... | 47 |
| **SUMMARY, CONCLUSION AND RECOMMENDATION**................................... | 52 |
| Summary…….…………….…………………………………………………… | 52 |
| Conclusion………………………….………………………………………….. | 53 |
| Recommendations………………….…………………………………….…… | 53 |
| **REFERENCES**.........................................……….………………………………… | 55 |
| **APPENDICES**........................................……….………………………………….. | 59 |
|  |  |
|  |  |

**LIST OF TABLES**

|  |  |  |
| --- | --- | --- |
| **Table** |  | **Page** |
| 1 | Data dictionary………….…………………………………………………... | 25 |
| 2 | Breakdown of Participants …………………………..……………………. | 27 |
| 3 | Rating Scale………………………………………………………………… | 28 |
| 4 | Implementation Plan…………………………………………….…………. | 48 |
| 5 | Summary of Evaluation Result from 100 respondents…………………. | 49 |

**LIST OF FIGURES**

|  |  |  |
| --- | --- | --- |
| **Figure** |  | **Page** |
| 1 | Conceptual framework of the study (IPO)............................................... | 8 |
| 2 | Agile software development life cycle……….......................................... | 20 |
| 3 | Use case diagram................................................................................... | 22 |
| 4 | Hierarchical Input Process Output……………..……………………..…… | 23 |
| 5 | Entity Relationship Diagram……………………….…………………….…. | 24 |
| 6 | Sign Up page……………………………………………………………........ | 32 |
| 7 | Sign In Page………………….………………………………………………. | 32 |
| 8 | Home Page…………………………………………………………………... | 33 |
| 9 | Resolution Page…………………………………………………………....... | 34 |
| 10 | Ordinances Page…………………………………………………………… | 34 |
| 11 | Dashboard…………………………………………………………………… | 35 |
| 12 | Manage User page………………………………………............................ | 35 |
| 13 | Manage User page Modal……..……………………................................. | 36 |
| 14 | Manage Resolution Page……………………….……………..…………… | 36 |
| 15 | Manage Resolution Page modal…………….....……………..………….... | 37 |
| 16 | Manage Ordinance Page……………………….……………..………….... | 37 |
| 17 | Manage Ordinance Page Modal……………..…...…………..…………… | 38 |
| 18 | Sign In page in React Native App………………………………………….. | 39 |
| 19 | Document Scanner Button.…………………..…...…………..………….... | 40 |
| 20 | Scanning of Document……………………............................................... | 41 |
| 21 | Crop and Rotate of Scanned Document..……….................................. | 42 |
| 22 | After Scanning Document..................................................................... | 43 |
| 23 | Dirt Remover Option…………………………….……………………..…… | 44 |
| 24 | Color Change Option……………………..…………………………….… | 45 |
| 25 | Delete Scan Photo………………………………...................................... | 46 |
| 26 | Checking Page / Submit button………….……...................................... | 47 |

**LIST OF APPENDIX TABLES**

|  |  |  |
| --- | --- | --- |
| **Table** |  | **Page** |
| 1 | Gantt Chart………….……………………………………………….……... | 60 |
| 2 | Data Dictionary………….…………………………….…………………... | 67 |
| 3 | Summary of Evaluation Result.………………………………………... | 68 |
|  |  |  |

**LIST OF APPENDIX FIGURES**

|  |  |  |
| --- | --- | --- |
| **Figure** |  | **Page** |
| 1 | Request Letter for System Evaluation …………………………………... | 61 |
| 2 | Evaluation Instrument…… …………………………..……………………. | 62 |
| 3 | Application letter for Final Defense…………………..…………………… | 64 |
| 4 | Request for Final Defense…………..………………………….…………. | 65 |
| 5 | Request Letter for the Client ……………………………………………… | 66 |
| 6 | System Requirements and Step 1………………………………………... | 70 |
| 7 | Step 2………………………………………………………………………… | 70 |
| 8 | Step 3………………………………………………………………………… | 71 |
| 9 | Step 4………………………………………………………………………… | 71 |
| 10 | Step 5………………………………………………………………………… | 72 |
| 11 | Step 6………………………………………………………………………… | 72 |
| 12 | Step 7………………………………………………………………………… | 73 |
| 13 | Step 8………………………………………………………………………… | 73 |
| 14 | Step 9………………………………………………………………………… | 74 |
| 15 | Step 10………………………………………………………………………. | 74 |
|  |  |  |

**LIST OF APPENDICES**

|  |  |  |
| --- | --- | --- |
| **Appendix** |  | **Page** |
| 1 | Timetable of activities (Gantt Chart)……………………………….……. | 60 |
| 2 | Request Letter for System Evaluation………………………………..… | 61 |
| 3 | Evaluation Instrument……………………………………………….……. | 62 |
| 4 | Application letter for Final Defense……………………………………… | 64 |
| 5 | Request for Final Defense……………………………………………….. | 65 |
| 6 | Request Letter for the Client…………………………………….………. | 66 |
| 7 | Data Dictionary……………………………………………….…………… | 67 |
| 8 | Summary of Evaluation Result………..…………………………………. | 68 |
| 9 | Operation Manual…………………………………………………………. | 70 |

**HYBRID DOCUMENT MANAGEMENT SYSTEM FOR SANGGUNIANG BAYAN**

**USING BOYER-MOORE ALGORITHM AND IMAGE TO CHARACTER**

**RECOGNITION**

**Gabriel Christian D. Arcega**

**Stanley C. Panag**

**Jhonvic B. Bonifacio**

A Thesis project submitted to the faculty of the Information Technology Department, Cavite State University Naic, Cavite in partial fulfilment of the requirements for the degree of Bachelor of Science in Computer Science with Contribution No. \_\_\_\_\_\_\_\_\_\_\_\_\_\_. Prepared under the supervision of Mr. Henry R. Balanza

**INTRODUCTION**

In our modern era, an efficient and effective document management system is vital to modernizing operations and boosting productivity for organizations. To efficiently handle the large growing volume of paper-based documents in an office/organization, would require migrating to an Electronic Document Management System.

The constant increase in the size of the organizations and what they contain of Documents, many showed a need and desire of these organizations to organize their data mechanism in a way to save time, effort and cost, which led to the escalation of the need for electronic documents that contribute to the management systems and organization of all special operations to deal with the data automatically, such as indexing, archiving, summary and search and retrieval (Said Kittanah et al., 2018).

The office of Sangguniang Bayan of Naic has massive amounts of paper-based documents that must be scanned to PDF, classify the content of the generated PDF to identify its content, and lastly, an efficient and user-friendly document search functionality. This research aims to innovate designing of a document management system that will be beneficial for local government unit offices like Sangguniang Bayan.

This proposed hybrid system will be designed as a web application and an Android-based application. The basis of the design of the system would be the business processes of the Sangguniang Bayan of Naic thus, the features and functionality of the system will be divided and assigned to either the platform Website or Android. In the development of the website platform, it will be developed by using current emerging technologies such as Deep Learning Image Analysis that uses Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) to be applied for many tasks such as processing of images or image classifications, object classification, and text recognition, etc. (Jiuxiang Gua, et al., 2017). The CNNs are utilized in an open-source library like “LayoutParser” which is used for streamlining the usage of Deep Learning in Document Image Analysis research and applications (Zejiang Shen et al., 2021). In developing the front-end and back-end of the proposed website will be built around the current trends of technologies in web development, such as ReactJs library, NodeJS for the environment, Express API for handling user requests, and Supabase, a cloud database. The technologies stated provide advantages in web development based on empirical studies. React Js is a modern way to build websites with many features to meet the requirements of uprising trends (Aggarwal, 2018). Additionally, the use of cloud computing databases like MongoDB is a viable solution for applications that require data to be scalable (Sherif Sakr, 2012). Dealing with a large dataset decelerates the performance of a search functionality thus, it requires identifying a proper search algorithm to be implemented. Boyer-Moore algorithm is a string search variation that jumps as far as possible. The Boyer-Moore algorithm compares patterns from left to right. A text is matched with a specific pattern to determine whether there is a pattern in the text matched. (E. G. Hartoyo et al., n.d.). In comparison to other string-matching algorithms, the Boyer-Moore algorithm includes the most efficient string-matching algorithm (Fitriyah, F., 2020). For the development of the mobile platform the functionality of image-to-PDF conversion will be implemented using Optical Character Recognition (OCR) which is a widely used method to digitize printed/paper-based materials (Li L. et al., 2020). OCR is recognizing text on scanned materials (like images) and converting that text into easily understandable encoded text.

However, there may be numerous ways of scanning images and conversions to PDF files. Such as the implementation of batch scanning of documents to pdf that requires hardware such as scanners and modern printers to store documents automatically by software in a website. This study is designed to employ innovation in document management systems with the use of a mobile phone and a personal computer.

**Statement of the Problem**

The employees of the office of Sangguniang Bayan lack a computerized library for the municipal resolutions, they still store documents on paper which takes up much space in the office. With the lack of a computerized library, it resulted in an inefficient and time-consuming search for resolution documents. Even if they have document management on paper, they still need to send copies of resolution documents by electronic mail or e-mail. There is a slow processing of amendments in the office of the Sangguniang Bayan. It is the outcome of the lack of digital access and modifications of the resolutions.

Specifically, the study aimed to provide solutions to the problems such as:

1. How to design and develop a hybrid document management system which will be capable of:
   1. Document scanning
   2. Document archiving
   3. Online registration
   4. Admin management
   5. Enhanced Searching
   6. Document Digitization
2. What are the algorithms that can be used in developing the hybrid document management system?
3. What is the level of conformity of the (software type/name) with the ITD Software Evaluation Tool?
4. How to implement the Hybrid Document Management system?

**Objectives of the Study**

The main objective of the study was to developed a Hybrid Document Management System, that aims to improve digitalizing and organizing documents, specifically, resolutions and ordinances using image text recognition technology with Boyer-Moore searching algorithm.

Specifically, the study aims to accomplish the following objectives such as:

1. To design and develop a hybrid document management system (DMS) called "HDMS" that will have the following features:
2. Document scanning – refers to classifying different data on an image file of a document through image to text extraction processes.
3. Document archiving – module that has search filters to efficiently retrieve files for users.
4. Online registration - to provide security and identification of users.
5. Admin management – a module to view, submit and edit documents based on amendments from government and assign roles to users.
6. Enhanced searching - that implements the Boyer-Moore Search Algorithm to improve the search for document title keywords or resolution numbers.
7. Document Digitalization - refers to conversion of paper documents into searchable digitized files through processes of image text extraction.
8. To identify the algorithms used in developing the hybrid DMS.
9. To develop the hybrid DMS in order to determine if it is compliant with the ITD Software Evaluation Tool.
10. To prepare an implementation plan for the deployment of the hybrid DMS

**Significance of the Study**

This research presents a Hybrid Document Management System for Sangguniang Bayan that includes the Boyer-Moore Algorithm and Image to Character Recognition that will help the SB Employees and their clients.

The system promises faster and more accurate document searches, which will help people to search or to find the resolutions more quickly. It also improves the speed of collecting information from images, so it can handle a wide range of content. The system makes managing documents easier, which saves time and effort. In real life, it means that the Sangguniang Bayan can improve overall workflow and business processing.

The system also meets regulatory requirements, which makes sure that rules are followed, and data integrity is maintained. The main importance of the study is based on its potential to change the way documents are managed by providing a simple solution that makes governance processes more efficient and accurate.

**Time and Place of the Study**

The collection of data started on May 2023 at Barangay Ibayo Silangan Naic Cavite. The researcher is currently collecting data for an ongoing study at the Municipality of Naic – Sangguniang Bayan, using online reviews, interviews, and literature analysis.

Ongoing data collection the researcher starts to analyze all the statement to find the best solution how to resolve the problem and to give them what are they needed, then the researchers start the documentation and plan how to make the system, in October 2023 at Ibayo Silangan Naic Cavite the researcher start to design and develop the system.

**Scope and Limitation of the Study**

The purpose of this research is to create a complete Document Management System (DMS) suited to the Sangguniang Bayan of Naic. The scope of the system includes the various document types commonly managed by the organization, such as resolutions and ordinances. The Optical Character Recognition (OCR) module extracts text content from images and scanned documents, and it supports common formats such as JPEG, PNG, and PDF. The Boyer-Moore Algorithm will allow for efficient searching within the extracted text, allowing for the fast retrieval of specific keywords, phrases, or relevant information. Administrators, document uploaders, and general users will be considered, each with their own set of permissions and access levels. Users will require little orientation because the user interface will be designed to be accessible or easy to learn. First, the system is intended for future integration with existing systems and databases.

There are some problems that need to be known about how this Document Management System is used. The quality of scanned documents or images, especially those with a lot of noise or low quality, may affect how well the OCR tool works. The OCR module's language support will mostly focus on English text recognition and not a Tagalog, though other languages may not work as well. The Boyer-Moore Algorithm's performance in string searching depends on the type of data it is used on. Better security features, like encryption and advanced access controls, will be added in the future. Performance may be different depending on the hardware and network infrastructure. For the system to work well in the long run, it will need to be constantly evaluated and changed to meet new requirements.

**Definition of Terms**

For a better understanding of the study, the following terms are defined in the context of this research.

**Algorithm** – is a set of instructions that will be used to achieve the desired goal of the system.

**Boyer-Moore Algorithm** – is a pattern searching algorithm to be used in the search engine of the system.

**Database** - refers to the virtual storage of data of the system for the Sangguniang Bayan Office of Naic, Cavite.

**Hybrid Document Management System (HDMS) –** refers to an organized document management architecture integrated in both mobile and website system for Sangguniang Bayan of Naic, Cavite.

**Node.js** – is a JavaScript runtime environment for the back end of the system.

**OCR (Optical Character Recognition)** - It is a technology that recognizes text within a digital image that will be used to scan documents for the image to text character recognition functionality of the system.

**Sangguniang Bayan (SB) –** refers to the client of the researchers for this project. It is a local legislature of the municipality government.

**React.js** - it is a JavaScript library that enclosed the web client-side of the Hybrid Management System, and it used functional components.

**Vite** - it is a powerful build tool that supports HRM (Hot Module Replacement) used to build the client-side of the Hybrid Management System for Sangguniang Bayan Office quicker.

**JSX** - stands for JavaScript XML that allows to incorporate HTML in building the front end of the Hybrid Document Management System.

**Conceptual Framework**

Based on the objectives and different significant requirements, the Researchers have come up with the conceptual framework shown below.

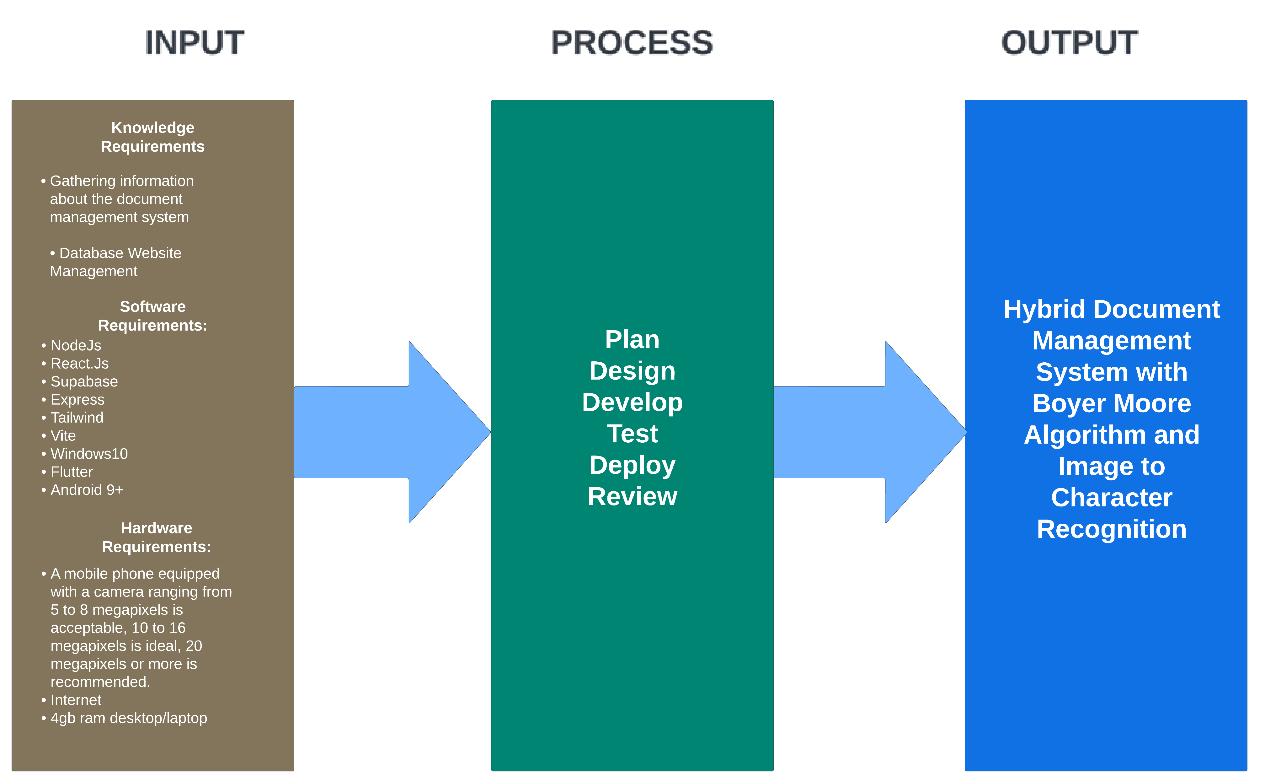


Figure 1: Conceptual Framework

As show ­­in figure 1 the input has 3 different type of requirement knowledge, software, and hardware. The knowledge is about the information of the study.

This information is gathered by doing research and interviews in another related topic. The software is use for managing and building the website in the helps of SupaBase for the database of the system, NodeJS and Vite for the backend of the system, Vite and React for the frontend of the website, Flutter and Dart for the mobile frontend, Visual studio code IDE use for coding. The hardware is the electronic computer part that use to develop the software, computer at least 4GB ram, internet connection, and a mobile phone equipped with a camera ranging from 5 to 8 megapixels is acceptable, while one with 10 to 16 megapixels is ideal, and a recommended choice would feature a camera with 20 megapixels or more.

In the process indicate the flow of the studies and the step-by-step procedure on how to finish this work, plan how to start the program, designing the- website for the interface, develop to start the program, testing to make sure it works properly, deploy to use the program and review to know what are needed to add on the program.

The output of the research study will be of Hybrid Document Management System with Boyer-Moore Algorithm and Image to Character Recognition

**REVIEW OF RELATED LITERATURE**

This chapter presents the relevant literature and studies that the researchers considered in strengthening the importance of this study. It also presents the synthesis of the art to fully understand the research for better comprehension of the study.

**Agile Methodology**

There are a variety of methodologies that we can use for software projects. Traditional methods are one of the most effective and yielded good results, but they are hindered by the ongoing increase in requirements brought about by the rapid advancement of science and technology as well as the digital transformation. Highly structured, traditional project management methods lack flexibility, awareness of change, adaptability to it, and independence. This led to the development of Agile methodology as it exists today, which has improved the team's knowledge and abilities. (Alzeyani, Szabo 2023).

A combination of waterfall and agile methodology was used to guide the application development activities. Despite the many debate on Waterfall methodology, Davis suggested the it is among the renowned and wellunderstood approach among system developer, hence offer a good transition to embed and leverage new techniques that agile methodology has. (Kartiwi et al., 2018)

**Cloud Computing**

Cloud computing is the delivery of IT resources through a cloud services platform via the Internet. It offers on-demand computing power, database storage, applications, and more. Cloud computing enables data storage, recovery, backup, application creation, software delivery, website hosting, and more. Users can access cloud services dynamically via the internet. (KamalaKannan et al., 2019)

Cloud computing refers to the collection of networks, allowing users to use various modalities without requiring physical infrastructure. Users pay only for the services they use, and the workload can be shifted to reduce it. Networks handle the service load, reducing hardware and software requirements. All users need a web browser like Chrome to use cloud computing (Srivastava & Khan, 2018). Cloud computing offers numerous benefits for Document Management System. It allows for efficient storage and management of large digital documents, with expandable storage. Cloud service providers offer strong security measures and reliable backup solutions, protecting documents from unauthorized access, data loss, or system failures.

**Cloud Database**

One of the Cloud service providers offers is a Cloud Database. According to Google Cloud, a cloud database is a database deployed, delivered, and accessed in the cloud, organizing and storing structured, unstructured, and semi-structured data like traditional on-premises databases. It offers benefits like speed, scalability, agility, and reduced costs. Cloud databases can be offered as managed DBaaS or deployed on a cloud-based virtual machine (VM) and self-managed by an in-house IT team.

Cloud services utilize a wide variety of database management systems, including relational and non-relational databases (Zohra M. & Nadia NT., 2019). Relational databases organize data in predefined relationships, while non-relational databases store and manage unstructured data. The exponential growth of data makes it difficult to integrate traditional approaches. Legacy databases face challenges, but cloud databases offer flexibility, reliability, security, affordability, and a solid foundation for modern business applications. They can adapt to changing workloads without increasing workloads, and provide better reliability, flexibility, and cost savings (Google Cloud, n.d.). Cloud databases are faster, more scalable, flexible, and cost less than traditional on-premises databases. Cloud databases are flexible, reliable, and safe, reducing operational costs.

