**ABridge: Text-type Instant Notes Generator using Natural Language Processing Algorithm**

A Thesis

presented to

The College of Computer Studies

Central Philippine university

Iloilo City

In Partial Fulfillment

of the Requirements for the Degree of

Bachelor of Science in Computer Science

By

Buñag, Rachelle Julie F.  
Draper, Sherry L.

Pari-an, Kent Bryle Q.

Ruiz, Christian Jake A.

#### **CHAPTER I**

**INTRODUCTION**

*Overview of the current state*

Education is vital to the youth and has affected billions of lives. Proper education opens doors to multiple opportunities and has been the cause of the greatest inventions and innovations, bringing efficiency and comfort to another level while fostering critical thinking skills and responsible global citizenship. Time management is a critical skill in education that directly influences students' academic success and overall productivity, which can affect a student's performance amid responsibilities and academic activities. According to Statista (2013), 86% of college students in the US struggle with time management. Time management and mental stability are crucial aspects of a student's life, especially considering their academic pressures and personal challenges.

Another important skill students must have is effective note-taking techniques, as it is considered a fundamental aspect of students' learning experiences and plays a significant role in their educational performance. Note-taking can be applied to numerous activities in educational spaces and professional workplaces. It is a method that allows data organization, resulting in efficient and effective ways of retaining information that can be used to gain clarity or a better understanding of certain topics. Students use this during lectures to record and process information as their study guide for assessments or examinations.

Recent research states that even the fastest note-takers only get around 80% of the key lecture themes, with students recording an average of about one-third. The result implies that many students have difficulty remembering important knowledge. (Glean, 2023). Taking notes during lectures can be intellectually challenging, and students may find it difficult to remember important details or may record them incorrectly. Remembering important details might result from student note-taking abilities and/or the instructor's content presentation, such as an inaudible or fast-paced speech (Stacy, E. M., & Cain, J., 2015). Effective note-taking helps students optimize their time by enabling them to focus on essential content. Note-taking practices' efficacy and impact on academic performance warrant important consideration.

Technology integration has revolutionized nearly every aspect of education, offering students and educators innovative tools and resources to enhance teaching and learning experiences. One area where technology has significantly impacted student learning is note-taking. Traditional methods of pen-and-paper note-taking are increasingly being supplemented or replaced by digital tools and platforms that offer a range of features to enhance the learning process and allow students to explore its impact on their learning outcomes. Especially during the pandemic, as students increasingly turned to personal computers, laptops, and mobile devices for online education, the landscape of note-taking experienced a significant shift towards technology integration. Students opted for digital platforms such as Word-type documents and Quizlet, a preferred application among peers, to effectively capture and organize their notes in this new educational paradigm.

Another educational technology application, like Knowt, offers file import systems, OCR and text summarization that can generate inaccurate notes where students cannot utilize the result.

Integrating technology with note-taking offers numerous benefits but presents several challenges and problems. Students may need help navigating complex software interfaces and customized settings. Traditional handwriting methods also pose a problem: it can be challenging to structure and categorize handwritten notes effectively, leading to difficulty reviewing and studying the material later.

Considering the advantages, challenges and implications of using traditional and technology-enhanced note-taking methods, this leads the team to develop a text-type instant note-generating application to allow students to upload their notes in either PDF or image format. The text data from the uploaded files will be processed, summarized and generated into notes. The student can then create quizzes or flashcards for their review.

This study's proposed subject or stakeholders are the Grade six students of the Elementary Department of Central Philippine University. The grade six level is critical in a student's cognitive and academic development. Students are transitioning from concrete to more abstract thinking at this age, which impacts their ability to take notes and comprehend scientific concepts effectively. Understanding note-taking problems at this stage can provide insights into cognitive processes and learning strategies that can be applied across various subjects. Initially, we only focus on the Science curriculum of the Grade six students, as the subject contains multiple terminologies and definitions. Science education in grade six often introduces foundational concepts in biology, chemistry, physics, and earth sciences. Effective note-taking skills are essential for students to grasp these fundamental principles, develop scientific literacy, and cultivate a curiosity for inquiry-based learning.

*Desired State of Technology*

Traditional note-taking methods often rely solely on pen and paper, overlooking the vast array of digital tools and resources available to enhance note-taking experiences. By acknowledging the potential benefits of technology in note-taking, such as increased efficiency, accessibility, and customization, the group seeks to address the gap between conventional and modern approaches to learning. Technology-based note-taking allows for easier organization, searchability, and access to notes across devices, providing greater convenience for students who prefer digital platforms. Integrating technology into traditional note-taking practices fills several significant gaps, enhancing the effectiveness and efficiency of note-taking.

The desired state of technology for the proposed web application, developed using Python, integrates various features such as text generation, summarization, quizzes, and flashcards. This application utilizes Natural Language Processing (NLP) algorithms for text classification and summarization. It uses both extractive and abstractive approaches to apply strategies like text classification for organizing material and text summarizing for rapid reading of long texts.

Furthermore, document embedding using the Bag of Words paradigm makes simple categorization tasks possible. Techniques for lemmatization and stemming decrease word variability, improving the effectiveness and stability of NLP algorithms and ensuring the accuracy of the system's results.

Based on popular educational technology applications, only Knowt utilizes OCR technology. The application uses Tesseract, an open-source optical character recognition (OCR) engine, for text recognition and document scanning. While text detection and character recognition separate and examine text sections, image preprocessing techniques increase the accuracy of OCR. Language modeling uses statistical language models and dictionaries to improve recognition accuracy. The identified text output is refined for better readability through post-processing procedures.

Overall, this web application developed with Python offers a complete platform to generate study notes, summarize information, make quizzes, and use flashcards, all improving users' learning outcomes.

*Statement of the Problem*

The study aims to address the following problems:

* The lack of time available for studying directly results from the significant time required to craft detailed notes thoroughly. This issue often leads to decreased quality study time and can negatively impact academic performance.
* Improper time management becomes a common struggle for students when they have to juggle studying multiple subjects, each demanding time and attention, ultimately leading to overloaded schedules and potential burnout.
* Disinterest in reading and utilizing textbooks in a school environment leads to reduced learning outcomes due to a lack of efficiency and student engagement in the traditional learning process.

*General Objective:*

This study uses machine learning, natural language algorithms, and OCR technology to develop a Text-Type Instant Study Notes Generator for the Central Philippine University Elementary Science Curriculum.

*Specific Objectives:*

* To develop an automated note-taking feature to reduce the time required for creating study notes
* To develop a time management tool that will help students learn how to manage their time between multiple subjects efficiently
* To develop a real time notes generator tool that instantly generates notes from textbooks dedicated to enhancing student engagement towards achieving successful learning outcomes.

