

PLAGIARISM DETECTION AND NOTIFICATION SYSTEM IN TERTIARY INSTITUTION IN NIGERIA**UKEJE OKORIE AUGUSTINA. M**DEPARTMENT OF COMPUTER SCIENCE, FEDERAL COLLEGE OF AGRICULTURE, ISHIAGU, IVO
LOCAL GOVERNMENT AREA, EBONYI STATE.

EMAIL: TINAUKEJE45@GMAIL.COM

ABSTRACT

Plagiarism involves the use or close imitation of the language and thought of another author with or without authorization and the representation of that authors work as one's own. Following the recent increase in the popularity of on-line services, offering an unlimited access to digital books and information to the general public, especially to students and researchers to aid in their researches and presentation has significantly increased plagiarism as they would copy all the content of the original work, and would not think of adding their own thought. Over the years different applications have been developed to curb this act but there is no application yet that plagiarized institutional repository, where student work can be checked if they actually plagiarized past student work or any of the lecturers' work, and no means that will capture the information of the student that plagiarized other content and send to the head of that department through SMS. These challenges led to the development Plagiarism detection system for tertiary institutions in Nigeria that will detect plagiarism after a user has uploaded his or her content to the application, the developed application detect plagiarism paragraph by paragraph and also send the information of the student which will include the name of the author whose book was plagiarized to the head of department through SMS and also there will be a module that detect plagiarism at zero time and effective manner and also where all plagiarized content will be stored in the database for easy retrieval and will enable the upload of books in soft copy to the database. The research will be done using Prototyping Methodology. The project will be implemented using Adobe Macromedia Dreamweaver CS6 as the integrated Development Environment (IDE) and MYSQL as database management system.

KEYWORDS: Plagiarism, Institutional Repository, SMS Notification and Head of department**1 INTRODUCTION**

Plagiarism is a very common phenomenon now days and it is center of discussion at various educational events. Plagiarism is defined as the practice of claiming or implying original authorship of (or incorporating material from) someone else's written or creative work, in whole or in part, into one's own without adequate acknowledgement [3].

E-learning programs has touched every corner of this small world due to the advanced in the Information Technology, this in turn led to easier availability of the research papers , books , technical and non-technical papers etc which are easiest source for making plagiarized documents. Copying and pasting of paragraphs or even entire essays now can be performed with just a few mouse clicks. [4]

[10] Researcher are focusing on inventing newer ways for secure information flow so as the confidential information can be transformed freely so as e learning programs not prove harmful for the books publishers and arise cases of copy rights violations. This seems to be a temporary solution for this problem because eventually we are trying to control the possible cases of plagiarism. [5] Most of

the research in this field has till now concentrated in the type of text extraction which in turn can be sentence or word chunking (Chunking is a method of presenting information which splits concepts into small pieces or "chunks" of information to make reading and understanding faster and easier.), no one has yet bi-furcated (divide into two branches) the extraction into sub parts such as rearrangement to explore the performance issues.

Current systems or tools are effective on the cost of time as if you need more performance system will work slowly even they give effective output till a threshold level based on the document size. [9]

[8] Plagiarism has received much attention from both the academic and commercial communities. This has been particularly true in academia as students turn to technology to fabricate texts which are not entirely their own work. Text reuse is defined as the activity whereby pre-existing written material is reused during the creation of a new text, either intentionally or un-intentionally. Many other examples of text reuse surround us today, including the creation of literary and historical texts, summarization, translation or revision of existing texts.

Many factors influence text reuse including translating an original text into a different language, restyling an original to fit different authorial or consumer needs (e.g. rewriting a scientific text to be readable by the layman), reducing or expanding the size of the original text and the competency and production requirements of the writer.

Recent advances in technology are making text reuse much easier. For example, the Google web search engine claims to index over 3 billion web pages¹ providing a large variety of source texts on a diverse range of topics in many different languages. Word processors have also become more sophisticated, enabling users to easily cut and paste, merge and format pre-existing texts from a variety of sources. This, coupled with the change of culture brought about by electronic 'cyber-space' has caused concern to authors surrounding the ownership of their written material. Either the owner (perhaps the publisher) has to protect their texts (e.g. using digital watermarks), rely on finding illegitimate copies, or even de-value their digital content in some way.

