

## SOFTWARE POLICE: AN EXPERT SYSTEM THAT DETECT, INVESTIGATE AND REPORT PLAGIARISM ON THE INTERNET

<sup>1</sup>*Chinwe.G. Ezekwe*, <sup>2</sup>*Njideka N. Mbeledogu*, <sup>3</sup>*Joshua J. Tom* and <sup>4</sup>*Chile-Agada Bob*

<sup>1</sup>Admiralty University of Nigeria, Ibusa, Delta State, Nigeria Department of Computer Science, <sup>2</sup>Nnamdi Azikiwe University, Awka, Anambra State, Nigeria, <sup>3</sup>Department of Cyber Security, Elizade University, Ilara Mokin, Nigeria, <sup>4</sup>Department of Computer Science and Robotics, Alvan Ikoku University of Education

<sup>1</sup>[ezekwechinweg@gmail.com](mailto:ezekwechinweg@gmail.com), <sup>2</sup>[nn.mbeledogu@unizik.edu.ng](mailto:nn.mbeledogu@unizik.edu.ng) <sup>3</sup>[joshua.tom@elizadeuniversity.edu.ng](mailto:joshua.tom@elizadeuniversity.edu.ng)  
and <sup>4</sup>[bobagada@gmail.com](mailto:bobagada@gmail.com)

---

### Abstract:

Software Police is an expert system that investigates, detects, and reports plagiarized documents, instances of plagiarism and the degree of plagiarism (stylometry) on the internet. Software police officers also known as the business software alliance is the watchdog that helps to trace and report people who commit crimes of document duplication, plagiarism and copyright infringement. The aim of this software is to help curb the rate of which people copy other people's work (document) without proper reference. The developed system will go through any document it comes across on the internet and test for instances of plagiarism and the degree of plagiarism (stylometry) with the content of other documents on cloud and generate alerts to the control panel in a case where the level of plagiarism is above twenty percent. Document plagiarism detection algorithm with semantic method was used for the design, PHP a server side scripting language for the development of dynamic web applications was used in implementing this system, Code igniter 3.1 frameworks, Sublime Text3 integrated development environment, and MySQL database management for back end support were also used. At the end, Software Police was implemented as a functional system that accurately detects plagiarized content of files at a document level on the internet.

---

**Key Words:** Software Police, plagiarism, plagiarism detection, plagiarism investigation, expert system.

### I INTRODUCTION

Software police is a proposed software program that will help curb crimes of Copyright infringement which broadly describes practices such as software piracy and plagiarism. This project focuses mainly on plagiarism which is majorly predominant and often times casually regarded, especially within the educational sector in Nigeria.

According to Bailey (2011), cited by Aboyade, (Aboyade, 2015), copyright is viewed from the fact that anything that is created is an extension of “self” and in that case it must be protected from general use by any one. That is to say, such creation should be secured from being used anyhow by any one for the purpose of self-interest. To these writers, copyright encourages the growth of writing, performing and creating of artistic works. In the absence of copyright there will be little or no encouragement for people to create anything, as most people will use creations of other people to enrich themselves illegally.

Aboyade et al., (2015) noted that students and teachers need vital information to excel in their academic quests. They need information to develop their social, economic and political experiences. It is no doubt that education promotes both economic and technological development of any nation, therefore ideas and insights for achieving these are contained in the works of people which can be tapped to achieve that result. This is because a nation that is educated will develop and a nation that is developed will have opportunity to educate her citizens.

Although plagiarism is often considered from the point of view of ethics while copyright violation is seen as a legal issue, we are of the view and rightly so, that the root of both cases lies in proper knowledge of intellectual property management. Because publishers who have invested resources with the expectation of recouping such with profit have pushed the issue of copyright violation to a legal point should not make us lose sight of the fact that every form of intellectual investment needs protection. This includes all essays for as long as they have an original creation.

In Nigeria to be precise, there are many cases of copyright infringement that manifest in form of unauthorized reprography, plagiarism, and piracy. This study is focused in the area of plagiarism which constitutes one of the commonest and yet little discussed copyright violations in Nigeria. Plagiarism is a plague that is ravaging our educational system at all levels and it is not peculiar to us in Nigeria alone but the world over. Software police is a proposed plagiarism detection system to be adopted to help curb the rate of plagiarism practices on the internet. The benefits of implementing this system are that it could save time for academic staff trying to detect plagiarism in student's project and mitigate plagiarism within the institution.