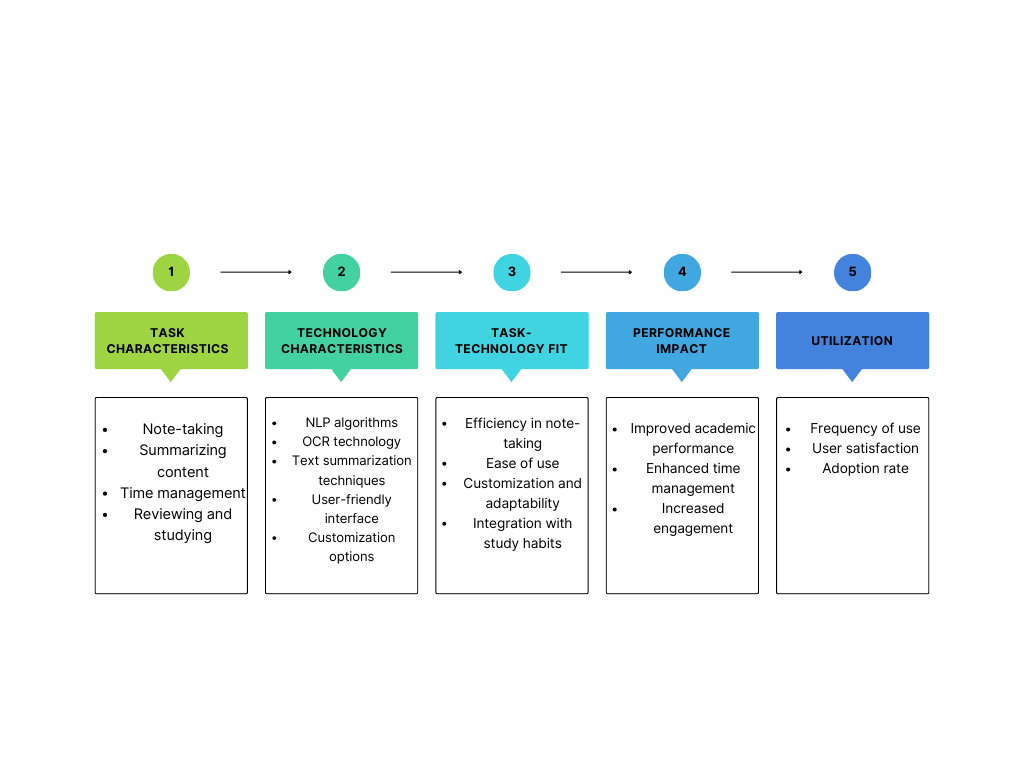
*Theoretical Framework of the Study*

Figure 1. Theoretical Framework of the Study

With the rapid progression of technology, combining technology and learning can build vital teaching and learning. According to Noesgaard et al. (2015), their research shows that E-learning has become another tool for learning inside and outside schools within the last five years. It also tackles that eLearning allows learners to access educational materials at any time and location, eliminating time and location constraints. It offers flexibility, allowing students to tailor their educational journey to their needs, pace, and preferences, resulting in a personalized learning experience. Using different platforms to generate materials helps the student to study and learn. (Garcia, 2017)

While the Philippines' e-learning environment is still in its early stages, prominent universities have already adopted and implemented it. E-learning environments stand to benefit greatly as the Internet becomes more resilient. Researchers and academics are expanding their studies to include more advanced courses and fields. They are investigating the factors influencing Filipino college students' willingness to use e-learning as part of their education. These considerations include integrated multimedia instruction, perceived quality of life, system interactivity, internet connectivity experience, perceived ease of use, perceived usefulness, and social media influence. (Garcia, 2017)

*Conceptual Framework*

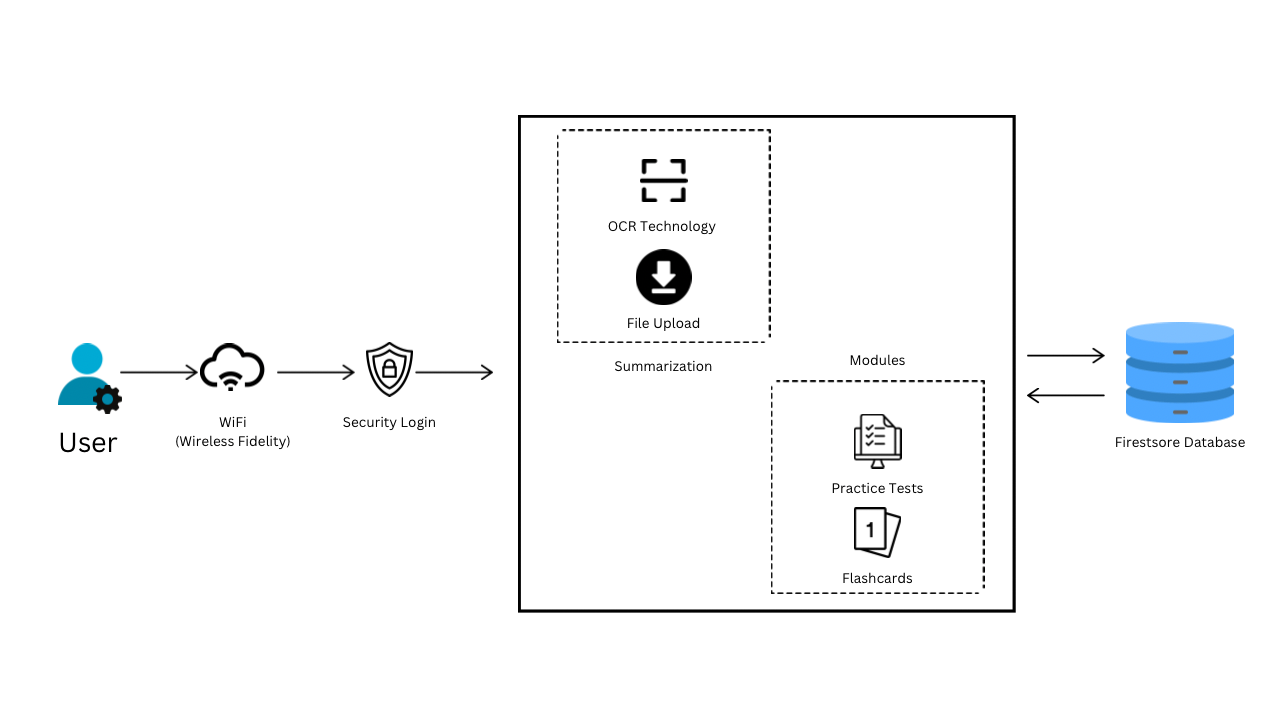
**

Figure 2. Conceptual Framework of the Study

Security

This framework ensures data and systems' confidentiality, integrity, and availability using logins and passwords. To register, users must input the required credentials to access the system. Registered accounts can log in directly using their email and password.

Wireless Fidelity

Wireless Fidelity (WiFi) is necessary for the web application to work for it to connect to the database of the history logs of past works of the users. WiFi technology significantly enhances the functionality and efficiency of summarizing tools by providing seamless internet connectivity. Wireless connectivity allows summarizing tools to quickly access vast amounts of online data, enabling real-time updates and processing.

Web Application

ABridge runs on a web server and is accessed by users through a web browser, providing design and usability of the web application, guided by UX/UI principles to ensure intuitive and accessible interfaces. The web application is an excellent platform for a summarizing tool due to its accessibility, allowing users to access it from any device with an internet connection and a web browser without needing software installation.

Summarization

This process involves identifying and extracting key points, main arguments, conclusions, and significant data from the original files and presenting them in an easy-to-understand and review way. Summarizing serves various purposes within the framework, such as facilitating decision-making, enhancing comprehension, and streamlining information retrieval. The process ensures that the summarised content is clear, accurate, and relevant by organizing extracted information thematically and choosing appropriate formats like abstracts, executive summaries, or bullet points.