**Hybrid Application Development**

Aside from cloud services reducing developing costs, Hybrid Application Development is also a solution for the higher cost and development time of developing applications that is compatible with many platforms like Android, iOS, Windows, etc. (Pinto, C. M., & Coutinho, C., 2018). While the traditional approach of developing applications in other platforms requires the developers to know specific programming languages that are native to the platform’s Software Development Kit (SDK) such as Java/Kotlin in Android Studio and Objective /C++ in Swift for Apple. Hybrid app development enables a more time-efficient development of Web Applications that can run on any device/platform or also known as Progressive Web Apps (PWAs). These PWAs are run by web technologies mainly HTML, CSS, and Javascript that are put inside the native shell of a device (Ali, A. A. A.., & Dauwed, M., 2022). PWAs require Hybrid Development Frameworks or researchers cite Cross-platform App Frameworks (CAF), which allow developers to write a unified code to deploy their app on every platform.

Although CAF benefits from time-saving development and unified code in multi-platforms, CAFs have the same drawback: the user experience. The app is supposed to look and feel native to the platform, as PWAs fail to achieve it as it was initially developed for the web application (Kuitunen, M., 2019). As Kuitenen also stated that this inconvenience was addressed by native scripting frameworks like React Native, which utilize an interpreter to execute code during runtime. React Native provides UI platform controls directly and its framework environment is the same as React, React web developers can also easily migrate and code cross-platform mobile applications with React Native. As the React catchphrase of learn once, write anywhere.

**Optical Character Recognition**

In either Web apps, Native Apps, and PWAs can implement Optical character recognition (OCR), it is a method for automatically identifying various characters from a recorded image. It also offers full alphanumeric recognition of printed or handwritten characters, text numerals, letters, and symbols in a computer-processable layout, such as ASCII, Unicode, and so forth. **(Ahmed, 2019)**. This research study will use OCR technology alongside branches of the process below.

During the image acquisition, coding, transmission, and processing stages, noise is always present in digital images. **(Kishan et al., 2019)**. For the analysis of images, image de-noising is a crucial task in image processing. There are numerous image de-noising algorithms. However, the optimal one should remove the noise from maintaining the features of the image. There are both linear and non-linear de-noising techniques. While non-linear methods are faster, linear methods do not keep the features of the photos, which are preserved by non-linear methods. A variety of methods, including blurring (averaging, gaussian filtering, median filtering, bilateral filtering), erosion, dilation, opening and closing, mean filtering, standard median filtering, switched median filtering, progressive switching median filtering, adaptive median filtering, decision-based algorithms, improved DBA, and trimmed median filters, can be used to remove noise.

The accuracy of the OCR system greatly depends on how successfully Preprocessing and Segmentation are carried out, making them the most crucial of all the OCR steps. **(Reddy, 2019)**. The primary goal of the preprocessing stage is to make it as simple as possible for the OCR system to differentiate a character or word from the backdrop.

OCR software uses a variety of techniques, but they all generally focus on one character, word, or block of text at a time. **(TechTarget Contributor, 2022)**. After that, characters are determined by one of two algorithms. Pattern Recognition, when comparing and identifying characters in the scanned document, OCR algorithms compare instances of text in different fonts and formats that are provided into them. Feature Detection, for characters in the scanned document to be recognized, OCR applications use rules pertaining to the characteristics of a particular letter or number. A character's characteristics might include things like how many angled or intersecting lines there are, or how many curves there are. Using the capital letter "A" as an example, two diagonal lines may be saved, meeting in the middle with a horizontal line.

Error detection and rectification are the two tasks that post-processing techniques focus on. **(Jatowt, et al. 2021)**. The purpose of post-OCR error detection is to locate false positives in the input text. Error detection helps human helpers quickly fix mistakes and makes it possible to highlight noisy data for further processing as needed. Additionally, error detection generates a list of errors that are found as the input for post-OCR error correction, which is meant to fix erroneous tokens. Typically, a list of potential words is formed in response to a mistake and assessed using the information at hand. In automatic correction procedures, the candidate with the highest ranking is then selected to fix the issue, or in semi-automatic correction approaches, the user is presented with the top n candidates.

Deep learning models called convolutional neural networks (CNNs) are very good at recognizing images. **(Everen, 2023)**. OCR systems that use CNNs are capable of learning and generalizing features from input photos, which allows them to handle a variety of text recognition settings. Contrary to Tesseract OCR, which uses LSTM networks and rule-based techniques, OCR employing CNNs can adapt to different fonts, sizes, and layouts with greater accuracy.

**Mobile Phone Camera-Based OCR**

The need for handheld cameras to capture OCR document images has been steadily increasing recently. Traditional scanners' primary drawback is that they can only handle text on paper, or print documents, whereas handheld digital cameras may capture scene photographs in addition to text on paper.

However, it is quite difficult to recognize characters taken by a camera since they differ from those in scanned manuscripts. Compared to typical, standard font, plain printed text documents, text identification and recognition in documents with several parts, complicated and varied backgrounds, mixed fonts, or other irregular properties are far more demanding. The output of captured photos is typically distorted and skewed due to the non-contact nature of digital cameras attached to wearable or portable devices. Complex backgrounds, blur, skew, non-uniform lightning, low resolution, and perspective distortion all have an impact.

Processing document images obtained by desktop scanners is, at its core, the fundamental purpose of OCR because it is way clearer and it can get a better text recognition when it starts to undergo on the OCR. This is explained by the fact that scanned images have a sufficient resolution. **(Nwokoma, et al. 2021).**

**Binarization**

A common task in computer vision and digital image processing is image segmentation. Binarization is a specific type of segmentation that involves grouping image pixels into two categories, typically referred to as background and foreground. Because the document is binary in nature, this case is relevant to the field of document image analysis. (Rivero, et al. 2021). The correctly binarized document may be the end product in some applications (such as archive systems), or it may only be the first step in preprocessing before moving on to other phases. The most popular of them is optical character recognition, or OCR, which transforms the characteristics of the document into text that can be edited and transferred.

**Skewing**

The segmentation of text lines is primarily challenged by the skewed documents. Skew correction is therefore regarded as one of the pre-processing's most crucial subtasks. This problem is addressed by numerous standard techniques. These techniques take a lot of effort, and they can only handle a small range of skew angles. A script-independent, quick, dependable, and border- and noise-resistant skew detection method would be ideal. It ought to function for a respectable variety of angles as well.

Skewing is a common problem in document scanning, which can negatively impact optical character recognition (OCR) performance. (Xu, et al. 2022). Scanned papers will always have some skew. The extraction of the region or regions of interest (RoI) and further processing, such as detection and classification on such RoI, become challenging if this skew is not addressed. (Ahmad, et al. 2021).

**Document Management System**

A document management system is a computer program that keeps track of and stores digital copies of paper documents as well as electronic versions. **(Abbasova, 2020).** This word somewhat corresponds with the ideas of content management systems, which are frequently viewed as a component of enterprise content management systems and are connected to digital asset management, document mapping, document management systems, and records management systems. The word "document" typically refers to a container of information that holds written or drawn material in structured form for a specific purpose.  A document is typically thought of as a piece of paper or a collection of papers, like a note or a letter. The document's basic concept is that it may be simply communicated, stored, and processed in its entirety. The definition of the word "document" has drastically changed within the last ten years. Information technology is partially responsible for this change. As a result, most documents handled in today's business environment are kept in distinct computer files and are regarded as components of operating and email systems.

**Boyer- Moore Algorithm**

Boyer-Moore is defined as a fast string-search algorithm that is considered one of the most efficient string-search and pattern-search algorithm **(Dawood S., 2020)**. It is widely used as a search algorithm for websites, digital libraries, inventory management systems, information systems, etc. **(Saputri, etc., 2018; Mulyati, I. A., 2019; Lin, L.L., Soe, M.T. 2021)**. Boyer-Moore Algorithm (BM) has two stages: pre-processing and searching for a given pattern from the rightmost followed by using to skip characters in the event of a mismatch. The following are the stages of the BM algorithm:

* Pre-processing: this includes processing the shift tables acquired from the text and pattern.
* Searching: it is the searching of matching of the characters in the Pattern on the Text string

**Good Suffix Rule**

Not only does BM algorithm can match characters but also can search substrings within the Pattern.

Boyer-Moore Algorithm is an efficient algorithm in which it has a time complexity in Pre-processing a text with length n and a pattern with length m takes *O(m + sizeOfAlphabet)* time. The best-case performance time complexity is *O(n/m),* and the worst-case performance time complexity is *O(nm)*.

**Review of Related System**

In the study titled Official Document Identification and Data Extraction using Templates and OCR **(Irimia, C. et al., 2022)**, focuses on developing an application that can create textual information data from a simple image. The goal is to enable the possibility of sharing digital versions of documents, making workload quicker. The key functionalities combined are OCR algorithms and identifying the type of documents. After uploading an image, the system checks whether the image is matched on one of the templates stored in the application. In this study, Driver's license and ID card (type not specified) are used as templates.

The OCR-Enhanced Digital Asset Management System: Prototype Design and Construction **(Surmieda, J. 2022)** developed a management system with an OCR as a tool inside it. Text extracted to an image will have a digitalized file copy that will eventually transfer to the database, while files will be saved on the local file storage within the server. The study concluded that OCR capabilities were effectively included in the creation of an open-source document retrieval tool.

The Document Management System using optical character recognition, clustering, watermarking, and QR coding algorithms **(Censoro, K., 2021)**, focused on giving a security feature to a document that undergoes on OCR. The scanned document will be accepted by the Central office and converted to text and clustered into supervised keywords for searching. The scanned document will then undergo authentication by imprinting a watermark image and a QR Code using the watermarking algorithm using Text Brush Embedding and a QR Code Model 2 matrix code.

The study entitled A Necessary Review on Optical Character Recognition (OCR) System for Vehicular Applications, **(Bansal, S, et al., 2020)** uses a vehicle's license plate as an image in a database. The primary goals of automatic number plate recognition are to keep an eye on traffic, traffic flow, Smart Parking system, driver safety, and driver surveillance.

The OCR-based Document Archiving and Indexing using Py Tesseract: A Record Management System for DSWD in Caraga, Philippines **(Jayoma, J., et al. 2020)**, utilized OCR as a main tool for the system. The program was created to achieve orderly document storing and archiving without much physical space, helping records officers and departments with safekeeping and quick retrieval of records.

**Synthesis**

According to the related literature and studies that the researchers found, the Document Management System and OCR combined are a powerful tool for making our lives easier especially for the people who work inside the office that deal with a lot of paperwork. In many aspects, studies and literatures related to ours but different in some ways are significantly given us a new wide branch of ideas that we can utilize in our future research or enhancement of this Hybrid Document Management System. Within a variety of different kinds of approaches using OCR integrated in Document management Systems, it only proved how flexible we use technologies nowadays.

**METHODOLOGY**

This section describes the research design, participants of the study, data gathering procedure, sampling technique, data analysis, statistical treatment of data, and research instrument.

**Research Design**

The researcher employs Agile Methodology by utilizing small, versatile teams comprising different skills that independently create small functional components regularly. This approach enables consistent client input and the flexibility to adjust when necessary.

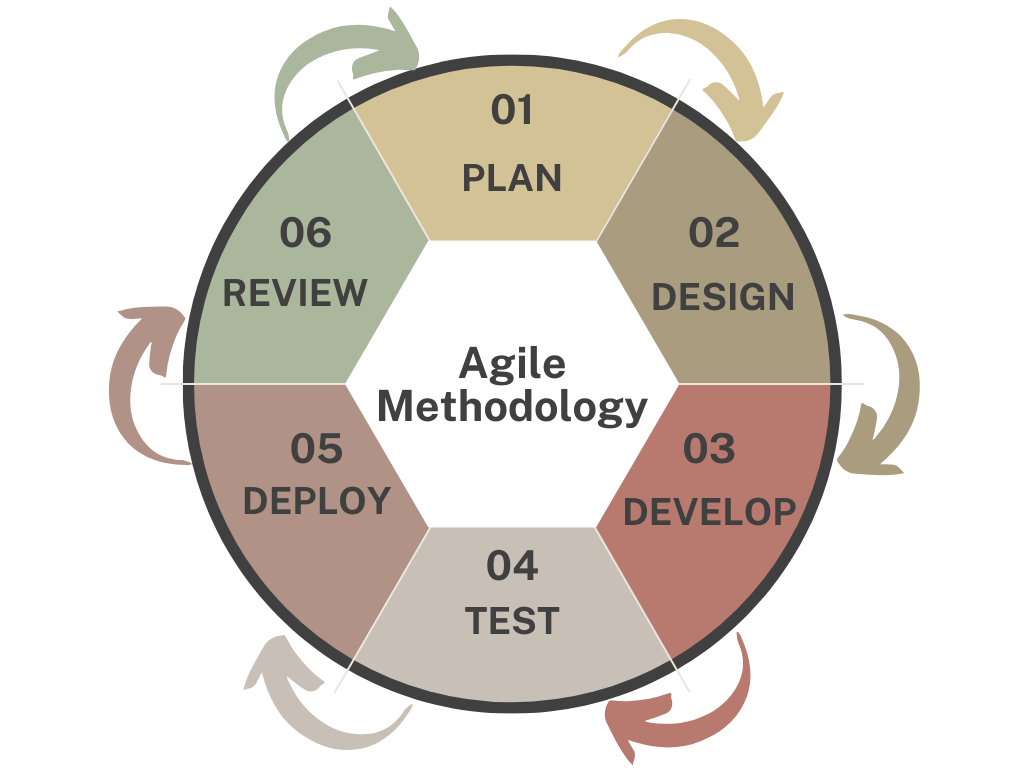


Figure 2. Agile Methodology

**Planning**

The researchers gathered information from the client to identify what features to be included to solve the problems they will encounter. The researchers need to make a management system that can help the Sangguniang Bayan Employees to digitalize resolution and ordinance documents efficiently. The researchers came up with the plan to build a Hybrid (App and Website) Management System which allows employees of SB to use their phones as a medium to convert physical documents to PDF using the app and then store it on the website that is accessible on the internet.

**Designing**

During this phase, the initial development of the Boyer-Moore Algorithm and Image to Character Recognition Hybrid Document System for Sangguniang Bayan is underway. The insights gained from this stage will pave the way for the creation of the system's database, which will store its data.

**Use Case Diagram**

This refers to a graphical depiction of a user with the system. Figure 3 shows the use case diagram presenting the features of the Hybrid Document Management System for Sangguniang Bayan with Boyer-Moore Algorithm and Image to Character Recognition.

A diagram of a network

Description automatically generated

Figure 3: Use Case Diagram

Figure 3 shows the features which can be accessed by the admin, registered users and site visitors. Admins can log in and log out, submit, edit, delete and view details documents, add document that will reflect to recently upload document, and view event logs. Registered users can login, logout, download documents, view document details also search documents. Site visitors can view details of the documents and register as users.

**Hierarchical Input Process Output**  
 The HIPO (Hierarchy Input Process Output) technique was used by the developers for planning and/or record a computer system. A HIPO model includes a hierarchical representation of the program's control structure and a collection of IPO (Input-Process-Output) charts that describe the functions (or processes) executed by each module and the inputs and outputs to those modules.

A diagram of a website

Description automatically generated

Figure 4. Hierarchical Input Process Output (HIPO) of HDMS

In this figure 4, shows the hierarchical input process of the HDMS website side. It demonstrates that the user will be redirected to either user home page or admin panel depending on the user’s assigned role. When logged in as an admin, the system has options that contain document request page and documents page. It contains the following features which are approve or decline document request, search, edit resolution/ordinance documents, and upload image of documents. When the users are logged in as user's role, the system has options that contain about page, resolution and request page. While the resolution and ordinance page have both features of search and request of documents. The about features contains all other social links of the SB.

**Entity Relational Diagram**

The Entity Relationship Diagram (ERD) of the Hybrid Document Management System for Sangguniang Bayan Using Boyer-Moore Algorithm and Image to Character Recognition of Naic, Cavite.

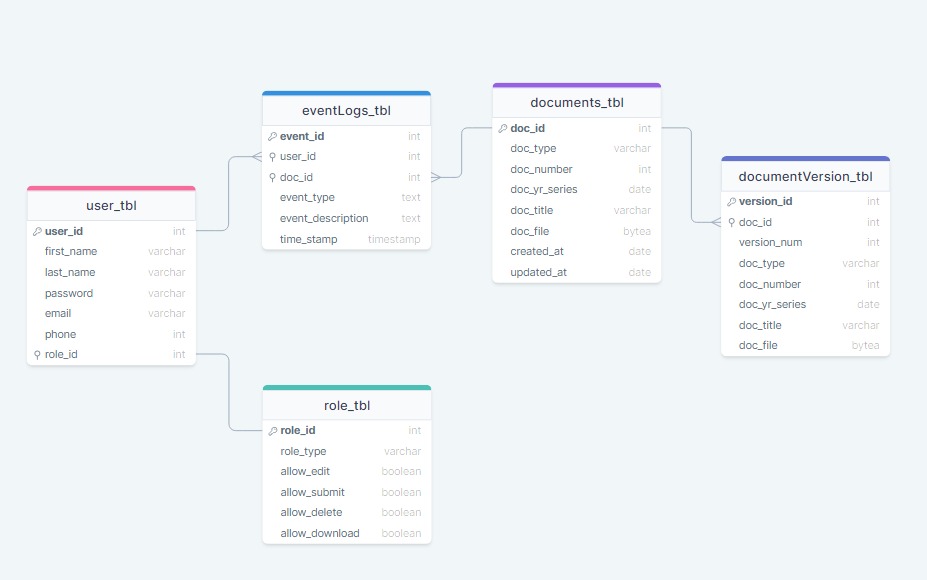
Figure 5: Entity Relationship Diagram

Figure 5 illustrates the connections between different entities. Within this diagram, there are five entities demonstrating their relationships with others within the system. In the tables of *user\_tbl* and *role\_tbl* is a one-to-one relation it indicates that the user can only have one role. For the relationship *of user\_tbl* and *eventLogs\_table* it indicates that an *eventLogs\_tbl* can be associated with many users. In the relationship of *document\_tbl* to *eventLogs\_tbl* and *documentVersion\_tbl*.

**Data Dictionary**

The data dictionary that the researchers used for the database development of Hybrid Document Management System for Sangguniang Bayan with Boyer-Moore Algorithm and Image to Character Recognition.

Table 1. Data dictionary for users\_tbl

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Name | Type | Field Size | Null | Default | Description | Example |
| user\_id | integer | 11 | no | n/a | Unique identifier or Primary key for each user in the system. | 110 |
| first\_name | varchar | 100 | no | n/a | The given or first name of the user. | Pedro |
| last\_name | varchar | 100 | no | n/a | The family or last name of the use. | Juanito |
| password | varchar | 255 | no | n/a | Hashed password for user authentication | AM9HkKk7fx |
| email | varchar | 100 | no | n/a | Email address associated with the user’s account. | user@gmail.com |
| phone | integer | 11 | yes | NULL | Phone address associated with the user’s account. | 09093334444 |
| role\_id | integer | 11 | no | n/a | Foreign key which is an identifier to role table that associates level of access for users | 110 |

Table 1 shows the *users\_tbl, users\_tbl* has seven fields*, user\_id, first\_name, last\_name, password, email, phone, role\_id. user\_id* which has 11 field size with a datatype of integer and is the Unique Identifier for the users of the system. *first\_name* has 100 field size; its datatype is varchar which represents the first name of the users in the system. *last\_name* has 100 field size; its datatype is varchar which represents the last name of the users in the system. password has 255 field sizes; its datatype is varchar which the value is hashed and will be used for authenticating users in the system. phone can be null and its datatype is integer with 11 field size, it’s an optional information for users in the system. *role\_id*, is a foreign key which is referenced at the role table that has information about the level of access granted to the users in the system.

**Developing**

The researchers employed various software programs such as Tailwind CSS, the React.js library, React Native android platform, and the NodeJS runtime environment within Visual Studio code for the software's development and design.

Regarding computer specifications, the researchers used an 3rd Gen Intel(R) Corre processor with 2 Logical Cores, specifically the i5-3320M @ 2.60GHz. Additionally, the system comprised 8GB of RAM with with 1600 MHz speed, and a 256GB SSD, and ran on Windows 10 Pro.

**Testing**

Following the coding phase, the system undergoes testing at the Sangguniang Bayan Office in the Municipality of Naic. This assessment ensures that all functions operate according to their intended purposes and achieve the anticipated outcomes.

**Deploying**

The responsibility of managing the release phase rests with the Information Technology school faculty. They will meticulously strategize and execute a comprehensive transfer plan to ensure the smooth and effective deployment of the intended release.

**Participants of the Study**

The participants of this study were the government institution named Sangguniang Bayan at the Naic, Cavite and their clients. They would be testing the entire document management system and how this will work. The researchers selected 8 IT Experts, 2 SB officers and 15 end users from the officers of the SB office of Naic, Cavite. There will be a total of 25 respondents. They will be selected according to their availability.

