Chapter I

INTRODUCTION

This chapter introduces the research topic and contains the purpose and description, project objectives, and its scope and limitation.

Project Context

This system is both Web-Based and Mobile Application and act as a repository for the research materials of Bulacan State University. It can check plagiarism content of the Materials (to the internet and to other research materials) and so as the Grammar.

Purpose and Description

This project provide a repository for the research materials of Bulacan State University. Also, to create a web-based system that can check if a research is plagiarized and also to have a grammar checking capability.

This system serve as an upgrade to an already made system the Bulacan State University uses. As this project have a faster plagiarism checking using a different algorithm, an android application version, grammar and spelling checker which the current system does not have.

Algorithms for plagiarism checker

Sherlock Algorithm

Sherlock Algorithm is the algorithm used in the existing system. Sherlock algorithm is an algorithm that checks for plagiarism, based on the similarity of sentences between documents. This algorithm uses the Rabin-Karp algorithm in a modified way. Sentences can have different sets of keywords means that the two sentences are not very similar and have different content. If the sentences being compared have the same set of keywords then the two sentences have the same content.

The first step of the algorithms calls for the documents to be processed by filtering sentences for stop words, followed by deleting duplicate words from each sentence , getting rid of punctuation and converting all letters to upper-case.

Next sentences are compared between each documents using the following scheme provided by,

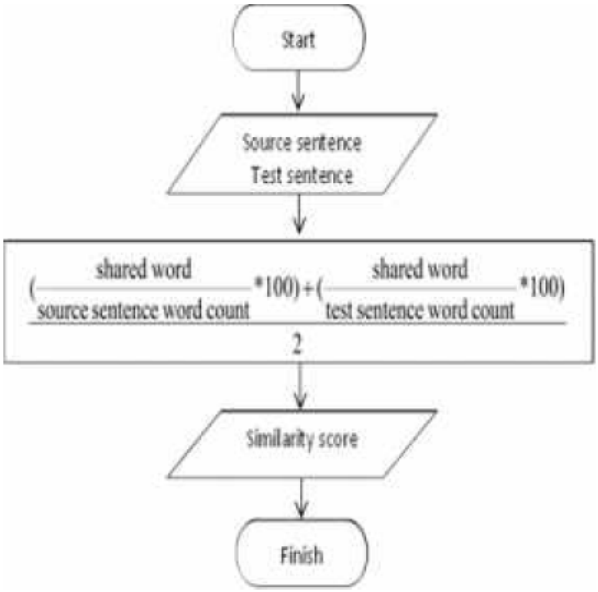


Figure 1 Sherlock Algorithm Flowchart

After calculating the similarity score, we can determine if documents have been plagiarized by comparing those values to the set thresholds. Similarity scores that resulted in a value less than 30% means that the suspicious documents being compared have very little plagiarism. Similarity scores that resulted in a value between 30% and 70% means that the suspicious documents have some moderate plagiarism. Finally, similarity scores that resulted in a value greater than 70% means that the suspicious documents have heavy plagiarism or are clearly plagiarized. Collectively the total number of sentences would be divided over the total number of sentences within a given document to return a similarity score for the overall document. The Rabin Karp Algorithm and Knuth Morris Pratt Algorithm were used to make string matches.

Rabin-Karp algorithm

The algorithm of our system uses Rabin-Karp algorithm. The Rabin-Karp algorithm was used in the Sherlock algorithm stated previously. This algorithm is a string matching algorithm which takes advantage of a hashing algorithm during the pre-processing phase.

The Rabin Karp matcher transforms the pattern P from the text T into a numeric value. The hashing algorithm multiplies each letter's index value in the alphabet t[i] by the index of that letter in the pattern starting from 1 then that value is multiplied by some large base value b which is 105 in our testing. This value can get very large so then we mod the value by some relatively large prime number. The matcher checks the pattern value to the values of the source text in string lengths of the pattern. If the pattern value does not match the source text string then it moves forward and checks again. If there is a match between the two values then the matcher checks to see if the strings exactly match. This is required because the pattern value and the source text may not be the same string but have the same hash value. The pre-processing phase is complexity O(m) and the matching is O(n) where the overall complexity is O(n+m) where the worst-case scenario is O(n-m+1).

KMP

Knuth Morris Pratt's algorithm for string matching was used for the second scheme to detect plagiarism. This method was developed by Donald Knuth and Vaughn Pratt who worked together. It was also developed independently by James H. Morris in 1977. This algorithm is known for its linear time exact matching. The algorithm compares the text from left to right and is able to shift more than one position. The algorithm is very clever in a sense that it is able to avoid trivial comparisons due to the pre-processing phase.

KMP works by pre-processing the pattern to be searched in a document. This process involves finding the largest prefix, P[0..j-1], that matches the largest suffix, P[1..j]. This value represents the number of shifts to be done when a mismatch, match, or partial match occurs. It indicates how much of the last comparison can be reused if the algorithm fails to find a match. Below shows an example of how the pre-process phase works by creating a table for the pattern "abacab". It is highly efficient because the number of comparisons for the pattern against the original text is minimized. In all, the algorithm has a complexity of O(n+m) where O(m) is the computation for the prefix function values and O(n) is the comparison of the pattern to the text. KMP is very advantageous because it never has to backtrack on a string which makes it efficient for large text files.

Jaccard Similarity

For the third scheme to detect plagiarism, Jaccard distance formula was used. In processing each individual document, there are a factors of thing to consider, such as white spaces, capitalization, punctuation, characters versus words, how large are the single, count replicate, and stop words. The first step of the process we decided to use was to process each sentence by changing all letter to lowercase, remove stop words, duplicates and periods. Next step was to generate a set of shingle. This takes consecutive words and group them as a single object. A k-shingle is a consecutive set of k words. A sample sentence might be "math is fun." A 1-shingle would be the set containing {[math], [is], [fun]}. While a 2 shingle would be the set containing {[math, is], [is, fun]}. We used a shingle size of 2 to test for plagiarism.

The Jaccard Similarity formula is defined as the intercept of document A and B divided by the intercept of A and B. For the purpose of comparing document the similarity score would be computed by comparing the singles generated by one document to another documents and the amount of shingle that exist in both document would be divided by the number of unique shingle generated total.

Difference of all algorithm(speed and accuracy)

The results of our testing showed interesting a promising results. Below shows the results of our testing using the three schemes as stated.

Sherlock/Rabin-Karp

Detection

<30% - little plagiarism.

55.8381503%

30% - 70% - moderate plagiarism.

19.9421965%

>70% - heavy plagiarism.

23.4682081%

Average Run Time

5.203879203 seconds

KMP

Detection

<30% - little plagiarism.

87.6230661%

30% - 70% - moderate plagiarism

08.7201125%

>70% - heavy plagiarism.

03.4810127%

Average Run time

1.378091645 seconds

Jaccard Similarity

Detection

<30% - little plagiarism.

99.1648244%

30% - 70% - moderate plagiarism.

00.5404078%

>70% - heavy plagiarism.

00.2947679%

Average Run time

0.032808049 seconds

The data above shows the level of detection of plagiarism of the 5000 comparisons made and the average run-time for each comparison (Not the run-time for all 5000 at once). Although the Sherlock and Rabin-Karp algorithm have the same speed and accuracy we often used them in the comparison as one combined algorithm that will compared to other algorithm available. What is interesting is that the Sherlock/Rabin-Karp matcher took the longest, but yielded

more detections in plagiarism where it found that 23.5% of the comparisons had heavy plagiarism. The KMP matcher ran fairly quickly with an average run-time of 1.38 seconds per comparison, but yielded poor results, 3.5% of heavy plagiarism detected, compared to the Sherlock/Rabin matcher. Finally, the final test was the Jaccard Similarity matcher. This resulted in the fastest run-time of the three algorithms with an average run-time of 0.032 seconds per comparison. Although the matcher was quick on comparison and detection, the detection efficiency failed miserably compared to Sherlock/Rabin and even compared to the KMP matcher. This matcher resulted in only 00.3% heavy plagiarism detected.

Results from our testing was very interesting. Some values we did not expect to be so different from what we initially hypothesized. The KMP matching heavy plagiarism value we did not expect to be so low compared to the Sherlock/Rabin matcher. We also did not expect the Sherlock/Rabin to be almost 5x the run-time from KMP. Some of these values may have been caused due to implementation or formatting. The Jaccard Similarity being so fast but so poor in detection was almost expected only using 2-shingles. In future testing we will be more cautious about formatting the text documents as well as try to improve on the Jaccard Similarity as the results may be very different if we were to use 3-Shingles.

Project Objectives

In this section, general objective be discussed as well as the specific objective of this research.

General Objective

To create a plagiarism and grammar checker web site for Bulacan State University.

Specific Objectives

1. To create an online public access catalog for the Bulacan State University.
2. To create a repository for the Bulacan State University research materials.
3. To enhance the plagiarism checker and to add a grammar checker web site for Bulacan State University.
4. To create a mobile version of the web site for offline browsing on the research materials.
5. To develop the system using the Extreme Programming Methodology
6. To implement the system using the user acceptance testing.

Scope and Delimitations

The scope of the system is to easily detect the plagiarism content of the research materials, provided the title of the research that has been plagiarize and even the web site that has been plagiarize. It can provide a graphical representation to easily determine the percent of the plagiarize content. The system has 3 levels of access: (1)Administrator level can Manage Account, Manage Colleges, Manage Published Journal, Manage Researches, use plagiarism checker, use image to text converter, view audit trail of the user, can generate reports. (2)Librarian level can Manage College, Manage Researches, view audit trail to monitor the changes made in his assigned College, can use image to text converter, can generate reports. (3)Researcher level can browse,view and download research materials abstract, use plagiarism checker(Abstract Only), can generate reports. All levels can use grammar checker.

It has a mobile application version of the system to browse all research materials abstract without internet.

Limitations

The mobile application could be spacious or heavy weight. The offline part of the mobile application is only for the browsing of the research material abstracts. The grammar checking is not 100% accurate. The system is for the Bulacan State University libraries and university research office only. Other campus is not covered in the system.

Chapter II

REVIEW OF RELATED LITERATURE

This part of the study presents the information, principles, and studies needed to conceptualize and to develop the proposed project. It includes review of related literature and related studies that will aid the researchers in conceptualization of ideas in guiding them in pursuing and in the development of the research and proposed system.

Related Literatures

Plagiarism

Plagiarism is the "wrongful appropriation" and "stealing and publication" of another author's "language, thoughts, ideas, or expressions" and the representation of them as one's own original work.

