-16-

DETECTING UNACKNOWLEDGED PLAGIARISM USING STRING MATCHING-BASED CONTENT FRAMEWORK

[1]Lawal O.O, [2]Ajibode, T.S, and [3]Adepoju, A.O
[1, 3]Department of Computer Science, D.S Adegbenro I.C.T Polytechnic, Itori, Ogun State
[2]Dept. of Computer Science, Federal College of Education (Tech) Akoka Lagos State.
Corresponding author: opepolawal@gmail.com, 08062160129

ABSTRACT
Plagiarismisanugly trend that posed global concerns and challenges to information resources and documents spanning the internet repository and academia. Most often, an individual or some authors in teamwork usually have parallel construction in written text, term paper, and research article which connotes similar phrases, vocabulary overlap, and collaboration between texts across multiple pages of the same document. In this paper, we present a functional mechanism for higher performance in plagiarism detection especially in a collaborative manner. An iterative development technique was adopted involving 'String Matching Based Content' to handle text parsing and document analysis from file to file. The system functionality was modelled from the users' perspective by providing an algorithm for internal control and to check text similarity.
Keywords: Algorithm, Detection system, Plagiarism, String matching

1. INTRODUCTION

Plagiarism is an attempt and/or act of claiming credit for written ideas, published text, and intellectual works of others in academia. It is a trait of academic dishonesty involving undergraduate and postgraduate students, as well as intermediate researchers in different fields of endeavor. Although, this has enhanced teaching and learning the havoc done to academia is high, one of which is plagiarism. Digital computers and the advent of the internet have made published text plagiarism-prone in every facet of life (Faloore, 2014). However, plagiarizing by use of substitutions to elude detention software has rapidly evolved as students and unethical academics seek to stay ahead of detection software (Grove & Jack, 2014).

Predicated upon an expected level of learning being achieved, all associated academic accreditation becomes seriously undermined if plagiarism becomes the norm (Cully, 2013). On the other hand, Plagiarism is the stealing of someone's idea and claiming to be one's original work which may be written ideas (Atayero et al., 2011). Plagiarism is the illegal presentation of someone's ideas, words, expressions, data, computer programming, or any other creative endeavor and claiming it to be one's original work without giving the author credit. Since the emergence of digital computers and the internet, documents can no longer be handled in paper and ink form as they were decades ago. This has led to the “copying and pasting” of textual documents since these menus cannot be disabled in application programs to curb plagiarism.

Obinna (2012) noted that plagiarism and poor writing skills are the bane of Nigeria's educational system. Plagiarism affects not only the integrity of the individual concerned but also the integrity of the institution associated with the individual. Students' level of knowledge and ability in their respective places of work after graduation is a result of their disciplinary engagement in achieving the integrity of their academic experience. Ignorance and unawareness of the legal implication of plagiarism are also responsible for the verbatim copying of scientific texts (Jamali et al., 2014). Recently, computer-aided methods of plagiarism detection are in use but not all already existing tool simplement completely new comparison algorithms (Atayero et al.,2011).

Nowadays,itisgenerallybelievedthatattributecountingisinferiortocontentcomparisonsinceevensmal lmodifications can greatly affect fingerprints (Atayero et al., 2011). Meanwhile, a clever plagiarist that knows the best way of paraphrasing can easily use word substitution to outsmart the detection system. Parsing and document analysis a requite germane to the performance of any plagiarism detection system. The functional mechanism may be a desktop-basedor web-based application for document comparison or text similarity. It appears that developing countries suffer more from severe and obvious cases of plagiarism since the scientific foundation of developed countries was long established and the fight against plagiarism started earlier (Jamali et al.,2014).

Inconsistencies within the written text itself such as changes in vocabulary, style, or quality of an article can also be attributed to plagiarism except that such cases are known or described as unacknowledged and collaborative kind of plagiarism. Atayero et al. (2011) described unacknowledged and collaborative plagiarism as the occurrence of similar phrases, quotations, sentences, or parallel construction in two or more pages of the same article. Consequently, situationsinwhichspellingmistakesorlexicalerrorsoccurbetweentext;overlappingone'sphrasesandco ntentwithin a document, and as well as preparing correctly cited and referenced write-ups from individual research or handling part of that work twice for separate purposes or publications like journal articles, therefore, a framework to provide detection mechanism using string matching based content in two or more documents for text similarity with unlimited text size to check for unnecessary tautology is of necessity.