Table 2: Participants of the study

|  |  |
| --- | --- |
| PARTICIPANTS | NUMBER OF PARTICIPANTS |
| IT experts | 5 |
| SB Officers | 2 |
| End users | 18 |
| TOTAL | **25** |

**Data Gathering Procedure**

The researchers gathered data by conducting surveys and interviews involving SB employees and clients from the Municipality of Naic. After collecting this data, the researchers analyzed the information obtained from the interviews. Subsequently, the system was developed and constructed following the in-depth analysis conducted by the researchers. Finally, participants were invited to explore the developed system and provide feedback through assessments to evaluate its performance.

**Sampling Technique**

The researchers employed the Purposive Sampling Technique, categorized under the non-probability sampling method, specifically known as judgment sampling. This method allowed the researchers to selectively choose participants for the study, ensuring they obtained the most relevant and informative data required for their research.

**Data Analysis**

The answers of participants regarding the survey are in the form of tables. These tables show the result of the data gathered through the use of relative frequency formulation. By applying this concept, the researchers can easily identify and determine if the system functions well.

Table 3: Rating Scale

|  |  |
| --- | --- |
| RATING | INTERPRETATION |
| 4.51-5.00 | Excellent |
| 3.51-4.50 | Very Satisfactory |
| 2.51-.3.50 | Satisfactory |
| 1.51-2.50 | Fair |
| 1.00-1.50 | Poor |

**Statistical Treatment of data**

The following procedure will be used in computing the result of the evaluation. The weighted mean is used to represent the average value of the results for each criterion.

The Formula used to obtain the weighted mean is

𝑥̄ = ∑x/N

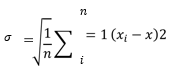
**Where:**

𝑥̄ ̄= mean;

 = sum of all scores; and

N = total number of participants

The standard deviation is a measure of how spread-out numbers are. The formula used to obtain the sample standard deviation is



**Where:**

𝜎 = standard deviation

xi = each value of the dataset

x = the arithmetic means of data

N = the total number of data points

∑ (xi - x2 = the sum of (xi - x2 for all data points)

**Research Instrument**  
 The researcher will utilize the evaluation form as the primary data acquisition instrument in this study. The assessment instrument designated for the ITD Software, adhering to the ISO 9126 standard, is the form that will be employed.

**RESULTS AND DISCUSSION**

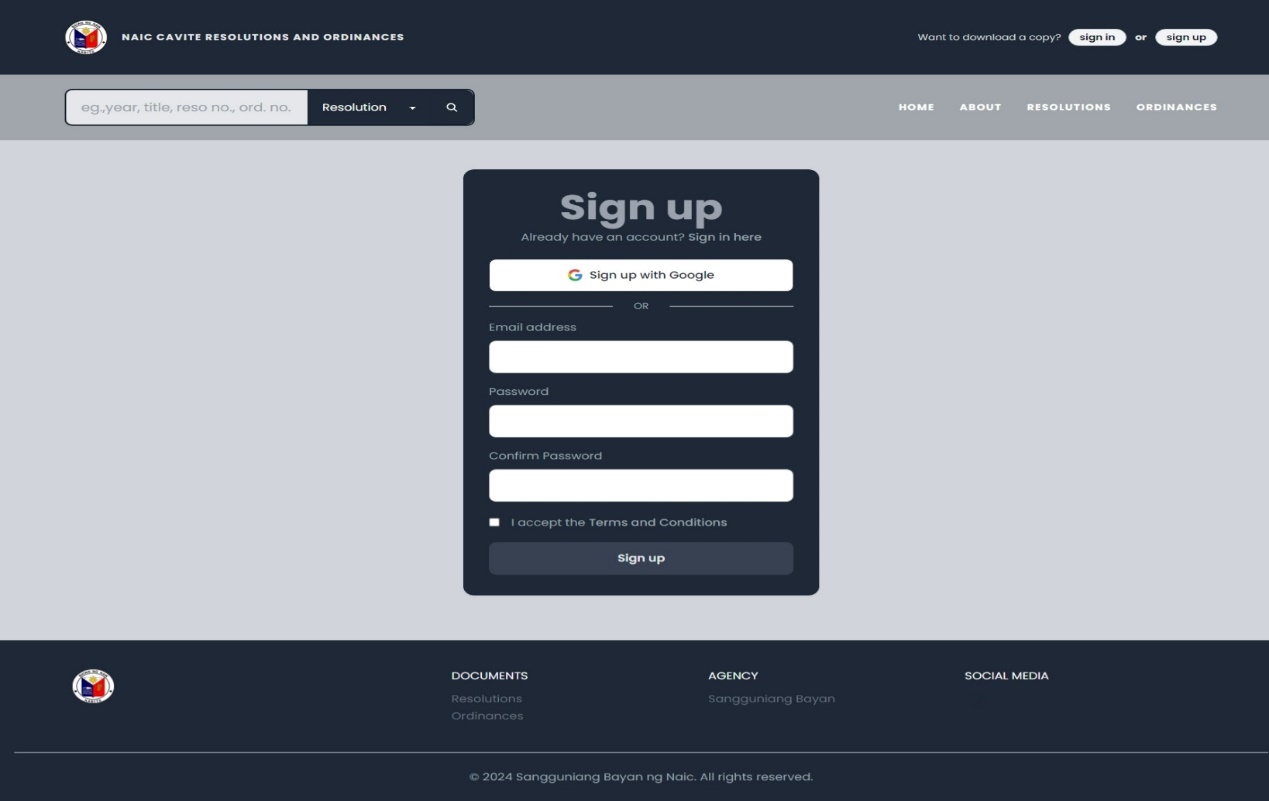
In this chapter, the researchers present the results of the tests and the evaluations conducted on the system development.

**Project Description**

The Hybrid Document Management System was developed for Sangguniang Bayan Office Municipality of Naic, Cavite. The system has a hybrid document management system that is capable of scanning, uploading, searching, and making the work in the office efficiently. It has a two system the one is for website and the one is for smartphone. It has two types of users: the admin or the employees of Sangguniang Bayan and the one is the users/public users. Admin is the employees who is working in resolution and amendments section who has an authority/access to the system. The system is developed using Supabase as database storage, for authentication is Powered by Google login, OCR (Optical Character Recognition). The mobile app is developed using react native document scanner plugin with the testing platform android version 12 snow cone. The system features login, scanning document using phone camera, uploading, searching, deleting, downloading.

**Project Structure**

Figure 6 to 25 illustrates the project structure of the Hybrid Document Management System for the Sangguniang Bayan, Municipality of Naic, Cavite.

****Figure 6: Sign Up Page

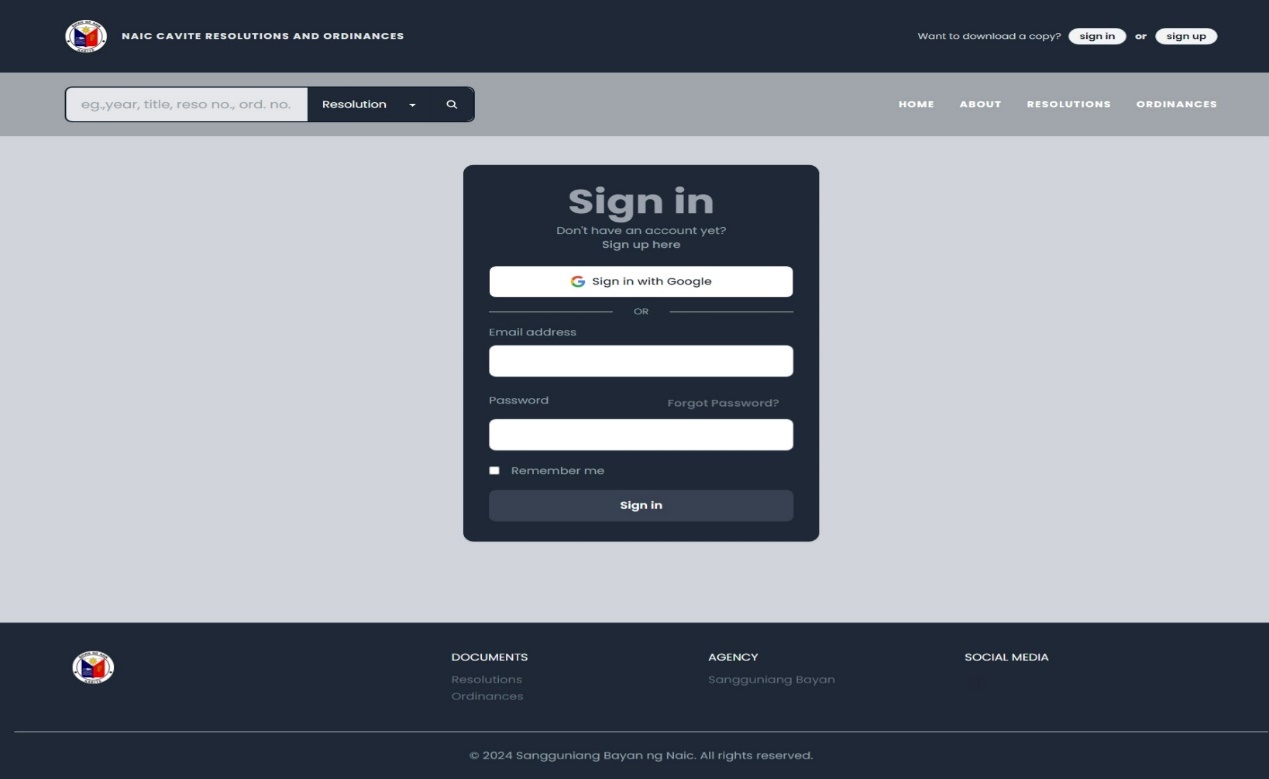


Figure 7: Sign In Page

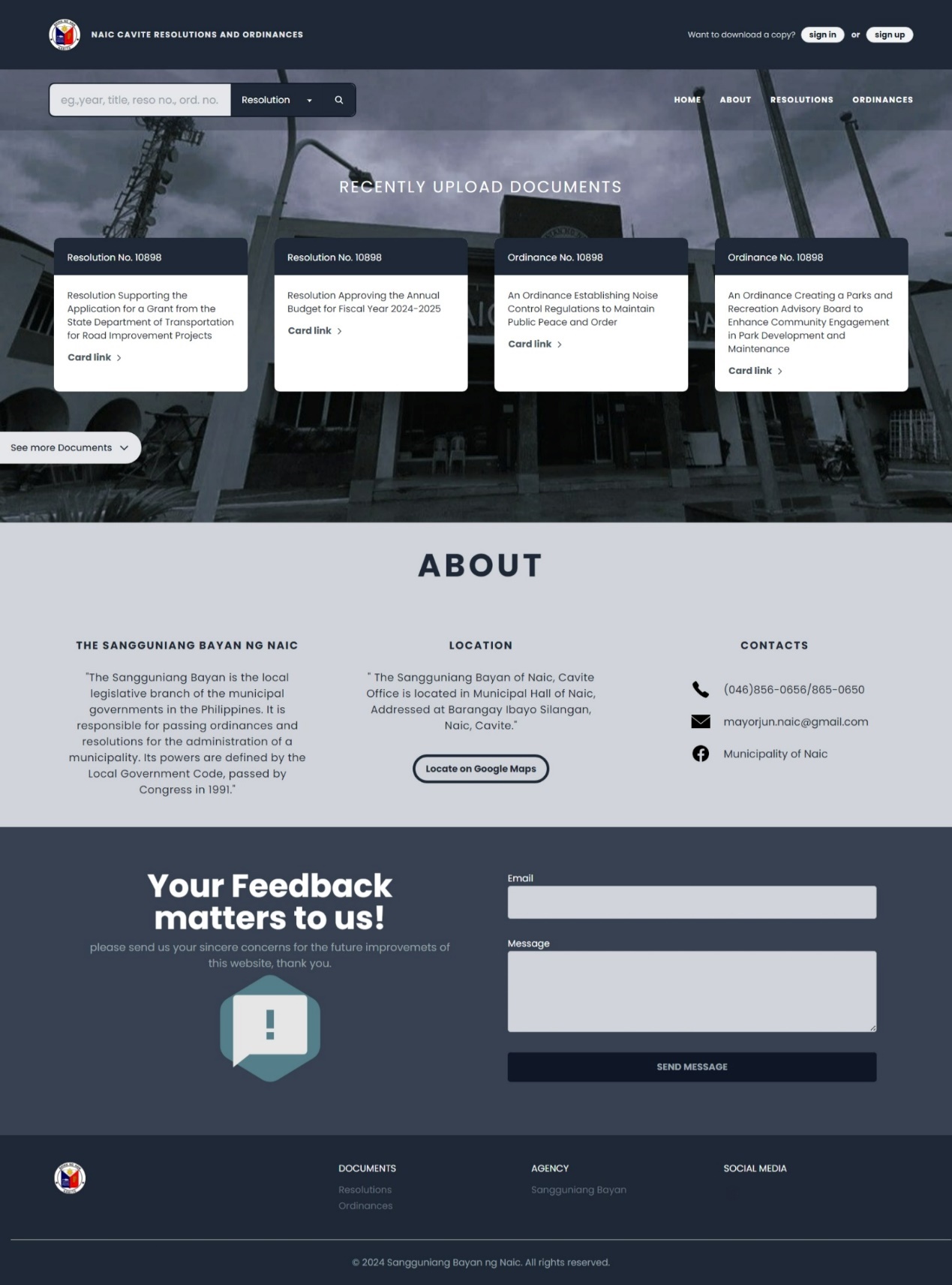


Figure 8: Home Page

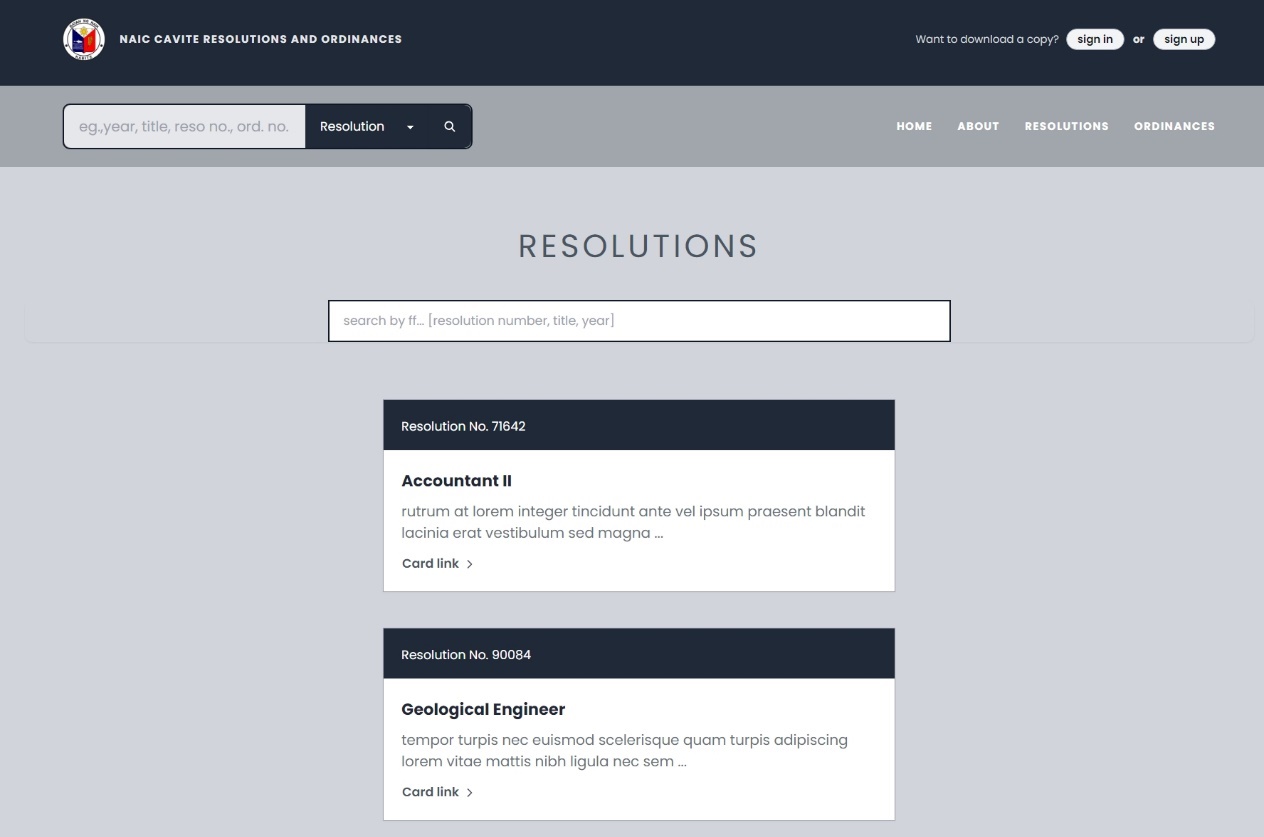
****

Figure 9. Resolution Page

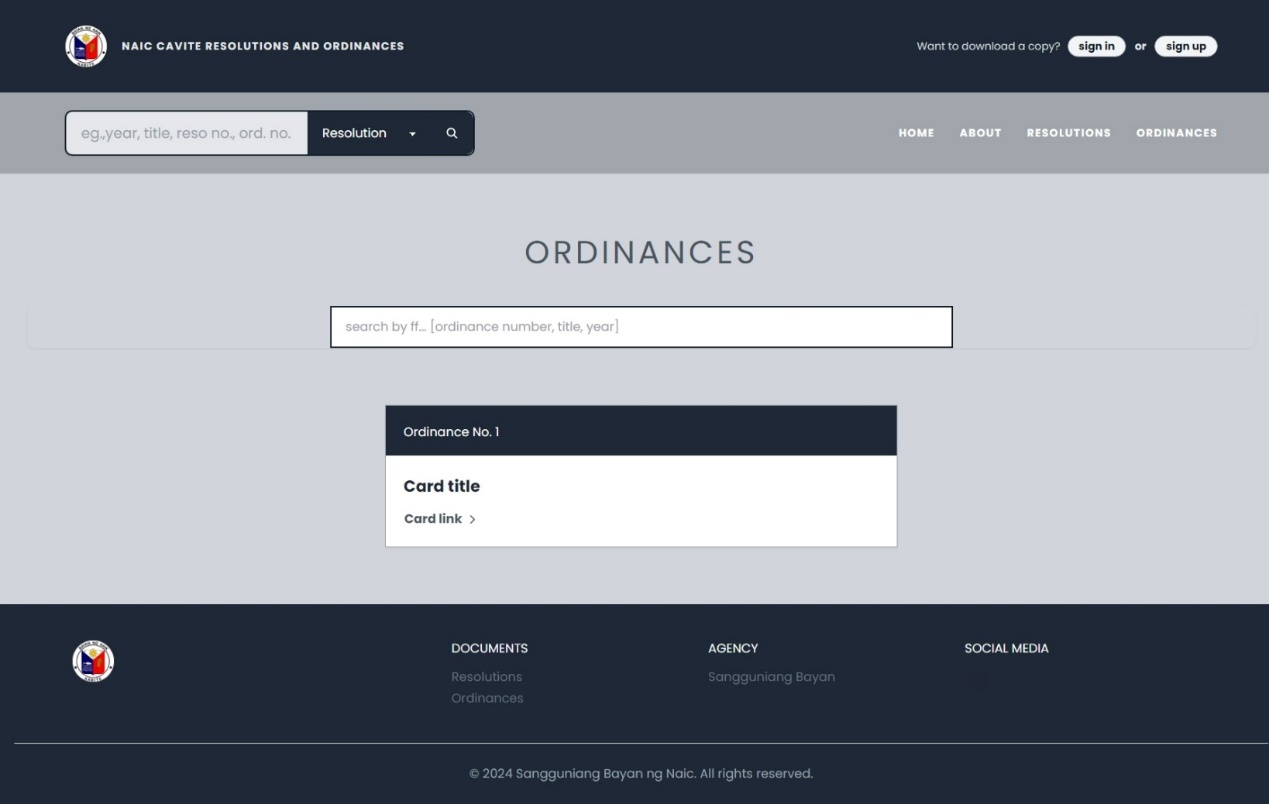
****

Figure 10: Ordinances Page

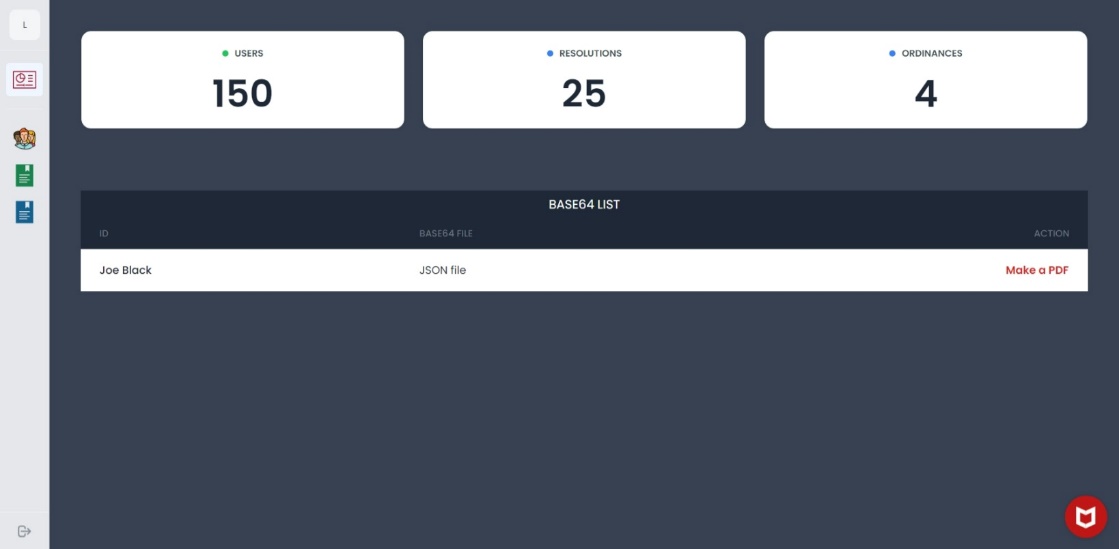
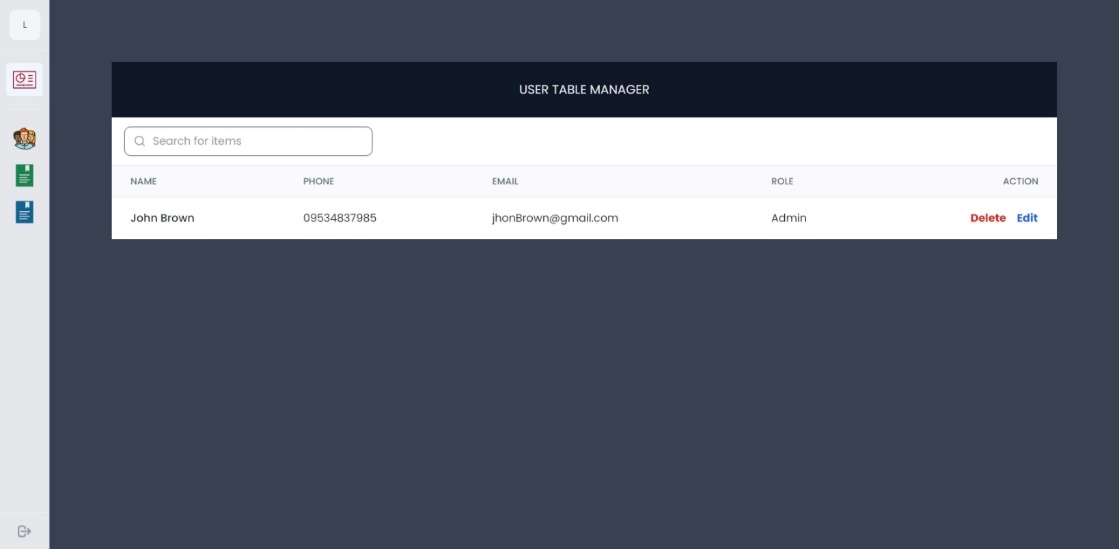
Figure 11: Dashboard