Plagiarism is considered academic dishonesty and a breach of journalistic ethics. It is subject to sanctions like penalties, suspension, and even expulsion. Recently, cases of "extreme plagiarism" have been identified in academia. The modern concept of plagiarism as immoral and originality as an ideal emerged in Europe in the 18th century, particularly with the Romantic movement.

Plagiarism is not in itself a crime, but can constitute copyright infringement. In academia and industry, it is a serious ethical offense. Plagiarism and copyright infringement overlap to a considerable extent, but they are not equivalent concepts, and many types of plagiarism do not constitute copyright infringement, which is defined by copyright law and may be adjudicated by courts. Plagiarism is not defined or punished by law, but rather by institutions (including professional associations, educational institutions, and commercial entities, such as publishing companies).

Plagiarism detection

Plagiarism detection is the process of locating instances of plagiarism within a work or document. The widespread use of computers and the advent of the Internet has made it easier to plagiarize the work of others. Most cases of plagiarism are found in academia, where documents are typically essays or reports. However, plagiarism can be found in virtually any field, including novels, scientific papers, art designs, and source code.

Detection of plagiarism can be either manual or software-assisted. Manual detection requires substantial effort and excellent memory, and is impractical in cases where too many documents must be compared, or original documents are not available for comparison. Software-assisted detection allows vast collections of documents to be compared to each other, making successful detection much more likely.

Algorithms used for string matching

There are some algorithms used for string matching, one of them is Aho Corasick algorithm, it is a kind of dictionary-matching algorithm that locates elements of a finite set of strings (the "dictionary") within an input text. It matches all strings simultaneously. The complexity of the algorithm is linear in the length of the strings plus the length of the searched text plus the number of output matches.

English grammar

English grammar is the way in which meanings are encoded into wordings in the English language. This includes the structure of words, phrases, clauses, and sentences, right up to the structure of whole texts.

There are historical, social, cultural and regional variations of English. Divergences from the grammar described here occur in some dialects of English. This article describes a generalized present-day Standard English, the form of speech and writing found in types of public discourse including broadcasting, education, entertainment, government, and news including both formal and informal speech. There are differences in grammar between the standard forms of

British, American, and Australian English, although these are minor compared with the differences in vocabulary and pronunciation.

Modern English has largely abandoned the inflectional case system of Indo-European in favor of analytic constructions. The personal pronouns of Modern English retain morphological case more strongly than any other word class (a remnant of the more extensive case system of Old English). For other pronouns, and all nouns, adjectives, and articles, grammatical function is indicated only by word order, by prepositions, and by the "Saxon genitive" (-'s).

Eight "word classes" or "parts of speech" are commonly distinguished in English: nouns, determiners, pronouns, verbs, adjectives, adverbs, prepositions, and conjunctions. Nouns form the largest English word class, with verbs being the second largest word class. Unlike many Indo-European languages, English nouns do not have grammatical gender (although many nouns refer specifically to male or female persons or animals).

Grammar Checker

A grammar checker, in computing terms, is a program, or part of a program, that attempts to verify written text for grammatical correctness. Grammar checkers are most often implemented as a feature of a larger program, such as a word processor, but are also available as a stand-alone application that can be activated from within programs that work with editable text.

The implementation of a grammar checker makes use of natural language processing.

Natural-language processing

Natural-language processing (NLP) is an area of computer science and artificial intelligence concerned with the interactions between computers and human (natural) languages, in particular how to program computers to fruitfully process large amounts of natural language data.

Related Systems

Extrinsic monolingual plagiarism checker for Bulacan State University Research materials using Sherlock Algorithm

The developers developed a technology solution that will lessen, if not eliminate, the plagiarism in any types of research papers inside Bulacan State University. This project focused on the development of Extrinsic Monolingual Plagiarism Checker for Bulacan State University Research Materials using Sherlock Algorithm that will benefit the University Research Office, University Librarians, and Researchers inside and outside the University. The developers gathered all essential research materials from different university libraries that will be used to examine text similarity of a document with other documents in the

database and justify those document comparisons with Sherlock Algorithm. Aside from plagiarism checker, the project will serve as a management system to both University Research Office and to the University librarians. Furthermore, it will also be an Online Portal which will showcase different research materials of the University. The developers have adopted the Scrum for the project. This method is an iterative and incremental Agile software development framework for managing software projects and product or application development of the website to be developed. The developers used different development tools such as: PHP, Codelgniter, JavaScript, AngularJS, and MySQL as its database management system. The system has the following significant features: user account management, research materials management, plagiarism checker, image to text converter, report generation, and online portal. The Evaluation Criteria for the project was the standardized instrument adopted by the College of Information and Communications Technology (CICT). The system has been evaluated by University Research Office's personnel and IT experts to determine its level of acceptability. The general weighted mean are as follows: Authority acquired 4.41; Purpose is 4.71; Coverage acquired 4.66; Currency is 4.75; Accuracy got 4.81; Functionality also got 4.60; Training and Documentation is 4.39. The overall acceptability of the system got a total weighted mean of 4.62 and found "Excellent" based on the pooled judgment of the evaluators and on the equivalent ratings presented in the Likert Scale.

This web site is an online public access catalog and a repository of the Bulacan State University research materials, it is capable of checking plagiarism and use image to text converter

Dupli Checker

It is is a free online plagiarism tool. You can use it to check content on your site against existing content. In fact, it will check the content you provide against every published page on the web.

Tools are necessary because it is not possible to check for plagiarism manually. You would have to review millions of website and billions of words of content to do it.

Using Duplichecker can help you be confident that any content you pay for is original. You will not have to worry about being accused of theft. You can even make payment for content contingent upon it passing a Duplichecker test.

It is simple to use. You can use it online or download it onto your computer or mobile device.

On the website, you have the option of checking your text one of two ways.

1. Upload a file in .doc or .txt format to Duplichecker.com
2. Paste up to 1000 words into the search box.

In either case, it will check your source content against all other online and return a result. The report you get will tell you if your work has been copied or duplicated online. It will also provide you with the website the material came from.

You can use the text box without signing up for an account. If you want to be able to verify longer documents (up to 1500 words), you will need to register. The tool has no charge to registered users, too.

With the free membership, you can scan up to 50 documents a day. For the owner of a website, it is important to have content that is plagiarism free. With the help of DupliChecker.com, you can check the originality of the content that you are going to use on your site. It will prevent any legal issues from arising, and increase your credibility as a business or website owner.

COPYLEAKS Plagiarism Checker

This works fully on the cloud, so the cloud-based functionality has given this tool much more power than you can imagine.

The tool has different options for education and business, so you can choose preferred one at any time. The education version has also different sections such as for schools, universities, and students. Business version also

offers two different options – one for the publishers and the other one for the SEO agencies. One cool thing about the tool is that it can be used on mobile platforms too. So, if you want to use it directly on your phone, you can do so just by installing its mobile app.

Detect plagiarism and paraphrased content using advanced AI technology. Confirm originality with sophisticated algorithms that scan and track textual content in every language.

Paper Rater

PaperRater.com is a free resource that utilizes Artificial Intelligence to help students write better. Our Paper Checker technology combines Natural Language Processing, Machine Learning, Information Retrieval, Computational Linguistics, and Data Mining to produce the most powerful automated proofreading tool available on the Internet today. PaperRater.com is used by schools and universities in over 100 countries to help students improve their writing and check for plagiarism. It has been our mission to offer tools that are both powerful and accessible to users regardless of income.

As part of the development process, we put together a team of computational linguists and subject matter experts to develop a core Natural Language Processing (NLP) engine using statistical and rules-based NLP to extract language features from essays and robustly translate that into statistical

models. The end product is a state-of-the-art system combining Automated Grammatical Error Detection, Automated Essay Scoring, Automated Proofreading, and plagiarism detection.

Plagiarisma

Plagiarisma.net is one of the free online plagiarism checkers. Their system allows you to check for plagiarism through your browser and they also offer you the option to download some software as well. Not to mention the fact that their current software supports 190 plus languages. So, users all over the world can take advantage of this plagiarism software.

Plagiarisma.net is one of the free online plagiarism checkers that deserves a five star rating. The rating comes into play because they have multiple ways that you can check for plagiarism and they best part is they are all free. Not to mention the fact that plagiarisma.net reviews have said nothing bad about this plagiarism checker in the nature that would lead one to believe that they are a plagiarisma.net fraud.

This is one of the free online plagiarism checkers that makes sure that their customers know exactly what plagiarism is as well as understand why they need to avoid it. Some plagiarisma.net reviews have compared this plagiarism checker with Turnitin.com – which is a plagiarism checker that is used by people in the academic world.

Plagiarism Checker

PlagiarismChecker.com allows you to search for several phrases from a student's paper at the same time without having to type quotation marks or special operators, which most search engines require if you're trying to look for exact copies of a student's writing. This site automatically adds the quotation marks and special operators for you.

In PlagiarismChecker.net, the scanner works by taking the first 18 words of any submission and running them through a search engine – by default this is Google, but Yahoo may also be selected. If the document has been plagiarised, and if the site the words have been plagiarised from has been indexed by Google, this will show up in the results. The results are listed as Google search results (if this has been selected) and highlight in bold any matching entries. There are a number of drawbacks with this online scanner. Firstly, as only the first 18 words of each submission are checked (because this is the maximum number of words which can be searched for in Google in one search) it would be possible to beat the scanner by taking an entirely plagiarised work and ensuring the first 18 words are original. As a result, multiple scans would be required to detect plagiarism, although there is no limit on the number of searches which may be carried out. Secondly, the purpose of the scanner is questionable as it relies entirely on search engines alone to detect plagiarism – copying and pasting text into the scanner is therefore the same as copying and pasting text directly into a search engine. The

‘Report Plagiarism’ link is helpful in outlining how to contact the relevant search engines to draw their attention to plagiarised work and also offers users the option of reporting plagiarism on their behalf.

PlagiarismChecker.com is free and works on any papers, whether students e-mail them to you or turn in hard copies.

Plagium

Plagium is one of the online plagiarism checker services that is available to individuals who need to have plagiarism checks performed via the web to insure that their work is 100% original. From the looks of their copyright it seems as though the company has been in business since at least 2009. Not to mention the fact that their plagiarism detection can be used as both a free and a paid plagiarism checker. However, they do have low prices in place for those who need more than what their free plagiarism check provides and they also have a button that invites those who wish to do so to donate money to their website.

This website for plagiarism checking deserves a 4 star rating. This rating comes into effect because some of the things that are mentioned on their website are a little confusing especially when it comes to their rates. Not to mention the fact that no one has left any Plagium.com reviews to even offer any type of explanation as to what to expect. However, obviously, they have been in business

for a little while and have avoided being listed as a scam/fraud/rip off so they must be doing something right.