[11] suggests that "the origin and ownership of all electronic documents is now peculiarly evanescent; one click of the 'Save As' button can give a whole new name and identity, instantly, to someone else's creation." of course not all text reuse is deemed a cardinal sin; there are certain areas where borrowing is perfectly acceptable.

For example, as [1] points out: "generally, borrowing is a tradition in literature and other art forms and more than a tradition: creativity feeds on what has gone before, new work is formed out of old." During the renaissance and romantic eras of literary writing, even the "great" authors would reuse the ideas, storylines and plots of others in their own literary creations. It was not considered immoral or unethical; rather it was seen as a stimulus for creativity. Text reuse was (and is) the epitome of literary recognition. Today, there are examples where even verbatim text reuse is not considered unethical; rather the norm. One such example is the reuse of newswire (or news agency) texts by journalists in the creation of newspaper articles [2]. As long as the journalist or the organization they write for are subscribers of the news agency, they are free to reuse agency text as and how they please, with or without attribution.

The goal of Plagiarism Detection System is to plagiarized institutional repository in order to detect if a student actually plagiarized past student content or other lecturers content. If plagiarism is detected then the information of the student including the name of the author that was plagiarized will be sent to the H.O.D of that department through SMS gateway and finally detect plagiarism at zero time and effective manner and also where all plagiarized content will be stored in the database for easy

retrieval and will enable the upload of books in soft copy to the database.

2 PROBLEM STATEMENT

There have been different applications developed for this same research, but still have some shortcomings, these shortcomings are enumerated below:

1. There is no application yet that plagiarized institutional repository, where student work can be checked if they actually plagiarized past student work or any of the lecturers work.
2. No platform for SMS gateway, which is where information of the student that plagiarized other author's content, can be sent to the H.O.D mobile phone.
3. The current method of detecting plagiarized content is very slow, as the time taken to compare a large amount of text takes a very long period of time.

3 AIM AND OBJECTIVES

The aim of this article is to propose a Plagiarism Detection System for tertiary Institutions in Nigeria. The objectives are to:

1. Create a module that will detect plagiarism by checking institutional repository database to know if the student copied from past student or any lecturer's work
2. Introduce an SMS gateway that will send the information of the student whose content was detected and the name of the author whose content was plagiarized to the head of that department.
3. Create a module that will detect plagiarism at zero time and effective manner and also where all plagiarized content will be stored in the database for easy retrieval and will enable the upload of books in soft copy to the database.

4 SCOPE OF THE STUDY

The scope of this study is to develop a Plagiarism Detection system for tertiary institution in Nigeria. The research would be limited to the development of a system that would compare large or small amount of text to check for any form of similarity and also send the name of the student who was found guilty and the name of the author of the original content to the head of department. This project is limited to tertiary institutions in Nigeria. This paper covers the following:

- i. There will be an admin that will register students and research coordinator and also will upload books or scholars work in the instructional repository database in softcopy.
- ii. The research coordinator check for plagiarism by entering the student registration number and also upload the student work while the system

automatically search the institutional repository database to check if plagiarism is detected, if detection level is fifty percent and above the system will send the name of the student and the name of the author whose book was plagiarized to the Head of that department via SMS.

- iii. The system uses MySQL to detect plagiarism at zero time and effective manner and also all plagiarized content will be stored in the database for easy retrieval and will enable the upload of books in soft copy to the database.

5 ANALYSIS OF THE EXISTING SYSTEM

The existing system is such that once students defend his or her research work both post graduate and undergraduate he or she will be ask to submit the softcopy of the work via an empty compact disk. A copy is sent to the library, exams and record and also to the school institutional repository. In some institution, plagiarism is usually checked by uploading the softcopy of the student work to Turnitin.com which will automatically check online to know if the work was plagiarized or not.

The existing system comprises both software and website which include both paid-for and free for use. However academically popular software for detecting plagiarism are Turnitin and iThenticate. Turnitin.com was designed by John Barrie and a group of his UC Berkeley colleagues. This website uses digital technology to conduct a meta-search of the internet to locate sources from where the document might have been plagiarized. Users receive an originality report that cites the degree of originality and links to Internet WebPages that help determine what Web resources their students have tapped.

PlagiarismDetection.org is a website that provides a user-friendly and innovative plagiarism prevention and detection online tool that can effectively help instructors, and students. It is designed to discover all chances for plagiarized text, and it runs against all the Internet resources, like online libraries, digital DBs, and websites

CheckForPlagiarism.net is also a website that had been created by dedicated instructors, students, and professionals to provide both offline and online based plagiarism.