Figure 12: Manage User Page

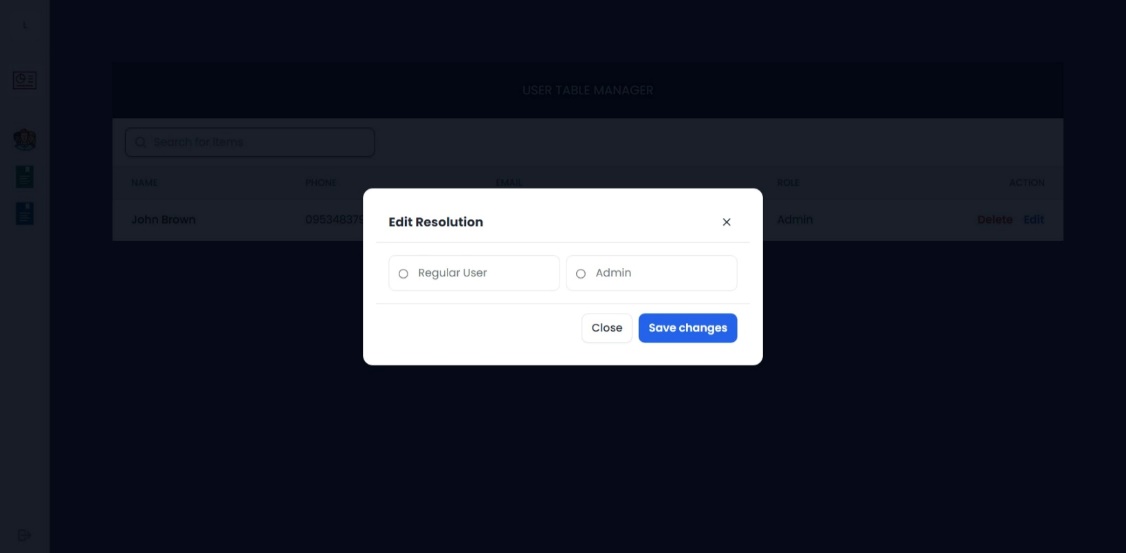
****

Figure 13: Manage User Page Modal

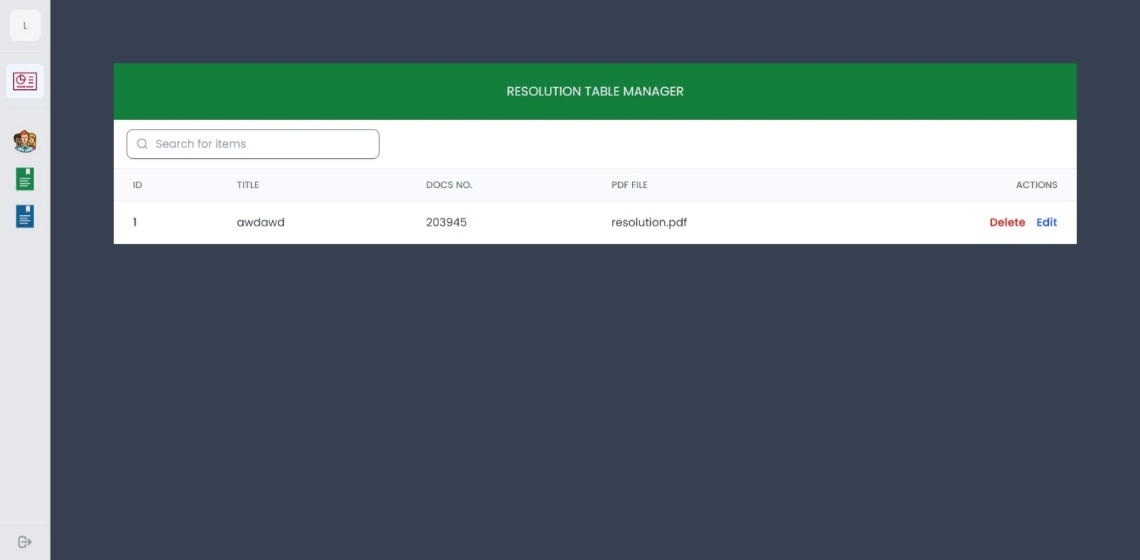
****

Figure 14: Manage Resolution Page

****

Figure 15: Manage Resolution Page Modal

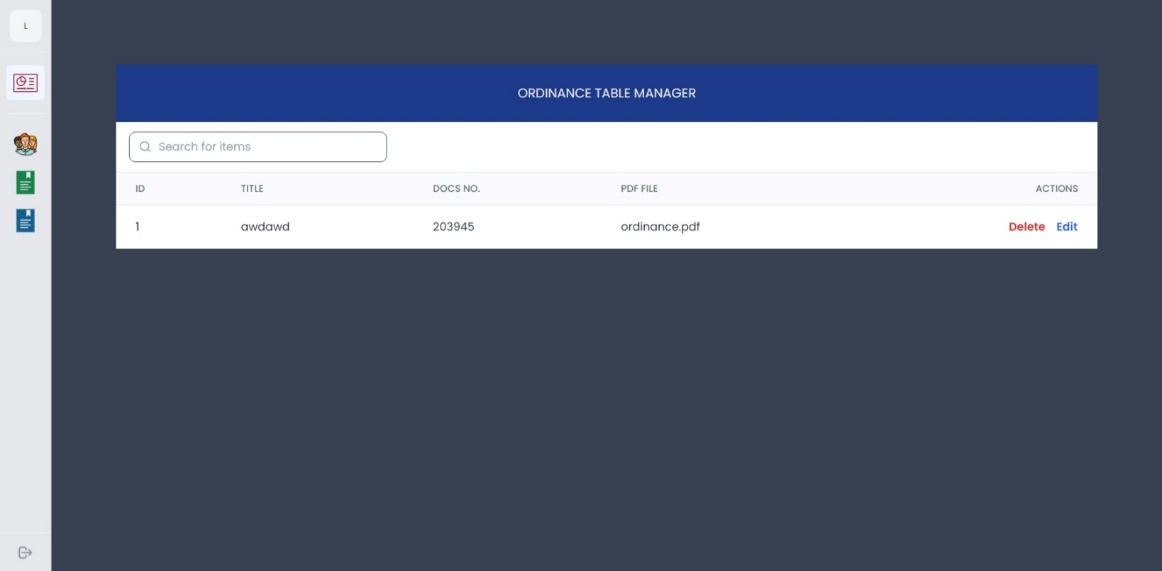
****

Figure 16: Manage Ordinance Page

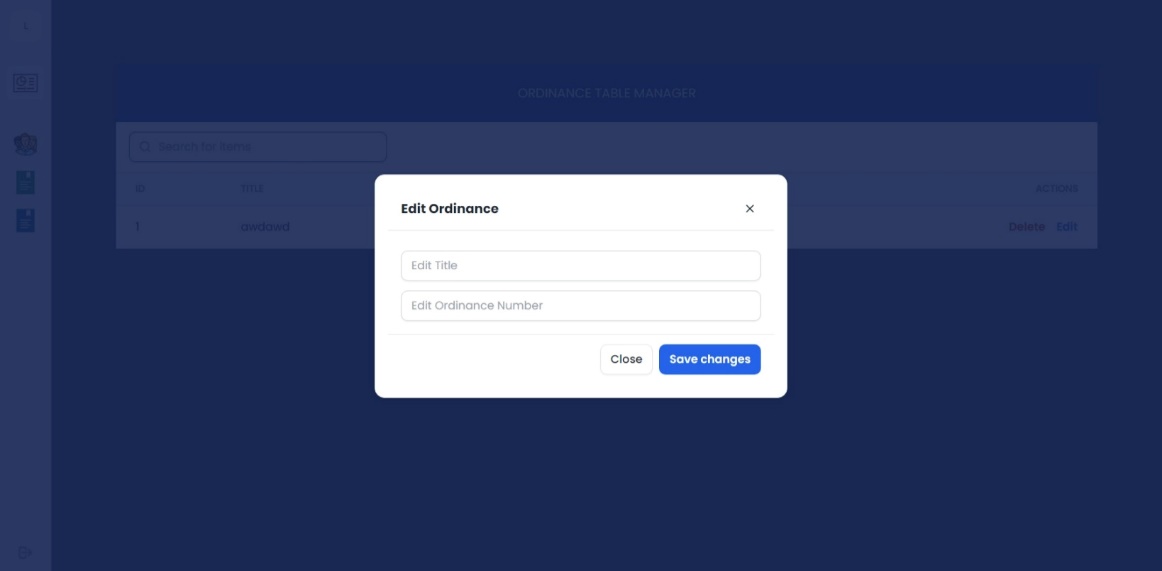
****

Figure 17: Manage Ordinance Page Modal

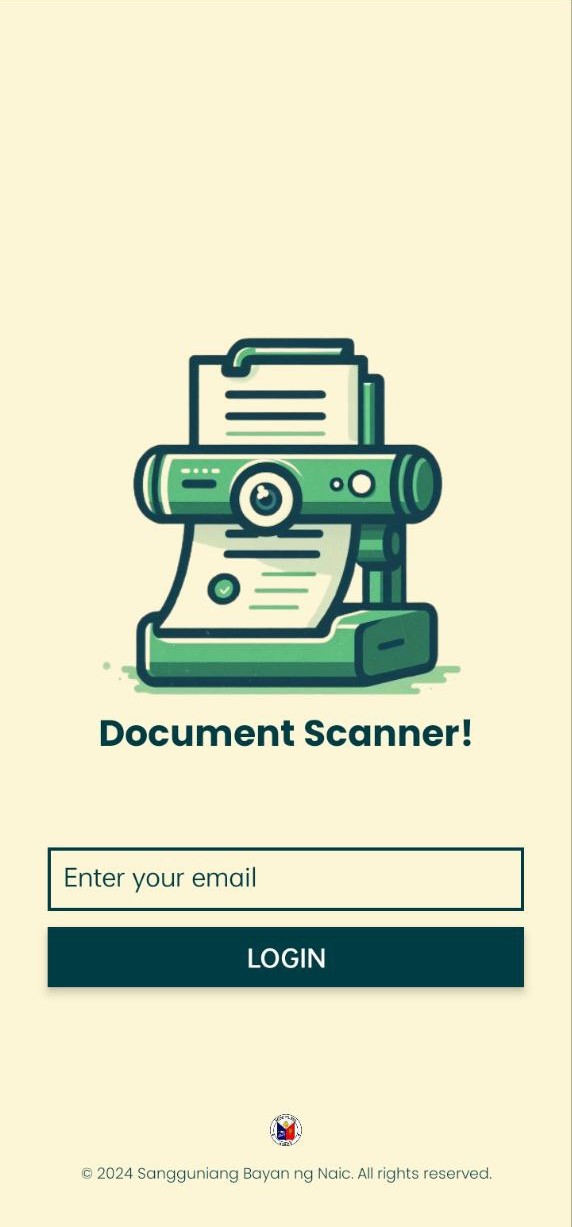
****

Figure 18: Sign In Page in Mobile Version

****

Figure 19: Document Scanner Button

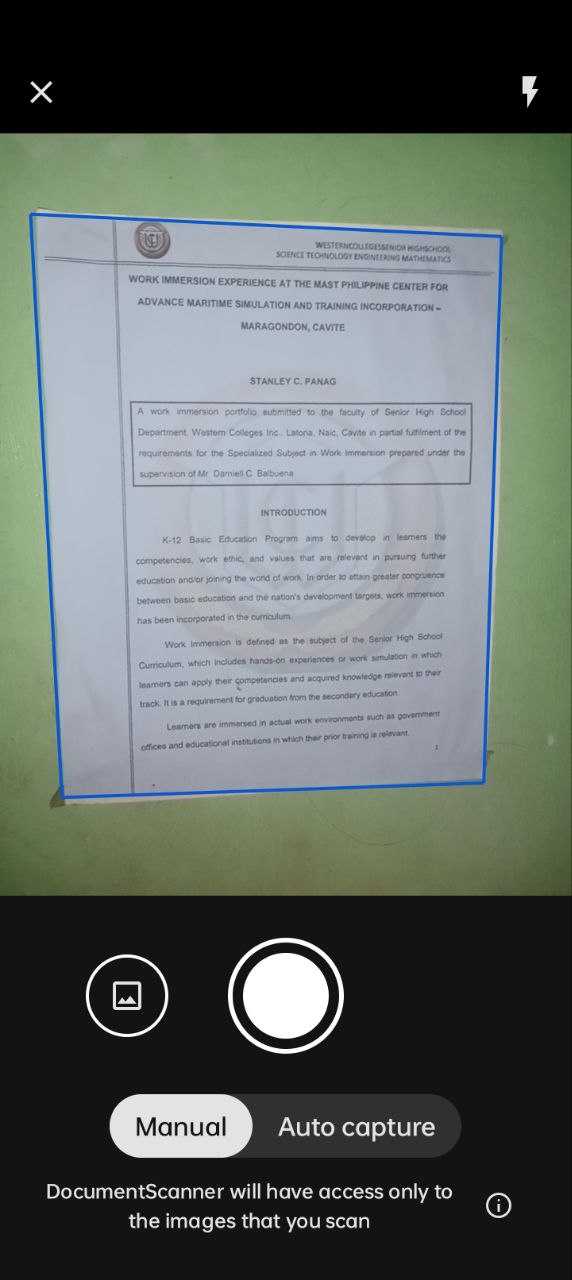
****

Figure 20: Scanning of Document

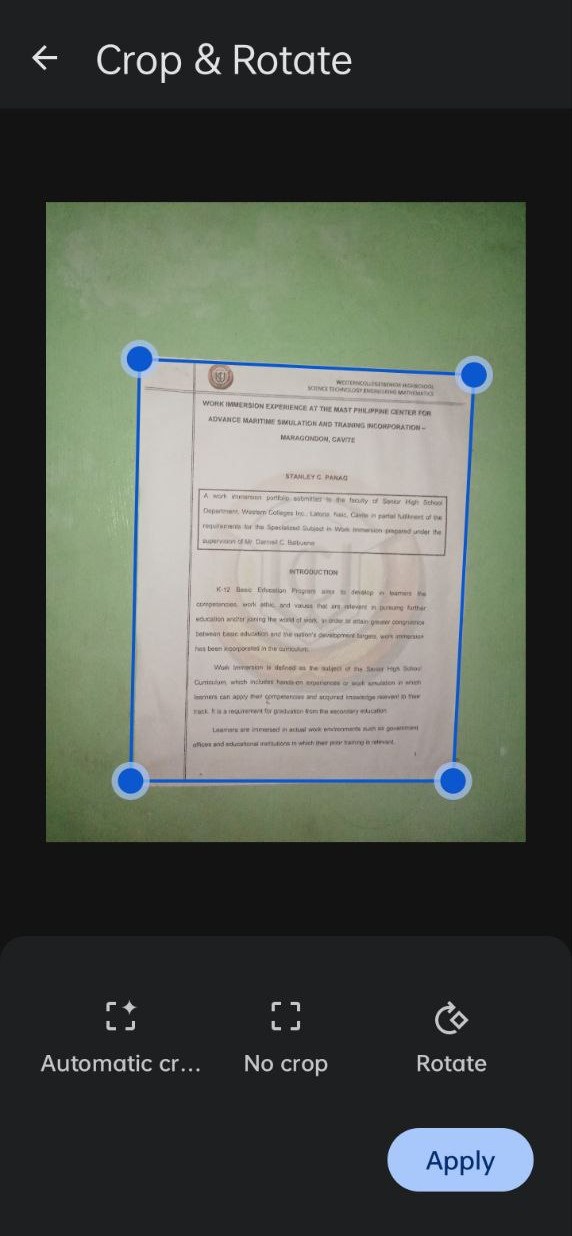
****

Figure 21: Crop and Rotate of Scanned Document

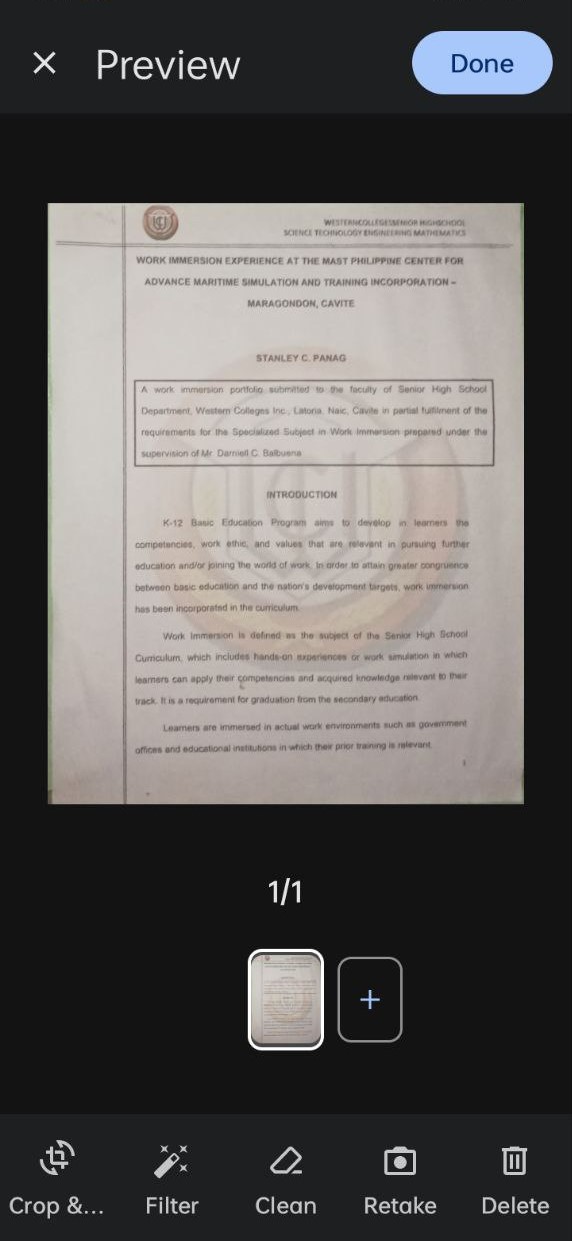
****

Figure 22: After Scanning of Document

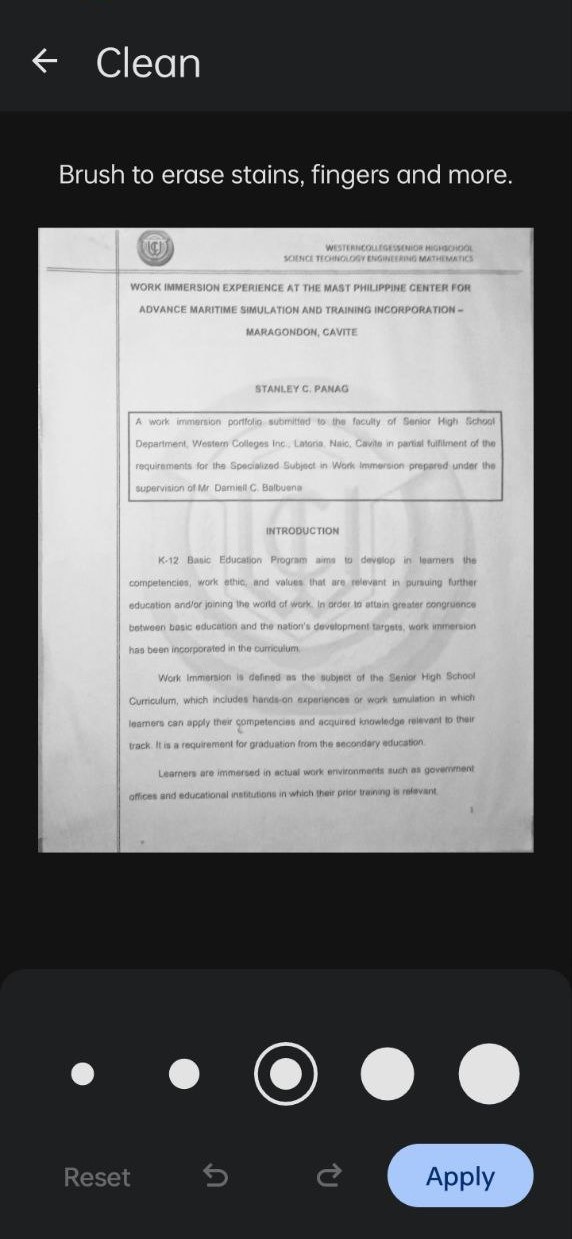
****

Figure 23: Dirt Remover Option

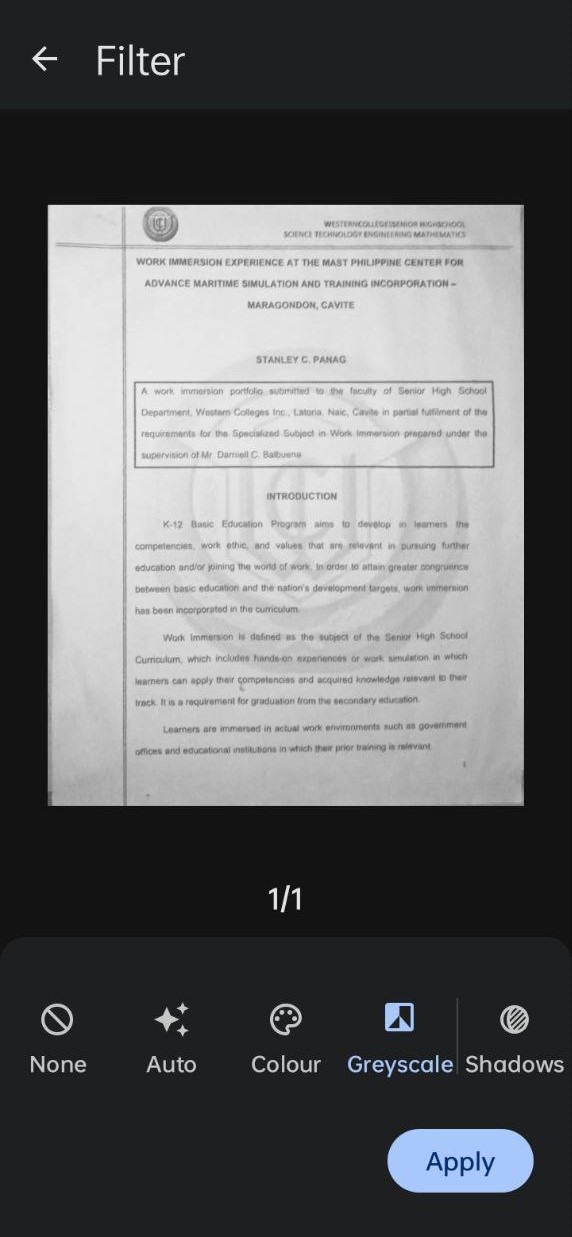
****

Figure 24: Color Change Option

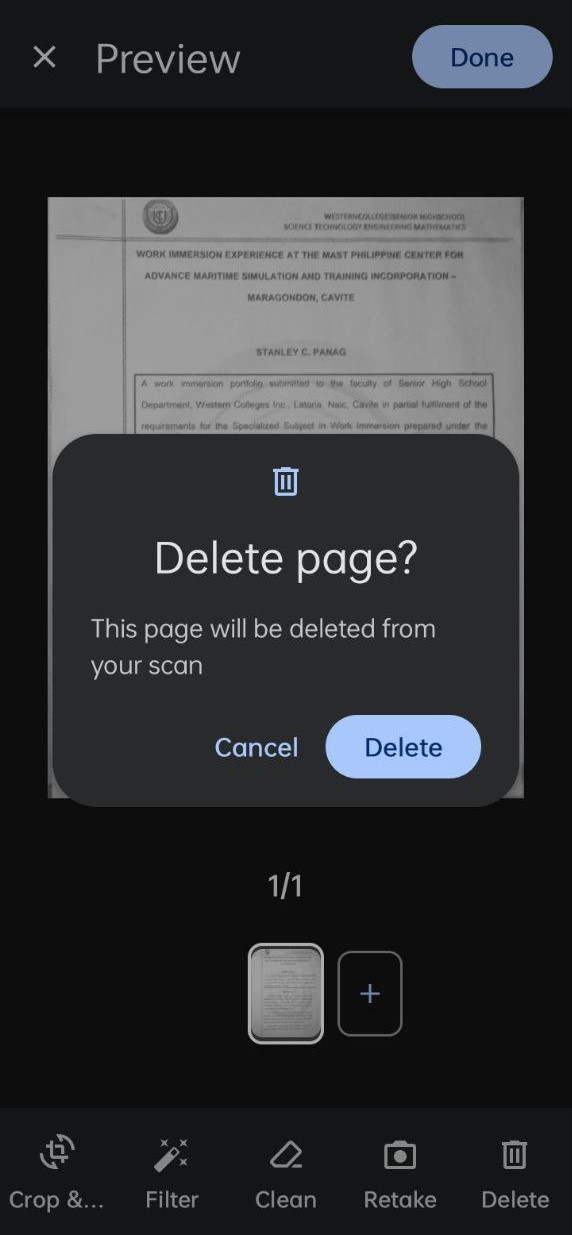
****

Figure 25: Delete Scanned Photo

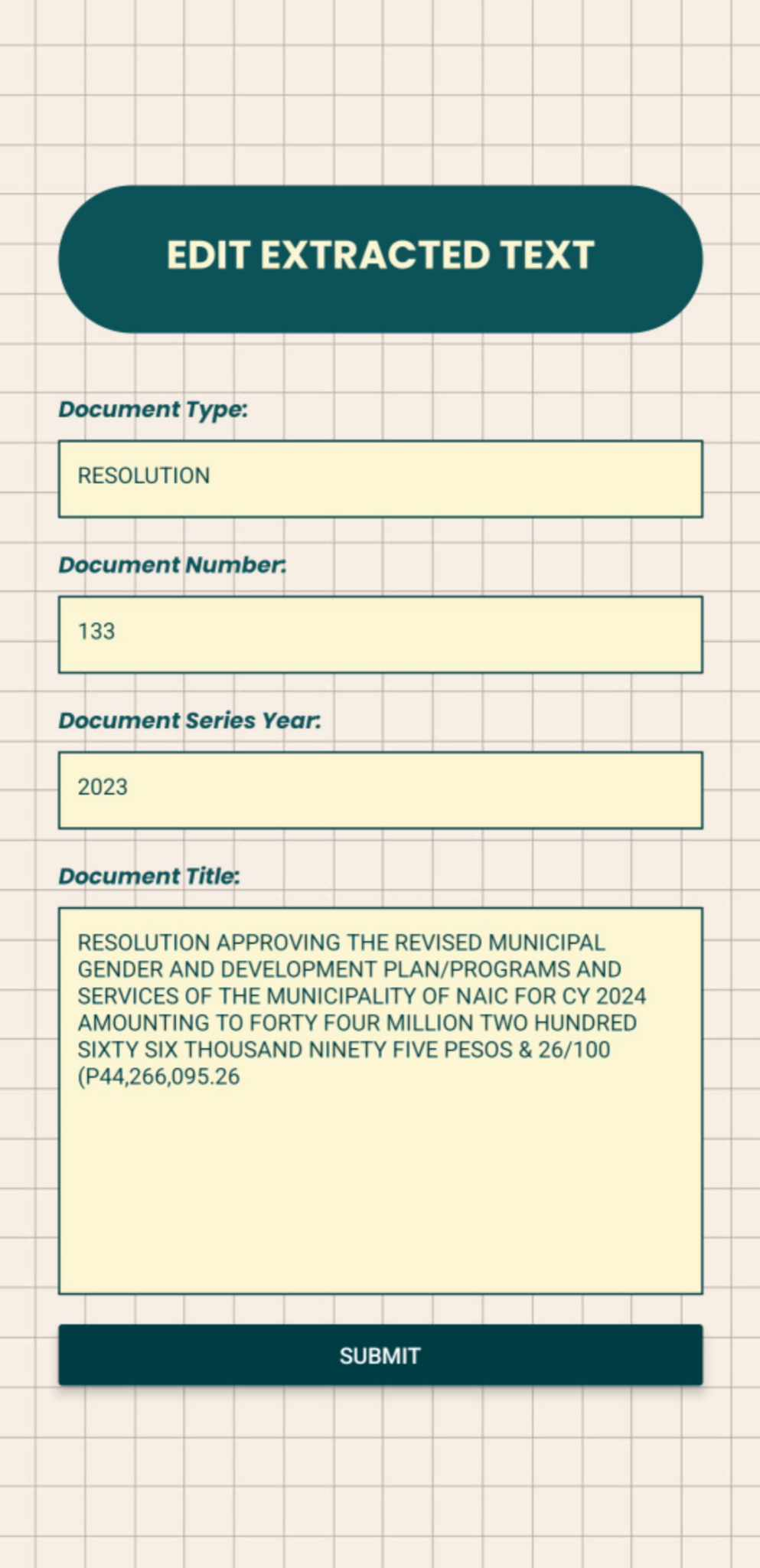
****

Figure 26: Checking Page / Submit button

**Implementation Plan**

Table 4: Implementation Plan

|  |  |  |  |
| --- | --- | --- | --- |
| **STRATEGY** | **ACTIVITIES** | **PESONS**  **INVOLVE** | **DURATION** |
| Approval from the Administrator | Letters for the  Administrator | Researchers and Client (SB Officers) | 1 day |
| System’s  Installation | Installation of the system and required software and  Hardware. | Researchers and Client (SB Officers) | 1 hour |
| Information  Distribution | Video Presentation | Researchers | 2 days |
| System Real world Exposure | Deployment | Client (SB Officers) | 1 day |

**Evaluation Results**

The system underwent several testing/evaluation procedures before it was distributed to the users. These evaluations are conducted to ensure the accuracy of the system. See the table below.

Table 5. Summary of Evaluation Result from Participants

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Software Quality Indicator** | | **Mean** | **Standard Deviation** | **Interpretation** |
|
| **FUNCTIONALITY** | | | | |
| 1 | The software provides appropriate functions | 4.12 | 0.60 | Very Satisfactory |
| 2 | The software shows accurate information and results. | 4.08 | 0.81 | Very Satisfactory |
| 3 | The software provides security measures. | 4.12 | 0.73 | Very Satisfactory |
| **Grand Mean** | | **4.11** | **0.71** | **Very Satisfactory** |
| **RELIABILITY** | | | | |
| 1 | The software shows no failures or bugs | 3.76 | 0.78 | Very Satisfactory |
| 2 | The integrity of data is maintained all throughout its operation. | 4.52 | 0.59 | Excellent |
| **Grand Mean** | | **4.14** | **0.69** | **Very Satisfactory** |
| **EFFICIENCY** | | | | |
| 1 | The software operates quickly and efficiently. | 4.12 | 0.60 | Very Satisfactory |
| 2 | The software requires less resources (memory/CPU/disk storage/internet bandwidth) | 4.28 | 0.68 | Very Satisfactory |
| **Grand Mean** | | **4.2** | **0.64** | **Very Satisfactory** |
| **PORTABILITY** | | | | |
| 1 | The software adapts to new specifications or operating environments. | 4.29 | 0.62 | Very Satisfactory |
| 2 | The software requires less effort to install. | 4.36 | 0.76 | Very Satisfactory |
| **Grand Mean** | | **4.33** | **0.69** | **Very Satisfactory** |

Continuation of Table 5. Summary

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **USABILITY** | | | | |
| 1 | The software functions are organized, clear, logical, and effective making it easy for the user to understand. | 4.24 | 0.60 | Very Satisfactory |
| 2 | The user can operate the software with ease. | 4.24 | 0.72 | Very Satisfactory |
| 3 | The user’s guide is available and clear. | 3.96 | 0.98 | Very Satisfactory |
| **Grand Mean** | | **4.15** | **0.77** | **Very Satisfactory** |
| **INFORMATION IN THE SOFTWARE** | | | | |
| 1 | The information is clear, concise, and informative to the intended user. | 4.44 | 0.51 | Very Satisfactory |
| 2 | The content is free from spelling and grammatical errors. | 3.92 | 0.70 | Very Satisfactory |
| 3 | The contents are covered in a comprehensive manner. | 4.32 | 0.69 | Very Satisfactory |
| **Grand Mean** | | **4.23** | **0.63** | **Very Satisfactory** |
| **TECHNICAL ASPECTS OF THE SOFTWARE AND MATERIALS** | |  | | |
| 1 | The software uses standard equipment that is reliable, widely available, and applicable to a variety of uses. | 4.32 | 0.48 | Very Satisfactory |
| 2 | Computer capabilities such as graphics, color are used for appropriate instructional reasons. | 4.28 | 0.61 | Very Satisfactory |
| **Grand Mean** | | **4.3** | **0.55** | **Very Satisfactory** |
|  | **Overall Mean** | **4.21** | **0.67** | **Very Satisfactory** |

Based from the result of evaluation obtained from the five (5) IT experts and eighteen (18) End User and two(2) SB Employees, the system offers all necessary and appropriate functions needed and it utilizes standard equipment that is reliable, widely available, and suitable for various applications, demonstrating its flexibility

The system has 4.11 in functionality, 4.14 in reliability, 4.02 in efficiency, 4.33 in probability, 4.15 in usability with an interpretation of very satisfaction. While, 4.23 in information in the software and 4.30 in technical aspects of the software and materials with an interpretation of excellent based on the rating scale.

**SUMMARY, CONCLUSION AND RECOMMENDATION**

**Summary**

The researchers developed the Hybrid Document Management System for Sangguniang Bayan Using Boyer-Moore Algorithm and Image to Character Recognition.

The creation of the Hybrid Document Management System for Sangguniang Bayan focused on optimizing the institution's processes for data collection, data searching and document management. By using the Boyer-Moore Algorithm and Image to Character Recognition technology, the system innovate features such as an online database, easily accessible documents, and improved data security.

