

A Framework for an Anti-Plagiarism System Across Nigerian Tertiary Institutions

*Oladeji, F.A., Ajayi, O. O., Okunoye, O.B. & Nwodo, N.

Computer Sciences Department
University of Lagos
Akoka, Lagos State, Nigeria
[*foladeji@unilag.edu.ng](mailto:foladeji@unilag.edu.ng)

Oyetunji O.M.

Department of Computer Science
Mountain Top University
Ibafo, Ogun State, Nigeria

Idowu P.A.

Department of Computer Science & Engineering
³Obafemi Awolowo University
Ile-Ife, Osun State, Nigeria

ABSTRACT

In Nigerian tertiary institutions as at today, most students' documentations are not subjected to proper archiving. While credible ones are dumped and remain unused most times; thus similar or exact works are replicated by other students in subsequent years or by students from other institutions. Repetition of the same argument, survey, and experiment is tantamount to re-inventing the wheels and affects the ability to restructure, re-innovate and contribute to existing models that could aid tackle emerging challenges in the industry. Electronic Theses and Dissertation (ETD) have the potential to solve this challenge. However for ETDs to be efficient, an Institutional Repository (IR) has to first be in place, as ETDs are to be housed in IRs. This paper follows up on previous survey carried out on the status of IRs across Universities in Nigeria. In this work, an extendable anti-plagiarism framework for Nigerian Universities is presented as well as a web-based working prototype showcasing certain aspects of the proposed framework.

Keywords: Anti-Plagiarism, Archiving, ETDs, Institutional Repositories, Nigerian Universities

Aims Research Journal Reference Format:

Oladeji, F.A., Ajayi, O. O., Okunoye, O.B., Nwodo, N. ,Oyetunji O.M. & Idowu P.A (2017): A Framework for an Anti-Plagiarism System Across Nigerian Tertiary Institutions. *Advances in Multidisciplinary & Scientific Research Journal*. Vol. 3. No.1, Pp144

1. INTRODUCTION

Plagiarism as defined by the Webster dictionary is the act of using another person's words or ideas without giving credit to that person. It is the theft of someone else's intellectual property and can thus be regarded as a form of theft on its own (Hexham, 2005). The problem of plagiarism in academia is a very serious one, copying and theft of other researchers' work simply runs contrary to the primary objective of setting up Universities in the first place which is to find solutions to challenges of mankind. This problem is now very prevalent in Nigerian Universities and institutions of higher learning, and is now considered one of the primary reasons why Nigerian tertiary Institutions chunk out half-baked graduates (Okeke, 2001).

On a positive note, this endemic has become a forerunner and stakeholders are even more aware of its menace now than ever. This has thus resulted in bodies such as the Nigerian University Commission (NUC) and the Committee of Vice Chancellors of Nigerian Universities (CVC) taking concerted steps to checkmate it in Nigerian institutions. In recent development, drastic measures have been taken by some Universities to tackle plagiarism; for instance some Universities have dismissed some lecturers including professors while others demoted members of staff on the grounds of plagiarism (Kayode-Adedeji, 2013, Kalu, 2016). The collaborations with TURNITIN, a software based plagiarism checker (Turnitin, 2013) to vet research papers and articles and the introduction and use of DSpace repository (Smith, et al. 2003) are other attempts being taken to mitigate plagiarism in Nigerian tertiary institutions.

Though numerous research works exist, many of these works focused on the causes of plagiarism and the techniques or tools that can be used to vet plagiarism in Nigerian institutions. The attempts to checkmate plagiarism though laudable, are far from a holistic approach. Most scholarly publications (theses, dissertations and journals) directly relating to students and researches in Nigerian Universities are printed on papers, bounded and arranged in library shelves or stores within the University. They remain unused and completely inaccessible to the likes of TURNITIN. There is therefore the need to create awareness and adopt Institutional Repository Systems (IRs) and Electronic Thesis and Dissertation (ETDs) to ensure that academic journals are digitally archived. Sequel to this, these repositories can then be linked and made available online; before tools like TURNITIN and other related anti-plagiarism can be deployed.