2. LITERATURE REVIEW

Plagiarism has been defined as an act of reproducing another individual' side as or any other form of information without giving credit to the source of the information. Plagiarism is a deliberate appropriation of others' ideas and statements without proper referencing (Smith et al., 2007). However, there are currently many measures designed towards combating it, such as the implementation of plagiarism detecting systems, as well as establishing penalties for the act. The advent of the Internet, the World Wide Web, and Social Media has undoubtedly taken our civilization a step higher, and especially, provided scholars with a useful avenue for the rapid spread and exchange of ideas and information. Centre for the Study of higher education said “Plagiarism in higher education can take many forms. Some of the more common forms are listed below, however, it

should be noted that definitions of plagiarism vary somewhat across the disciplines following differences in knowledge of authorship conventions and traditions.

Accordingly, the 'who is who' among American high school students- a survey conducted on American best students revealed that eighty percent (80%) of them plagiarized to get to the top of their class, as corroborated by another study in which fifty-five percent (55%) of the respondents admitted to having plagiarized (Wanetal.,2011). Dupree and Sattler (2010) pointed out that seventy-four percent (74%) of students in their report admitted to having plagiarized in the previous year. In addition, students and faculty admitted that plagiarizing via the Internet is fairly common. One of the causative factors of plagiarism could be unawareness. Wanetal. (2011) on engineering students studying in Malaysia revealed that most of the students know very little about plagiarism and what makes it a serious offense.

Furthermore, the research revealed that the students had not been given a very formal and concise orientation about it by the school. Although, at the moment, several plagiarism detection tools have been developed which effectively detect plagiarism. Alietal.(2011) reviewed these tools, pointing out their merits and demerits. Their findings show that none of the tools are hundred percent (100%) efficient, but considering the different features of the existing detection tools, they argue that a hybrid system could raise the capability of detection. In Nigeria, there are disturbing reports about a plague of plagiarism at different levels of the academic sector (Onuoha & Ikonne, 2013). Studies from the westernregionofNigeriaonuniversitystudentsrevealedvaryingacademicdishonestyamongrespondents,ranging from those that plagiarize from the Internet to those that copy from their colleagues (Babalola,2012).

This fault in the educational background of some Nigerians contributed to their failure to perform optimally abroad where academic rules are taken seriously (Orimetal.,2013). Confusion between plagiarism and paraphrasing among students is another influencing factor of plagiarism. Quite a significant number of students are unaware of the rules guiding paraphrasing. In fact, this is common when students are confronted with paraphrasing paragraphs from unfamiliar subjects or technical jargon. Students fall prey to unintentional plagiarism due to their inability to decipher the thin line between paraphrasing and plagiarism. Of utmost importance also is the place of poor writing skills of students among the various factors and reasons students plagiarize. Faculty members in related courses in a department must help students develop strong writing skills.

3. METHODOLOGY

Given the problems inherent in the existing system, it is necessary to seek an improvement because a hash function computes digital fingerprints for each chunk which are inserted into a hash table: a collision of hash codes within the hash table indicate matching chunks. Meanwhile, a clever plagiarist can easily maneuver by smooth paraphrasing; a detection mechanism with a string-matching algorithm as a functional framework is bound to perform better. Most of the works in plagiarism detection are meant for academic purposes where cheating, rewording, rephrasing, or restating

without referencing are strictly prohibited. In this regard, numerous plagiarism detection systems have been developed.

