



International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 5, May 2018

Online Service for Engineering Project Title Verification Based on Content Analysis

Shubham Podhade¹, Saily Deore², Mohini Pandey³, Padmanabh Saoji⁴, Prof M. L. Bangare⁵

Student, Department of Information Technology Engineering, Smt. Kashibai Navale College of Engineering,
Pune, India^{1,2,3,4}

Assistant Professor, Department of Information Technology Engineering, Smt. Kashibai Navale College of
Engineering, Pune, India⁵

ABSTRACT: In a world where copying is as simple as a couple of clicks of a mouse, digital content is worth a lot less than its physical counterpart, because there is no way of establishing authenticity we need to know what is authentic or not otherwise there will be no trust. We should not be using the internet simply as our shop window or to send somebody email. The true potential of the Internet and our digital world is so much more all we need is the technology to unlock this potential. Hence the content verification has got lot importance. Title as well content verification and duplication avoidance has become an important issue in colleges. Is it possible to live in the information age and not practice plagiarism? But in this process, are students gaining or losing. How seriously are their assignments taken in terms of their own contribution and originality in the work. Content checking has got lot of importance. In our college in existing system everything is being done manually. It's too manual process and takes too much time for checking. Normally, duplication happens in traditional work. This project focuses on preventing repetition of project and motivates students to implement new concepts. Proposed system store well organized repository of the previous year's projects in order to access them easily.

KEYWORDS: Verification, Extraction, Data Mining.

I. INTRODUCTION

BASIC CONCEPT: The algorithm which will detect the similar project automatically with the help of the data stored in the database is used. To store all previous implemented project data with synopsis in the system, the student register details first. The details are name, email, password, roll no, branch and year. To verify the title student has to give the project name, keywords and abstract as an input. Then based on content analysis and plagiarism, system will process a result. The result is in the form of percentage also known as threshold value, which shows how much content is matched with the database. Threshold value decides whether to reject or to accept the project. This saves time in the long term because there is no need to re-organize, re-format or try to remember details about projects. It also increases research efficiency since both the data collector and other researchers will be able to understand and use well-annotated data in the future.

PROJECT OBJECTIVE : This system used in colleges for project title verification and prevent repeated project management issues. This online portal offer more sources, such as large databases that includes project of every domain. Thousands of reports exist in the database or repository to refer by the students. Keep a well-organized reservoir for the project. Reduces manual work done by teachers and saves time and prevents redundancy of the project. This software also gives percentages of similarity in the project. Checking your paraphrasing abilities is another objective of our software to develop. Entire process will be fully automated as right from guide allocation till final report submission will be done online. It is a centralized portal which will be accessed by anyone in college area.



International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijirccce.com

Vol. 6, Issue 5, May 2018

II. RELATED WORK

The current^[1] plagiarism detection system was found to be too slow and takes too much time for checking. The matching algorithms are also dependent on the text's lexical structure rather than semantic structure. Therefore, it becomes difficult to detect the text semantically. The big challenge is to provide plagiarism checking with appropriate algorithm in order to improve the percentage of finding result and time checking. Due to this availability, users can easily create a new document by copying and pasting from these resources, as well as sometimes they tried to just replace the word with their synonyms. 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. 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.

In academics or research papers most of the students plagiarize mostly due to five reasons:

- (1) Lack of experience
- (2) Greed
- (3) Desperation
- (4) Fear
- (5) Ignorance.

SR. NO.	AUTHOR	PUBLICATION	DESCRIPTION
1.	Hemlatha A.M, M.Shubha	Nov 2014	Plagiarism checking by using data mining technique.[1]
2.	KamlakarPallela, SnehaTalari	Oct 2016	Plagiarism as a ethical issue.[2]
3.	Joshi O D, Pudale A H, Ghorpade R D	March 2016	Text extraction technique used for plagiarism detection as well as semantic analysis of the content.[3]
4.	Dr.Vijayarani,Ms R Janani	Nov 2016	Colussi String matching algorithm used for retrieving information.[4]
5.	Dr.Shasayeee, Dr Thelambai	2015	Comparison between Knuth Morris Pratt ,Naive and Moore algorithm implemented in python based on their execution time for different text and pattern length.[5]
6	Zhou Tong1 and Haiyi Zhang2	Aug 2016	A document exploring system on lda topic for Wikipedia articles [6]
7	Raymond J. Mooney and Un Yong Nahm	March 2017	Text Mining with Information Extraction [7]
8	AmrutaPatil, Nikhil Bomanwar	April 2017	Survey on Different Plagiarism Detection Tools and Software's [8]
9	NingZhong, YuefengLi and Sheng Tang Wu	January 2015	Effective pattern discovery for text mininization. [9]