Software police is proposed software that will help detect and report people committing these crimes on the internet: Software piracy, plagiarism, copyright infringement and so on. Further research is progressing on this expert system in other to curtail so many internet crimes.

## II LITERATURE REVIEW

### a) Plagiarism and Source Deception Detection Based on Syntax Analysis

According to Anzelmi *et al*, (2011), in this research, the shingle algorithm with Jaccard method is employed as a new approach to detect deception in sources in addition to detect plagiarism. Source deception occurs as a result of taking a particular text from a source and relate it to another source, while plagiarism occurs in the documents as a result of taking part or all of the text that belong to another research, this approach is based on Shingle algorithm with Jaccard coefficient , Shingling is an efficient way to compare the set of shingle in the files that contain text which are used as a feature to measure the syntactic similarity of the documents and it will work with Jaccard coefficient that measures similarity between sample sets . In this proposed system, text will be checked whether it contains syntax plagiarism or not and gives a percentage of similarity with other documents. As well as research sources will be checked to detect deception in source, by matching it with available sources from turned in report of the same research by using shingle algorithm with Jaccard coefficient. The motivations of this work is the discovery of literary thefts that occur on many researches and the deception that occurs in the sources.

## **b) Plagiarism and Self-plagiarism**

Taken from Digbijay 2013 plagiarism is seen as an unauthorized appropriation of other people's ideas, processes or text without giving correct credit and with intention to present it as own property. Appropriation of own published ideas or text and passing it as original is denominated self-plagiarism and considered as bad as plagiarism. The frequency of plagiarism is increasing and development of information and communication technologies facilitates it, but simultaneously, thanks to the same technology, plagiarism detection software is developing. Within academia, plagiarism by students, professors, or researchers is considered academic dishonesty or academic fraud, and offenders are punished by sanctions ranging from suspension to termination, along with the loss of credibility and perceived integrity. When we talk about self-plagiarism avid B. Resnik clarifies, self-plagiarism involves dishonesty but not intellectual theft. (Gipp, 2014) offers a useful classification system including four types of self-plagiarism: A duplicate publication of an article in more than one journal partitioning of one study into multiple publications, often called salami-slicing; - text recycling; and- copyright infringement. In cases of proven plagiarism and academic self-plagiarism consequences may include; the author is obliged to withdraw the disputable manuscript which is already published or in different pre-publication stages.- In the event of co-authorship, the co-author must approve of publication withdrawal, even if the misconduct is not related to them.- Publications proved to be false by the Commission are erased from authors bibliography or marked appropriately.- The procedure for detraction from academic degrees (MSc or PhD at the University is initiated if obtained based on false thesis or dissertation.- The procedure for detraction from scientific and educational titles is initiated by a relevant body if based on false publications or other.

## **c) Plagiarism A Noble Misconduct**

According to Md. Rezwanaur Rahman, (2015), serious misconduct, penalty depends on the severity of plagiarism. It ranges from formal disciplinary action (apology letters, retraction of the published article to criminal charges (suspension and prosecution of authors due to the lack of knowledge on plagiarism or awareness among the authors, editors, reviewers, and educational institutions some types of plagiarized articles are allowed to be publish unknowingly. All writers must check for the text duplication unintentionally by using plagiarism detection software before submitting to any journal office. Reviewers also should use plagiarism detection tools in order to avoid false publication practice and finally the editor of the journal should finalize the fate of the article based on the extent of plagiarism by using powerful plagiarism detection software. Academic plagiarism is more easily detected by the software as Turnitin and Safe Assign and scientific plagiarism with CrossCheck and eTblast software. The software consists of algorithms to detect similarities, associated databases and web sites by which it compares the article. Sometimes simple Google Search also helps in detecting plagiarism. It is very easy to find information on a topic that needs to be explored, but it is not always easy to add that information to own work and not to create a plagiarism. There are several ways to avoid plagiarism: paraphrasing - Important information written in own words. Quote - It is literally the wording of certain authors and the sentences are always placed in quotes. Citation - Citing is one of the effective ways to avoid plagiarism. This usually entails the addition of the author(s) and the date of the publication or similar information. Standard document formatting guidelines i.e. APA, MLA, Chicago, etc. are used.