But, the functional mechanism of the proposed framework is stand-alone with the following operational modules and algorithms:

1. Load files to be checked
2. Ignore white spaces and special characters
3. Compare the files line byline
4. Detect similar phrases
5. Extract and store in memory the similar phrases
6. Check the online server for similar phrases
7. Displays the links having similar phrases.

```
Int match          //value to return int I, j,k;
match = -1; j=1; k=1; i=j;
while (end Text (T,j)==false) if (k>m)
match=I                    //match found break;
if(t; ==p)j++; k ++;
 else                      / / b a c k    u p
          overmatched characters int backup
          =k-1;
          j=j-backup;
          k=k-backup;  //slide  pattern  forward,
          start over j++; i=j;
 return match;
```

System design

An iterative development technique was adopted in providing a functional mechanism of the proposed frame work which was code-named “plag-detect”; blueprints were presented through the use of visual aid like procedure, data flow, activity, and use case diagram. Meanwhile, the top-down approach was chosen for the design to build the system from the main module and later incorporate all other sub-programs as shown in Figure 1 and Figure 2below.

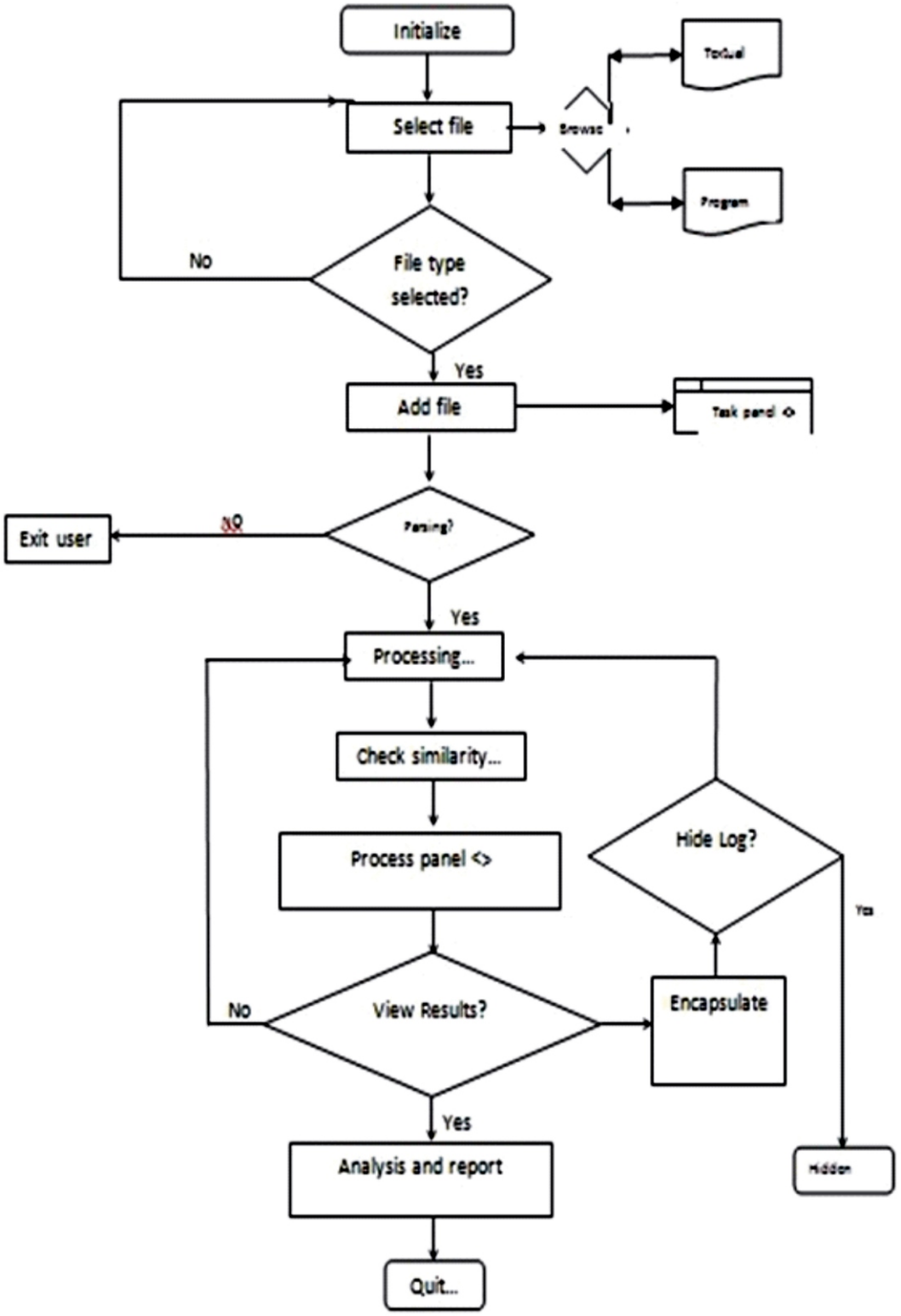


Fig. 1 Procedure and data flowdiagram

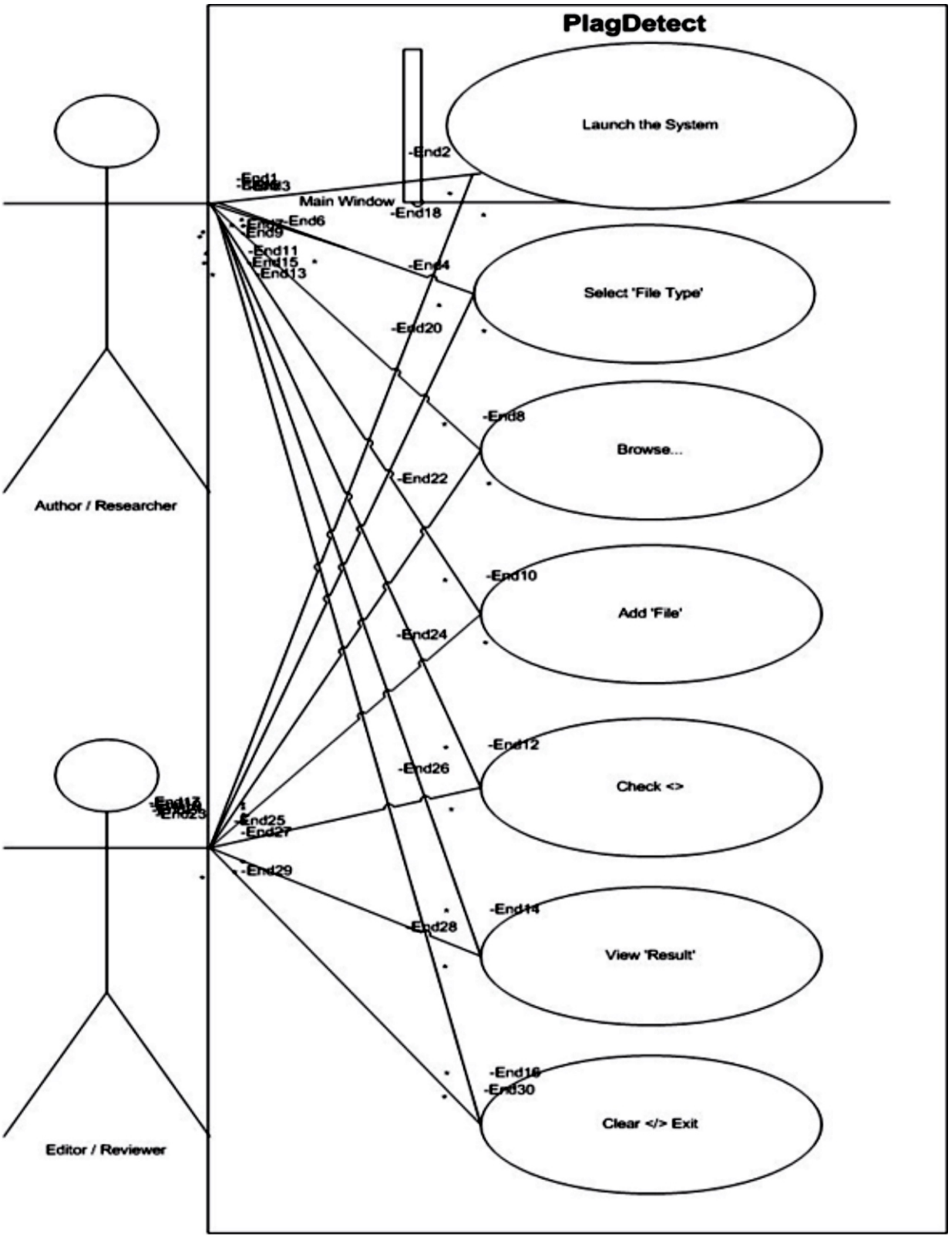


Fig. 2 Use casediagram

1. CONCLUSION

Natural language processing is of great importance to the functionality and modern construction of plagiarism detection systems as detecting plagiarized phrases in the given document becomes easier after the semantic analysis of the document. The algorithm has