International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijirccce.com

Vol. 6, Issue 5, May 2018

EXISTING SYSTEM: Currently for handling such kinds of problems no such system is available at our department level. All the tasks are performed by the project coordinators and the guide manually hence in order to apply the engineering knowledge and try to solve these real life time problems that department faces is done by this proposed system. The detailed reports are available at this repository. Hence in order to solve this problem which exists and solution to this is not any automated portal or the application.

- i) No such System
- ii) Everything done manually
- iii) Time consuming process
- iv) No repository Available
- v) Domain wise analysis is time consuming

III.IMPLEMENTATION DETAILS

3.1 WHAT IS TO BE DEVELOPED: Engineering and technology always deals with the problem solving in real life time and hence it's important to try and give solutions to the problem that we face in our day to day life. In our department we found management of the project is being a difficult task for the teachers as well as to the project coordinators and to the head of the department. In this era where everything is being available on the tip of fingers so we find it interesting that to develop software which will analyze the project titles and verifies them according to the content.

So we found this interesting and to develop such a management software for the department which will reduce burden on teachers and the project coordinators as well as on the head of the department. It will also help students to check out previous year's repository in well-organized manner.

3.2 TECHNOLOGY USED

3.2.1 JAVA : Java programming language was originally developed by Sun Microsystems which was initiated by James Gosling and released in 1995 as core component of Sun Microsystems' Java platform (Java 1.0 [J2SE]). The latest release of the Java Standard Edition is Java SE 8. With the advancement of Java and its widespread popularity, multiple configurations were built to suit various types of platforms. For example: J2EE for Enterprise Applications, J2ME for Mobile Applications. The new J2 versions were renamed as Java SE, Java EE, and Java ME respectively. Java is guaranteed to be Write Once, Run Anywhere.

3.2.2 MySQL RDBMS :MySQL is a fast, easy-to-use RDBMS being used for many small and big businesses. MySQL uses a standard form of the well-known SQL data language. MySQL works on many operating systems and with many languages including PHP, PERL, C, C++, JAVA, etc. MySQL supports large databases, up to 50 million rows or more in a table. The default file size limit for a table is 4GB, but you can increase this (if your operating system can handle it) to a theoretical limit of 8 million terabytes (TB). MySQL is customizable. The open-source GPL license allows programmers to modify the MySQL software to fit their own specific environments.

3.2.3 APACHE TOMCAT SERVER:Apache Tomcat is an open source Web server tool developed by the Apache Software Foundation (ASF). It is one of many Apache-related open source products used by IT professionals for various tasks and objectives. Apache Tomcat allows the implementation of Java Servlets and Java Server Pages (JSP) to promote an effective Java server environment. Users can also access resources for configuration and use extensible mark-up language (XML) to configure projects. Successive versions of Apache Tomcat have solved different problems by applying software patches and other solutions. Some experts characterize Apache Tomcat as a product offering a runtime shell for Java Servlets. Users can also set up Java virtual machines (JVM) to configure virtual hosting.

MAIN COMPONENTS (TOMCAT.CORE)

REQUEST/RESPONSE :This is the tomcat representation of a request. It acts as an "Adapter" for the underlying application's representation - in the case of a web server it is the java representation of (request_rec *) for example, if