The project successfully integrated the Boyer-Moore Algorithm for efficient text searching within documents, significantly reducing search times compared to traditional methods. Image to Character Recognition technology was employed to digitize physical documents, enabling quick and accurate conversion of printed text into editable and searchable digital formats. These technological advancements contributed to a more organized and accessible document management system.

The system was developed using an agile methodology. Key technologies used were Boyer-Moore for algorithms, JavaScript and React Native OCR for character recognition, Google Cloud for database management, and Express for the web application. Together, these tools provided a strong and scalable solution.

User feedback helped improve the system's features and interface. During testing, Sangguniang Bayan users evaluated its functionality, efficiency, reliability, and usability. The system received a high user satisfaction rating of 4.21 showing it effectively met user needs and improved document management.

**Conclusion**

The development of the Hybrid Document Management System for Sangguniang Bayan, integrating the Boyer-Moore Algorithm and Image to Character Recognition, represents a major advancement in modernizing document management practices for local government units. With careful planning and execution, the system effectively tackles the challenges of manual data processing and document retrieval within the Sangguniang Bayan.

By using advanced algorithms and OCR technology, the system simplifies data gathering processes and improves document management accuracy. Its new features, such as an online database and easy document access, provide a more efficient and secure way to handle municipal records.

The successful implementation of the system suggests it could be a model for other Sangguniang Bayans and similar institutions looking to upgrade their document management practices. Positive user feedback confirms its effectiveness in meeting the needs of local government units and supporting better decision-making processes.

In conclusion, the Hybrid Document Management System signifies a significant step in improving the efficiency, accuracy, and accessibility of document management for Sangguniang Bayans. Its deployment shows the value of technology in addressing the changing needs of local governance, resolving the problems initially identified.

**Recommendations**

For further development and enhancement of the "Hybrid Document Management System for Sangguniang Bayan, integrating the Boyer-Moore Algorithm and Image to Character Recognition," the following recommendations are proposed:

1. Algorithm Improvement: Continuously refine the Boyer-Moore Algorithm and Image to Character Recognition technology. This includes optimizing performance, improving accuracy, and ensuring compatibility with various document types for efficient processing.
2. Data Security Enhancement: Strengthen data security with measures like encryption, access controls, and regular audits. Work with cybersecurity experts to identify and address vulnerabilities, safeguarding municipal records' confidentiality and integrity.
3. Collaborative Tools Integration: Integrate the Document Management System with tools used by Sangguniang Bayan like for better collaboration, document sharing, and version control. This enhances efficiency and transparency
4. Scalability and Interoperability: Ensure the system scales effectively and integrates smoothly with other municipal systems or databases.

By following these suggestions, those involved can make the Hybrid Document Management System even better, improving how easy it is to use, how well it works, and how safe it keeps important documents. This ensures that it keeps helping Sangguniang Bayans manage documents well, leading to better governance and services for everyone.

**References**

Abbasova, V. (2020). *Main Concepts of the Document Management System Required for Its Implementation in Enterprises.* [(PDF) MAIN CONCEPTS OF THE DOCUMENT MANAGEMENT SYSTEM REQUIRED FOR ITS IMPLEMENTATION IN ENTERPRISES (researchgate.net)](https://www.researchgate.net/publication/339674095_MAIN_CONCEPTS_OF_THE_DOCUMENT_MANAGEMENT_SYSTEM_REQUIRED_FOR_ITS_IMPLEMENTATION_IN_ENTERPRISES)

Ahmed M. et al. (2019). *Review On Optical Character Recognition.* [(PDF) REVIEW ON OPTICAL CHARACTER RECOGNITION (researchgate.net)](https://www.researchgate.net/publication/334162853_REVIEW_ON_OPTICAL_CHARACTER_RECOGNITION)

Ahmad, R. et al. (2021). *Efficient skew detection and correction in scanned document images through clustering of probabilistic hough transforms.* https://www.sciencedirect.com/science/article/abs/pii/S0167865521003408

Ali, A. A. A.., & Dauwed, M. (2022). Development of A Hybrid Mobile App for Student Management System. Journal of Advanced Sciences and Nanotechnology, 1(1), 37–42. <https://doi.org/10.55945/joasnt.2022.1.1.37-42>

Alzeyani, E. , Szabo, C. (2023). *A Study on the Effectiveness of Agile Methodology Using a Dataset.*https://www.researchgate.net/publication/376366629\_A\_Study\_on\_the\_Effectiveness\_of\_Agile\_Methodology\_Using\_a\_Dataset

Bansal, S, et al. (2020). A Necessary Review on Optical Character Recognition (OCR)  
 System for Vehicular Application.  
 <http://dx.doi.org/10.1109/ICIRCA48905.2020.9183330>

Censoro, K. (2021). *Document management system using optical character recognition,*  *clustering, watermarking and QR coding algorithms.* [Document management system using optical character recognition, clustering, watermarking and QR coding algorithms (cpu.edu.ph)](https://repository.cpu.edu.ph/handle/20.500.12852/1519)

Dawood S., Sarhan. (2020). Empirical Performance Evaluation of Knuth Morris Pratt and Boyer Moore String Matching Algorithms. The Journal of Duhok University. 23. 134-143. 10.26682/sjuod.2020.23.1.14.

Everen, D. (2023). *Tesseract OCR vs. CNN-based OCR: Which is Right for You?*  [Tesseract OCR vs. CNN-based OCR: Which is Right for You? » Veryfi](https://www.veryfi.com/technology/tesseract-ocr-vs-cnn-based-ocr/#:~:text=Deep%20Learning%20Approach-,What%20is%20OCR%20using%20CNN%3F,range%20of%20text%20recognition%20scenarios.)