Plagium is a service of Septet Systems Inc. Septet Systems Inc. focuses on the development of innovative information search solutions for consumers, enterprises, government, academics, and healthcare. Much of our work is based on Septet's proprietary TX Miner engine, which employs advanced search technology for deep mining of documents on the public World Wide Web or within private repositories.

PlagScan

PlagScan is a plagiarism checker that certainly has its share of fans. In Dr. Weber-Wulff's most recent round of plagiarism tests, PlagScan was listed as "partially useful", the highest honor those tests awarded. Overall, PlagScan placed fourth.

That placed it well above better-known services, including both Copyscape and Plagium, both of which are more widely used by webmasters wanting to track their content.

Also, PlagScan has earned the trust of dozens of academic institutions and businesses, most of which are in the company's native Germany.

PlagScan originally saw life as SeeSources, a free plagiarism detection tool that focused on language patterns in plagiarism detection. However, to continue development, the people behind it took the product commercial and created PlagScan, which operates now as a professional and academic plagiarism detection service.

Plagscan works on the Yahoo!BoSS-API, the same API many similar services are built upon, and works by either uploading a document or by copying and pasting the text you want to check. PlagScan then analyzes the text involved and then returns the results in an email that includes PDF, plain text and docx formatted results. The results are also stored in your account.

Stealing intellectual property is considered theft just as much as stealing physical property. Often, people aren't aware of the fact that ideas and words can be owned by someone, which causes both intentional and unintentional plagiarism. The copyright owner deserves protection.

At PlagScan, we value copyright ownership and try to protect intellectual property to the best of our ability.

PlagTracker

PlagTracker is a Ukrainian-based online plagiarism detection service that checks whether similar text content appears elsewhere on the web. It was launched in 2011 by Devellar.

PlagTracker is used by content owners (students, teachers, bloggers, researchers) to detect cases of "content theft", in which content is copied, without the permission of the author or owner, from one site to another. Many content publishers also use it to detect cases of content fraud, in which old content is repackaged and sold as new original content.

Initially the URL or text of the original content is transferred on the webpage; PlagTracker returns a list of web pages that contain similar text to all or parts of this content.[6] The matching text is highlighted on the found web page.

For using, the program may take more time to develop a report based on the length of the paper and the amount of plagiarized content. Once the report appears, any potentially plagiarized portions of the paper are highlighted in red to show the plagiarized content. Clicking on the highlighted portions of the paper will reveal a list of the site or sites the same content may have come from. PlagTracker is used in abundance in United States, Finland, United Kingdom and Asian countries such as Bangladesh, Indonesia, Philippines, India, Pakistan, and others.

PlagTracker also provides a premium service which allows the user to upload documents instead of copying and pasting text. For premium users, a stronger plagiarism checking and professional editing assistance is available. The premium plan is faster than the free version. Users can also perform grammar checks and can download and upload pdf files.

Quetext

Quetext is a leading plagiarism detection software and citation assistant that combines DeepSearch™ technology with clear feedback to detect duplicate content and prevent plagiarism.

With over 1 million users worldwide, Quetext has been helping teachers, students, bloggers and professional writers to improve their writing.

We've listed Quetext in our ever-updated list of Top 20 best plagiarism checker tools because it offers a unique approach to detect copied content based on its proprietary DeepSearch™ technology.

DeepSearch™ allows it to check your writing for traces of contextual plagiarism and not simple word or phrase matches.

Now that you've a clear idea of what Quetext is and what it offers, let's head towards our detailed Quetext review section where we'll access all of its features, what it offers, its pricing model and if Quetext is safe or legit to use.

Our goal at Quetext is to empower writers with a sense of security in their work. We are technology enthusiasts, and passionate about pushing the limits in new and innovative ways. We have a strong belief that our users should maintain all of the rights to their work, with no exceptions. So when you use Quetext, rest assured knowing your work will never end up in the hands of someone else.

Viper Plagiarism Checker

Viper is a type of anti-plagiarism software that makes it easy for teachers as well as other users to check text for plagiarism. The software scans documents that are uploaded into the system and notifies users if they detect sentences, paragraphs or entire documents that are present anywhere on the internet.

One of the great things about Viper is that it works very quickly and efficient. Users are provided with a login name and password that they use to gain access to the Viper system. The software only takes up around one megabyte of space and documents that uploaded can be matched against local databases as well as the Web. Results of the plagiarism check are emailed to users and any parts of the text that appears in other documents will be highlighted.

Teachers who are tasked with regularly checking students? Assignments should make sure that they check out Viper. The fact that this software is completely free to use means that there is nothing to lose by taking it for a spin and this is a great way to check homework assignments and a wide range of other types of documents to make sure that they are completely original. However, this software is only compatible with older versions of Windows and those who have later versions of Windows on their computer will have to search for another solution.

Grammarly

Grammarly is a nice tool that can help you avoid typos when working online. It's not quite good enough to keep those who aren't confident in their writing abilities from making big mistakes in their content. It also adds a few extra steps to your writing process, and there is a slight delay while the system checks your work.

As a professional writer, I found the more advanced checker useful, especially since I work on several different projects at once and generate thousands of words each week. I liked having another set of eyes on my work, and using the program gave me extra confidence before I sent my writing to my editor.

While it's nice not to miss a comma in an email or while posting on social media, I'm not sure the program is worth its hefty price tag month-to-month, especially if you're a novice looking to learn more about writing.

Chapter III

TECHNICAL BACKGROUND

This chapter presents the research methods to be used, the methods of collecting data, the process of developing the system, the selection of the sample or respondents, the research instrument, and the process of analysing the collected data with regards to the development of the application.

Conceptual Framework

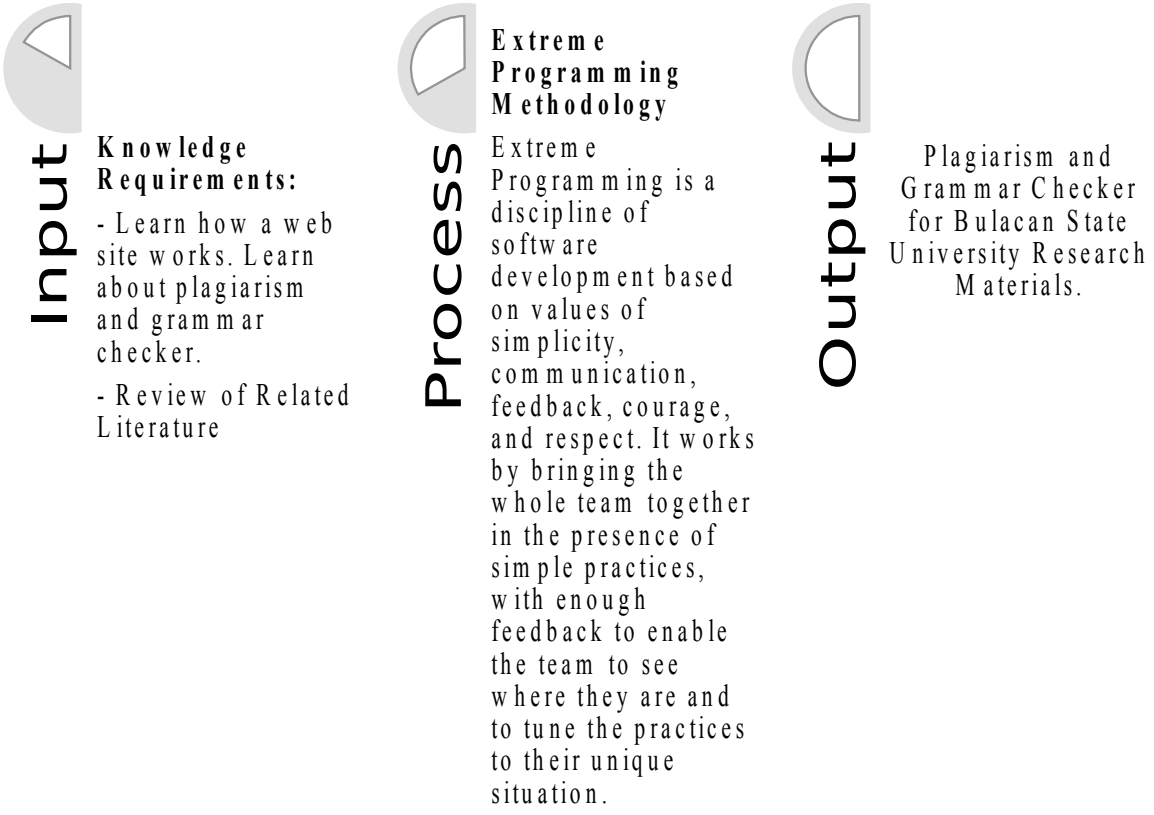


Figure 1.0 IPO Diagram

This figure 1.0 shows the input, process and output that represents the effect of the development to be conduct until the project is concluded.

The first input are the proponents to make the project successful. Second process is the procedure that will perform to meet the desired output. Lastly the output is Plagiarism and Grammar Checker for Bulacan State University Research Materials.

Development Methodology

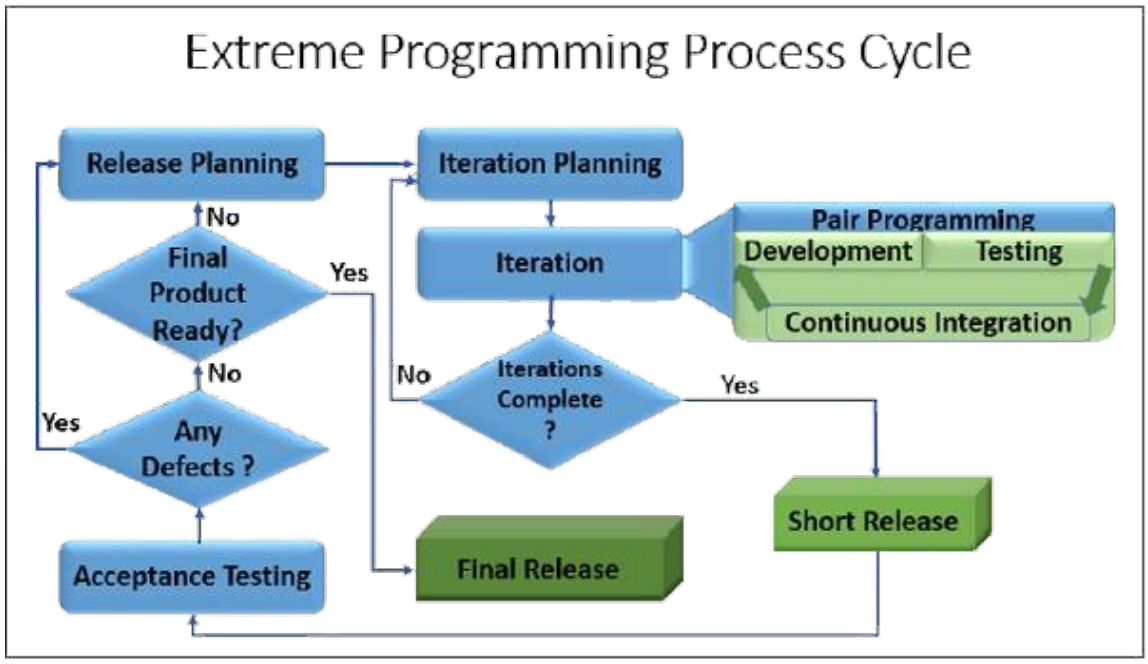


Figure 2.0 Extreme Programming Methodology

In the initial stage the customer formulates his requirements to the final product. He must gather the user stories and prioritize them to know exactly what he needs. After

that, he must set the term of project performance. Usually, all XP projects have quite short terms. That is one of the reasons for the name of this software development method.