SafeAssign is software that works by comparing submitted text (works) against a set of identified sources to find areas of overlap (similarities), between the existing works and submitted work.

PlagScan is a browser-based service that confirms the documents authenticity. Files could be attached with all common formats such as PDF, XLS, MS Word, and many more. PlagScan users can

alternatively paste any text directly into it and check the text authenticity easily.

Plagrisma.net is another website used in checking for plagiarized content in this website there is a platform where users can download the application for windows, this website work by pasting the content to be plagiarized and selecting desired search engine after which the user will click check for plagiarized content then the website automatically query the database and display the result.

There are also mobile applications developed like plagiarism check app found in play store in every android phones. In this mobile application the user will paste the text he or she want to check and also there is a platform where the user can select the search engine he want to use after which the user clicks on check duplicate content the application automatically displays the result paragraph by paragraph and also displays the total number of characters, and the result. This application also has a platform where users can upload content in different format.

From the analysis of all the available tools use for plagiarism detection explain above there has not been any application that checks institutional repository to know if students actually copy from their predecessors or lecturers work and also there is no platform that send SMS message which contain the name of the student, the name of the author whose book or work was plagiarized to the head of the department mobile phone for further action. This then form this research work.

6 DESIGN METHODOLOGY

Having understood how the existing system functions, Object-Oriented Analysis and Design Software Development Approach will be deployed to analyze and design the proposed system. The reason for choosing this methodology is because object-oriented systems focus on capturing the structure and behavior of information system in little module that encompass both data and processes.

7 ANALYSIS OF THE PROPOSED SYSTEM

The proposed system will work in a way that there will be a central database representing institutional repository where all scholars works will be stored which includes lecturers and student research. The administrator will register research coordinator and students by entering their names, their program and also enter their head of department's phone numbers, the administrator will also be in charge of uploading scholars' research which includes student project, seminar, M.sc research and Ph.d research works to the central database representing the institutional repository. To check for plagiarism the research coordinator will key in the student registration number and uploads his or her research content in digital format and clicks submit, while the application will automatically check through the

central database for similar work and display the result paragraph by paragraph, total number of words checked, and the percentage of the check. If the percentage of the check is fifty percent and above the system automatically send the name of the student, the author's name whose work was plagiarized to the head of the department or research coordinator for

further action to be taking. Finally once the HOD upload the student work, the application will automatically detect plagiarism at zero time and effective manner and also store all plagiarized content in the database for easy retrieval and will enable the upload of books in soft copy to the database.