## **d) Academic Plagiarism Detection: A Systematic Literature Review**

Rephrased from Tomas Foltynek, Norman Meuschke and Bela Gipp, (2019), this article summarizes the research on computational methods to detect academic plagiarism by systematically reviewing 239 research papers published between 2013 and 2018. To structure the

presentation of the research contributions, it proposes novel technically oriented typologies for plagiarism prevention and detection efforts, the forms of academic plagiarism, and computational plagiarism detection methods. We show that academic plagiarism detection is a highly active in research field. Over the period of review, the field has seen major advances regarding the automated detection of strongly obfuscated and thus hard-to-identify forms of academic plagiarism. These improvements mainly originate from better semantic text analysis methods, the investigation of non-textual content features, and the application of machine learning. We identify a research gap in the lack of methodologically thorough performance evaluations of plagiarism detection systems. Concluding from our analysis, we see the integration of heterogeneous analysis methods for textual and non-textual content features using machine learning as the most promising area for future research contributions to improve the detection of academic plagiarism further.

To structure the presentation of the research contributions, we propose novel technically oriented typologies for plagiarism prevention and detection efforts, the forms of academic plagiarism, and computational plagiarism detection methods. We show that academic plagiarism detection is a highly active research field. Over the period we review, the field has seen major advances regarding the automated detection of strongly obfuscated and thus hard-to-identify forms of academic plagiarism. These improvements mainly originate from better semantic text analysis methods, the investigation of non-textual content features, and the application of machine learning.

#### **e. Fostering academic competence or putting students under general suspicion? Voluntary plagiarism check of academic papers by means of a web-based plagiarism detection system**

In view of what Kohl (2011) says, the increasing number of cases of plagiarism and the ease of use of online published texts, universities are faced with a considerable challenge to prevent and take action against plagiarism in academic student papers. In reaction to plagiarism, web-based plagiarism detection systems (PDSs) are increasingly used to check submitted papers - this checking entails various problems, for example the percentage of plagiarism found is only an indication of the actual extent of plagiarism and not all types of plagiarism can be identified. To cope with this problematic situation the voluntary plagiarism check (VPC), an alternative preventive university didactic concept, was developed at the University of Education, Freiburg Germany. It focused on the development of individual skills. Students were able to submit their academic papers example an undergraduate paper, final thesis anonymously. These were then tested with the PDS Ephorus. Following interpretation and summary of the findings by the project team - plagiarism as well as referencing mistakes - we advised the students on a suitable approach to academic writing based on their own typical mistakes. The VPC was conducted as a three-semester research project and was later evaluated. About 500 academic papers were tested. In 90% of the undergraduates' work incorrect and/or missing citations were found. This high percentage decreased among students in later semesters. Instances of plagiarism were detected in about 40% of the papers when the texts of advanced student's semester were tested. At the same time the length of the plagiarized texts decreased. Around half of the students stated that it was acceptable to copy single sentences or short passages from other sources without citation; they did not consider plagiarizing on a limited scale as cheating. A similar number of students admitted to having doubts about whether they could write a good paper without plagiarizing.

### **III. STATEMENT OF THE RESEARCH PROBLEM**

Literature, both recent and past has established that copyright laws are very important for better creativity on the part of the authors and protection of intellectual content of their work. However, it is not certain if information users actually have a reasonable level of knowledge of these copyright laws as a means of protecting the intellectual creativity of respective authors. They might

have reasonable level of knowledge of these but still infringe on the laws. It is also believed that in Nigeria, the copyright law is not as effective as expected since no proper mechanism is put in place to curb infringement. Various forms of infringement are still practiced on a reasonable scale in higher institution of learning despite the copyright laws. Presently, there are web-based plagiarism detection systems (PDSs) that users can call up for in order to check the degree of plagiarism in a work, this checking entails various problems, for example the percentage of plagiarism found is only an indication of the actual extent of plagiarism and not all types of plagiarism can be identified and this checking also is at the convenient of the users.