Google Cloud. (n.d.). What is cloud computing? Retrieved from <https://cloud.google.com/learn/what-is-cloud-computing>.

Irimia, C. et al. (2022). *Official Document Identification and Data Extraction using Templates and OCR.* [Official Document Identification and Data Extraction using Templates and OCR - ScienceDirect](https://www.sciencedirect.com/science/article/pii/S1877050922010973)

Jatowt, A. et al. (2021)*. Survey of Post-OCR Processing Approaches*. [Survey of Post- OCR Processing Approaches (acm.org)](https://dl.acm.org/doi/fullHtml/10.1145/3453476#:~:text=Post%2Dprocessing%20approaches%20consist%20of,for%20reprocessing%20them%20if%20necessary.)

Jayoma, J., et al. (2020). *OCR based Document Archiving and Indexing using Py Tesseract: A Record Management System for DSWD Caraga, Philippines* [OCR-Based-Document-Archiving-and-Indexing-Using-PyTesseract-A-Record-Management System-for-DSWD-Caraga-Philippines.pdf (researchgate.net)](https://www.researchgate.net/profile/Elbert-Moyon/publication/350998841_OCR_Based_Document_Archiving_and_Indexing_Using_PyTesseract_A_Record_Management_System_for_DSWD_Caraga_Philippines/links/60ca50eb299bf1cd71d5217b/OCR-Based-Document-Archiving-and-Indexing-Using-PyTesseract-A-Record-Management-System-for-DSWD-Caraga-Philippines.pdf)

KamalaKannan, T0 C., & Devi, M. R. (2019). Study on cloud storage and its issues in cloud computing. International Journal of Management, Technology and Engineering, 9(1), 976-981. Retrieved from <https://www.researchgate.net/profile/KsharmilaSenthil/publication/333673809_Study_on_Cloud_Storage_and_its_Issues_in_Cloud_Computing/links/5cfe30744585157d15a00f28/Study-on-Cloud-Storage-and-its-Issues-in-Cloud-Computing.pdf>

Kartiwi, M., Gunawan, T. S., Anwar, A., & Fathurohmah, S. S. (2018, November). Mobile Application for Halal Food Ingredients Identification using Optical Character Recognition. In 2018 IEEE 5th International Conference on Smart Instrumentation, Measurement and Application (ICSIMA) (pp. 1-4). IEEE.

Kishan, R. et al. (2019). *Intelligent Image Segmentation for OCR using contour.*  [A2105030106.pdf (iosrjournals.org)](http://iosrjournals.org/iosr-jce/papers/Vol21-issue5/Series-3/A2105030106.pdf)

Kuitunen, M. (2019). Cross-Platform Mobile Application Development with React Native (Bachelor's thesis). <https://urn.fi/URN:NBN:fi:tty-201902111234>

Lin, L.L., Soe, M.T. (2021). Searching Process Using Boyer Moore Algorithm in Digital Library. In: Vasant, P., Zelinka, I., Weber, GW. (eds) Intelligent Computing and Optimization. ICO 2020. Advances in Intelligent Systems and Computing, vol 1324. Springer, Cham. <https://doi.org/10.1007/978-3-030-68154-8_50>

Mulyati, I. A. (2019). Searching Process using Bayer Moore Algorithm in Medicinal Information Media. International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8 Issue-2S7. [https://www.ijrte.org/wp-](https://www.ijrte.org/wp-content/uploads/papers/v8i2S7/B10070782S719.pdf) content/uploads/papers/v8i2S7/B10070782S719.pdf

Pinto, C. M., & Coutinho, C. (2018). From Native to Cross-platform Hybrid Development. 2018 International Conference on Intelligent Systems (IS). doi:10.1109/is.2018.8710545

Reddy, S. (2019). *Pre-Processing In OCR.* [Pre-Processing in OCR!!!. A basic explanation of the most widely… | by Susmith Reddy | Towards Data Science](https://towardsdatascience.com/pre-processing-in-ocr-fc231c6035a7)

Saputri, Wahyu & Dhenabayu, Riska & Febrinita, Filda. (2018). APPLICATION OF STRING MATCHING USING BOYER MOORE ALGORITHM IN THE APPLICATION OF DICTIONARY LIVESTOCK TERMS. JOSAR (Journal of Students Academic Research). 3. 11-21. 10.35457/josar.v1i01.614.

Said Kittanah, K., Saharfalehawadabujarour, S., & Amman-Jordan. (2018). The Impact of Document Management Systems on Organizational Performance. Global Journal of Management and Business Research, 16(1),1-2. <https://globaljournals.org/GJMBR_Volume16/1-The-Impact-of-Electronic-Documents.pdf>

Srivastava, P., & Khan, R. (2018). A review paper on cloud computing. International Journal of Advanced Research in Computer Science and Software Engineering, 8(6), 17-20. Retrieved from <https://www.researchgate.net/profile/Rizwan-Khan-27/publication/326073288_A_Review_Paper_on_Cloud_Computing/links/5b3c841fa6fdcc8506eef9de/A-Review-Paper-on-Cloud-Computing.pdf>

Surmieda, *J. (2022). OCR-Enhanced Digital Asset Management System: Prototype*  *Design and Construction*. [35\_PhJLIS2022issue02.pdf](file:///C:/Users/user/Downloads/35_PhJLIS2022issue02.pdf)

TechTarget Contributor (2022). *OCR (optical character recognition).* [What is OCR (optical character recognition)? | Definition from TechTarget](https://www.techtarget.com/searchcontentmanagement/definition/OCR-optical-character-recognition)

Zelic, F. et al. (2021). *How to OCR with Tesseract, OpenCV and Python*. [Tesseract OCR in Python with Pytesseract and OpenCV (nanonets.com)](https://nanonets.com/blog/ocr-with-tesseract/)

Zohra Mahfoud, & Nadia Nouali-Taboudjemat. (2019). Consistency in Cloud-based database systems. Informatica, 43(3). <https://doi.org/10.31449/inf.v43i3.2650>

Nwokoma, F. et al. (2021). *Camera-based OCR scene text detection issue: A review*. https://www.researchgate.net/publication/357453432\_Camera-based\_OCR\_scene\_text\_detection\_issues\_A\_review

Rivero, S. et al. (2021). *Assessing the Relationship Between Binarization and OCR in the Context of Deep Learning-Based ID Document Analysis*. https://www.researchgate.net/publication/355885450\_Assessing\_the\_Relationship\_Between\_Binarization\_and\_OCR\_in\_the\_Context\_of\_Deep\_Learning-Based\_ID\_Document\_Analysis

Xu, X. et al. (2022). *A Novel Adaptive Deskewing Algorithm for Document Images*. https://www.researchgate.net/publication/364426821\_A\_Novel\_Adaptive\_Deskewing\_Algorithm\_for\_Document\_Images

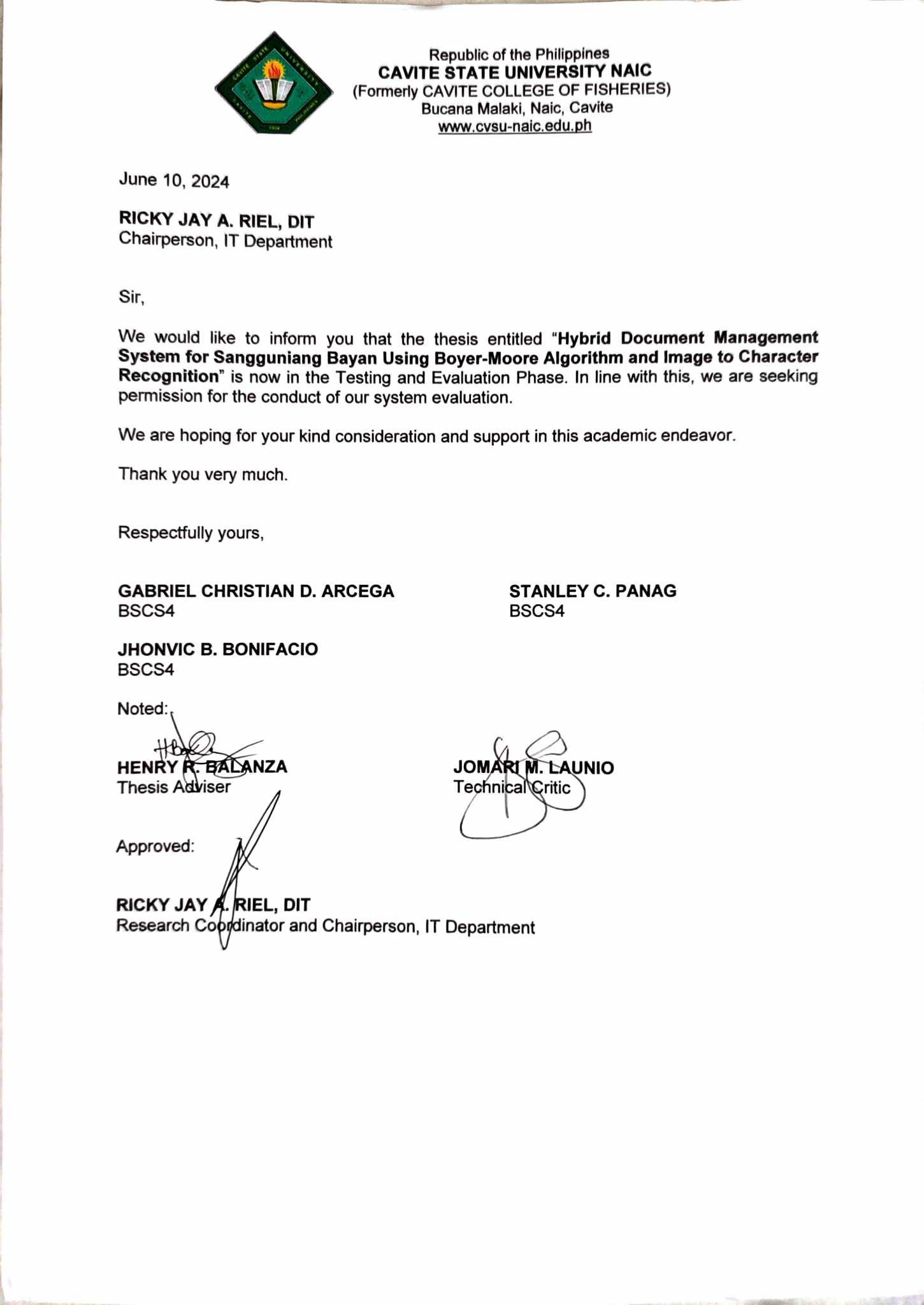
**APPENDICES**

**Appendix 1. Gantt Chart**

Appendix Table 1. Gantt chart

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| PHASE | MONTH | | | | | | | | | | | | | | | |
| 2023 | | | | | | | | | 2024 | | | | | | |
| May | June | July | Aug | Sept | Oct | Nov | Dec | Jan | | Feb | Mar | Apr | May | June |
| Client Consulting |  |  |  |  |  |  |  |  |  | |  |  |  |  |  |
| Planning |  |  |  |  |  |  |  |  |  | |  |  |  |  |  |
| Gathering Resources |  |  |  |  |  |  |  |  |  | |  |  |  |  |  |
| Prototyping |  |  |  |  |  |  |  |  |  | |  |  |  |  |  |
| Coding |  |  |  |  |  |  |  |  |  | |  |  |  |  |  |
| Deployment |  |  |  |  |  |  |  |  |  | |  |  |  |  |  |
| Evaluation |  |  |  |  |  |  |  |  |  | |  |  |  |  |  |

**Appendix 2. Request Letter for System Evaluation**



Appendix Figure 1. Request Letter for System Evaluation

**Appendix 3. Evaluation Instrument**

A software evaluation form with text

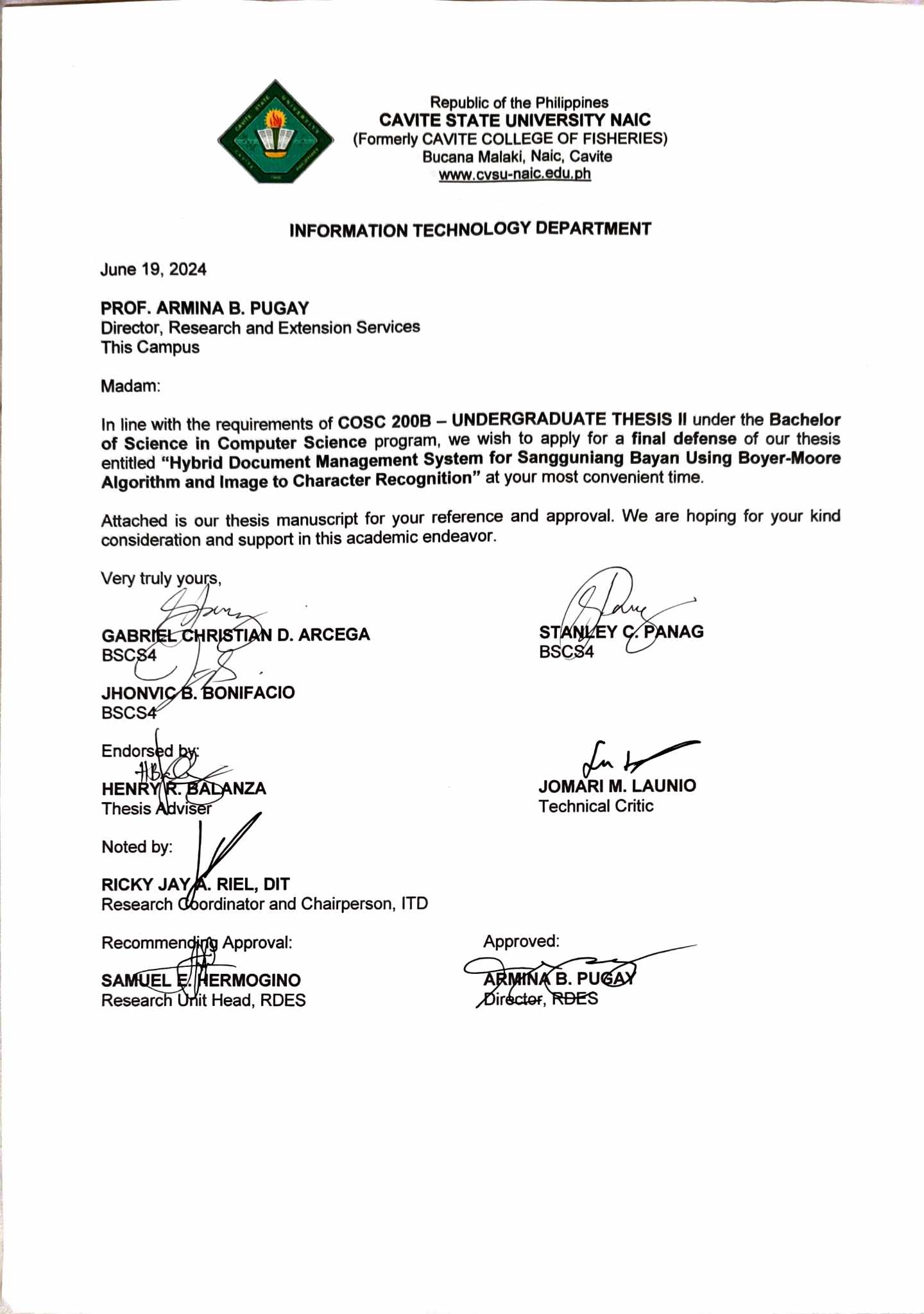
Description automatically generated

**A document with text and numbers

Description automatically generated**

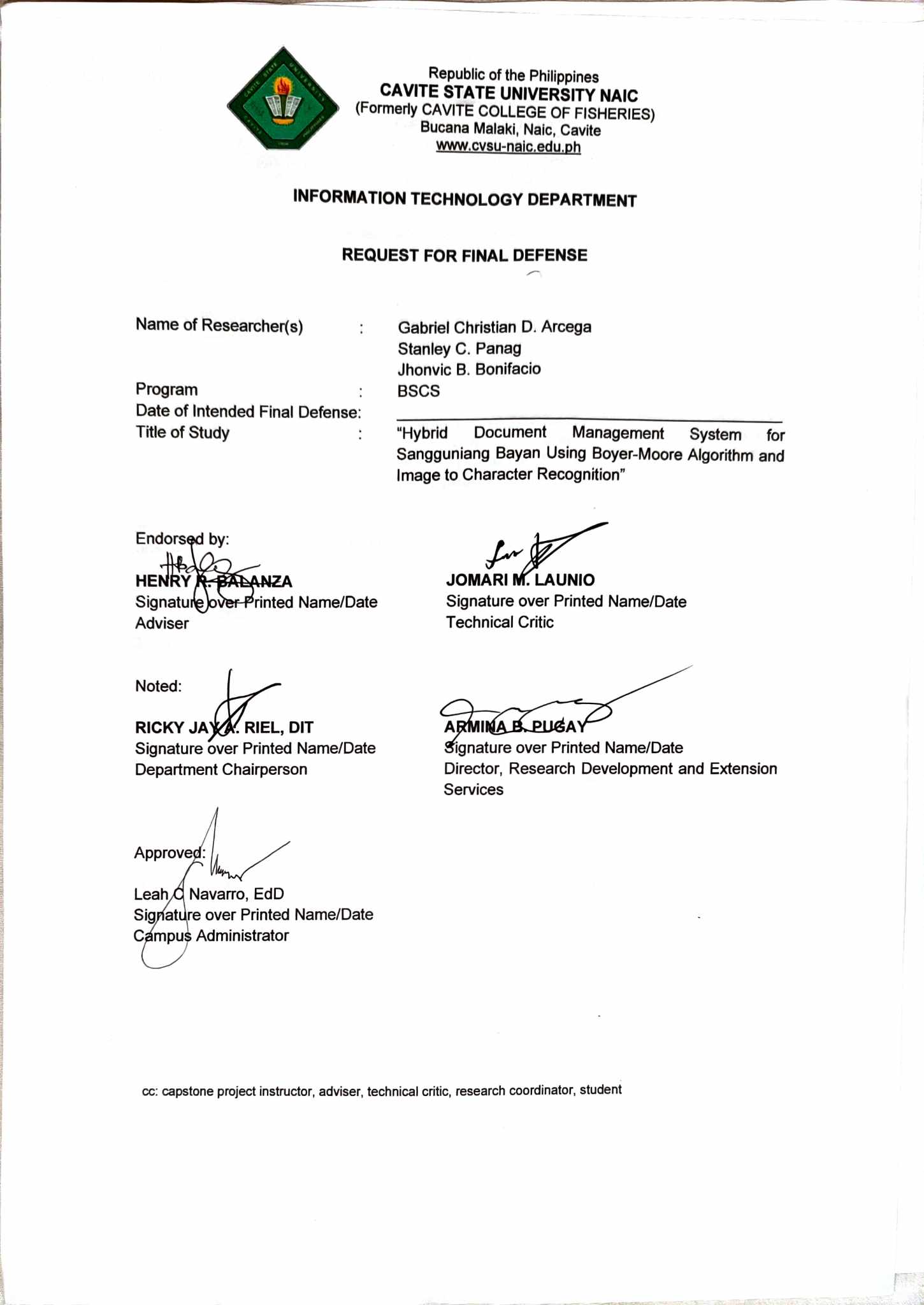
Appendix Figure 2: Evaluation Instrument