The second stage is the stage of planning. The process of planning in Extreme Programming projects is usually conducted during a special Planning Game. It is actually a meeting where the customer announces all his requirements to the final product and prioritizes them. After that, the XP team estimates them in accordance with the time and cost criteria.

The third stage is the stage of performing. After the plan of the project is ready, the team starts to perform it. Usually, teams of Extreme Programming developers work in short iterative cycles. Each cycle is dedicated to a certain part of the final product. Such distances may last from 1 to 2 weeks. The intermediate result is provided to the customer after each cycle. The plan of the cycle may be changed according to the customer's wishes.

After all the cycles are finished and all the parts of the product are tested, it is provided to the customer. If he is satisfied, the XP project comes to its end.

Requirement Analysis

In this system the developer used JavaScript/NodeJS for the main tool of developing the web site, ExpressJS is a NodeJS web framework used to fasten the development phase of the web site.

Technical Requirement

Lists of Software Used

Software	Uses
JavaScript/NodeJS	Is a web programming language, main tool on developing the web site.
ExpressJS	Is a NodeJS framework.
ReactJS	Is a JavaScript library.
ReactNative	Is a JavaScript framework used to develop mobile apps

Table 1.0

Table 1 display the lists of software and Hardware used for the development of the application software.

Recommended Hardware and Software Requirements for desktop and mobile device

Category	Desktop Specification Hardware Configuration
Memory	4 GB RAM
Storage	Atleast 5 GB available space storage
Processor Speed	Quad core
Operating system	Windows 7 or newer version of windows

Table 2.0

Category	Mobile Specification Hardware Configuration
Memory	4 GB RAM
Storage	Atleast 5 GB available space storage
Processor Speed	Quad core
Android Version	Lolipop up to latest

Table 2.1

Table 2 shows both software and hardware specifications required for the user to operate the application.

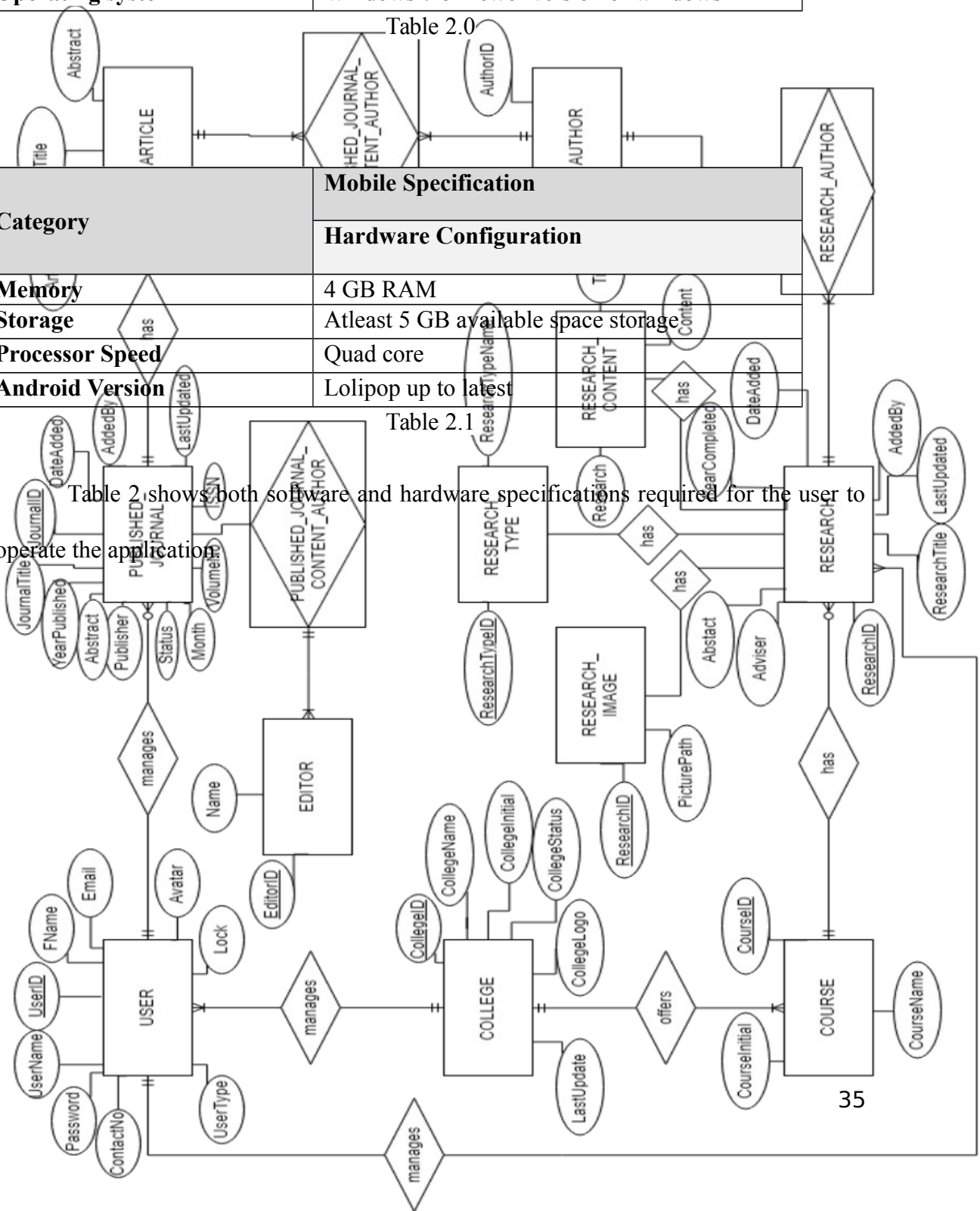


Figure 3.0 Entity-Relationship Diagram

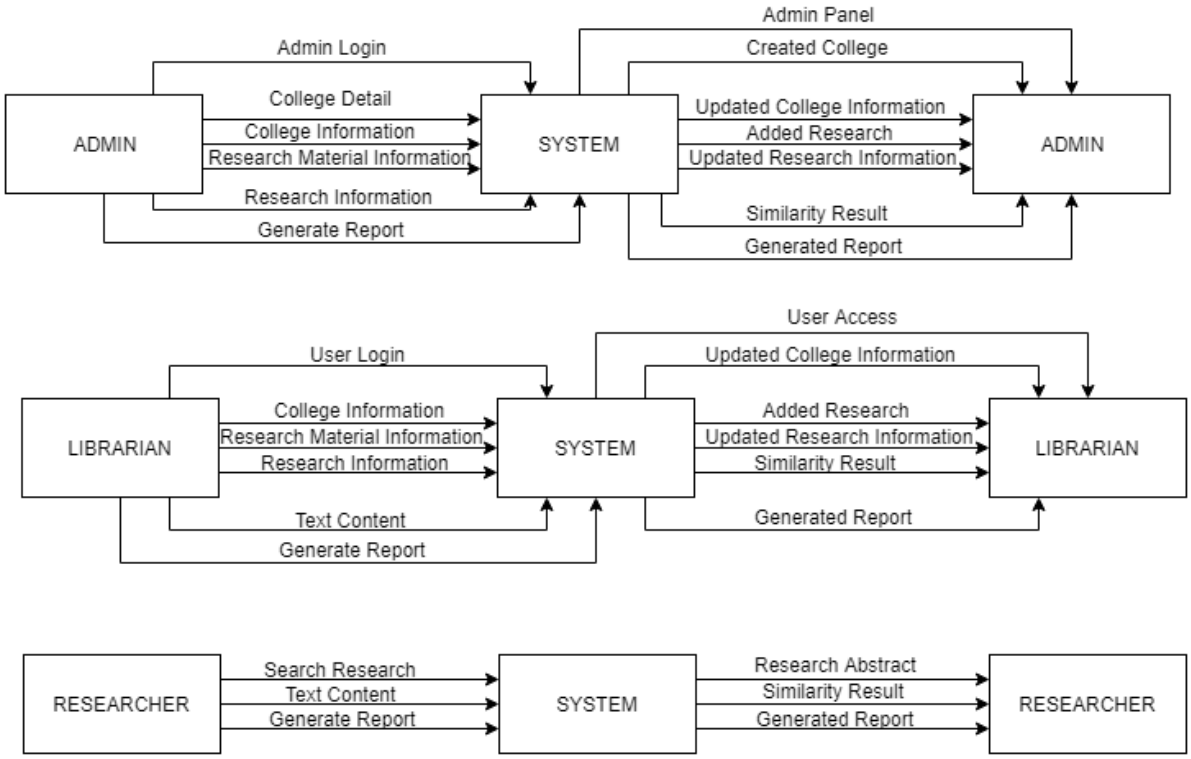
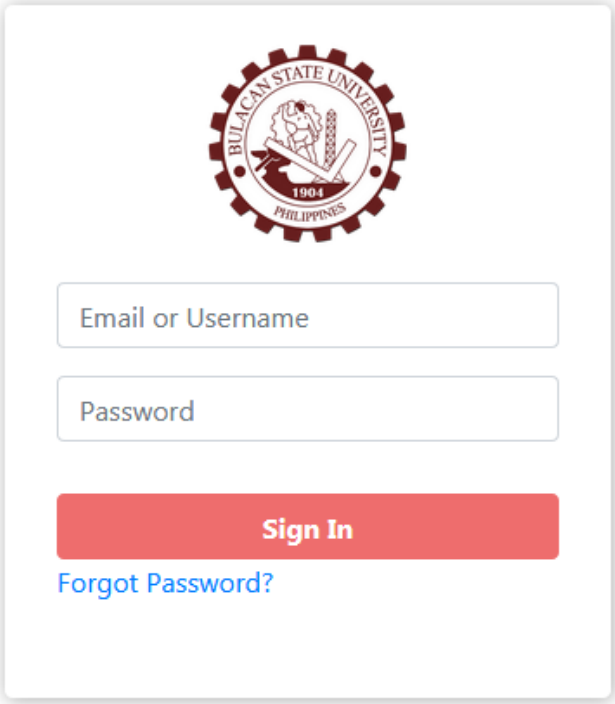
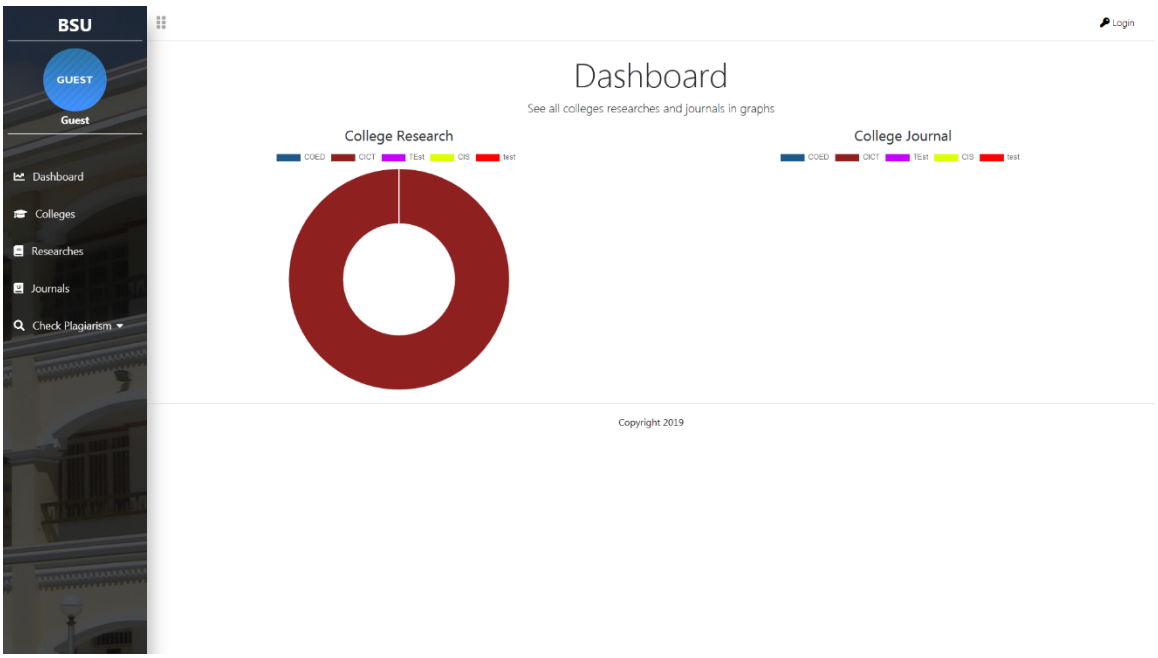