There are a number of potentially negative consequences of plagiarism:

- a) Plagiarism is a well-known and growing issue in the academic world. It is estimated to make up a substantial part of the total number of serious deviations from good research practice.(Titus S.L., 2008)
- b) For some journals it is indeed a serious problem, with up to a third of the published papers containing plagiarism. (Zhang, 2010)
- c) If plagiarism remains undiscovered, then the negative effects are even more severe. Plagiarists can unduly receive research funds and career advancements as funding agencies may award grants for plagiarized ideas.

#### **IV. RESEARCH OBJECTIVES**

In order to alleviate the problems encountered with the existing system of users calling up and checking for plagiarism at their convenient and only when necessary, the following objectives were considered:

- a) Propose an AUTO expert plagiarism checker on the internet.
- b) Propose an expert system (software police) that detects, and report plagiarized documents, instances of plagiarism and the degree of plagiarism (stylometry) to the control panel on the internet.
- d) Develop the proposed system.
- e) Evaluate the system performance capability in detecting and reporting plagiarized documents, instances of plagiarism and the degree of plagiarism (stylometry) to the control panel on the internet.

#### **V. SYSTEM DESIGN**

For the development of the proposed Software police that will be used to combat plagiarism, we have the block diagram in figure 1, the unified modeling language diagram in figure 2, figure 3 and 4 are the software police interfaces. The methodology adopted for this project work is OOADM (object oriented Analysis and Design methodology).

- a) The system block diagram

Figure 1 shows the system block diagram with four modules of the application; the investigating tools that was designed with Document plagiarism detection algorithm with semantic method that scrutinize the overall document in comparison with the content of document in the cloud through the machine-readable knowledge base, bring out the level of plagiarism in it and sent result to the

reporting module where the summary of the investigation are kept before they are moved to the control panel.