**Appendix 4. Application letter for Final Defense**



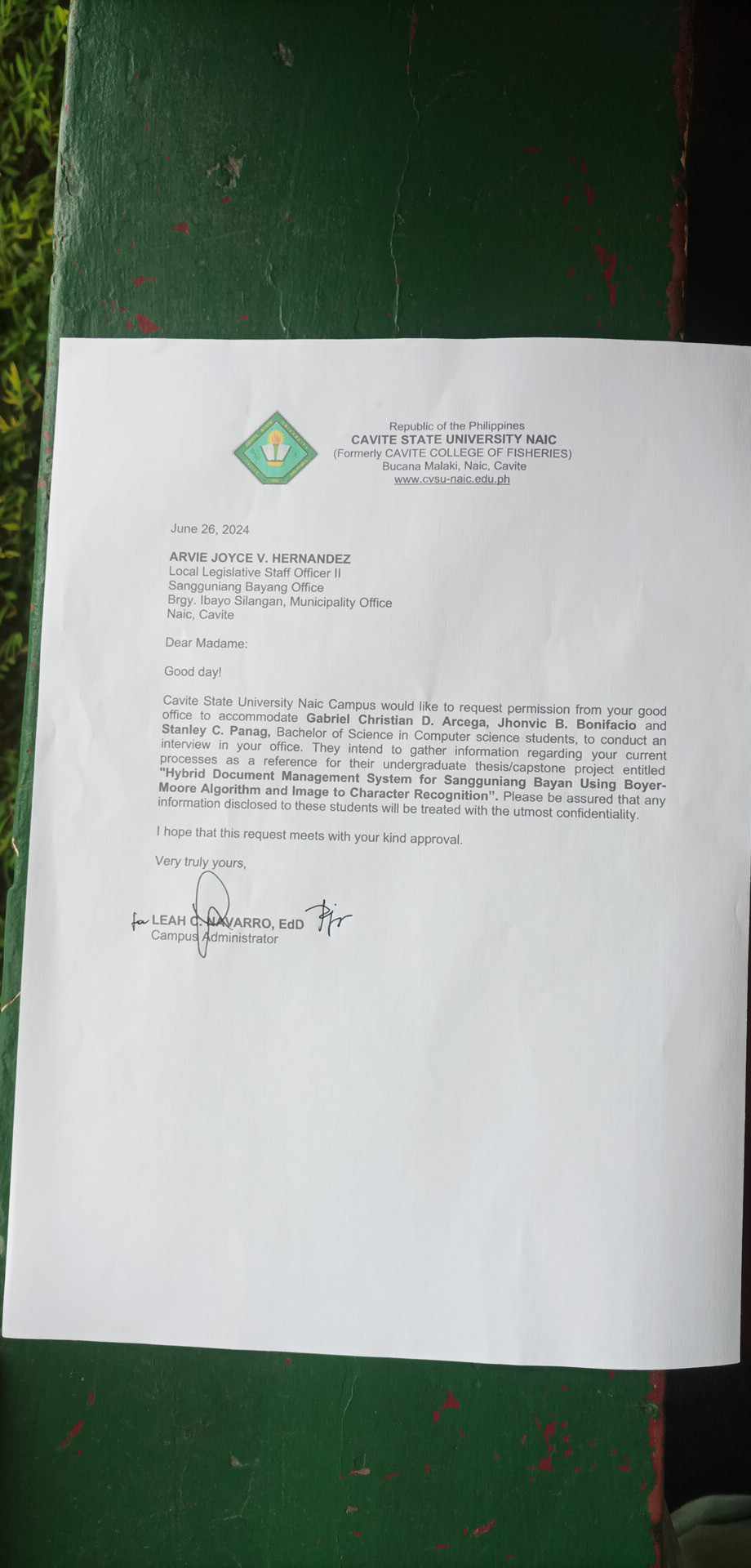
Appendix Figure 3. Application letter for Final Defense

**Appendix 5. Request for Final Defense**



Appendix Figure 4. Request for Final

Defense

**Appendix 6. Request Letter for the Client**

Appendix Figure 5. Request Letter for the Client

**Appendix 7. Data Dictionary**

Appendix Table 2. Data Dictionary

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Name | Type | Field Size | Null | Default | Description | Example |
| user\_id | integer | 11 | no | n/a | Unique identifier or Primary key for each user in the system. | 110 |
| first\_name | varchar | 100 | no | n/a | The given or first name of the user. | Pedro |
| last\_name | varchar | 100 | no | n/a | The family or last name of the use. | Juanito |
| password | varchar | 255 | no | n/a | Hashed password for user authentication | AM9HkKk7fx |
| email | varchar | 100 | no | n/a | Email address associated with the user’s account. | user@gmail.com |
| phone | integer | 11 | yes | NULL | Phone address associated with the user’s account. | 09093334444 |
| role\_id | integer | 11 | no | n/a | Foreign key which is an identifier to role table that associates level of access for users | 110 |

**Appendix 8. Summary of Evaluation Result**

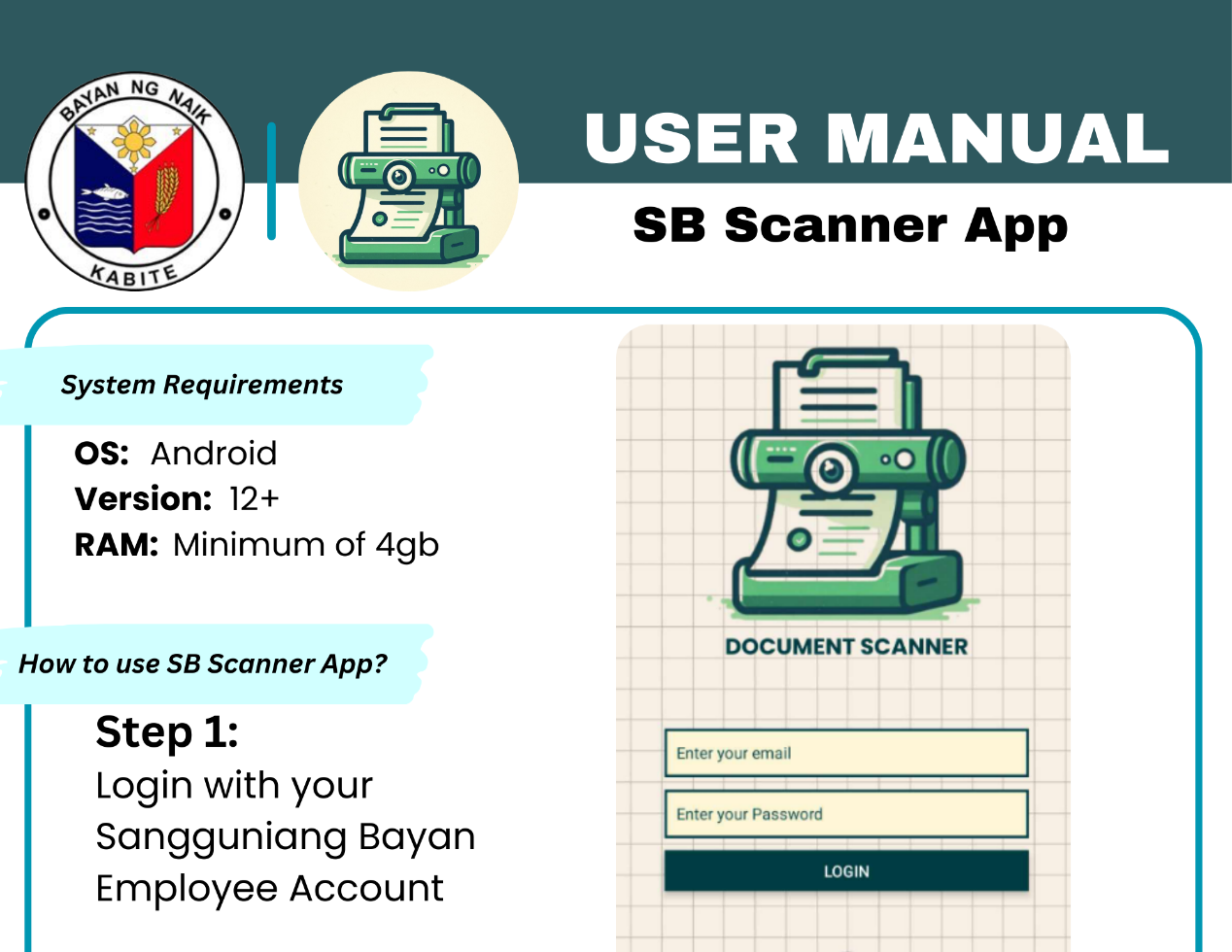
Appendix Table 3. Summary of Evaluation Result

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Software Quality Indicator** | | **Mean** | **Standard Deviation** | **Interpretation** |
|
| **FUNCTIONALITY** | | | | |
| 1 | The software provides appropriate functions | 4.12 | 0.60 | Very Satisfactory |
| 2 | The software shows accurate information and results. | 4.08 | 0.81 | Very Satisfactory |
| 3 | The software provides security measures. | 4.12 | 0.73 | Very Satisfactory |
| **Grand Mean** | | **4.11** | **0.71** | **Very Satisfactory** |
| **RELIABILITY** | | | | |
| 1 | The software shows no failures or bugs | 3.76 | 0.78 | Very Satisfactory |
| 2 | The integrity of data is maintained all throughout its operation. | 4.52 | 0.59 | Excellent |
| **Grand Mean** | | **4.14** | **0.69** | **Very Satisfactory** |
| **EFFICIENCY** | | | | |
| 1 | The software operates quickly and efficiently. | 4.12 | 0.60 | Very Satisfactory |
| 2 | The software requires less resources (memory/CPU/disk storage/internet bandwidth) | 4.28 | 0.68 | Very Satisfactory |
| **Grand Mean** | | **4.2** | **0.64** | **Very Satisfactory** |
| **PORTABILITY** | | | | |
| 1 | The software adapts to new specifications or operating environments. | 4.29 | 0.62 | Very Satisfactory |
| 2 | The software requires less effort to install. | 4.36 | 0.76 | Very Satisfactory |
| **Grand Mean** | | **4.33** | **0.69** | **Very Satisfactory** |

Continuation of Appendix Table 3. Summary of Evaluation Result

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **USABILITY** | | | | |
| 1 | The software functions are organized, clear, logical, and effective making it easy for the user to understand. | 4.24 | 0.60 | Very Satisfactory |
| 2 | The user can operate the software with ease. | 4.24 | 0.72 | Very Satisfactory |
| 3 | The user’s guide is available and clear. | 3.96 | 0.98 | Very Satisfactory |
| **Grand Mean** | | **4.15** | **0.77** | **Very Satisfactory** |
| **INFORMATION IN THE SOFTWARE** | | | | |
| 1 | The information is clear, concise, and informative to the intended user. | 4.44 | 0.51 | Very Satisfactory |
| 2 | The content is free from spelling and grammatical errors. | 3.92 | 0.70 | Very Satisfactory |
| 3 | The contents are covered in a comprehensive manner. | 4.32 | 0.69 | Very Satisfactory |
| **Grand Mean** | | **4.23** | **0.63** | **Very Satisfactory** |
| **TECHNICAL ASPECTS OF THE SOFTWARE AND MATERIALS** | |  | | |
| 1 | The software uses standard equipment that is reliable, widely available, and applicable to a variety of uses. | 4.32 | 0.48 | Very Satisfactory |
| 2 | Computer capabilities such as graphics, color are used for appropriate instructional reasons. | 4.28 | 0.61 | Very Satisfactory |
| **Grand Mean** | | **4.3** | **0.55** | **Very Satisfactory** |
|  | **Overall Mean** | **4.21** | **0.67** | **Very Satisfactory** |

**Appendix 9. Operation Manual**



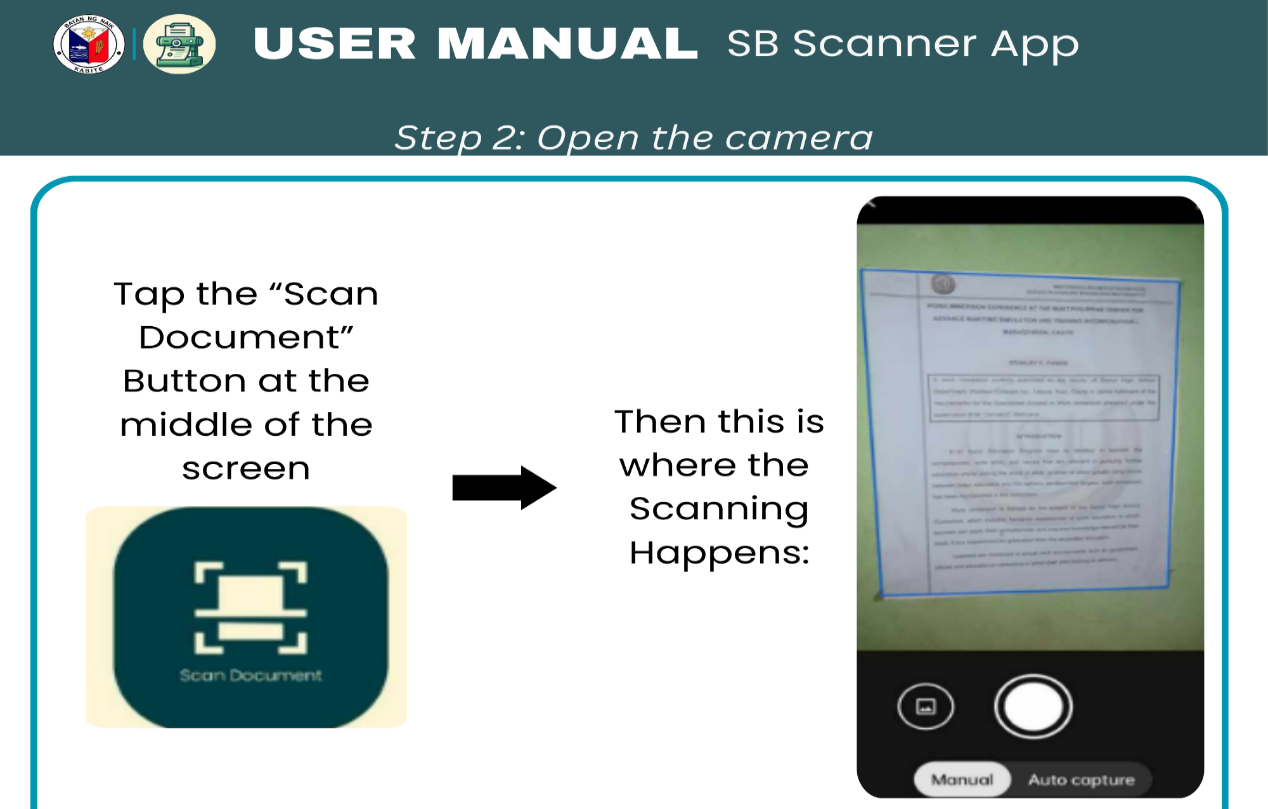
Figure 6. System Requirements and Step 1