OCR Technology

Optical Character Recognition (OCR) technology converts documents, such as scanned paper documents, PDF files, or images captured by a digital camera, into editable and searchable data.

OCR technology advances efficiency, productivity, and accessibility across industries by streamlining document-centric workflows and reducing manual data entry efforts.

Modules

Modules contain practice tests that refer to exercises or assessments designed to evaluate understanding and applying the concepts outlined within the framework. In education settings, these come in multiple choices or flashcards that support integrating theory and application, facilitating deeper understanding and mastery of the subject matter.

Firestore Database

Firestore, or Cloud Firestore, is a flexible, scalable database from Firebase for mobile, web, and server development and operates within the NoSQL database paradigm. With the leverage of Firestore, it provides a flexible and scalable solution for storing and synchronizing data in real-time across multiple clients, with robust security features and seamless integration.

Python

Python's conceptual framework revolves around its readability, versatility, robust standard library, support for multiple paradigms, and strong community support. Using Python's rich ecosystem of NLP libraries provides ease of use, strong community support, access to pre-trained models, integration capabilities, and performance optimizations, making it an ideal choice for developing a summarization tool using NLP techniques.

*Scope and Limitations of the Study*

This research is about developing an application that provides students aid and assistance in reviewing by generating text-type summarized versions of the inputted material that is editable by the user. The data collection will be conducted on Grade 6 students in the Elementary School of Central Philippine University.

This study is focused on the Science Curriculum of Grade 6 and will not extend to other subjects and grade levels. The respondents should be enrolled in the 2023-2024 Grade 6 school year at Central Philippine University. Technical constraints and limitations in processing capabilities may affect the app's accuracy and reliability. Moreover, the login credentials set up by parents may cause anomalies and complications in the system.

*Significance of the Study*

This research will offer the following benefits:

**Teachers**: The instant generation of study notes can save time for teachers as they can use the tool to generate resources for their lesson plans and offer accurate and concise lectures. Teachers may learn how to incorporate study note generation into their lessons and encourage a positive classroom environment by monitoring the effectiveness of the Instant Study Notes Generator for their students.

**Students**: Instant study notes generator can be tailored to cater specifically to their level of understanding. It can provide simplified explanations, visual aids, and interactive elements to engage young learners effectively. It can also encourage active learning by providing interactive content such as quizzes, diagrams, and simulations, potentially promoting better study habits and time management skills.

**Parents**: Parents could utilize the generated notes as a supplementary resource to solidify their child's understanding of the learned concepts offering easier study ethics at their homes. The generator will also allow parents to specify learning objectives or areas of difficulty; it could personalize the study experience and address specific knowledge gaps for their children.

Researchers:

Studying such a tool can enhance students' learning experience, alleviate teacher workload, and contribute to the development of innovative educational technology.

#### **CHAPTER II**

**Review of Related Literature**

**SynthNotes: A Generator Framework for High-volume, High-fidelity Synthetic Mental 79UI JN8 Health Notes**

SynthNotes is a cutting-edge framework designed to produce large-scale, privacy-preserving, and realistic mental health progress report notes. It addresses the challenge of insufficient training data for AI/Machine Learning tasks in the context of Big Data and AI. With the need to explore, evaluate, and train computational methods for the emerging mental health crisis in the US, SynthNotes influences literature and public statistical models to generate free text notes that accurately represent patient mental health characteristics. SynthNotes creates text notes using information from literature and public statistical models, providing an authentic portrayal of a patient's mental health traits. It also simulates licensed mental health professionals' writing style and expressive features. The framework is versatile, accommodating different conditions and model variations.

Furthermore, it offers flexibility in structure, content, and corpus size. Our paper presents the initial usage and performance of SynthNotes, along with ongoing work to incorporate content planning and deep learning techniques trained on real data (Begoli Edmon, ORCiD logo, Brown Kris, Srinivasan Sudarshan, Tamang, Suzanne, 2018).

#### **Automated Question Generator using NLP**

Automated Question Generator with Natural Language Processing (NLP) is a system that uses machine learning and artificial intelligence to generate question papers automatically. This system employs NLP techniques to analyze a given text, identify key concepts, and generate questions based on them. The generated questions are diverse and adhere to a predefined format. This system eliminates the need for manually creating question papers, which can be time-consuming and error-prone. The proposed system can help educators quickly generate question papers and assess student performance. This system works by analyzing the course material's topic, subject, and complexity level to generate a set of questions to assess students' comprehension. The system uses NLP. (Tejas Chakankar, Tejas Shinkar, Shreyash Waghdhare, Srushti Waichal, Mrs. M. M. Phadtare, 2023).

**AI Text Summarization System.**

AI text summarization system that uses GPT, a powerful language model, to create concise and meaningful summaries of long text documents. The system comprises four modules: user, android application, GPT API, and GPT server. The user interacts with the system using the Android application, which serves as the user interface. The GPT API bridges the Android application and the GPT server, which houses the GPT model and manages the text summarization process. The system generates summaries using cutting-edge natural language processing techniques that preserve contextual coherence and salient information. The system's summarization abilities are evaluated using metrics such as Rouge and F1 scores, demonstrating its effectiveness in capturing key information from various text documents. The system's integration with Android platforms allows for easy access for mobile users, making it ideal for applications like news summarization, document summarization, and content curation. The system's modular architecture promotes scalability and flexibility, allowing for future improvements and extensions. The AI text summarization system, which uses GPT, offers a promising approach for automatically summarizing large volumes of text, providing users with time-saving and meaningful summaries. The system has potential applications in various domains and can serve as a foundation for future research in text summarization and natural language processing. ( Swapnil Kuyate, Omdeep Jadhav, Pratik Jadhav, 2023)

**PhotoOCR: Reading Text in Uncontrolled Conditions**

PhotoOCR is a machine learning-based system designed for extracting text from images while focusing on improving text recognition from smartphones. This area often struggles to recognize text from the images. Leveraging advancements in machine learning isolated character classification and incorporated datacenter-scale language modeling. The PhotoOCR demonstrates reliable text recognition even in complicated image conditions, for example, blur, low resolution, low contrast, high noise, and distortion. PhotoOCR processes images quickly, with a mean time of 600 ms, outperforming previous text extraction benchmarks. PhotoOCR is currently in use in many applications at Google, and it's also available as a user input modality inTranslate for Android, demonstrating the system's effectiveness, dependability and value to users looking for a quick and accurate method of extracting text from images. (Alessandro Bissacco, Mark Cummins, Yuval Netzer, Hartmut Neven, 2013)

**NLP-Cube: End-to-end raw text processing with neural networks**

NLP-Cube is an end-to-end framework that excels in CoNLL" Multilingual parsing from raw Text to Universal Dependencies 2018 Shared task. NLP-Cube is free, and it uses a Python-based NLP system that performs various tasks, including sentence splitting, tokenization, compound word expansion, lemmatization, tagging, and parsing. Its use of neural networks enables accurate and efficient NLP processing for various applications. The performance of lexicalized and delexicalized features or text extraction degrades when using morphological features due to predicted features being used as input during runtime. To solve the issue, the NLP-Cube uses only lexicalized features as input for all modules, except for the lemmatizer, which requires morphological information. The NLP-Cube gives an accurate and efficient NLP processing for a wide range of applications. Its strong performance in the CoNLL shared task demonstrates its effectiveness in handling complex NLP tasks. (Tiberiu Boros, Ruxandra Burtica, Stefan Daniel Dumitrescu, 2018)

**SynthNotes: A Generator Framework for High-volume, High-fidelity Synthetic Mental 79UI JN8 Health Notes**

SynthNotes is a useful framework that addresses the challenges of creating realistic and authentic mental health progress report notes using AI/ML techniques. It presents a promising solution to the problem of insufficient training data and lays the groundwork for future research and development in this field.