This work presents a framework for checking plagiarism in Nigerian higher institutions of learning. In this work, it is assumed that each institution has an IR system in place and publications are being archived to it. This proposed framework builds on this premise and presents a system that allows researchers and students alike to inquire of various topics of interest and is presented with results of similar publications existing in various IRs across the nation. The rest of this paper is broken down as follows: a survey of the status of IRs in major Nigerian tertiary institutions is done in section two. In section three the proposed Anti-plagiarism framework is presented. Section four shows an implementation of some aspect of the framework. Section five discusses open challenges to the implementation, while conclusion and future works are in section six.

2. RELATED WORKS

According to Okiy (2008) a lot of libraries in Nigeria are now being digitized both for the purpose of preservation of knowledge and for easier global access to scholarly materials. One such effort is that of the Association of African Universities (AAU) with her Database of African Theses And Dissertations (DATAD) programme, which was pioneered by the University of Jos and Obafemi Awolowo University (OAU) in 1980 and 1995 respectively. (AAU-DATAD, in Baro, et al., 2014). Though as at the time of writing it is difficult to ascertain if this programme is still operational.

Iwudu (2012) discussed the use of DSpace an Online Institutional Repository (OIR) at the University of Jos, Nigeria (UNIJOS). The author introduced and emphasized on the need for an OIR as the modern day library. The authors conducted a survey on the awareness levels, challenges and institutional benefits of DSpace at UNIJOS. Some of the reported challenges included poor or limited Internet access, limited technical know-how and possible litigation by publishers. It was concluded that despite the advantages of IOR, adoption rate amongst academics is still very minimal, principally because of copyright and safe guarding of intellectual properties. In the work of Onuoha and Ikonne (2013), the authors viewed Internet as the main catalyst for plagiarism in Nigeria. It was inferred that the proliferation of Internet enabled devices as one of the major factors that has aided the growth of plagiarism in Nigeria. Authors simply log onto the Internet, obtain related publications and copy aspects of these works as part of their own work without duly acknowledging or citing the original authors. Poor time management, defiance and not understanding the implication of plagiarism were also listed as factors aiding plagiarism. Finally, approaches to combating plagiarism were discussed with the authors proposing a multi-level approach of curbing at the state / governing level and at the class level by lecturers and advisers.

Idiegbeyan-ose, et al., (2016) reported that ignorance, skill deficiency, and academic pressure were the primary causes of plagiarism. The authors proposed a framework based on the theories of awareness and perception and tested this on post-graduate students in Ogun state, Nigeria. The authors' results show that students in private schools have a higher level of awareness about plagiarism versus their public school counterparts. This notwithstanding, the average awareness level perceived is still below acceptable standards. The author concluded by calling for an introduction of compulsory courses that focus on teaching ethics at the University level as well as enforcing the use of Turnitin by the National Universities Commission. In Oladeji, et al. (2016), a survey of the status of IRs systems in place in some tertiary institutions in Nigeria was done. This survey is as summarized in Table 1.