Figure 4.0 Data Flow Diagram Level 0

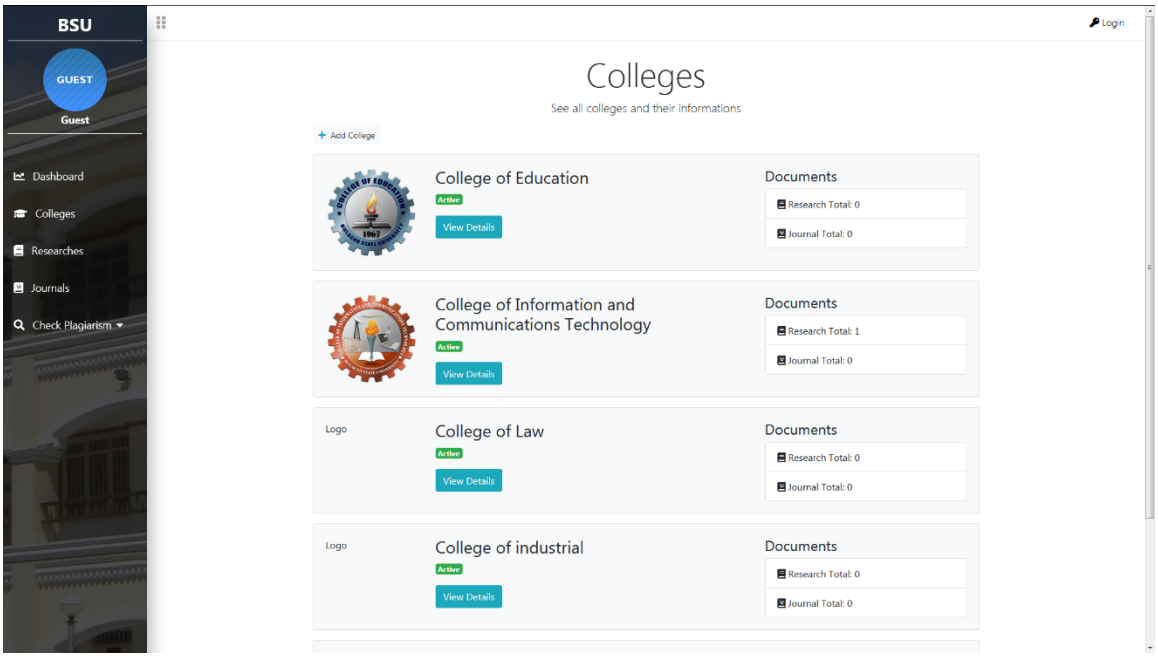
Description of the Prototype



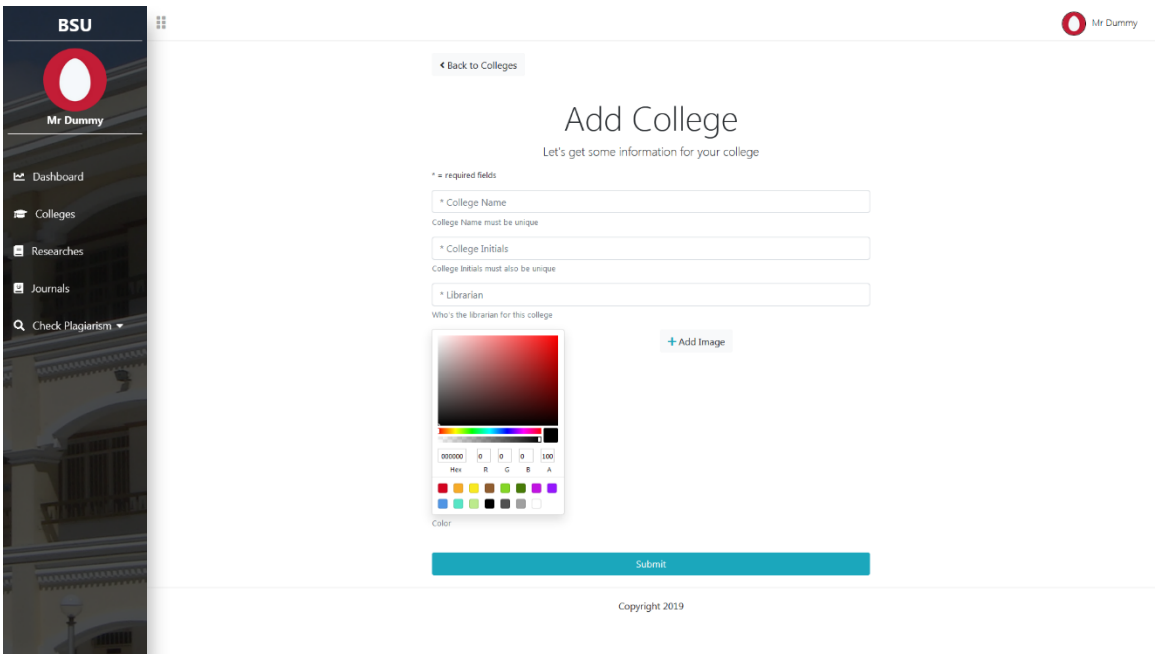
Login interface



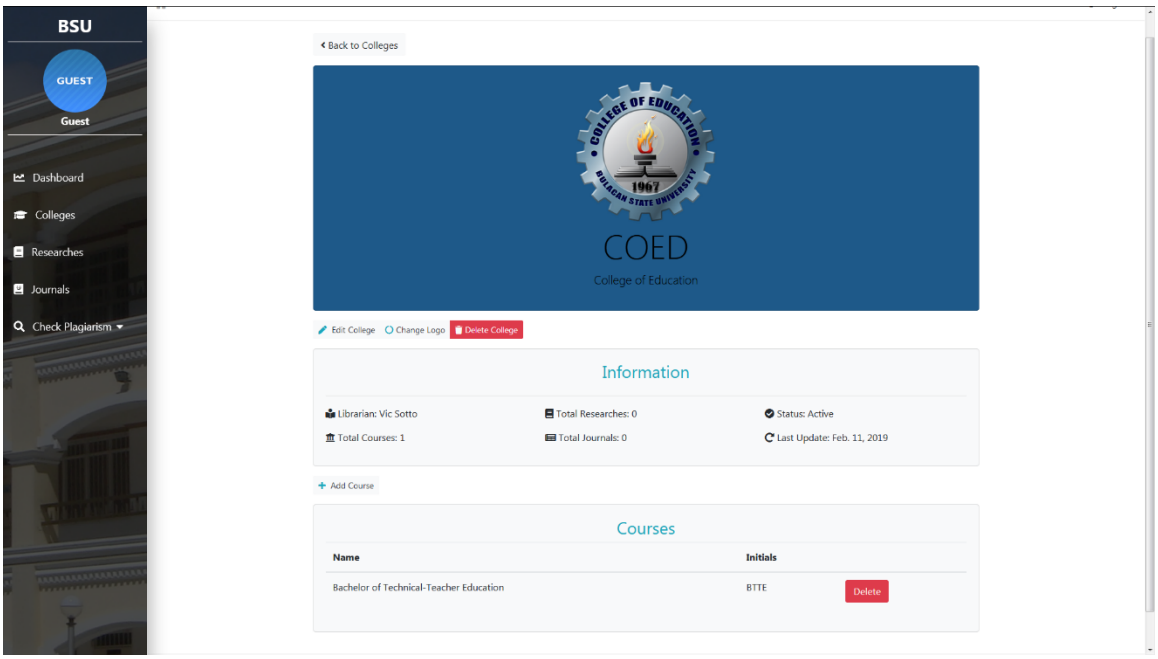
Dashboard Interface



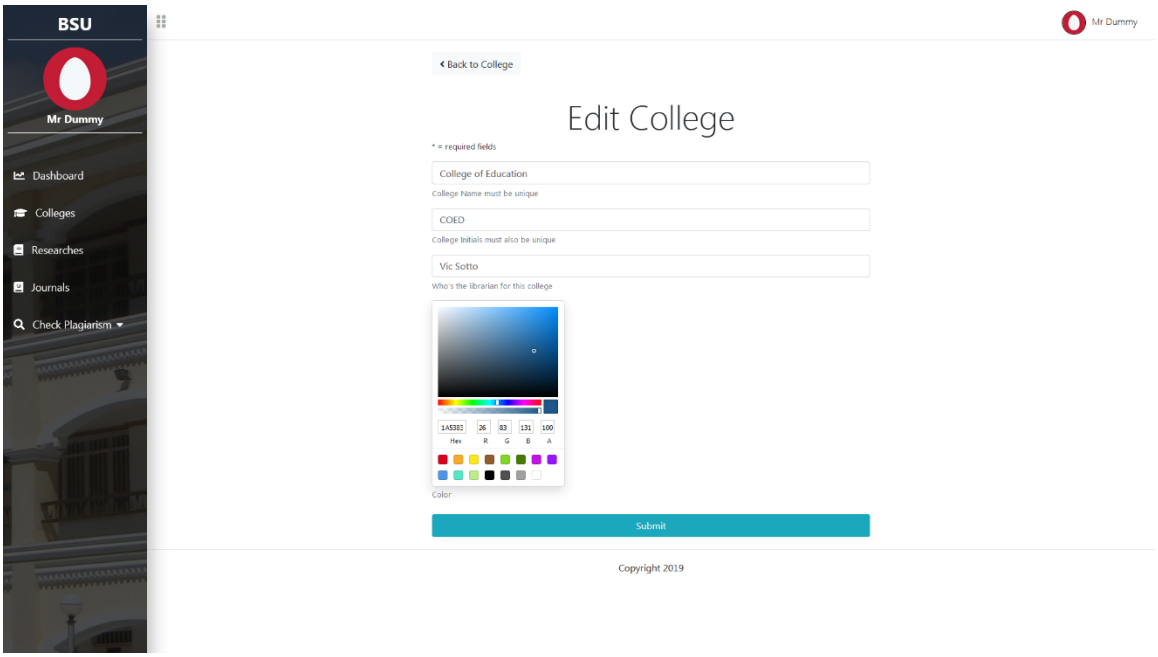
List of Colleges



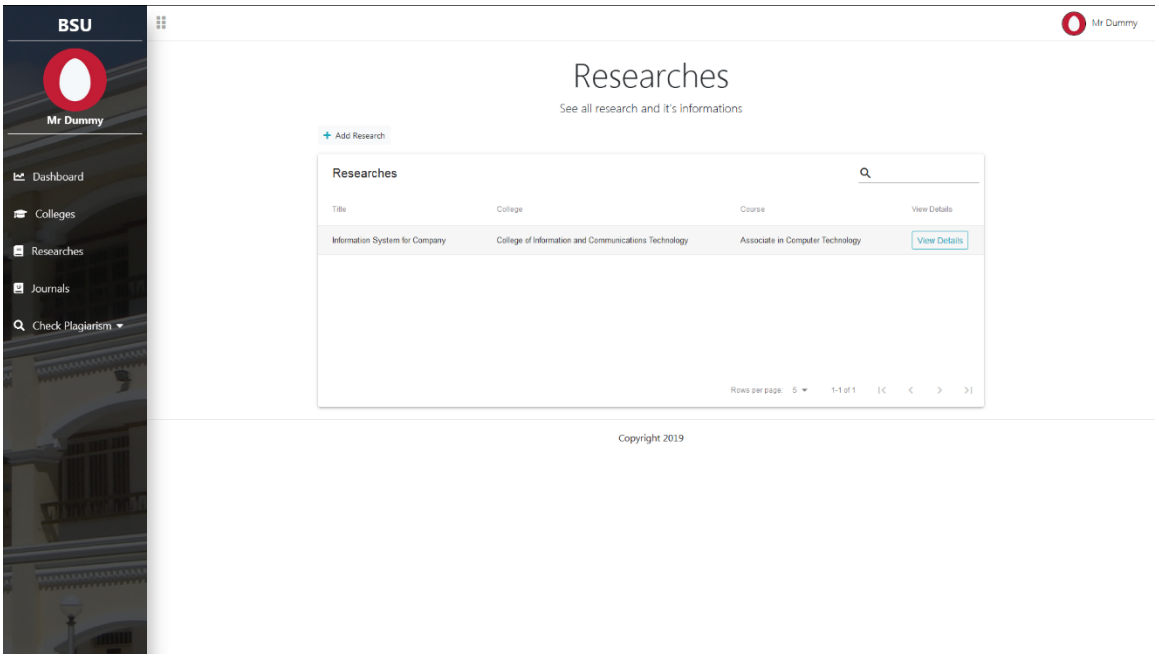
Add College



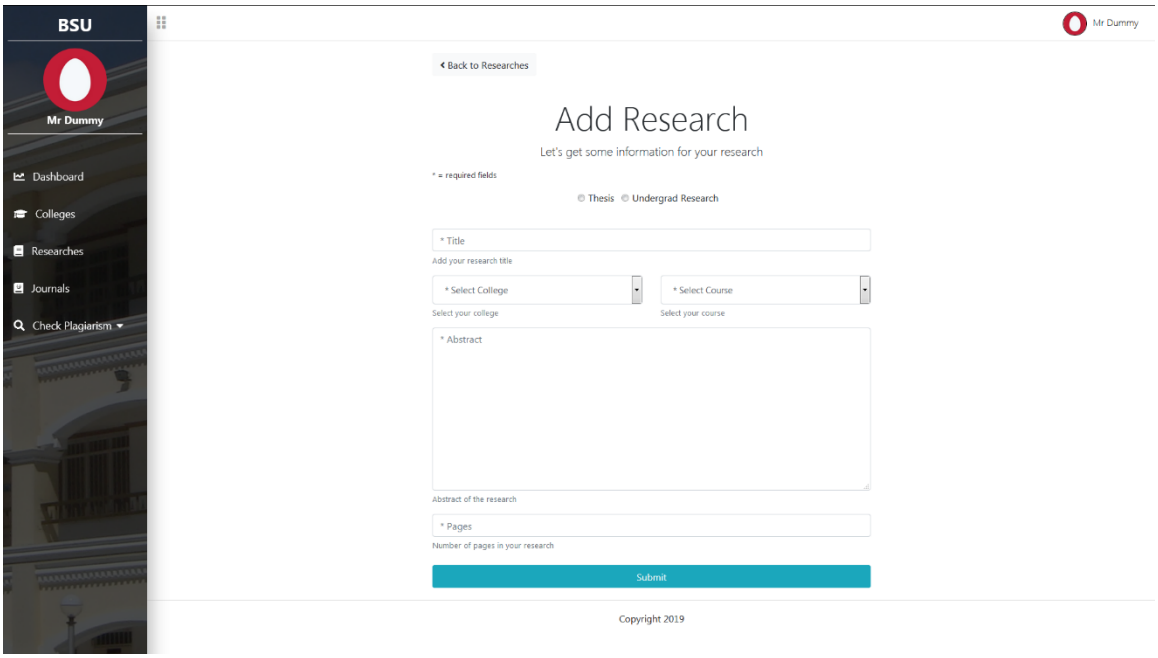