#### **Automated Question Generator using NLP**

Automated Question Generator with Natural Language Processing (NLP) is a system that generates question papers using machine learning and artificial intelligence. This system uses natural language processing (NLP) techniques to analyze a given text, identify key concepts, and generate questions from them. The system eliminates the need to manually create question papers, which can be time-consuming and error-prone, while also assisting educators in quickly generating question papers and assessing student performance.

**AI Text Summarization System.**

AI text summarization system that uses GPT, a powerful language model, to create concise and meaningful summaries of long text documents. The AI text summarization system utilizes a modular architecture comprising four main components: the user interface via an Android application, the GPT API acting as a mediator, the GPT server housing the language model, and advanced NLP techniques. It effectively generates concise and coherent summaries from lengthy text documents, evaluated by metrics like Rouge and F1 scores.

**PhotoOCR: Reading Text in Uncontrolled Conditions**

The PhotoOCR system enhances text recognition from images, particularly on smartphones. It utilizes advanced machine-learning techniques such as isolated character classification and large-scale language modeling. The PhotoOCR is integrated into Translate for Android, giving users a quick and accurate method of extracting text from images.

**NLP-Cube: End-to-end raw text processing with neural networks**

NLP-Cube is a free Python-based NLP system with sentence splitting, tokenization, compound word expansion, lemmatization, tagging, and parsing features. It utilizes neural networks for precise and efficient NLP processing in various applications. The system's success in the CoNLL shared task demonstrates its ability to manage complex NLP tasks.

The studies display the different applications of text summarization and image extraction.

ABridge offers a text-type instant note-generating application that allows students to upload their notes in either PDF or image format. Text data gathered from the uploaded files will be processed, summarized and generated into notes. The user then is given the option to create quizzes or flashcards to allow efficient and effective studying for the user.

Table 1. Matrix of Studies and System

|  | A | B | C | D | E | F |
| --- | --- | --- | --- | --- | --- | --- |
| Login |  |  | ✔ |  |  | ✔ |
| Register |  |  | ✔ |  |  | ✔ |
| Profiling |  |  |  |  |  | ✔ |
| User Interface |  |  |  |  |  | ✔ |
| Quizzes |  | ✔ |  |  |  | ✔ |
| Text generator | ✔ | ✔ | ✔ |  |  | ✔ |
| Flashcard |  |  |  |  |  | ✔ |
| Summarization tool |  | ✔ | ✔ |  |  | ✔ |
| NLP Algorithm |  | ✔ |  |  | ✔ | ✔ |
| Machine Learning | ✔ |  |  | ✔ | ✔ | ✔ |

Table no. 1

Legend:

1. SynthNotes: A Generator Framework for High-volume, High-fidelity Synthetic Mental 79UI JN8 Health Notes

#### Automated Question Generator using NLP

1. AI Text Summarization System.
2. PhotoOCR: Reading Text in Uncontrolled Conditions
3. NLP-Cube: End-to-end raw text processing with neural networks
4. Text-Type Instant Study Notes Generator for Central Philippine University Elementary Science Curriculum

#### **CHAPTER III**

**Methodology**



Figure 3. Agile Method

Agile Methodology

The Agile methodology is an iterative approach to software development that emphasizes flexibility, collaboration, and customer feedback. It diverges from traditional waterfall methods, where development progresses linearly through phases like requirements gathering, design, implementation, testing, and maintenance.

Planning

This phase involves assessing the project's scope and objectives in developing a text-type summarizing tool. The researchers outlined the functions to attain effective time management, willing assessments, and reliable resource allocation. These objectives set a clear trajectory for development, ensuring alignment, clarity, and efficiency.

Design

The researchers utilized Figma to lay out the application modules, including the registration, the home, the dashboard and the repository for the user to access. An activity diagram was used to map the flow of activity, processes and workflow when using the app. A use case diagram was designed to illustrate the interactions between the user and the system as they navigate the functionalities provided. The sequence diagram was used for visualizing the flow of messages and behaviour during each executed process.

Develop

During this phase, the team learned to utilize the NLP algorithm. Then began coding on the system's back-end, such as registration and login, user account modules, and uploading files for summarization. Registration will include providing the school email, password, user full name, and user password. The Registration module will allow the user to access the system. Summarization will let you have options whether the user will upload a file or the system will scan images using OCR technology. Users can make practice tests or flashcards and store them in their dashboards, where the history log of past projects will be directed.

Testing

The researchers utilized a systematic approach by implementing repetitive unit testing procedures to verify the precision and dependability of the system, ensuring its overall quality and performance. Each function within the system operated flawlessly and as intended, ensuring the system's overall integrity and performance.

Deployment

During this phase, the user deploys the application via the web application. Summarized content will contain bulleted forms of important and highlighted information from the uploaded file. The user may wish to edit and alter the summarized output. After finishing, the user may access their previous works on the dashboard.

Review

Retrospective deliberations and sighting feedback from unit testing are conducted to identify areas for improvement and refine the development approach to incorporate in future iterations. Discussions allow the researchers to reflect and foster continuous enhancement of the summarizing tool's functionality and performance.

#### **CHAPTER IV**

**ABridge: Text-type Instant Notes Generator using Natural Language Processing Algorithm**

*System Overview*

ABridge is a text-type instant note-generating web application using natural language processing algorithms, providing students with more efficient ways of learning. The application allows students to convert their study materials into notes and are given the option to choose whether they want a quiz or flashcard as their note format. The user would then be allowed to save their uploaded material and edit the contents if they wish to do so in the dashboard.

The application will use Optical Character Recognition (OCR) to export the text material from the source. The material sources may differ from textbook images, Portable Document Format (PDF) or Word documents when importing to generate the notes the user wishes to create. The application features the summarization function for their uploaded material. Natural Language Processing (NLP) uses the Bag of Words paradigm to get accurate summarization results. It also features the generation of quizzes and flashcards, which the toolkit provided by Python will support.

The system is intended to be used by students on their web browsers, providing them with more options and easy access to review their lessons. The system uses SQLAlchemy, an open-source Python SQL toolkit, to store the data gathered from the application. As an Object Relational Mapper, it provides an efficient way to maintain or manage processed information that can also be utilized upon separating user data.

*System Objectives*

The Web Application is capable of the following:

* A page that will allow users to register before using the application.
* A page will allow users to import materials and convert them into notes, including flashcards or quizzes.
* A page allows users to edit their uploaded notes after conversion.
* A page that allows users to search for other materials created by other users.

System Functions

*Web Application*. The Python-based application serves as a summarization tool to assist students with academic performance. The application offers text-type summarization using the NLP algorithm and OCR technology, practice tests and flashcards that let users edit to their liking. The application includes a repository for past works a user can access afterwards.

*Login Page*. This module provides users access to the application. Users will input their email address and password. Users of the system are then welcomed to the home page.