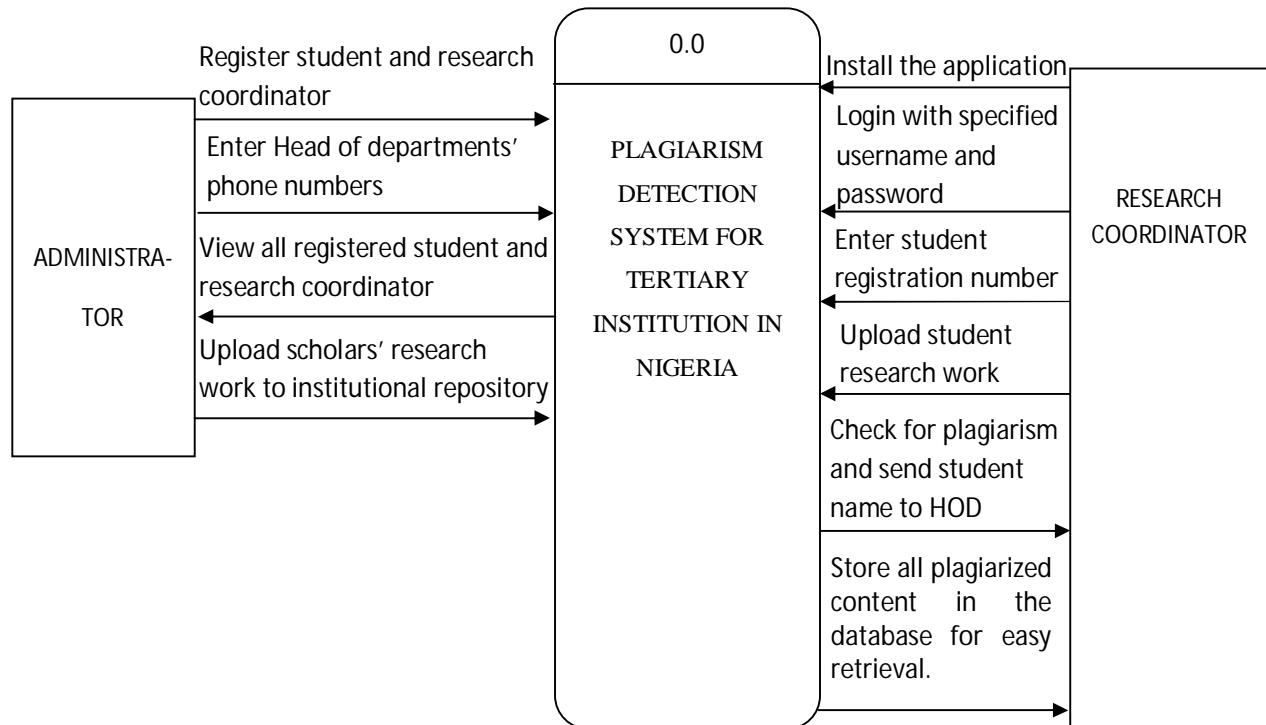


Figure 1: Context Diagram of Proposed System

9 FLOWCHART OF THE PROPOSED SYSTEM

Flowchart is a graphical representation of a computer program in relation to its sequence of functions (as distinct from the data it processes). For system users

to understand the complete function of a program it has to be explained through a symbolic diagram called flowchart. The two flowcharts below describe the complete flow of information from various entities.