Table 1: Analysis of ETD / IR platforms across Nigerian Universities

INSTITUTION	IR PLATFORM	IR SIZE
Afe Babalola University	EPrints	1,000 – 1,999
Ahmadu Bello University, Zaria	OpenDoar	10,000 – 20,000
America University of Nigeria, Yola	N/A	< 500
Bayero University, Kano	DSpace	5,000 – 10,000
Covenant University, Ota	OpenDoar	2,000 – 5,000
Ebonyi State University,	OpenDoar	2000 – 5000
Federal University of Technology, Akure	OpenDoar	2,000 – 5,000
Federal University of Technology, Minna	OpenDoar	2,000 – 5,000
Federal University of Technology, Ndufu-Alike	OpenDoar	1,000 – 1,999
Federal University of Technology, Owerri	ENDNOTE / DSpace	2000 – 5000
Federal University of Technology, Oye	OpenDoar	1,000 – 1,999
LandMark University	OpenDoar	1,000 – 1,999
Nnamdi Azikiwe University, Awka	DSpace	1,000 – 1,999
Obafemi Awolowo University, Ile-Ife	DSpace	2,000 – 5,000
University of Benin, Benin City	DSpace	2000 – 5000
University of Ibadan	DSpace	5,000 – 10,000
University of Ilorin	OpenDoar	500 - 1000
University of Jos	DSpace, OpenDoar	1,000 – 1,999
University of Lagos	DSpace, OpenDoar	5,000 – 10,000
University of Nigeria, Nsukka	DSpace, OpenDoar	Above 20,000

Table 1 depicts the IR platform and the IR size (number of scholarly materials hosted) for each of the institutions considered. The table shows that the University of Nigeria, Nsukka has the highest number of scholarly materials at over 20,000; followed by the Ahmadu Bello University, Zaria and the Universities of Ibadan and Bayero. The America University of Nigeria, Yola had the smallest amount of hosted materials at less than 500 publications.

From the survey, it can be inferred that OpenDoar is the most used IR platform in Nigeria, being used by twelve (12) of the twenty institutions considered. DSpace comes second, being exclusively used by eight (8) of the twenty and jointly used with OpenDoar by three (3) other institutions. Other outlier platforms are EPrints used in Afe Babalola University and ENDNOTE used at the Federal University of Technology, Owerri; though this is being migrated to DSpace as reported in (FUTO Library, n.d).

3. PROPOSED MODEL: A CONCEPTUAL FRAMEWORK FOR ANTI-PLAGIARISM

The proposed framework is depicted in Figure 1 and is made up of two main parts: the Institutional Repositories Network (IRN) and the Anti-Plagiarism Engine (APE). These entities and their respective components are described as follows:

- a. The Institutional Repositories Network: This is a network of Institutional Repositories housing their various ETDs. The IRN specifies a standardized formatting style for which all subscribing institutions should adhere to. In order to prevent re-invention of the wheel, the IRN could simply be OpenDoar or DSpace as these are already in existence and are in by many of the tertiary institutions in Nigeria as shown in table 1.
- b. The Anti-Plagiarism Engine (APE): The APE interfaces between the users on one end and the IRN on the other. For instance, once users submit project or research titles inquiries to the APE and the APE returns exact or similar titles that exist and have been done in other institutions to the users. The APE consists of the following components:
 - i. The Web Portal: This is simply a browser-based interface through which users interact with the system.
 - ii. Harvester: This component crawls through the IRN searching for projects or researches with similar or exact titles to that requested by the user. Retrieved information are stored on the APE's database.
 - iii. Meta-data Translator (MDT): Though there is the option of using standardized IRN platforms, such as OpenDoar or DSpace, not all institutions uses these – as is the case with Afe Babalola University. To cater for such institutions, a MDT is required. The MDT on receiving data received from the various IRs formats these data into that which the APE can understand. This translation is done with the help of the meta-data and data schema provided by the IRs. All crawling operation done by the Harvester go through the MDT, however should the data be retrieved from an IR using an APE compatible schema, the MDT simply switches to transparent mode and does not interfere in the data collection process.
 - iv. The Database: This is the APE's internal information repository. For all data retrieved by the Harvester, certain aspects are stored on this database; these include the title, author(s), date, institution and source repository of the retrieved data. For subsequent user requests, the APE first queries the database, if no record is found, it then invokes the Harvester.
 - v. The Plagiarism Checker: This component searches through the database for titles similar to that requested by the user through the Web Portal. The checker uses a cluster based text detection system that is able to detect similar titles even if they have been rephrased. The Plagiarism checker replies the user with details of all similar publications via the web portal.
- c. The Users: These are individual authors, researchers, students or anyone interested in anti-plagiarism.