*Register Page*. The registration process in this module requires users to provide their full name, email address, and password to log in to the application. To finish the registration process, users must complete the necessary information before hitting the register button.

*Home Page*. The home page serves as the main screen of the system. The home consists of different categories of basic functions the system offers, such as summarization, dashboard, and practice tests.

*Summarization Page*. In this page the user can choose whether to input a file, pdf or an image to scan and summarize.

*Upload button*. The upload button lets the user receive the desired summarised product output.

*Practice Test Page*. The practice test page allows users to create and edit quizzes and flashcards for better-enhanced studying.

*Quizzes Page*. The quizzes page allows users to create quizzes with options for multiple choices, true or false, and identification.

*Flashcards Page*. This page will allow the user to create flashcards to their liking.

*Dashboard Page*. This page contains the history log of the user's past works. The dashboard allows the user to view and access all created tests and flashcards and their summarized outputs.

*Edit Name*. This button will allow the user to edit their full name.

*Edit Picture*. This page allows users to edit their profile picture.

*Enter password*. On this page, users may confirm the changes to their information by entering their password.

*Save/Cancel button*. After editing their credentials, the users can decide whether to save it using the save and cancel option.

*Log-out.* The user can exit the application by using the logout feature.

Physical Environment

Hardware Specification

Table 2 contains the minimum hardware requirement for the server side.

Table 2. Hardware Requirement for Server Side

|  |  |
| --- | --- |
| Server Side | |
| CPU Type/Model | Intel Core i7-10510Y Processor |
| Input Devices | Mouse and Keyboard |
| Operating System | Windows 10 and 11 |
| Output Devices | Monitor and Printer |
| Storage Device | 4TB (SSD, Solid State Drive) |
| Memory | 8GB or Higher |

Table 3 contains the minimum hardware requirements for the client.

Table 3. Hardware Requirement Client Side

|  |  |
| --- | --- |
| Client Side | |
| CPU Type/Model | Intel Core i3-10510Y Processor |
| Input Devices | Mouse and Keyboard |
| Operating System | Windows 10 and 11 |
| Output Devices | Monitor and Printer |
| Storage Device | 1TB (SSD, Solid State Drive) |
| Memory | 8GB |

Software Specification

Table 4 contains the software tools used to develop the system.

Table 4. Software Specification

|  |  |
| --- | --- |
| Operating System | MacOS |
| Back-end | Python using Flask |
| Front-End | Html, Bootstrap, JavaScript |

*Back-end: Python*

Python is a dynamic programming language renowned for its versatility across various applications. Its readability, characterized by a clear and intuitive syntax, facilitates easier comprehension and maintenance of code. Moreover, Python's full modularity allows developers to organize and structure their programs efficiently, enhancing scalability and maintainability. Additionally, Python offers a rich ecosystem of libraries and extension capabilities, providing access to many tools and functionalities to streamline development processes further and expand the language's capabilities. (python.org).

Python uses an interpreter to execute the final code, like many other programming languages. Additionally, Python boasts scalability as a notable feature. Scalability means that a program built with Python can maintain its structure seamlessly, even as it transitions from a simple application to a comprehensive corporate program (Whitman, 2021).

Python is favored in web development due to its abundance of frameworks like Django and Flask, which streamline development. Its simplicity and readability reduce coding time and debugging efforts. Additionally, Python's extensive library and compatibility with other technologies make it an efficient choice for building robust web applications.

Flask, recognized as a lightweight micro-framework developed using Python, serves as a highly practical tool for crafting web applications. Its lightweight nature ensures minimal overhead, making it particularly suitable for projects with paramount efficiency and simplicity. Despite its simplicity, Flask offers a versatile set of features that can accommodate the needs of both small-scale and more complex applications. Its intuitive design allows developers to start building web applications quickly. At the same time, its robust scalability ensures it can effectively handle the demands of larger and more intricate projects as they grow (palletsprojects.com).

The researchers aimed to gain essential coding skills for machine learning and thus sought a programming language and framework with a manageable learning curve. Python was chosen for its wealth of available resources. Additionally, the research team selected the Flask framework because of its perceived ease of learning and usage and its lightweight nature compared to alternative frameworks.

*Front-End: HTML*

HTML (Hypertext Markup Language) is the foundational language for crafting most web pages and online applications. Within HTML, hypertext denotes text that directs users to other related pieces of content, while markup language encompasses symbols instructing web servers on document style and structure. Developers utilize HTML code to dictate the appearance and arrangement of web page components like text, links, and multimedia elements within browsers. (Sarosa, 2022).

HTML is essential for creating the user interface and displaying data analysis results. It enables access through web browsers, broadens the audience reach, and ensures a standardized format for presenting information, enhancing user readability and comprehension.

*Front- End: Bootstrap*

Bootstrap, developed by Mark Otto and Jacob Thornton at Twitter, stands out as the most popular framework for building responsive and mobile-friendly websites. This framework, comprising HTML, CSS, and JavaScript components, provides design templates for various web elements such as typography, forms, buttons, and navigation. Bootstrap was released as open source on GitHub in August 2011 and includes JavaScript plugins for creating responsive designs, thus simplifying and accelerating web development processes (What is Bootstrap - JavaTpoint).

Bootstrap is favored for website development because it's easy for anyone with basic HTML and CSS skills to use, enables the creation of responsive websites, and works smoothly on popular browsers like Chrome, Firefox, Internet Explorer, Safari, and Opera.

*Front-End: JavaScript*

JavaScript, an open-source scripting language, develops and controls dynamic web pages, allowing elements to move on the screen seamlessly without requiring a browser refresh. As the most extensively utilized programming language globally, JavaScript offers considerable flexibility (Sufiyan, 2021).

JavaScript proves to be an invaluable asset in constructing a web-based student performance prediction system. It generates dynamic and interactive data visualizations like charts and graphs. Additionally, JavaScript facilitates the integration of user interface functionalities such as pop-ups and form validation.

Cost of Development

Table 5 contains the cost associated with the hardware requirement.

Table 5. Hardware Development Cost

|  |  |
| --- | --- |
| Equipment | Amount |
| 1 System Unit | P 30,000.00 |
| 1 Monitor | P 4,000.00 |
| 1 Keyboard and Mouse | P 1,500.00 |
| **Total** | P 35,500.00 |

Table 6 contains the estimated cost of the software developed.

Table 6. Software Development Cost

|  |  |
| --- | --- |
| Researcher Role | Amount |
| Project Leader | P 35,000.00 |
| Programmer | P 30,000.00 |
| Researcher | P 15,000.00 |
| System Admin | P 25,000.00 |
| Analyst | P 20,000.00 |
| GUI Designer | P 20,000.00 |
| **Total** | P 145,000.00 |

System Depreciation

Hardware Depreciation

Table 7. presents the hardware depreciation of the developed study. Hardware Depreciation rate (Rate of Depreciation = Cost/Year)

Hardware Cost = Php 35,500.00

Depreciation Year = 5 Years

Rate of Depreciation = Php 7,100.00

Table 7. Hardware Depreciation

|  |  |
| --- | --- |
| Year | Amount |
| 0 | P 35,500.00 |
| 1 | P 28,400.00 |
| 2 | P 21,300.00 |
| 3 | P 14,200.00 |
| 4 | P 7,100.00 |
| 5 | P 0.00 |

Software Depreciation

Table 8. presents the software depreciation of the developed study. Software Depreciation rate (Rate of Depreciation = Cost/Year)

Software Cost = Php 145,000.00

Depreciation Year = 10 Years

Rate of Depreciation = Php 14,500.00

Table 8. Software Depreciation

|  |  |
| --- | --- |
| Year | Amount |
| 0 | P 145,000.00 |
| 1 | P 130,500.00 |
| 2 | P 116,000.00 |
| 3 | P 101,500.00 |
| 4 | P 87,000.00 |
| 5 | P 72,500.00 |
| 6 | P 58,000.00 |
| 7 | P 43,500.00 |
| 8 | P 29,000.00 |
| 9 | P 14,500.00 |
| 10 | P 0.00 |

*Tangible and Intangible Benefits*

Tangible Benefits

* Reduce the use of paper and writing materials.