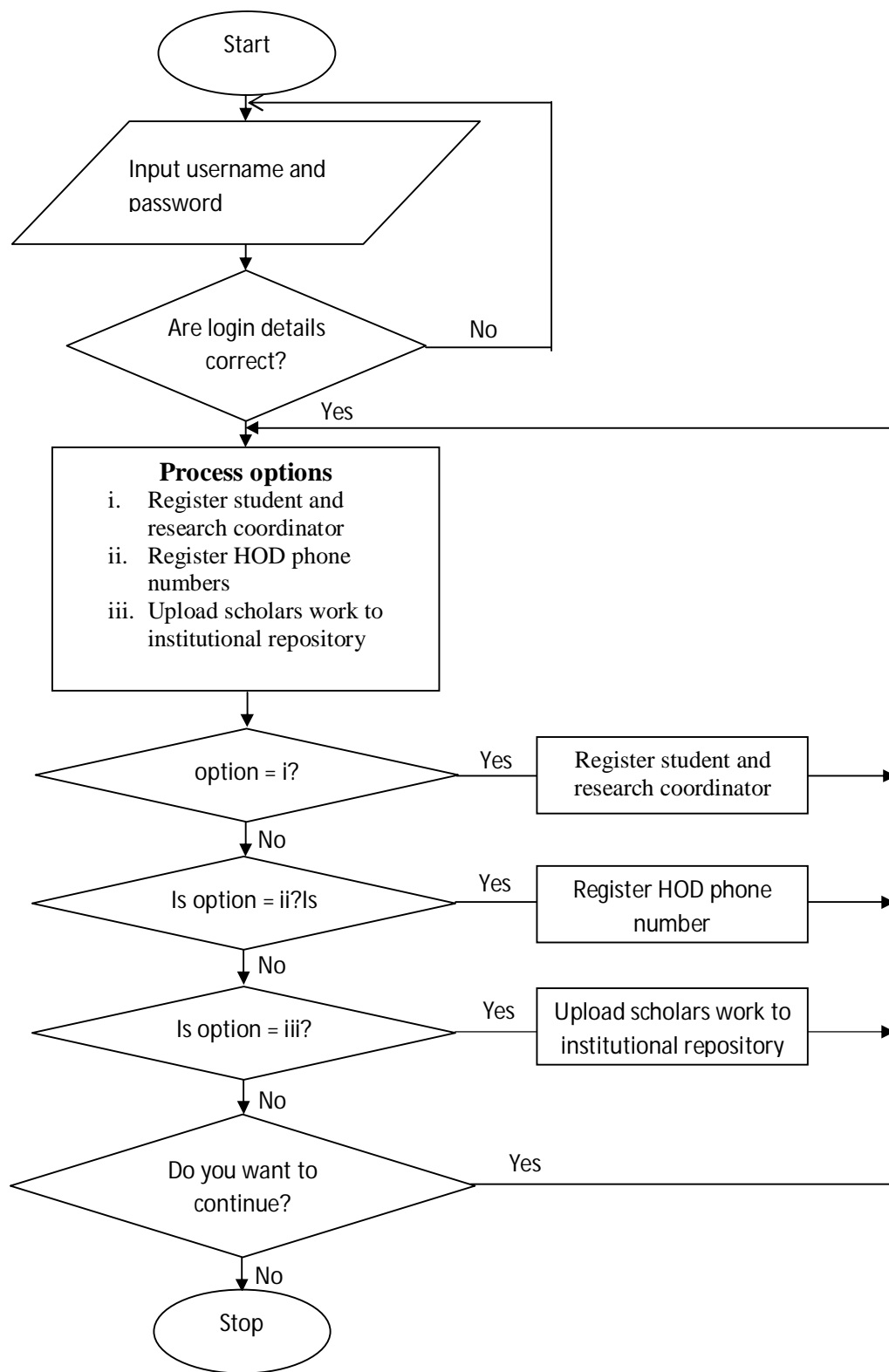


Figure 2: Flowchart for Administrator of the Proposed System

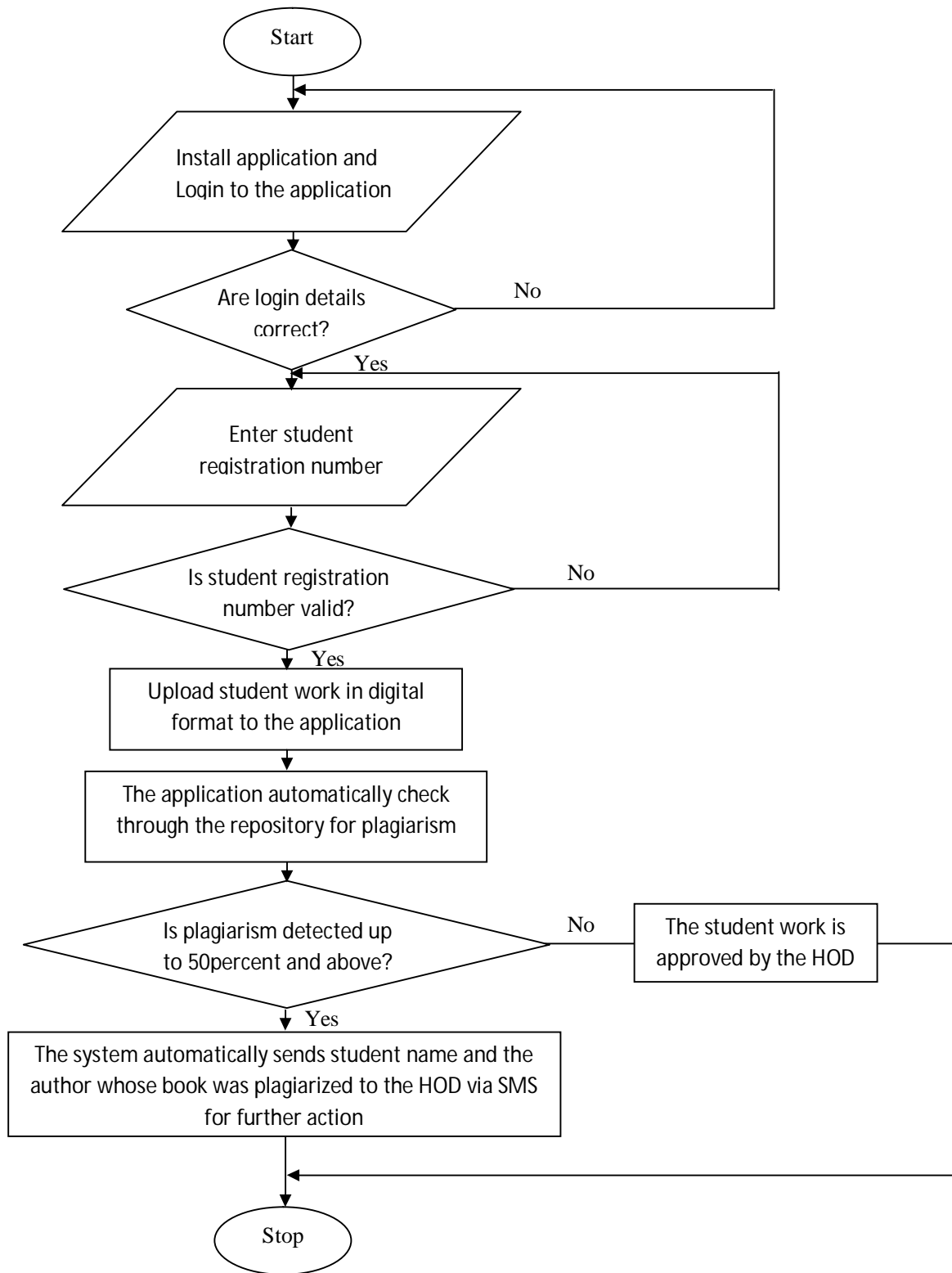


Figure 3: Flowchart for Research Coordinator of the Proposed System

10 USE CASE MODEL OF THE PROPOSED SYSTEM

A use case is a list of actions or event steps typically defining the interactions between a role (known in the Unified Modeling Language as an actor) and a

system to achieve a goal. The actor can be a human or other external system. The Use Case Model describes the proposed functionality of the new system.