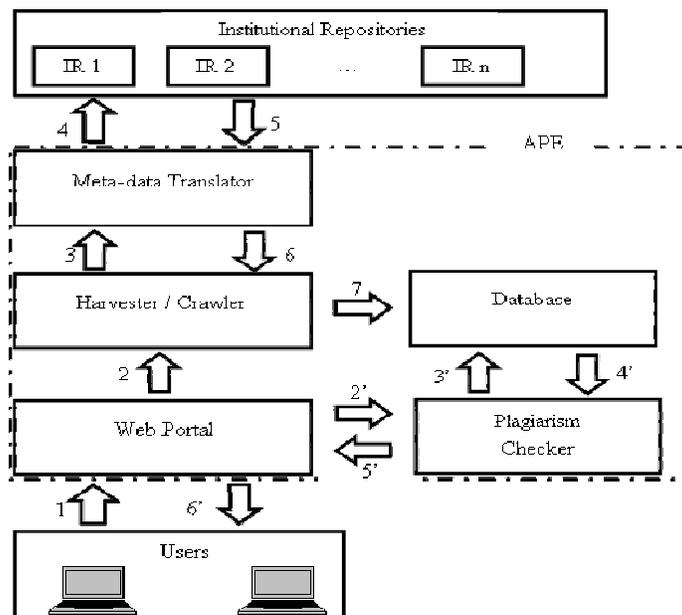


Figure. 1: Conceptual framework for Anti-plagiarism

Implementation

In this section, implementation of certain aspects of the anti-plagiarism framework is done, in particular the plagiarism checker. For easy deployment and open accessibility, a web-based version was implemented, hence does not require the installation of the system's components on users' computer. All that is required from the user, is a web browser and an active Internet connection. To this end, implementation of the web portal was done using HTML, JavaScript and PHP while the database was designed using MySQL.

A. Text Matching

Text matching was done using the Aho-Corasick algorithm (Hasib et al., 2013), which is shown in Algorithm 1. It consists of two phases - the Preprocessing and Searching or Matching phases and has a running time of $O(n)$ and complexity of $O(P+S+O)$ where: P = length of patterns, S = length of searched text, O = length of output matches. The preprocessing a Trie of all words is built. The trie has three functions: Go To, Failure and Output. While the matching phase, traverses breadth first over the given texts to find matching words

Algorithm 1: Aho-Corasick

```

Procedure enter( $a_1, \dots, a_m$ )
begin
 $s=0; j=1;$ 
While goto[ $s, a$ ]  $\neq \emptyset$  do // follow existing path
     $s = \text{goto}[s, a];$ 
     $b. j = j+1;$ 
For  $p=j$  to  $m$  do // add new path (states)
     $\text{news} = \text{news}+1;$ 
     $\text{goto}[s, a_p] = \text{news};$ 
     $s = \text{news};$ 
Output[ $s$ ] =  $a_1, \dots, a_m$  then end
begin
 $\text{news} = 0$ 
For  $i=1$  to  $k$  do enter(  $P_i$  )
    For  $a \in \Sigma$  do
If goto[ $0, a$ ] =  $\square$  then goto[ $0, a$ ] =  $0$  ;
end
Fail function
queue =  $\emptyset$ 
For  $a$  to  $\Sigma$  do
    if goto[ $0, a$ ]  $\neq 0$  then
        enqueue( queue, goto[ $0, a$ ] )
        fail[goto[ $0, a$ ] ] =  $0$ 
While queue  $\neq \emptyset$ 
     $r = \text{take}( \text{queue} )$ 
    For  $a$  to  $\Sigma$  do
        if goto[ $r, a$ ]  $\neq \emptyset$  then  $s = \text{goto}[ r, a ]$ 
            enqueue( queue,  $s$  ) // breadth first search
        state = fail[ $r$ ]
        While goto[state,  $a$ ] =  $\emptyset$  do state = fail[state]
        fail[ $s$ ] = goto[state,  $a$ ]
Output[ $s$ ] = output[ $s$ ] + output[ fail[ $s$ ] ]
Output function
Input: Text  $S[1..n]$  and an Aho-Corasick automaton  $M$  for pattern set  $P$ 
Output: Occurrences of patterns from  $P$  in  $S$  (last position)
state =  $0$ 
For  $i = 1..n$  do
While (goto[state,  $S[i]$ ] =  $\emptyset$ ) and (fail[state] != state) do
    state = fail[state]
    state = goto[state,  $S[i]$ ]
    If (output[state] not empty )
        then report matches output[state] at position  $i$ 
    
```