Intangible Benefits

* Data Confidentiality.
* Enhanced Productivity
* Accessibility

*Architectural Design*

*Use Case Diagram of the Proposed System*

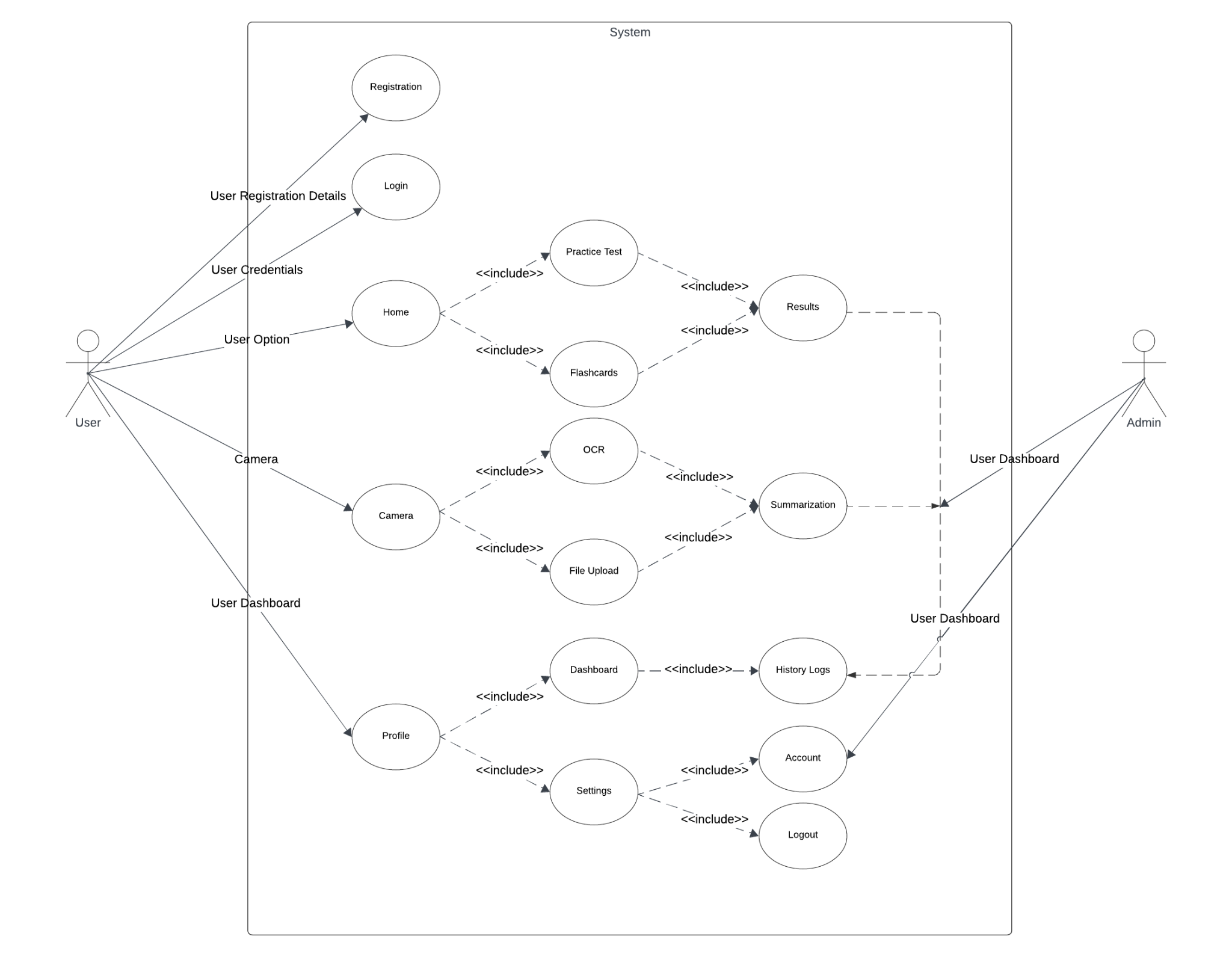


Figure 4. Use Case Diagram of the Study

Figure 2 shows the use case diagram of the proposed system for the users and the Admin. The users can access and interact with the systems while the admins use the application's back-end.

The users are required to input their information for them to get registered and have access to the system.

The administrators and programmers are privileged to access the server housing the accumulated data and the application's source code. With this access, they can modify and review the collected data comprehensively.