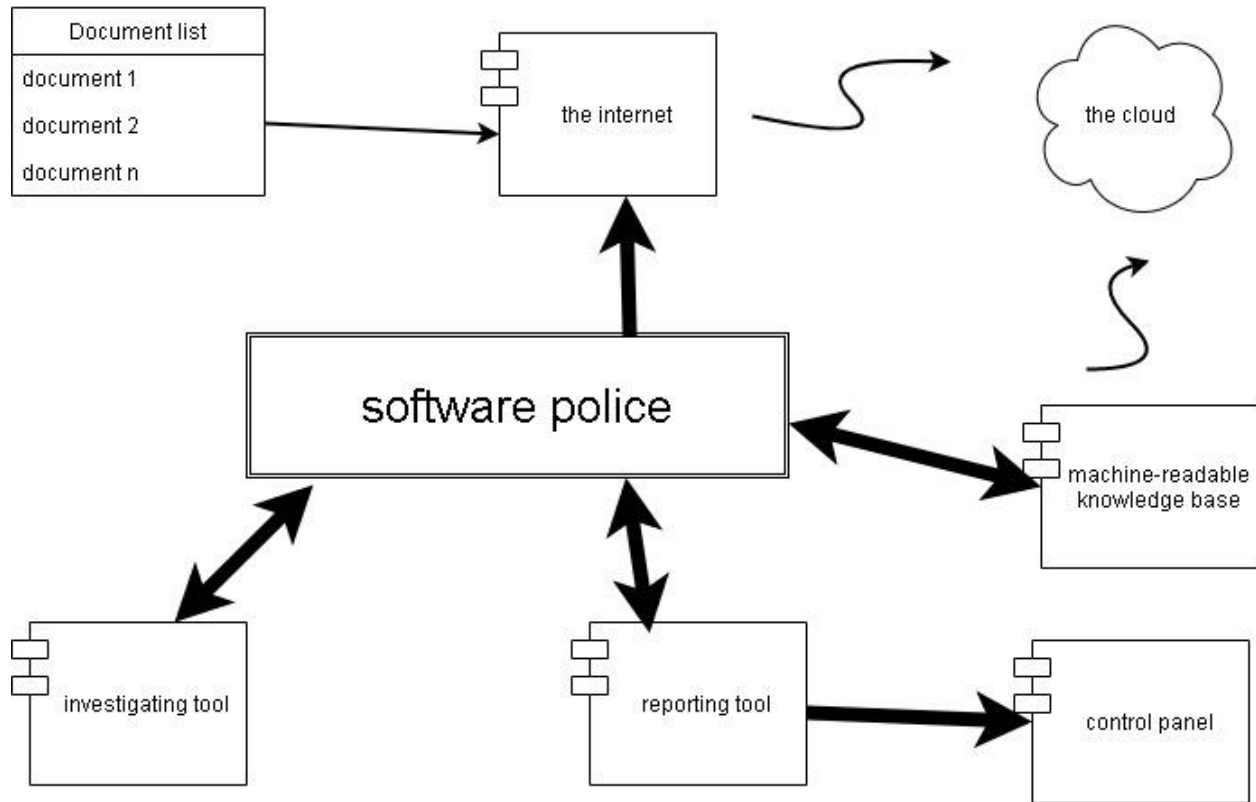


Figure 1: System block diagram

b) The system unified modeling Language diagram

The unified modeling Language diagram for this research work consists of four modules that are integrated for the development the application. The first module is the document detection and gathering module that detects all available documents on the internet for plagiarism checking. The second module is the investigation tool that is the plagiarism detector that was designed with Document plagiarism detection algorithm through semantic method. It investigates, detects, and reports plagiarized documents, instances of plagiarism and the degree of plagiarism (stylometry), it also compiles the status of all the document it has scrutinized and send to the reporting tool. The third module is the reporting tool that separate the status of the document whether plagiarized or not and sends the details of the status to the control panel. The last module is the control panel that can be deployed and monitored only by cyber security personnel, from the control panel, plagiarized document can be blocked and circulate on the internet for viewers. Documents that are not plagiarized are sent back to the internet with recommendation reports. With this system attached to the internet as plagiarism watchdog, writers, Journal editors and reviewers will be careful with what they publish on the internet for public consumption.