B. Functional Requirements

1. Authentication
2. The System must be able to check for matches in a document against several references of local contents efficiently

C. Quality Metrics (Non-Functional Requirements)

- ❖ Access Control: The System will have two access levels, administrator and user. Administrators have access to all aspects of the system, while users' access are restricted.
- ❖ User Friendliness: The System's interface should be easy to understand and use.
- ❖ Security: Data protection should be through the use of passwords to prevent unauthorized users to access.
- ❖ Error Handling: Reaction to special cases and undesired circumstances should be put in place to guarantee that the system works is fail safe.

D. User Classes and Characteristics

1. Administrator (Master User): An administrator is responsible for: Adding new users / admins, uploading new documents as reference corpus, deleting records, checking documents for Plagiarism and viewing plagiarism detection reports for all documents.
2. User: Can checking documents for Plagiarism, viewing plagiarism detection reports for all documents.

4. RESULT AND DISCUSSION

Figure 2 shows the interface of the landing page. This is the first page a user sees on launching the application. Here, the user is presented with three options: Login, Register or Demo. Registered users can login, new users have to register before they can get access, and the demo is for quick plagiarism check without logging in.

Figures 3 and 4 show the user registration and login pages, while figure 5 shows a text area in which content is copied (or typed) into for a quick plagiarism check under the demo option. The result of the plagiarism check illustrated in Figure 5, is shown in Figure 6, with a 75 % match in this instance.

Open Challenges

From literature, some of the major challenges experienced with IR are funding, lack of IT proficiency, unwillingness of authors to upload onto and use the IR systems, poor and epileptic Internet connectivity and power supply as well as copyright infringement. A survey of these challenges was presented in the previous work and summarized on Table 2

Table 2: Top Challenges to Deploying ETDs in Nigerian Universities

CHALLENGES	RANKING
IT proficiency	1
Internet Connectivity and Infrastructure	2
Poor / Limited Funding	3
Power outages	4
Copyright laws	5
Lax ETDs policies	6

Table 2 shows that IT proficiency is the main challenge to the deployment and usage of IR systems (Fabunmi, et al. 2006, and Warraich and Tahira, 2009). Though this challenge had earlier been reported, it important to emphasize that this does not directly apply to authors but to other personnel that interact with the IR system. From the authors' perspective however, Internet connectivity and power outages are the leading concerns, as it hampers the ability to readily check if a prospective work has already been done or if a plagiarized work is being submitted for assessment. Hosting IR platforms locally is a major challenge in most Nigerian tertiary institutions due to the incessant power outages. Thus for the IR systems to remain online and accessible always, the institutions either have to host their IRs abroad or run on generators. The cost of running and maintaining generators is also a big deterrent (Eke, 2011, Ezema and Ugwu (2013). Eke (2011) has reported that funds are a major challenge to deploying IR systems. Apart from being capital intensive project in itself, there is also the need to recruit or train members of staff as well as digitize hardcopies of archived publications. As is the case in many other tertiary and research institution globally, copyright remains a major challenge to the use of IR/ETDs, especially with respect to the protection of intellectual property rights.