*Sequence diagram of the proposed system*

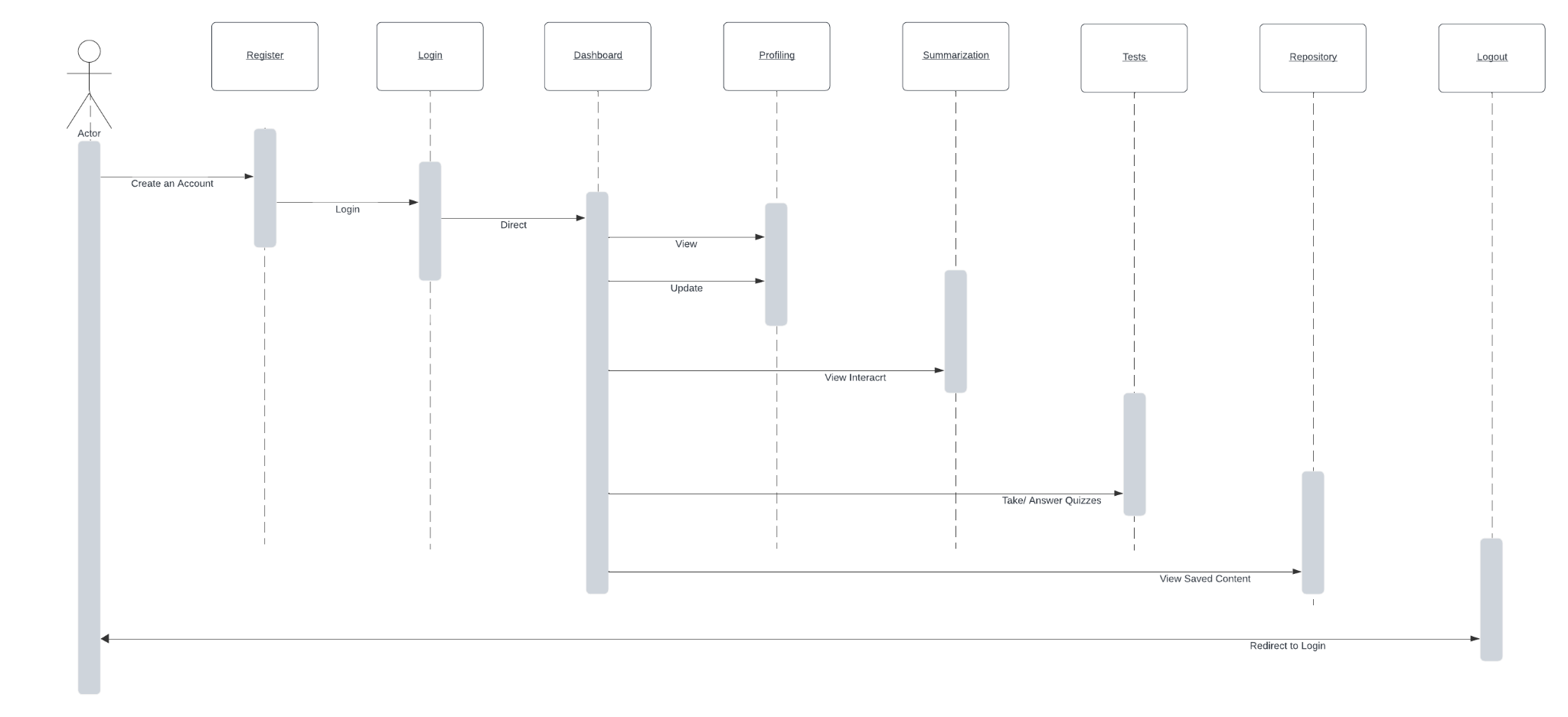


Figure 5. Sequence Diagram of the Study

Upon entering their login credentials, the user will be directed to the dashboard interface. Here, they will be given the choice to make summarized content via OCR or file upload or check the repository containing the history logs of their past works.

*Activity Diagram of the Proposed System*

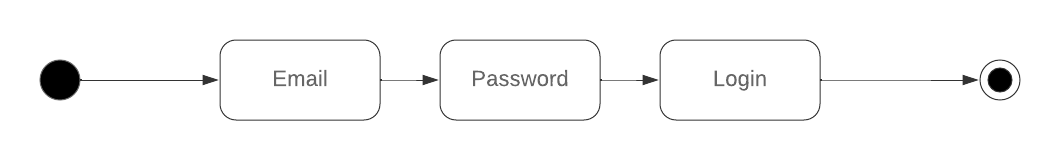


Figure 6

Figure 7 illustrates the procedure to access the system. It commences after putting in the login credentials. Information about users with log history from the past. A different procedure is shown for users who have not yet registered.

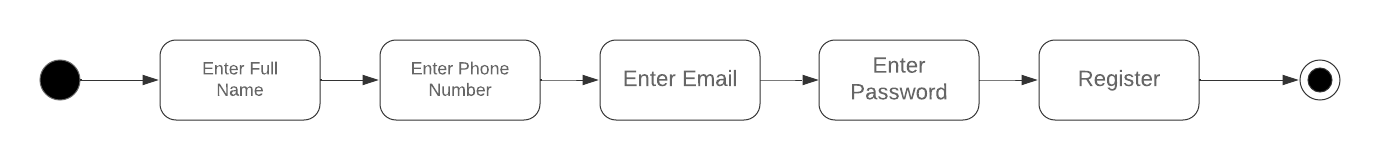


Figure 7

The system will ask the user to input their full name, email, and password. Once complete, the user can now have access to the application.

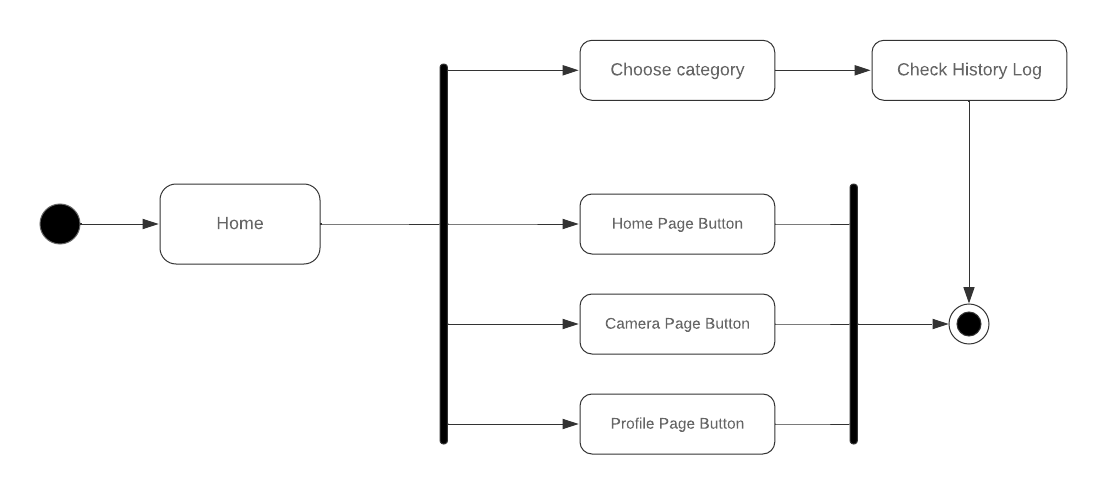


Figure 8

Figure 8 encompasses the application's main page containing the home page button, the camera page button, the profile page button, and the dashboard, which contains the history logs of past projects.

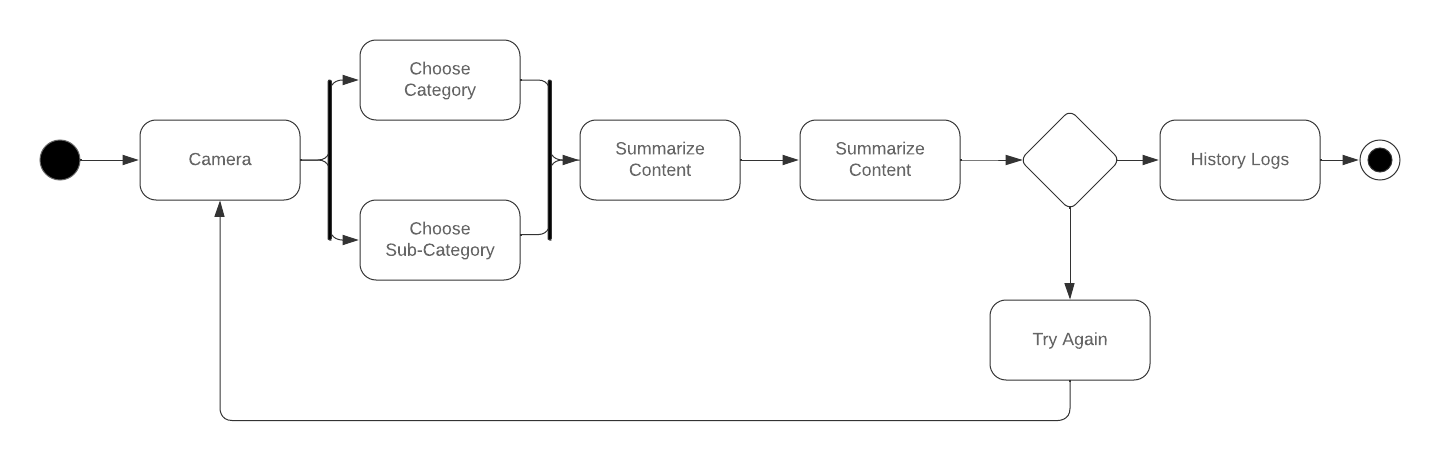


Figure 9

By accessing the camera, the user can either use OCR technology or upload their files to summarize the content. Failure to scan will result in trying again. Successful summarization will be stored in the repository and accessed via history logs.