College Information



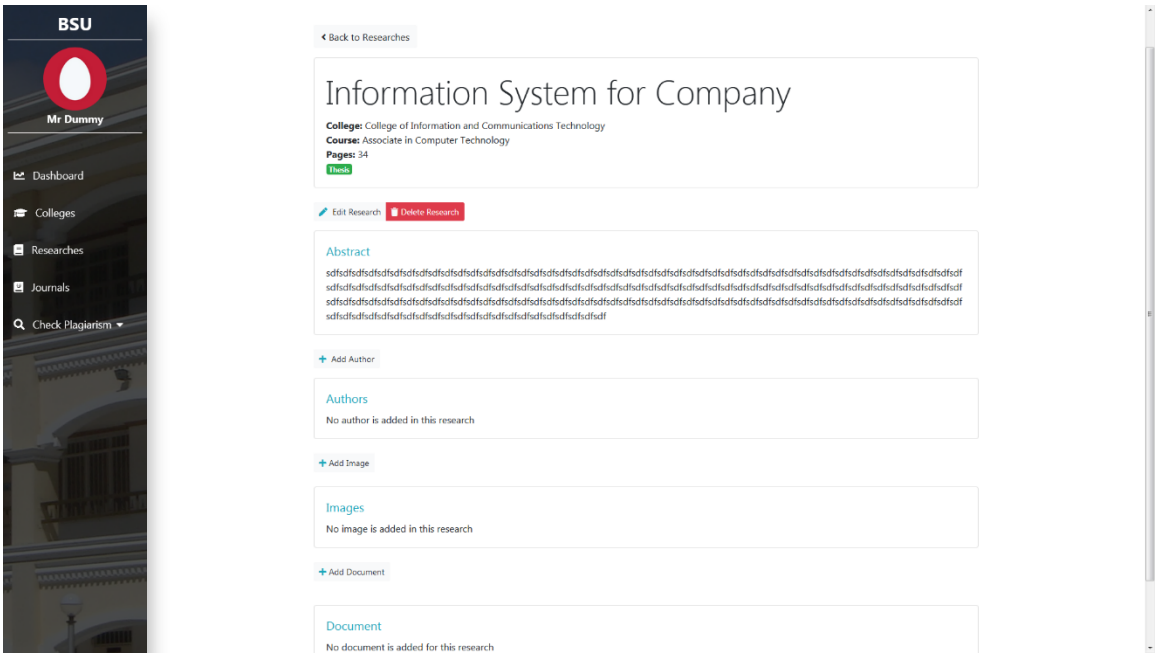
Edit College



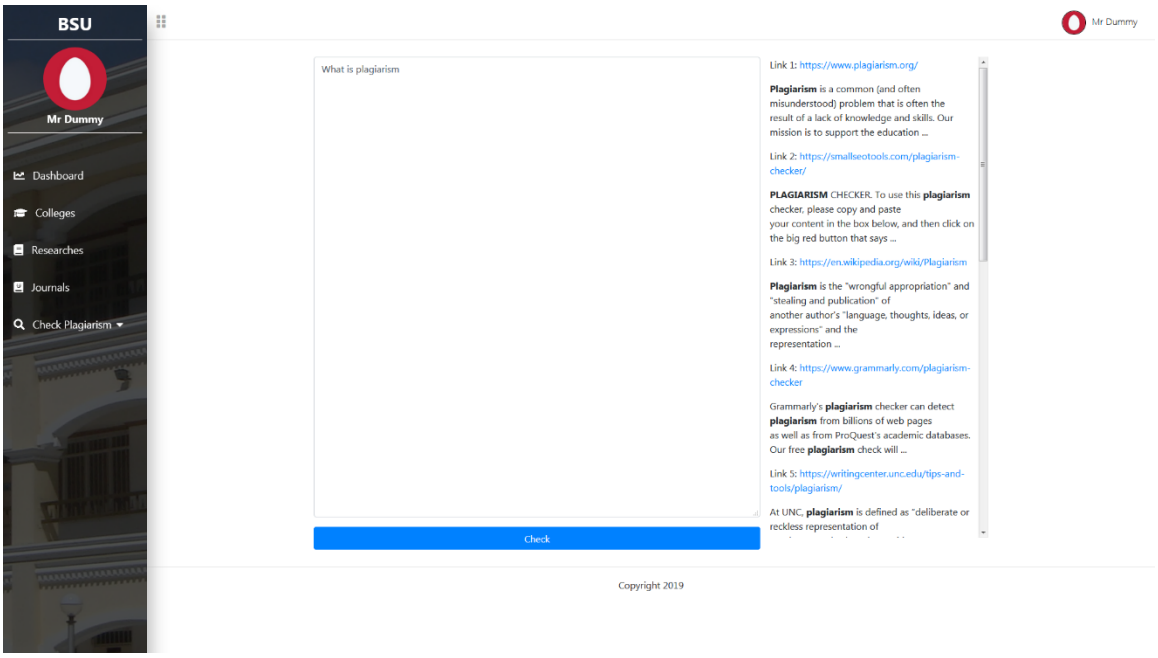
Research List



Add Research

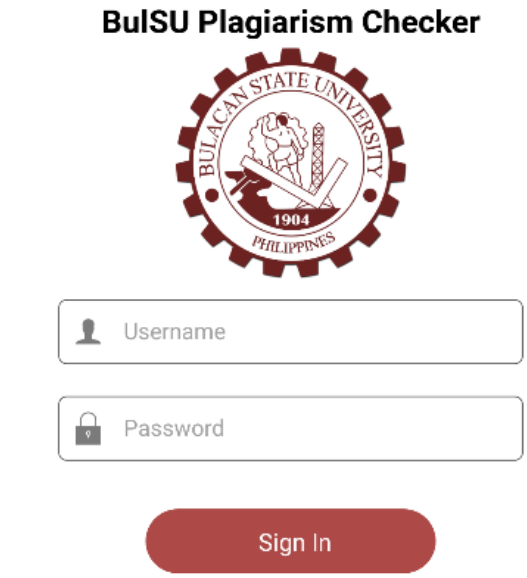


Research Information

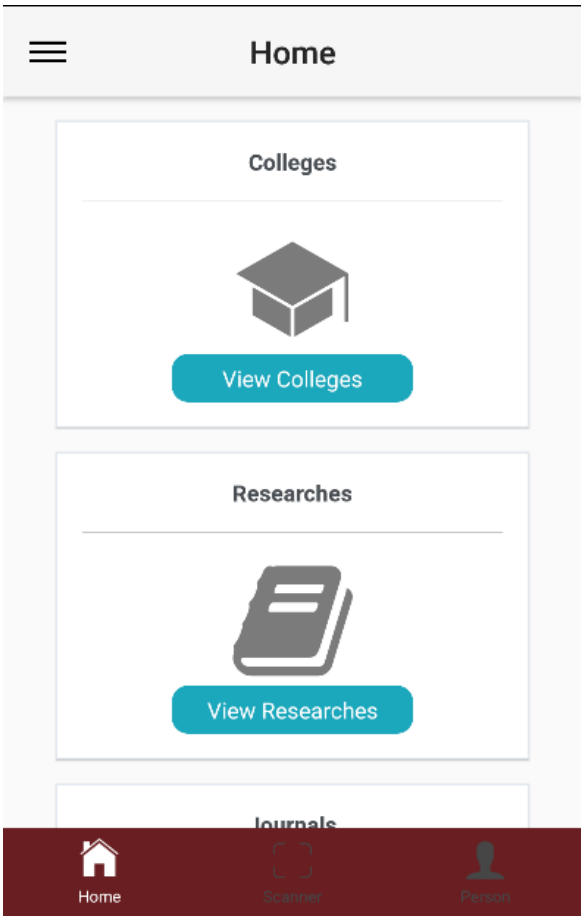


Online Plagiarism Check

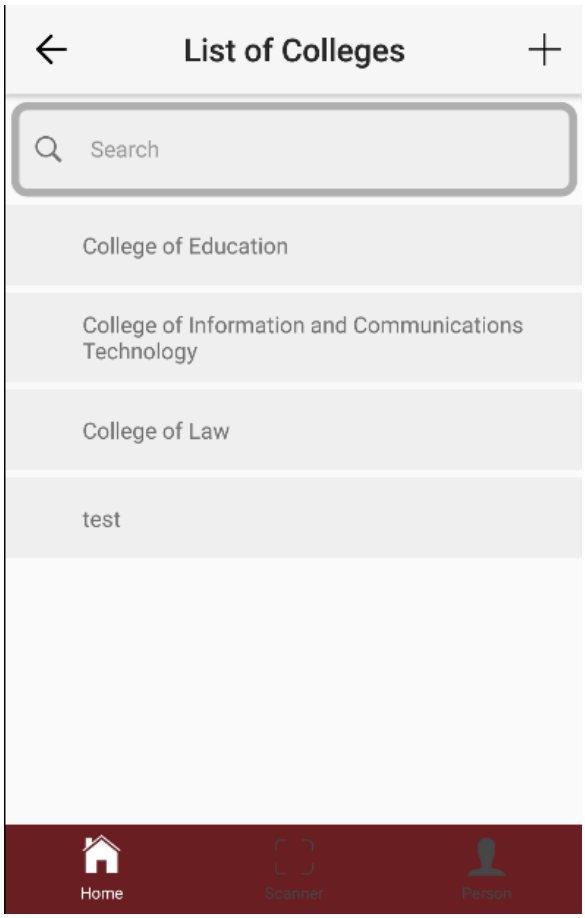
User Interface (Android)