With respect to deploying the proposed anti-plagiarism framework, the issue of uniformity of content is a major challenge. With no means of compelling individual institutions to adapting a particular IR, a situation of disperse IRs with often proprietary database schemas ensues. Integrating these into a uniform model is a major challenge and might require having access to numerous schemas in order for the proposed Meta-data translator to work efficiently.

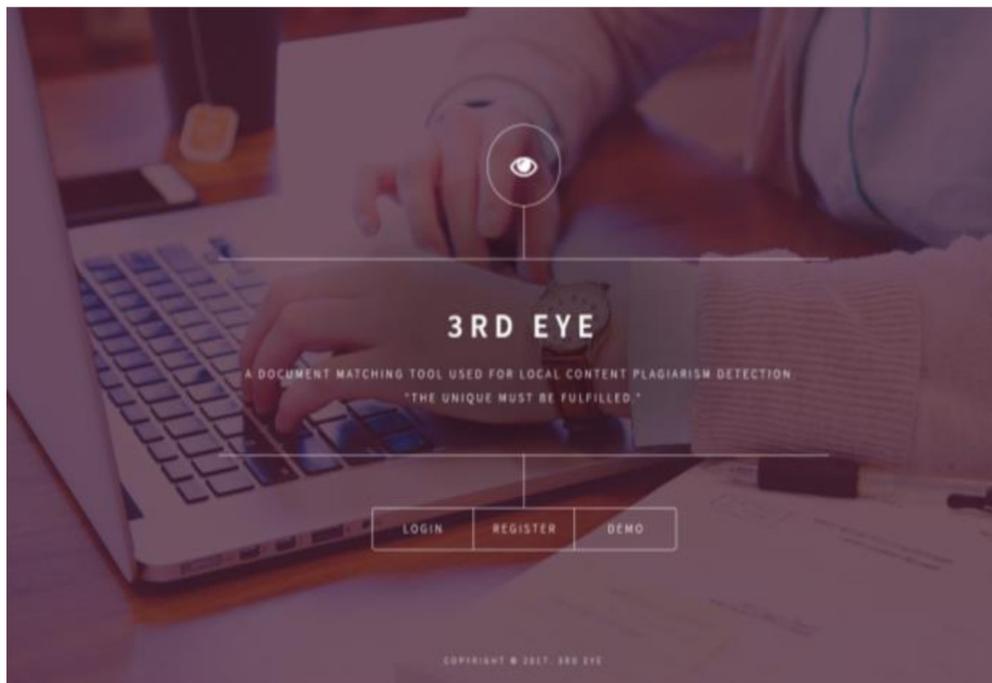


Figure 2: Landing page

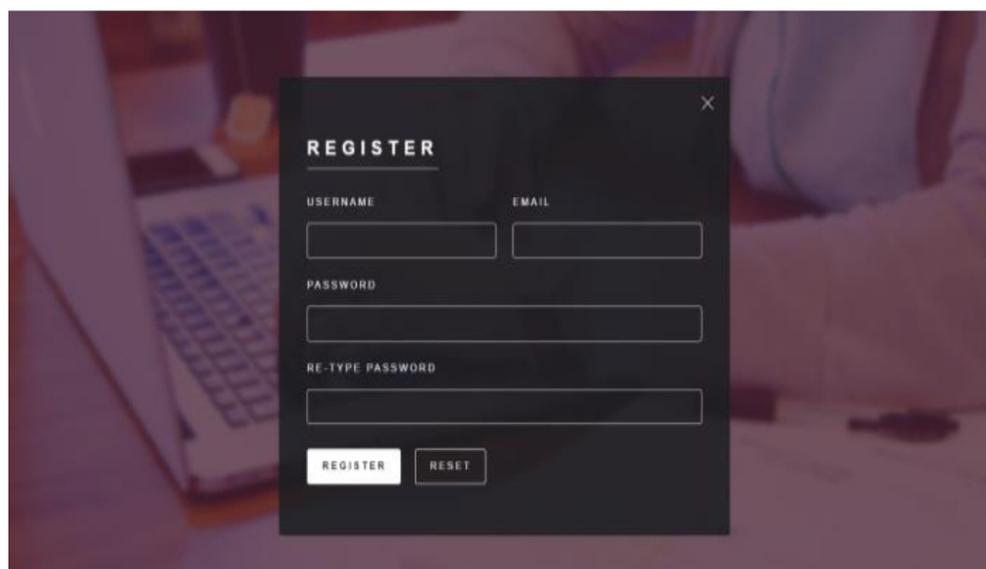


Figure 3: New user registration page

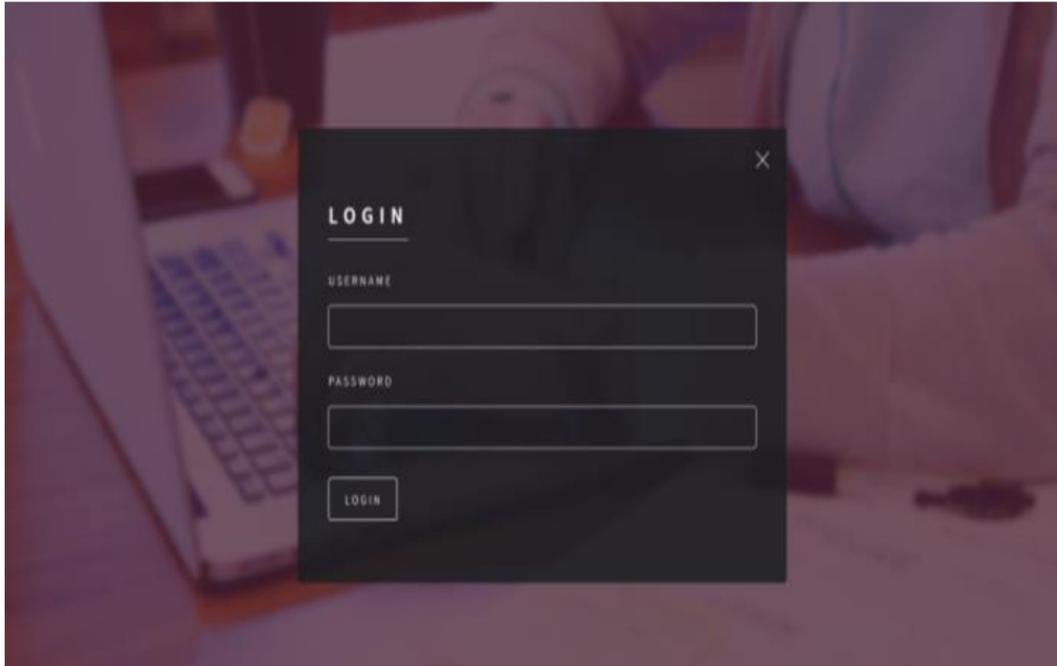


Figure 4: User login page



Figure 5: Demo plagiarism page

```
'android's file system layout consists of device storage, portable sd card, and device root' was found  
on page -- 1 of 504Assignment.pdf  
  
'secure erase on a managed device' was found on page -- 2 of 504Assignment.pdf  
  
'using mobile device management, a system admin can install or remove profiles, remove passwords, and  
begin ' was found on page -- 2 of 504Assignment.pdf  
  
Plagiarism Percentage: 75%
```

Figure 6: Result of plagiarism check

5. CONCLUSION AND FUTURE WORKS

ETDs and IRs have been identified as a potential solution to curbing plagiarism. They are the first steps that need to be taken in order to provide global visibility to scholarly materials produced within Nigerian Universities; after which anti-plagiarism software and tools can be deployed. In this paper, a survey of the status of IRs and ETDs in place within some Universities in Nigeria was done. A scalable anti-plagiarism framework for Nigerian Universities was presented as well as snapshots of implementation of certain aspects of the framework. The top challenges to the deployment and running of IRs systems were also presented. In future works, the authors seek to implement all aspects of the proposed anti-plagiarism framework. Lastly, research work still need to be done in the aspect of meta-data compatibility and the attempt to fuse multiple database schemas from numerous IR systems into a single one. This is a vital aspect of the proposed framework.