been provided with a formal representation as a detection mechanism using string-matching-based content for better performance. Thus, handles file-to-file similarity checks within textual search and program codes, like copy detection in large comparison from electronic resources. Enterprise-level plagiarism check software would normally compare input documents against databases.

2. FUTUREWORK

Computational intelligence is of great importance to the modern construction of plagiarism detection tools and their functionality. The application of natural language processing (NLP) to similarity tests and document analysis will be considered as part of the areas for further study.

Its features can leverage pattern matching to handle ambiguity and unconstrained vocabulary to resolve lexical overlap in large files. Multi-lingual capability can also handle language issues in detecting exact copied percentages in part or wholly, translating documents in different languages irrespective of the writing style of the author.

REFERENCES

Atayero,A.A.,Alatishe,A.A.,Sanusi,K.O.2011.DevelopingiCU:APlagiarismDetectionSoftware. *International Journal of Electrical & Computer Sciences.* 11(4), 27-32.

Ali, A.E.T, Abdulla, H.D., Snasel, V. 2011. "Survey of plagiarism detection methods". Proceedings of 5th Asia Modeling Symposium, Manila,39-42.

Bar, D., Zesch, T., Gurevych, I. 2015. Composing Measures for Computing Text Similarity. Technical report of the language lab., Germany.1-30.

Babalola, Y. T. 2012. Awareness and incidence of plagiarism among undergraduates in a Nigerian private university. Gale-Bult. 22(1),53.