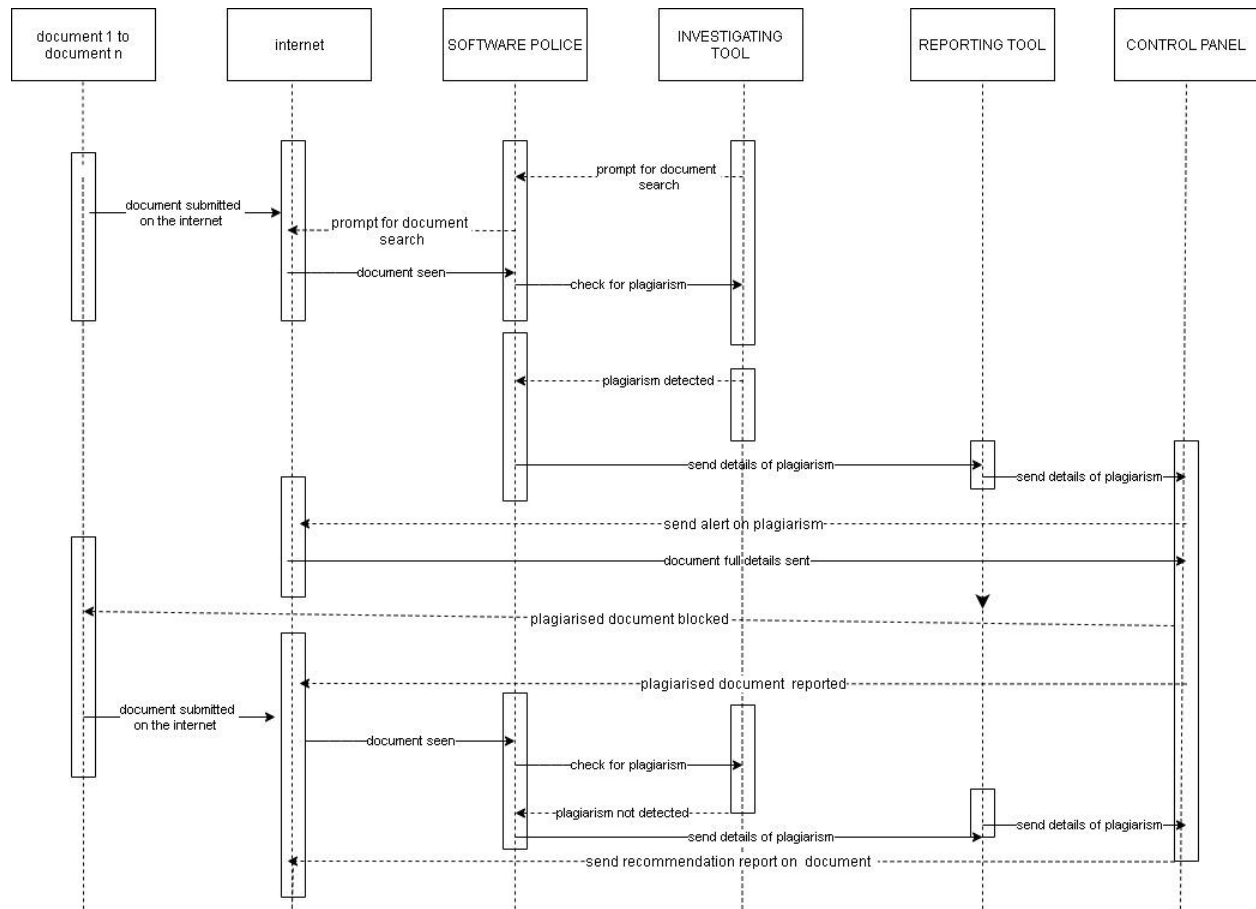


Figure 2: the overall System processes

c) The program design

The interfaces of the application will be highly confidential, so for this paper, we will show only two interfaces at the control panel end as shown in figure 3 and figure 4. This consists of the registration of the cyber security personnel with their biometric authentication (iris or blood authentication). From the control panel, the three modules (document detection and gathering, investigation tool and the reporting tool) can be initiated.

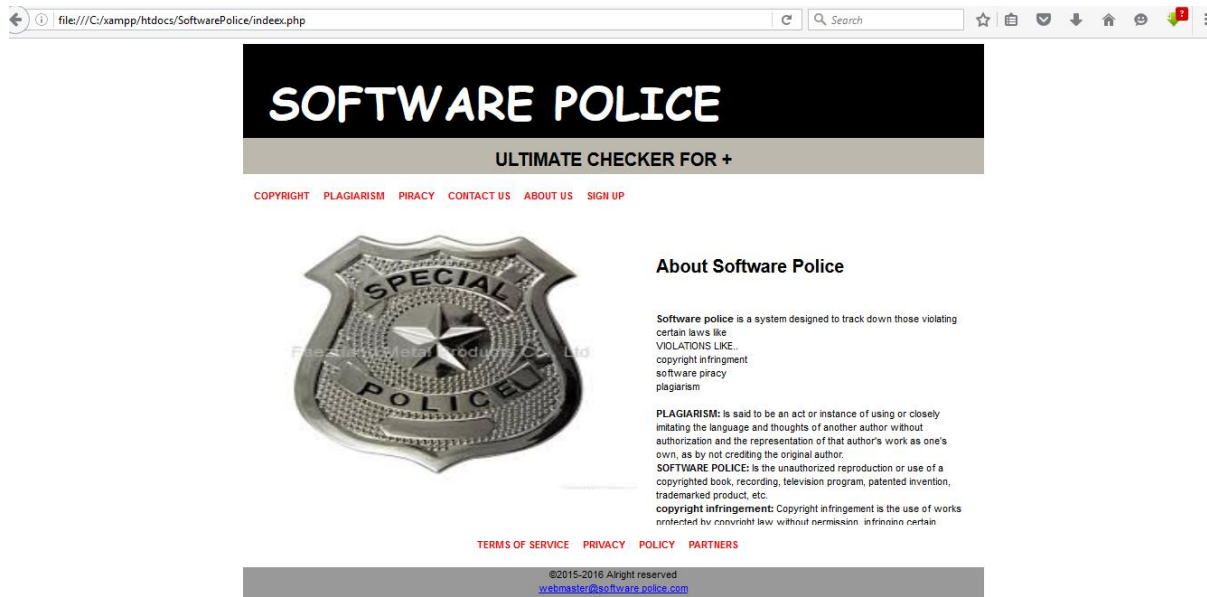


Figure 3: the control panel



Figure 4: the control panel login

## VI. CONCLUSION

The design and implementation of this expert system called Software police on the internet will curb the rate of plagiarism globally and force writers, Journal editors and reviewers to conduct plagiarism check on all document before public consumption. With this system attached to the internet as plagiarism watchdog, writers, Journal editors and reviewers will also be careful with



what they publish on the internet. The system will be able to automate plagiarism checking of document on the internet. When deployed, the overall objectives will be achieved.