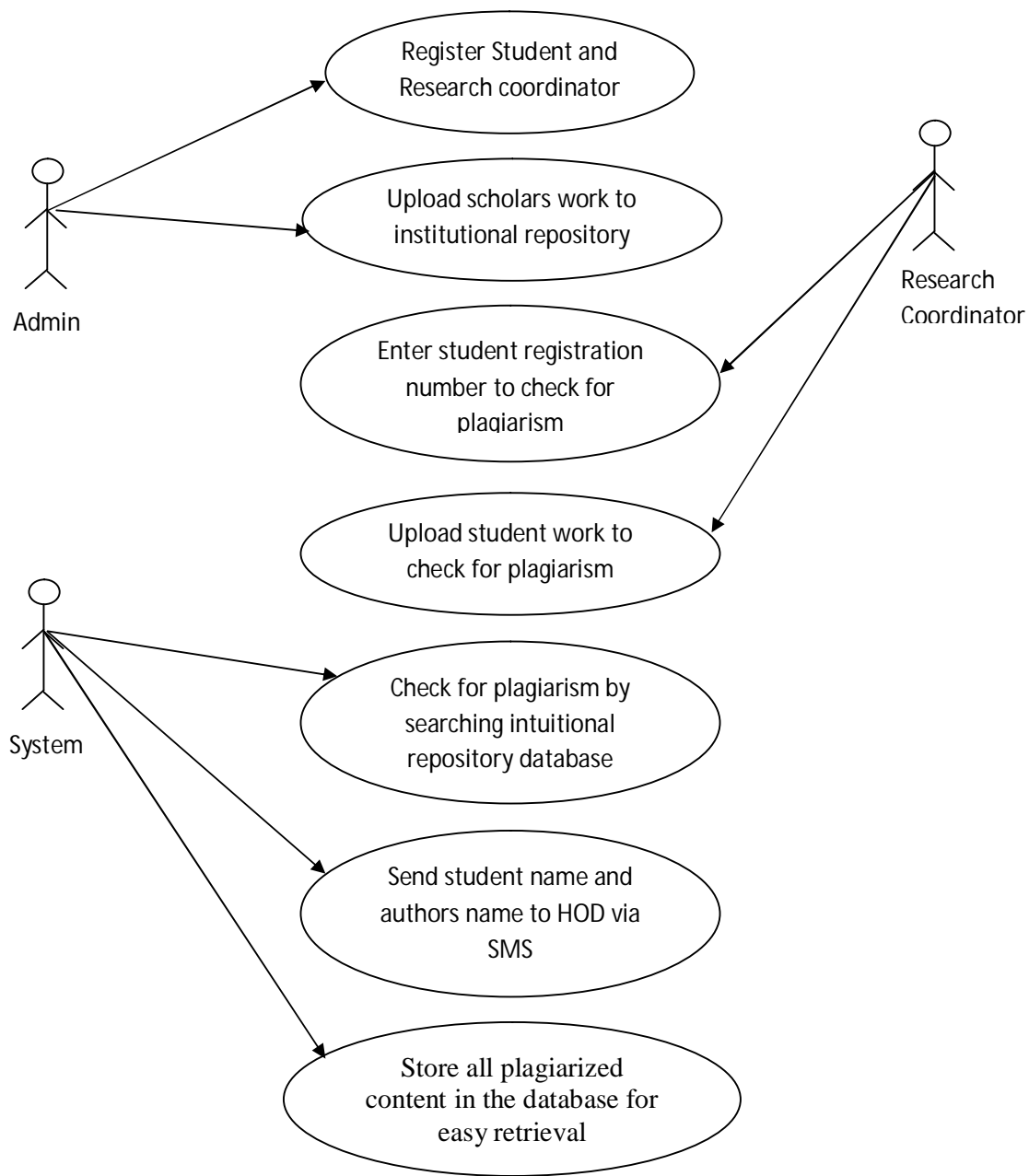


Figure 4: Use Case Diagram of the Proposed System

11 IMPLEMENTATION TOOLS

In the design of this system, Adobe Macromedia Dreamweaver CS6 tool will be used for the complete interface of the design, JavaScript will be use for validation, JQuery will be use for popping out menu, Application program interface (API) will be use for the SMS alert and MYSQL will be use for the database management system.

12. EXPECTED RESULT

1. Uploading of Scholars research work: here administrator will be able to upload scholars research work to the institutional repository database.

2. Check for Plagiarism: the develop application will be able to search the institutional repository to check if any of the uploaded content matches with the work in other to check for plagiarism.

3. Plagiarism Analysis: the develop application will be able to calculate the total number of words checked, and the percentage of detection.

4. SMS alert: here if the plagiarism detection level is up to fifty percent and above then the system automatically send the name of the student, the name of the author whose name was plagiarized to the head of that department for further action.

5. Content Storage: there will be a module where the HOD of the department can upload student work while the system will detect plagiarism at zero time and effective manner and all plagiarized content can be stored in the database for easy retrieval and will enable the upload of books in soft copy to the database.

13. CONCLUSION

The plagiarism in institutions internally is a universal problem and we attempted to analyze it and integrate it using Mobile based and SMS notification. It is a detection system for academic activities in institutions internally. This research covers any research work from undergraduate to post graduate levels. With the developed application students will sit upright and do their researches work without plagiarizing other author's work. Till date no application have been developed for plagiarism detection using institutional repository database to check if a student copied from their predecessors work or lecturers work, which is a big problem in institutions where case of same work is detected even after the student have defended and marks awarded.

REFERENCES

- [1] Clough P., (2018), Plagiarism in natural and programming languages: an overview of current tools and technologies.
- [2] Devlin, M. (2019). "Policy, preparation, and prevention: Proactive minimization of student plagiarism." *Journal of Higher Education Policy & Management*, 28(1): 45-58.
- [3] Engels V. L. S., and Craig M., (2017). , "Plagiarism detection using feature-based neural networks.," in *Proceedings of the Thirty-Eighth SIGCSE Technical Symposium on Computer Science Education*, Covington, Kentucky, March 2017, pp. 34-38.
- [4] Federica R. M., Mandreoli, Paolo and Tiberio, (2018) "A document comparison scheme for secure duplicate detection," *Int J Digit Libr* vol. Springer-Verlag 2018.
- [5] Hart M., and Friesner, T, (2017)"Plagiarism and poor academic practice - A threat to the extension of e-learning in higher education?," *Electronic Journal of e-Learning*, vol. 2, 2017.
- [6] Hughes, J. M. C., and McCabe, D. L. (2016). "Understanding academic misconduct." *Canadian Journal of Higher Education* 36(1): 49-63.
- [7] Kraus, J. (2017). "Rethinking plagiarism: What our students are telling us when they cheat." *Issues in Writing* 13(1): 80-95.
- [8] Liles J. A., and Rozalski, M.E., (2018) "It's a Matter of Style: A Style Manual Workshop for Preventing Plagiarism," *College & Undergraduate Libraries*, vol. 11, pp. 91-101, 2018.
- [9] Lukashenko, R., Graudina, V., Grundspenkis, J.: (2017). *Computer Based Plagiarism Detection Methods and Tools: an Overview*. *CompSysTech '07: Proceedings of the 2017 International Conference on Computer Systems and Technologies*, New York, ACM Press.
- [10] Scherbinin V. and Butakov S., (2019) "Using Microsoft SQL server platform for plagiarism detection," in *Proc. SEPLN*, Donostia, Spain, pp. 36–37.
- [11] Parker K.M and Hamblen J. O., (2019) "Computer algorithms for plagiarism detection," *IEEE Trans. Educ.*, vol.32, no. 2, pp. 94–99, May 2019.