Cully, P. 2013. Plagiarism avoidance in academic submissions. Dublin Institute of Technology, Full PDF available at: http://arrow.dit.ie/bescharcoth/4/. *Retrieved on 15/03/2016.*

Dupree, D., Sattler,S.2010.Texas Tech University Academic Integrity Survey Report. Retrieved from<http://www.depts.ttu.edu/provost/qep/docs/Dupree and Sattler Academic_Integrity_Report _Cover.pdf

Faloore,O.O.2014.TowardsaMoreEnduringPreventionofScholarlyPlagiarismamongUniversityStudents in Nigeria. *Global Journal of Human-Social Science: Sociology and Culture.* 14(6), 1-7.

Grove, M., Jack, W. 2014. "Roget would blush at the crafty cheek Middlesex lecturer gets to the bottom of meaningless phrases found while marking essays". URL<http://www.msn.com>.

Jamali, R., Ghazinoory, S., Sadeghi, M. 2014. "Plagiarism and Ethics of Knowledge": Evidence from International Scientific Papers. *Journal of Information Ethics*. 23(1),101-110.

Orim, S.I., Davies, J.W., Borg, E. 2013. Exploring Nigerian postgraduate students' experience of plagiarism: A case study. *International Journal for Educational Integrity.* 9(1),20–34.

Obinna, C. 2012. "Plagiarism, bane of Nigeria's educational development". Retrievedfrom <http://www.vanguardngr.com/2012/09/plagiarism-bane-of-nigerias-educational-devt-provost/> on April 24th, 2016.

Onuoha, U.D., Ikonne, C.N. 2013. Dealing with the Plague of Plagiarism in Nigeria. *Journal of Educationand Practice.* 4(11),1-11.

Smith, M., Ghazali, N., Minhad, S.F. 2007. Attitudes towards plagiarism among undergraduate accounting students. Malaysian, *Review of ACCA.*

Wan, R., Nordin, S. E., Halib, M.B., Ghazali, Z.B. 2011. Plagiarism among Undergraduate Students in an Engineering University: An Exploratory Analysis. *European Journal of Social Sciences.* 25(4),537-549.