REFERENCES

1. Baro, E., Godfrey, V. and Eze, M. (2014), Electronic Theses and Dissertation Initiatives: A Survey of University Libraries in Nigeria, *International Information & Library Review*, 46(1-2), pp.41-50.
2. Eke, H. (2011), Digitizing resources for University of Nigeria repository: Process and challenges, *Webology*, 8(11), Article 85. Retrieved from <http://www.webology.org/2011/v8n1/a85.html>
3. Ezema, I. and Ugwu, C. (2013), Electronic Theses and Dissertations in Nigeria University Libraries: Status, Challenges and Strategies, *The Electronic Library*, 31(4), pp. 493-507
4. FUT0 Library (n.d.), The University Library, Retrieved from <http://library.futo.edu.ng/index.php/8-pages/4-joomla>
5. Hasib, S., Motwani, M. and Saxena, A., (2013), Importance of Aho-Corasick String Matching Algorithm in Real World Applications, *International Journal of Computer Science and Information Technologies*, 4(3), pp. 467-469
6. Hexham, I. (2005), The plague of plagiarism: Academic plagiarism defined. *Calgary: University of Calgary*.
7. International School of Management (2015), Academic Integrity and Culture Sensitivity, Centre for Academic Integrity Research and Anti-Plagiarism, White paper.
8. Idiegbeyan-ose, J. Nkiko, C. and Osinulu, I. (2016), Awareness and Perception of Plagiarism, of Postgraduate Students in Selected Universities in Ogun State, Nigeria, *Library Philosophy and Practice (e-journal)*, paper 1322. Retrieved from <http://digitalcommons.unl.edu/libphilprac/1322>
9. Iwudu, N. (2012), Utilization of DSpace Open Source Institutional Repository Software in University of Jos, Nigeria, MLS dissertation, University of Jos, Nigeria.
10. Kalu, N. (2016), Two UNICAL Teachers Face Sack for Plagiarism, *The Nation Newspaper*, Retrieved from <http://thenationonlineng.net/two-unical-teachers-face-sack-plagiarism/>
11. Kayode-Adedeji, D. (2013), Nigeria's Agriculture University, FUNAAB, Sacks Professor, Two Other Lecturers For Plagiarism, *Premium Times Newspaper*, Retrieved from <http://www.premiumtimesng.com/news/135438-nigerias-agriculture-university-funaab-sacks-professor-two-other-lecturers-for-plagiarism.html>
12. Oladeji, F., Ajayi, O., Ajetunmobi, R., Okunoye, B., Okiki, C., Uwadia, C. (2016), A Survey of the Status of Nigerian Universities in Curbing Plagiarism, 11th University of Lagos Research Conference & Fair Proc., pp. 564-572
13. Okeke, B. (2001), Quality Management and National Goal Attainment in Education: The Case of Nigeria, *Nigerian Universities Inaugural Lecture Series*, National Universities Commission, pp.263-295.
14. Okiy, R. (2008). Digitizing and archiving information resources in libraries: Reactions to global initiatives from Nigeria. *IFLA ITS Newsletter*, December 2008, pp. 5-6.
15. Onuoha, U. and Ikonne, C. (2013), Dealing with the Plague of Plagiarism in Nigeria, *Journal of Education and Practice*, 4(11), pp. 102 - 106
16. OpenDoar, Directory of Open Access Repositories, Retrieved from <http://www.opendoar.org/countrylist.php#Nigeria>
17. Smith, M., Barton, M., Bass, M., Branschofsky, M., McClellan, G., Stuve, D., Tansley, R. and Walker, J., (2003). DSpace: An open source dynamic digital repository.