## REFERENCES

- Aboyade, W. A. (2015). Copyright infringement and photocopy service among university students and teacher in Nigeria. *International Journal of Art and Sciences.*, 463.
- Anzelmi, D., Carlone, D., & Rizzello, F. (2011, 1 1). *Plagiarism Detection Based on SCAM Algorithm*. Retrieved 03 20, 2022, from <https://www.forskningsdatabasen.dk/>
- Bailey, J. (2011). *PlagiarismToday*. Retrieved April 27, 2022, from <https://www.plagiarismtoday.com/2011/03/03/the-3-uses-for-plagiarism-detection-tools/>
- Digbijay Guha, R. R. (2013). Cyber Crimes- Challenges & Solutions. *International Journal of Computer Science and Information Technologies, IV(5)*, 4.
- Gipp, B. (2014, 1 1). *Citation-based plagiarism detection detecting disguised and cross-language plagiarism using citation pattern analysis*. Retrieved March 20, 2022, from <http://cdsweb.cern.ch/>
- Kakkonen, T., & Mozgovoy, M. (2010, 1 1). *Hermetic and Web Plagiarism Detection Systems for Student Essays--An Evaluation of the State-of-the-Art*. Retrieved 03 20, 2022, from <https://www.science.gov/>
- Kohl, K. E. (2011). *Directory of Open Source Access Journals (SWEDEN)*.
- Maxim Mozgovoy, T. K. (2010). Automatic Student Plagiarism Detection: Future Perspectives. *Research Gate*.
- NORMAN MEUSCHKE, a. B. (2019). Academic Plagiarism Detection: A Systematic Literature Review. *ACM*, 52(6).
- Perkins, S. (2022). *How Technology has Changed Crime*. Retrieved Apr 17, 2022, from <https://study.com/academy/lesson/how-technology-has-changed-crime.html>
- R., F. Z. (2016, 6 1). *Detection of bilingual plagiarism*. Retrieved 3 20, 2022, from <https://doaj.org/search#.V6PA8bgrK70>
- Ranal. (2018). *UK Physical Sciences Centre- Briefing Paper*. Retrieved March 2022, from [https://www.heacademy.ac.uk/system/files/new\\_plagiarism.pdf](https://www.heacademy.ac.uk/system/files/new_plagiarism.pdf)
- Shababi, M. (2012). *International journal of computational linguistics*. Retrieved 2022, from <http://www.cscjournals.org/manuscript/Journals/IJCL/Volume3/Issue1/IJCL-33.pdf>
- Titus S.L., J. W. (2008). Repairing research integrity.
- Ugo Chidera Chinedu, D. C. (2020). Plagiarism Detection Systems . *International Journal of Scientific and Research Publication, 10(3)*.
- UKEssays. (2018). Retrieved April 1, 2022, from <https://www.ukessays.com/essays/computer-science/proposed-system-plagiarism-detection-9544.php?vref=1>

University of Oxford. (2018). Retrieved 2022, from  
<https://www.ox.ac.uk/students/academic/guidance/skills/plagiarism>

WIKEPEDIA. (2022). *Plagiarism*. Retrieved march 2022, from  
<https://en.wikipedia.org/wiki/Plagiarism>

Williams, J. B. (2002). *THE PLAGIARISM PROBLEM: ARE STUDENTS ENTIRELY TO BLAME?* Retrieved 2022, from  
[https://www.researchgate.net/publication/27481233\\_The\\_Plagiarism\\_Problem\\_Are\\_Students\\_Entirely\\_To\\_Blame](https://www.researchgate.net/publication/27481233_The_Plagiarism_Problem_Are_Students_Entirely_To_Blame)

Yousef K, S. M. (2017). THE PLAGIARISM DETECTION SYSTEMS FOR HIGHER EDUCATION- A CASE STUDY IN SAUDI UNIVERSITIES. *International Journal of Software Engineering & Applications (IJSEA)*, 8(2), 17.