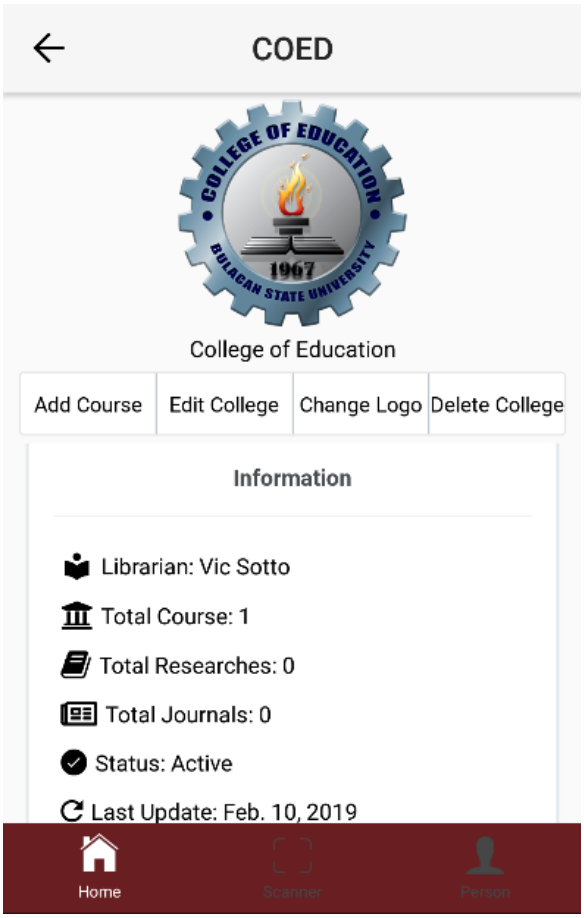
Login Screen



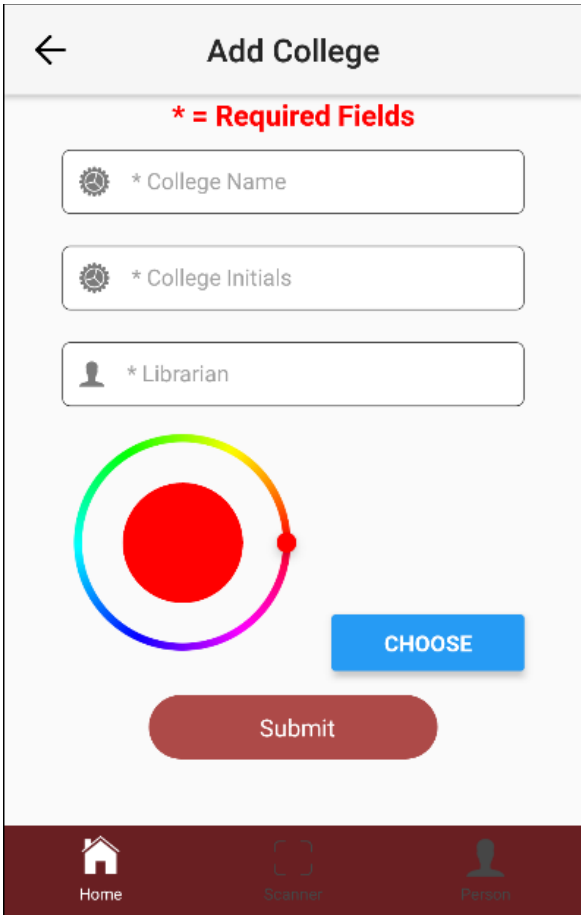
Home Screen



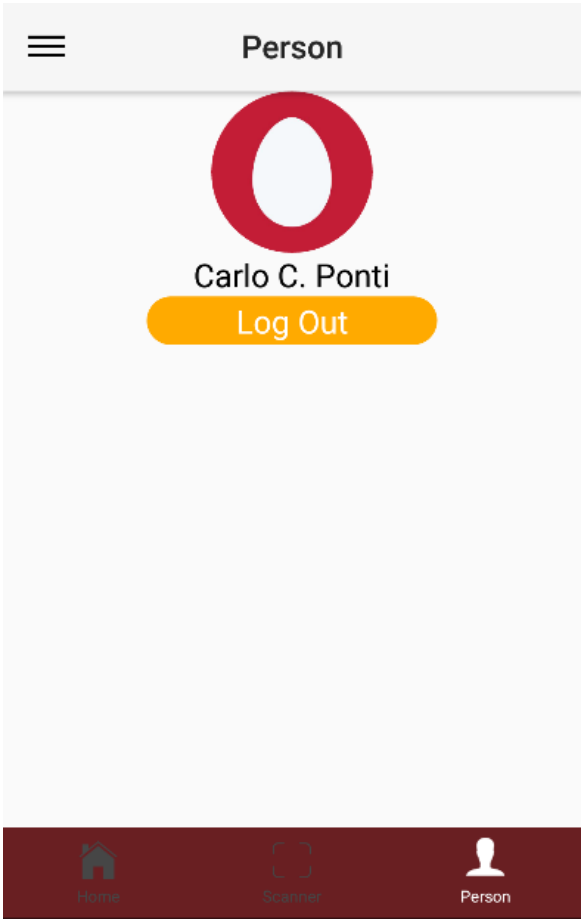
List of Colleges



College Information



Add College



Account

User Requirement

In this section it tackles about what should the user expect from the whole system and what the system able to do. Here are the user requirements for the system:

1. First you will have a default Researcher level of access in the system, you will see the dashboard containing the number of the colleges, researches, journals and visitors.
2. You can login as an Administrator or Librarian in the system that will widen your access in the system.
3. In Administrator level you can Manage Account, Manage Colleges, Manage Published Journal, Manage Researches, use plagiarism checker, use image to text converter, view audit trail of the user and can generate reports.
4. In Librarian level you can Manage College (only his/her college), Manage Researches, use image to text converter, view audit trail of the user and can generate reports.
5. In Researcher level you can view and download research abstract, use plagiarism checker (abstract only) and generate reports.
6. This system can also check grammar errors.
7. The system has a mobile version, that has an offline feature, it can view the abstract of the research materials of the Bulacan State University.

References

https://en.wikipedia.org/wiki/Plagiarism

https://en.wikipedia.org/wiki/Plagiarism_detection

http://www.iaescore.com/journals/index.php/IJEECS/article/download/6097/6128

https://en.wikipedia.org/wiki/English_grammar

https://en.wikipedia.org/wiki/Grammar_checker

https://en.wikipedia.org/wiki/Natural-language_processing

http://people.uncw.edu/tagliarinig/Courses/380/F2014PPs/paperTranShi.docx.

Carl Justine M. De Guz

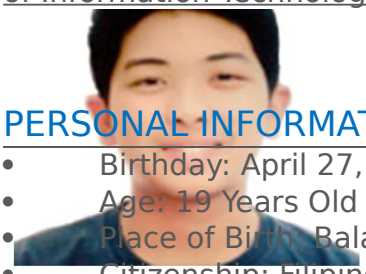
Dakila, Malolos,
Bulacan | +639224365530 | carljustinedg@gmail.com

Jenaro Lopez

Student

OBJECTIVE

Seeking a position in life to utilize my skills and abilities ology and helps me to learn more by the everyday and achieve professional growth while being resourceful, innovative and flexible. To add valuable assets to your esteemed organization as an active member in the field of Information Technology.



PERSONAL INFORMATION

- Birthday: April 27, 1999
- Age: 19 Years Old
- Place of Birth: Balagtas, Bulacan
- Citizenship: Filipino
- Civil Status: Single
- Sex: Male

Personal Info

Address
0490 St. Joseph, Street, Tibag, Pulilan, Bulacan.

Phone
09162431422

E-mail
jenjenlope237@gmail.com

Date of birth
1999-01-10

Skills

Java	●●●●○
Python	●●●●○
C++	●●●○○
Php	●●●○○
HTML and CSS	●●○○○
Microsoft Office (Word, PowerPoint, Excel)	●●●●○
Adobe Photoshop	●●●○○

Languages

English	●●●●○

SKILLS & ABILITIES

Knowledgeable in Java, HTML, CSS, JavaScript, PHP
Highly Organized and Efficient. Ability to work independently or as a part of team

EXPERIENCE

Aug. – Dec. 2018 Internship, *DICT – Luzon Cluster II*

EDUCATION

2015 - 2019 BSIT, Malolos Bulacan, *Bulacan State University*

2018-09 - 2018-11 **OJT**
NLEX Corporation, Traffic Systems Engineering
Assisting in installation of tollbooth and perimeter cameras as well as configuring it in the database. Creating speed report that includes the captured speeding motorist and apprehended ones. Assisting in capturing drone shots on the new equipment installed and also the project segment 10 extension of the NLEX from Mindanao Avenue, Quezon City to the C-3 (Circumferential Road 3) in Caloocan City and to Commonwealth Avenue in Quezon City. Assisting the maintenance of the patrol vehicle cameras. Assisting in Server room and the Traffic Control room for the databsase and viewing of cameras.

Education

2015 - present **Bulacan State University, Information Technology, Associate in Information Technology**
Guinhawa St., City of Malolos, Bulacan
Associate in Computer Technology

2005 - 2011 **Segundo Esguerra Sr. Memorial Elementary, Elementary, Kindergarten**
3004 Esguerra Street, Pulilan, Bulacan
Mobile Number - +63915 577 8124
Founded in 1979

2011 - 2015 **Mary Chiles College of Arts and Sciences, High School**
154 Sandico Street, Lumbac, Pulilan, Bulacan
Telephone No. 910-12-25
School Head - Mr. Geronimo G. Durante, Sr.

Certificates

2015-10	The Fifth I.T. Congress: "Features the Future"
2015-11	The 6th I.T. Congress: "IT's How We Connect"
2017-10	Game Development: "Creating Game in Unity 3D"
2017-10	Stay Connected
2017-11	7th I.T. Congress: "Wa[illegible]
2017-12	JUMPSTART: "Maximise Your Careeer Success"

References

Engr. Evelyn C. Samson, Bulacan State University, Instructor, 09208698367

Engr. Gibson Repomanta, NLEX Corporation,
Supervisor, https://www.facebook.com/gibson.repomanta

De Guzman, Carl Justine M.
Applicant Signature

Jenaro S. Lopez
Applicant Signature

Karl Arceo Martin

Buisan, Bustos, Bulacan| +639052547935
| karlmartin226@gmail.com

OBJECTIVE

Seeking a position in life to utilize my skills and abilities and achieve professional growth while being resourceful, innovative and flexible. To add valuable assets to your esteemed organization as an active member in the field of Information Technology.

PERSONAL INFORMATION

Age	: 19 y/o
Date of Birth	: January 22, 1999
Place of Birth	: Liang, City of Malolos, Bulacan, Philippines
Religion	: Roman Catholic
Height	: 5’9’’
Weight	: 220 lbs
Civil Status	: Single
Citizenship	: Filipino

SKILLS & ABILITIES

Ability to work independently or as part of a team

Good Knowledge in Microsoft Office (MS Word, Excel, PowerPoint, etc.)

EXPERIENCE

Aug. – Nov. 2018	Internship, *Channel Technologies Inc., Quezon City, Philippines*

EDUCATION

2015 - 2019	BSIT, Malolos Bulacan, *Bulacan State University*

Karl A. Martin
Applicant Signature

Carlo Jesus Cruz Ponti

0687 Sta. Elena Hagonoy, Bulacan
+639558154672
cj.dcponti @gmail.com

OBJECTIVE

To obtain a position that will enable me to use my strong organization skills, educational background and ability to work well with people.

PERSONAL INFORMATION

Name	:	Ponti, Carlo Jesus C.
Birth Date	:	May 14, 1999
Birth Place	:	Hagonoy, Bulacan
Age	:	19 years old
Nationality	:	Filipino
Religion	:	Catholic
Weight	:	185 lbs
Height	:	5'9"
Marital Status	:	Single

TECHNICAL SKILLS

- Knowledgeable in Java, C++, VB.net
- Knowledgeable in Javascript, AngularJS, ReactJS, React Native and Redux,
- Good knowledge in Microsoft Office (MS Word, Excel, PowerPoint, etc.)
- Good knowledge in Adobe Photoshop, Microsoft SQL Server, MongoDB
- Basic knowledge in Python and NodeJS

PERSONAL SKILLS

- Can easily adapt to new environment.
- Ability to work independently or as part of a team.
- Flexible in any kind of work.
- Responsible for given task.

CHARACTER REFERENCES

Engr. Evelyn C. Samson
Instructor
Bulacan State University

I hereby declare that all the information provided in my resume is true to the best of my knowledge.

Krishield Kyle R. Santoyo

Erlo Jesus Cruz Ponti
Applicant

BACHELOR OF SCIENCE IN INFORMATION TECHNOLOGY

CAREER OBJECTIVE

Seeking a challenging career with a progressive organization that provides an opportunity to capitalize my technical skills & abilities in the field of information technology (IT).

TECHNICAL SKILLS

- Java
- Servlet/JSP
- Maven
- Spring MVC
- HTML/CSS
- ReactJs
- React Native
- Redux
- C++
- MongoDB

PERSONAL

- Willing to learn and work under pressure.
- Highly organized and efficient
- Ability to work independently or as part of a team

EDUCATION

Bachelor in Industrial Technology (CT) (2013 – 2015)
Bulacan State University – Matungao Bulakan, Bulacan
Bachelor of Science in Information Technology (2015 – Present)

Address

692 Waling waling St.
San Nicolas Bulacan, Bulakan.

Contact

Phone: 0906-103-6001

E-mail

krishieldkyle@gmail.com

GitHub

https://github.com/KrishieldKyle

Portfolio

krishieldkyle.github.io/my_portfolio/

Date of Birth

May 17, 1997

REFERENCES

Available upon request

PRE-PROFESSIONAL EXPERIENCES

Intern – Pamzeth Computer Dynamics. (November, 2014 - February, 2015)

- Handle other clerical task given by the technician.

Intern – Palm Solutions (August, 2018 - November, 2018)

- Act as a developer for our operations team. Developing admin dashboard (web site). Attending to external clients to present our product.

Sales Utility Clerk(SUC) – Star Appliance Center Inc.

- Assist Customers to their desired item and also suggest other item.
- Checking the inventory every end of the day.