Figure 7. Step 2

Continuation of Appendix 9. Operation Manual

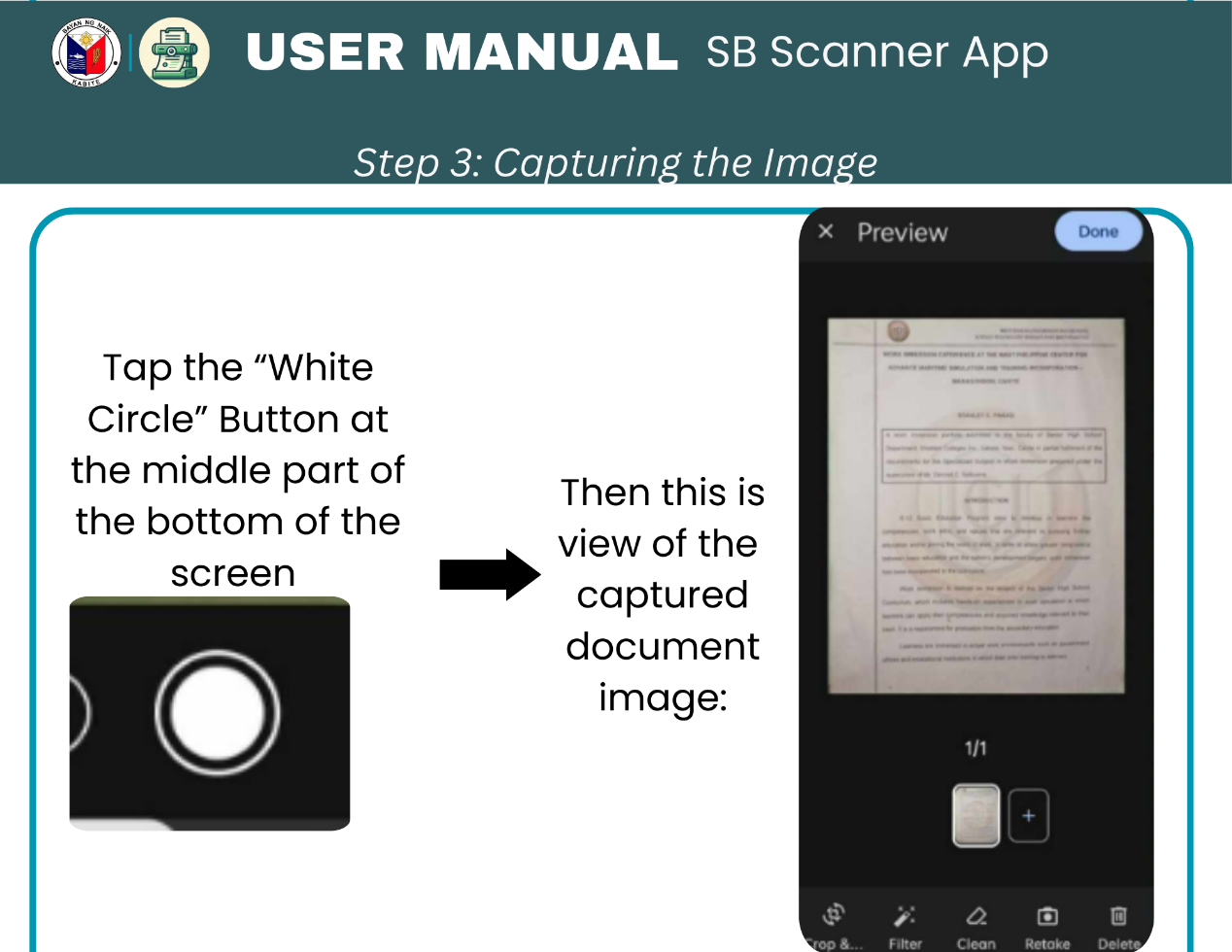


Figure 8. Step 3

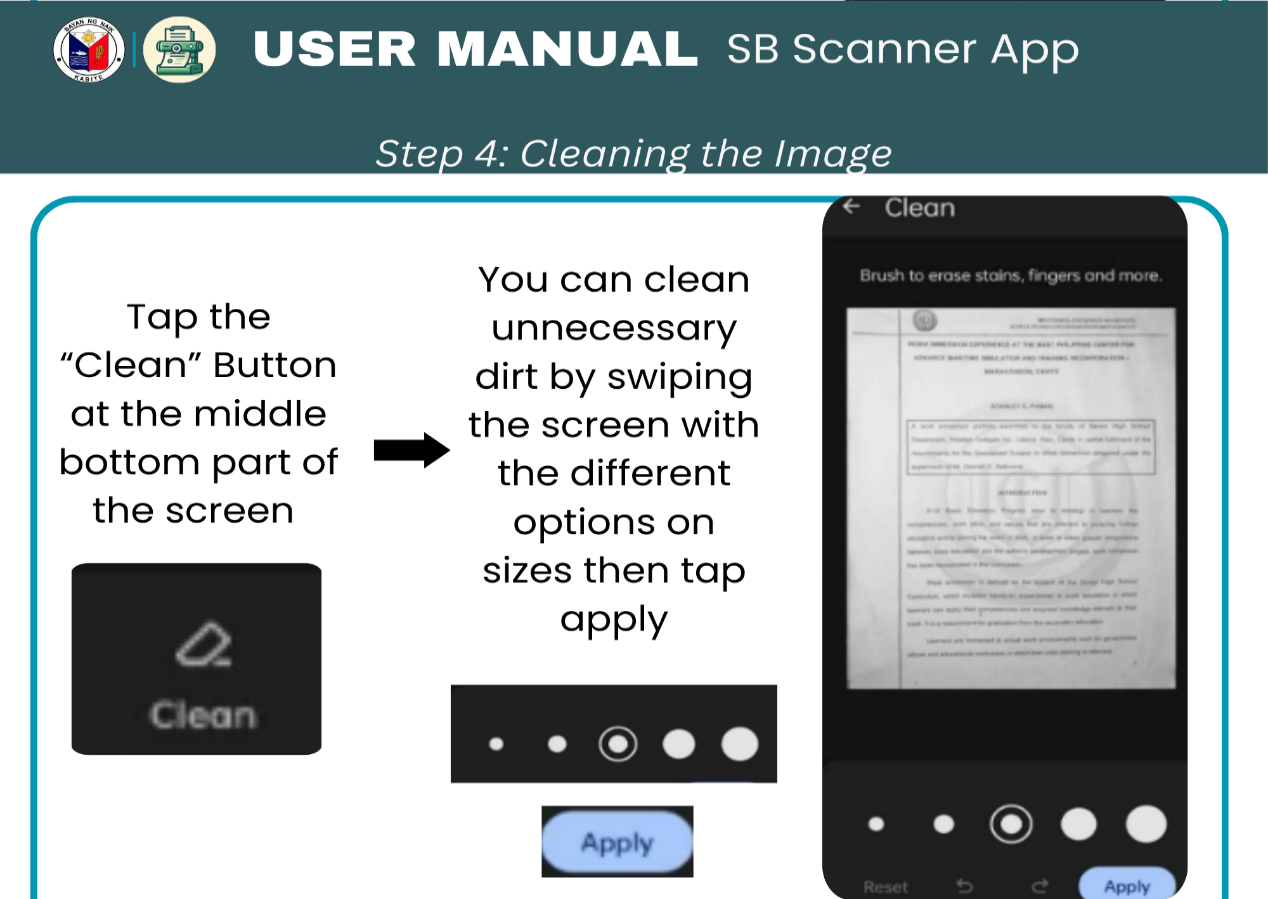


Figure 9. Step 4

Continuation of Appendix 9. Operation Manual

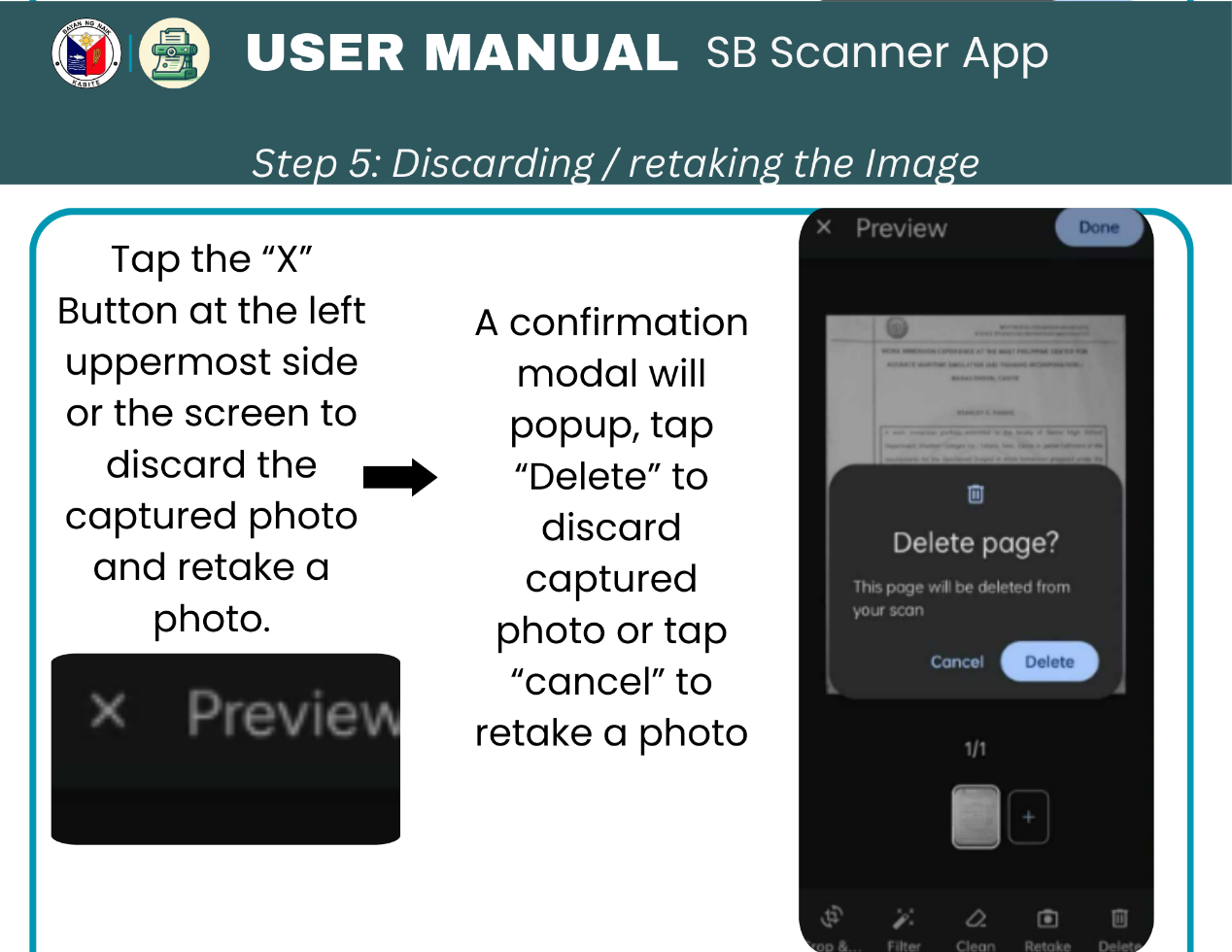


Figure 10. Step 5

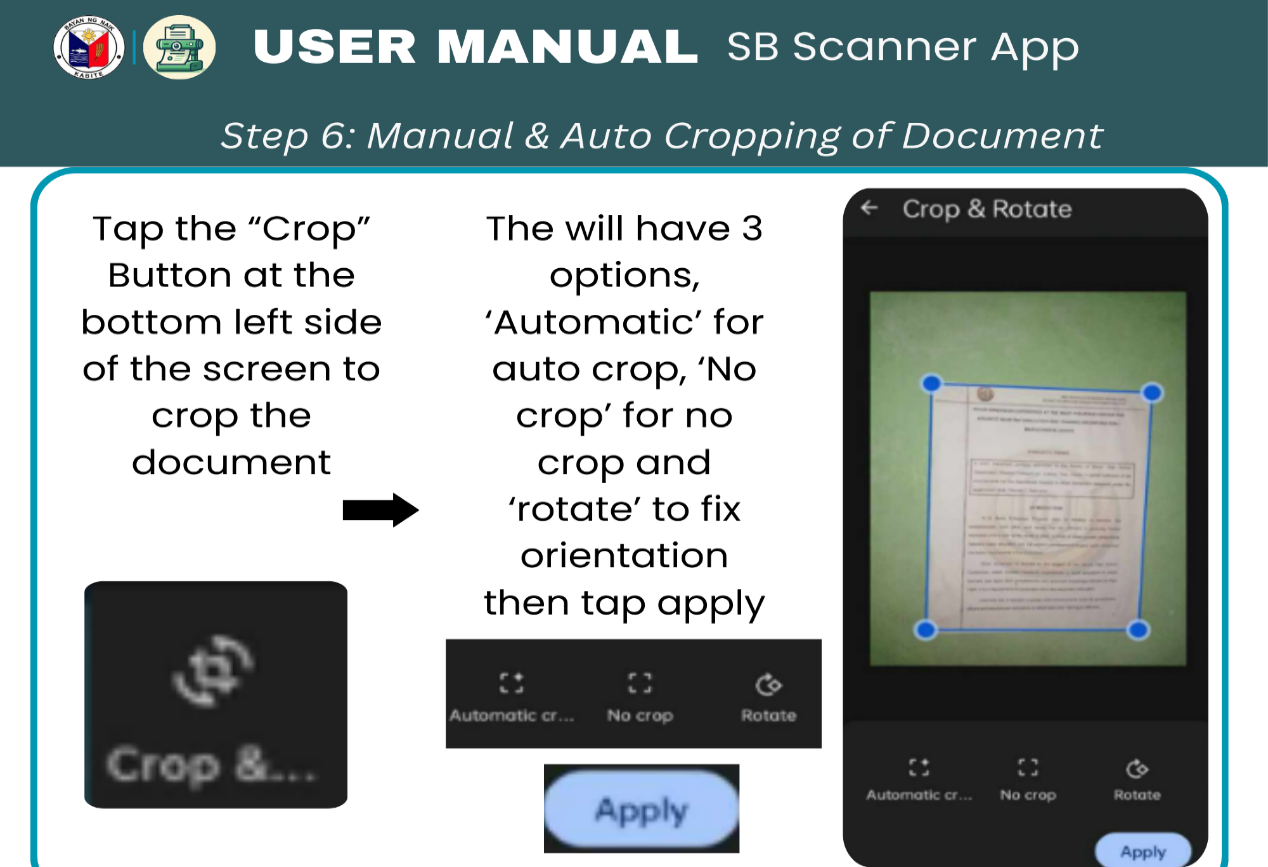


Figure 11. Step 6

Continuation of Appendix 9. Operation Manual

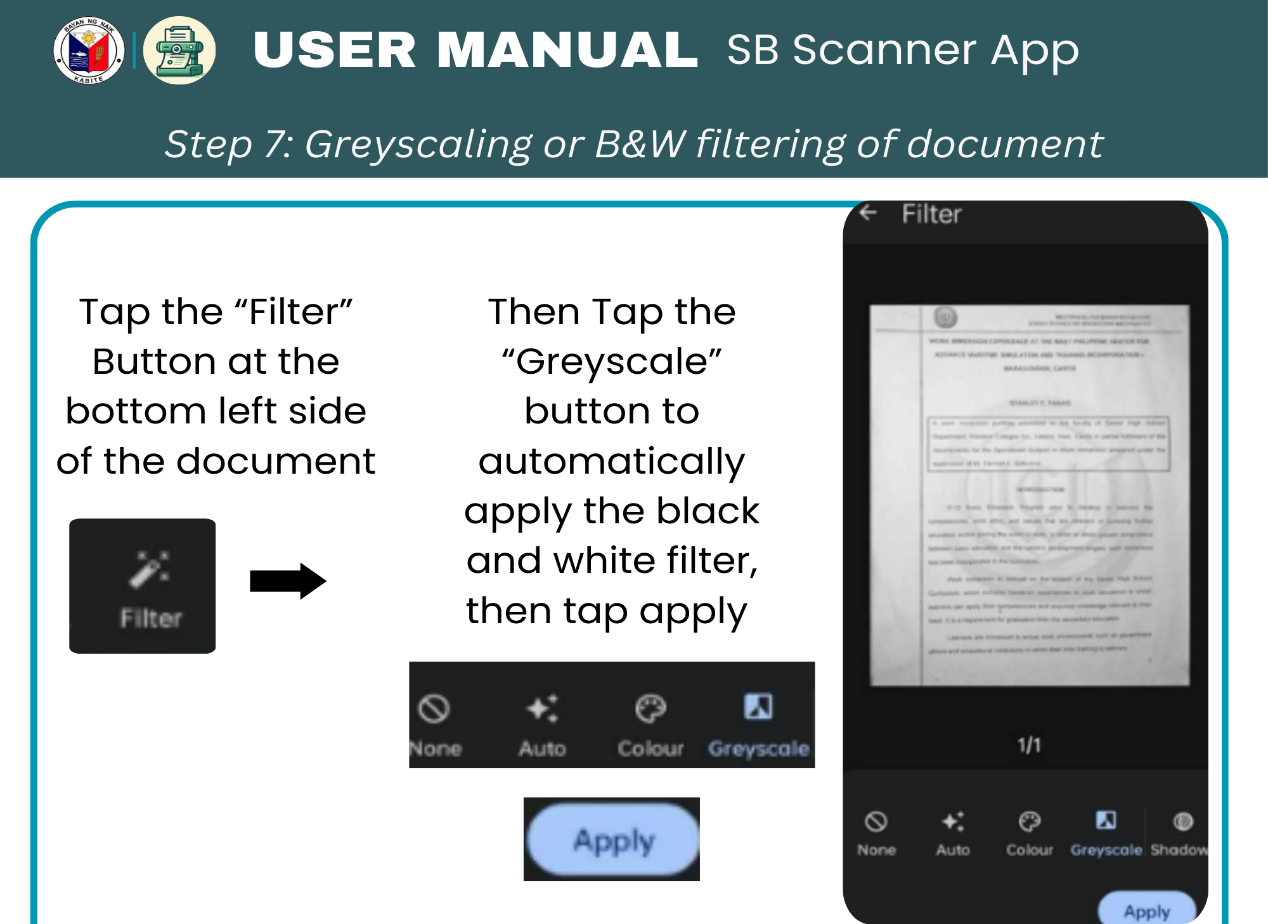


Figure 12. Step 7

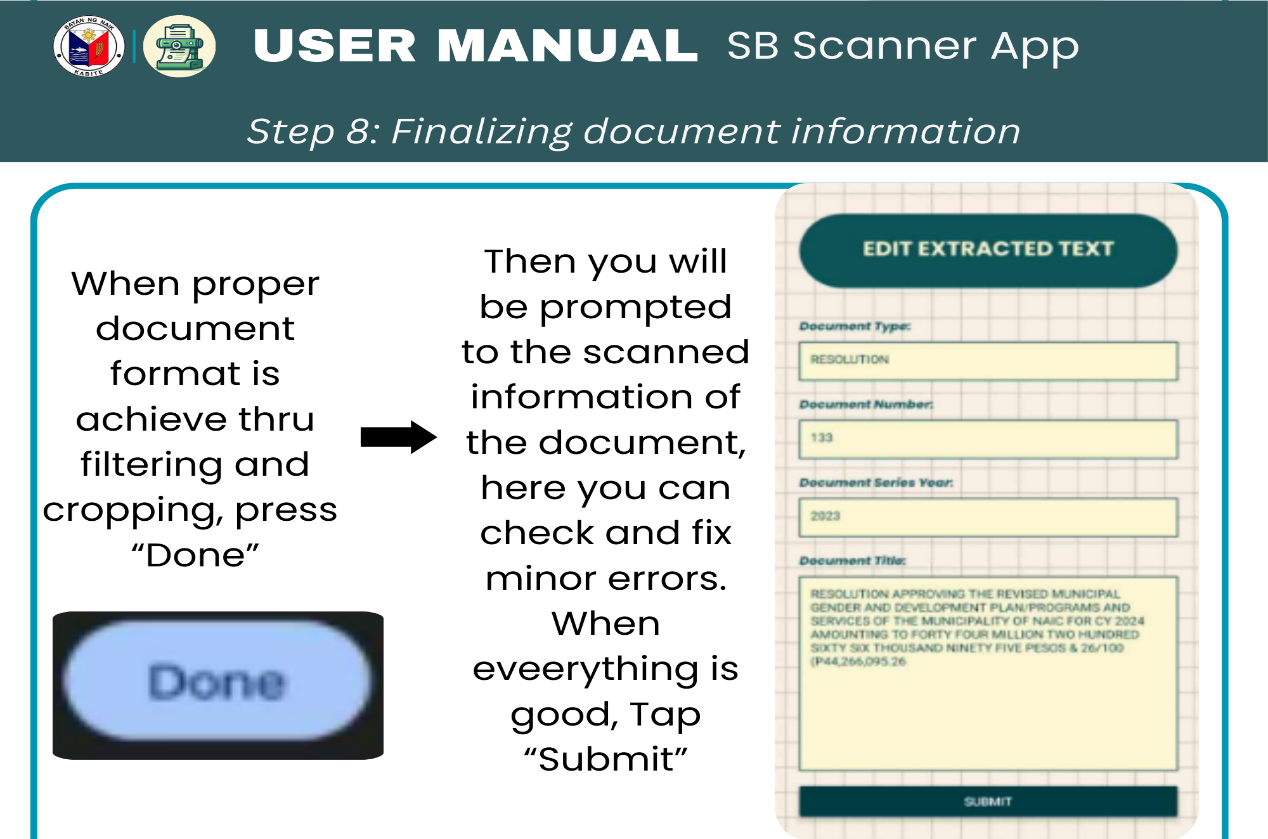


Figure 13. Step 8

Continuation of Appendix 9. Operation Manual

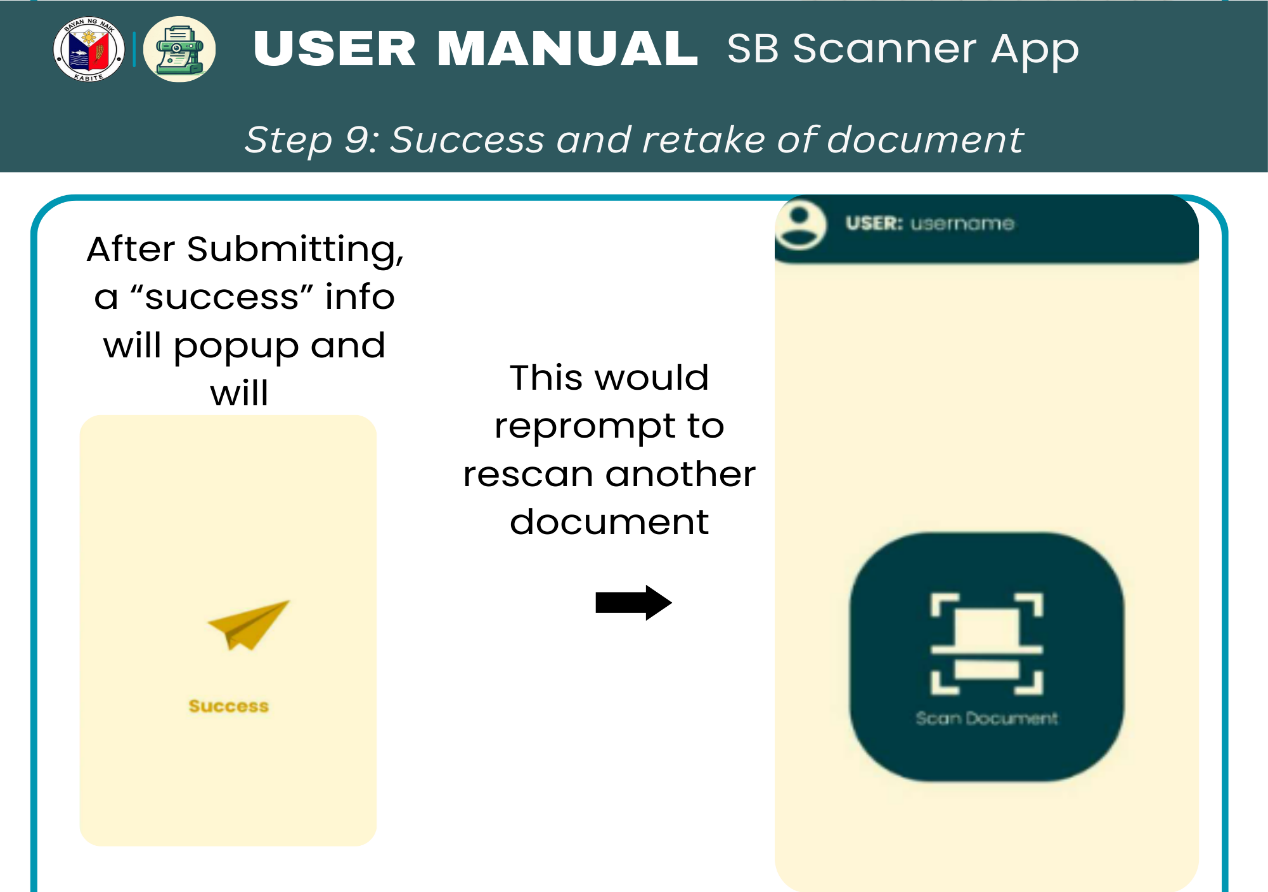
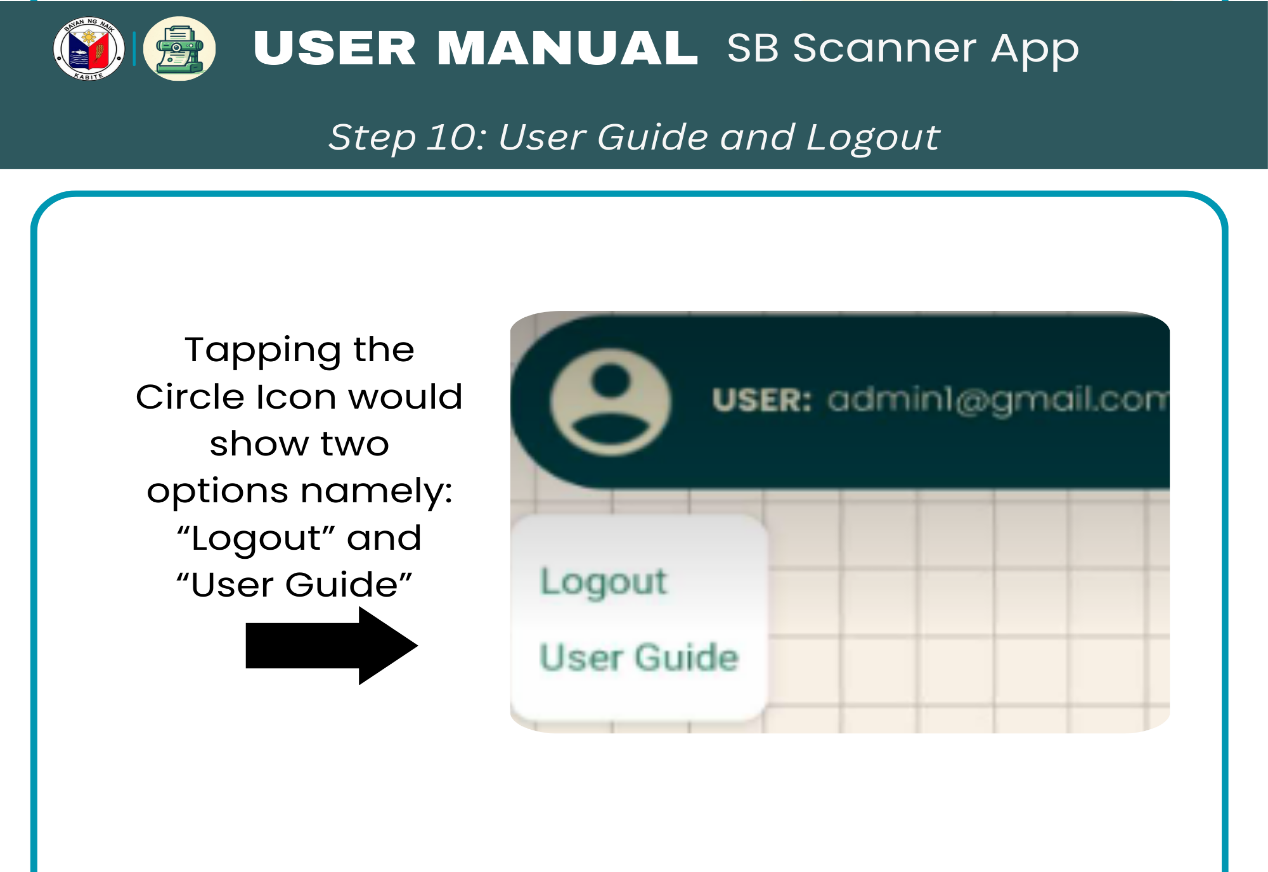


Figure 14. Step 9



Appendix Figure 15. Step 10