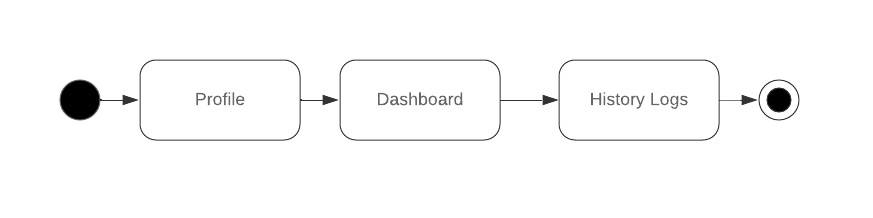


Figure 10

The dashboard is located by accessing the user profile, where all past contents are stored. Quizzes and tests the user makes are accessed here and previous summarized outputs.

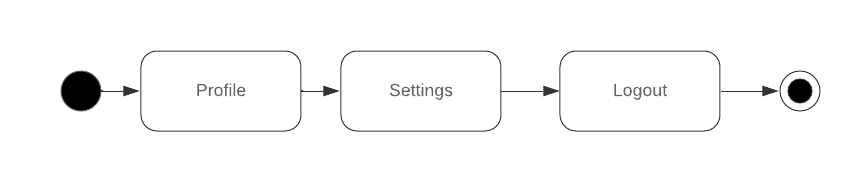


Figure 11

The Figure illustrates the process of logging out of the application.

*Database Structure*

Database Name: database.db

Table No: 1  
Foreign Key: User\_Name, Email

Description: This database includes the basic information of the user.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Key | Field Name | Data Type | Size | Description |
| FK | User\_Fullname | VARCHAR | 100 | Complete Name of User |
| FK | Email | VARCHAR | 100 | Email Address of User |

Database Name: database.db

Table No: 2

Foreign Key: Flashcards, Quizzes, Scores, Date

Description: This database contains all the tests the user has created.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Key | Field Name | Data Type | Size | Description |
| FK | Flashcards | VARCHAR | 100 | Flashcards |
| FK | Quizzes | VARCHAR | 100 | Quizzes |
| FK | Scores | INT | 11 | Scores |

Database Name: database.db

Table No: 3

Foreign Key: Summary, Date

Description: This database contains all the summarized contents the user had created.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Key | Field Name | Data Type | Size | Description |
| FK | Summary | VARCHAR | 100 | Summarized File |
| FK | Date | Date | 20 | Date Created |

Database Name: database.db

Table No: 4

Foreign Key: User\_Name, User\_Email, Password

Description: This database contains the signup details of the user.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Key | Field Name | Data Type | Size | Description |
| FK | User\_Fullname | VARCHAR | 100 | Complete Name of User |
| FK | User\_Email | VARCHAR | 100 | Email Address of User |
| FK | Password | VARCHAR | 100 | Password |

*Network Topology*

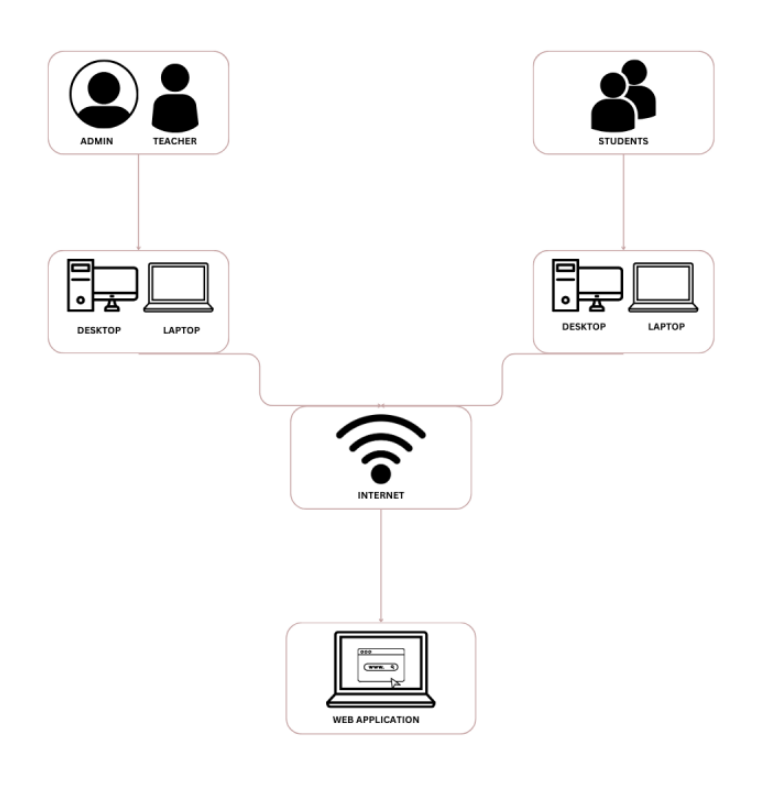


Figure 12 Network Topology

Figure 12 depicts the planned network layout for the system. The system can be accessed via a web application by administrators, teachers, and students using desktops and laptops.

Design and Implementation Issues

ABridge: Text-type instant notes generator using a natural language processing algorithm is a web-based application comprising various modules aimed at generating instant study notes in text format tailored to the elementary science curriculum of Central Philippines University.

The instant notes generator will benefit the students by lessening the time of making notes of the user's lesson with the help of NLP and OCR which will be the algorithm of the web application using Python programming language.

The NLP is for text classification and summarization. It utilizes extractive and abstractive approaches to implement strategies such as text classification for material organization and text summarization for rapid reading of lengthy texts. Also, combining document embedding with the Bag of Words model simplifies categorization tasks. Using lemmatization and stemming techniques reduces word variability, increasing the efficiency and consistency of NLP algorithms and ensuring the accuracy of system outcomes.

The Tesseract OCR will be responsible for text recognition and document scanning. Text detection and character recognition separate and analyze text segments, while image preprocessing techniques improve OCR accuracy. Language modeling improves recognition accuracy by combining statistical language models and dictionaries. Post-processing procedures improve the identified text output to ensure readability.

The system features a quiz module that the user can create and quizzes with a given option of what type of test the user wants. Also, one feature is a flashcard module that will let the user make flashcards to their liking, which will help them study.

One of the issues observed during the implementation process was that the module intended for generating notes was not generating a standard of 95% accuracy. To resolve the issue, the team examined and fixed the problem with the help of resizing the image to convert it to grayscale format to remove the noise pixels and make it clearer. Also, the first and most important step in improving the accuracy of the NLP model is to use high-quality datasets. High-quality data is distinguished by relevance, completeness, consistency, and dependability. It is critical to collect enough data that covers the domain, language, and style relevant to the task. In addition, clean, preprocess, and normalize the data to remove noise, errors, and inconsistencies. Furthermore, the data should be labelled, annotated, and divided into training, validation, and test sets following the standards and guidelines for the specific task.