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijirccce.com

Vol. 6, Issue 5, May 2018

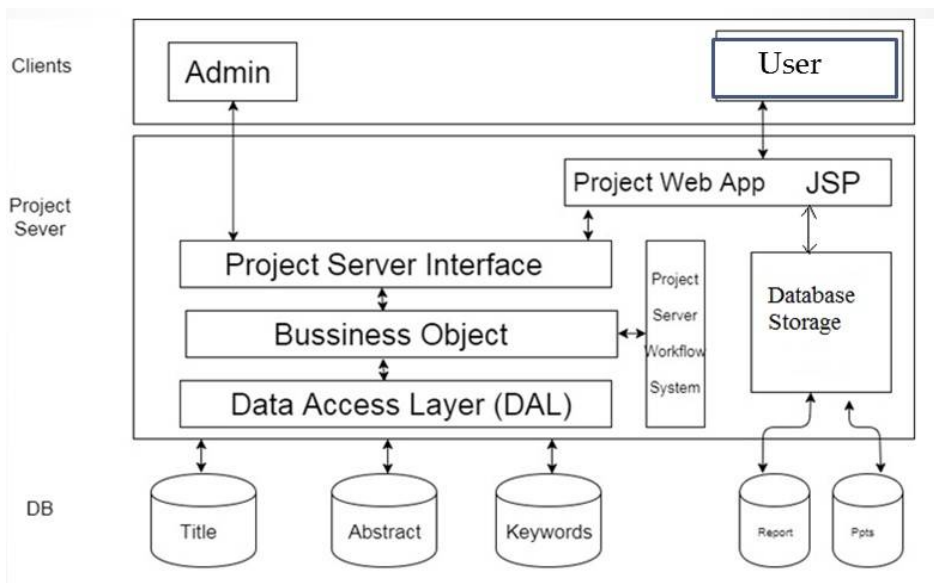
tomcat is embedded in a non-web application it represent the application's notion of request. This is a passive object, with all operations delegated to modules. It is important to make as much as possible "lazy" evaluations, and keeps the code very simple. The objects are not a replication of HttpServletRequest and HttpServletResponse.

OUTPUTBUFFER:

The idea is to have full control over tomcat buffering and over char/byte translation, as this is one of the most expensive operations (according to empirical tests). It should allow to plug external byte/char convertors.

MESSAGEBYTES: This is a very special and important component of tomcat3. The first problem it tried to solve was the conversion from byte to chars. The charset is known very late (either when Content-Type header is parsed or even later, when the user sets it in servlet 2.3), and converting the received bytes to Strings can't be done (correctly) until the encoding is determined.

3.3 ARCHITECTURE DIAGRAM



Architecture Diagrams of the System

3.4 ALGORITHM USED: Authors have used following algorithm for the computation of the data in database and those toentered by the user.

```

1 public class StringSimilarity
2 {
3 public static double similarity(String s1, String s2)
4 {
5 String longer = s1, shorter = s2;
6 if (s1.length() < s2.length()) { // longer should always have greater length
7 longer = s2; shorter = s1;
8 }
9 intlongerLength = longer.length();
10 if (longerLength == 0) { return 1.0; /* both strings are zero length */ }

```



International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijirccce.com

Vol. 6, Issue 5, May 2018

```
11 return (longerLength - editDistance(longer, shorter)) / (double) longerLength;
12 }
13 public static int editDistance(String s1, String s2)
14 {
15     s1 = s1.toLowerCase();
16     s2 = s2.toLowerCase();
17     int[] costs = new int[s2.length() + 1];
18     for (int i = 0; i <= s1.length(); i++) {
19         int lastValue = i;
20         for (int j = 0; j <= s2.length(); j++) {
21             if (i == 0)
22                 costs[j] = j;
23             else {
24                 if (j > 0) {
25                     int newValue = costs[j - 1];
26                     if (s1.charAt(i - 1) != s2.charAt(j - 1))
27                         newValue = Math.min(Math.min(newValue, lastValue),
28                             costs[j]) + 1;
29                     costs[j - 1] = lastValue;
30                     lastValue = newValue;
31                 }
32             }
33         }
34         if (i > 0)
35             costs[s2.length()] = lastValue;
36     }
37     return costs[s2.length()];
38 }
39 public static String printSimilarity(String s, String t) {
40     String match = String.format("%.3f", similarity(s, t), s, t);
41     System.out.println(match);
42     return match;
43 }
44 }
```

3.5 MODULES OF WEB PORTAL:

Student Module

1. Student Login
2. Student Registration
3. Old Projects Details Department Wise
4. Verify Title of the project and contents.
5. Select Guide.
6. Enter New Selected Project Details

College Faculty Module

1. College Faculty Login
2. College Faculty Registration
3. Old Projects Details Department Wise
4. View all student requests.



International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijirccce.com

Vol. 6, Issue 5, May 2018

Admin Module

1. Admin Login
2. View Student Data Department Wise
3. View Staff Data Department Wise
4. Old Projects Details Department Wise

IV. CONCLUSIONS

4.1 SUMMARY AND CONCLUSION

The proposed idea will surely help at the college level to avoid the manual process of project title verification done by the teachers as well as reduce the project management issues and avoid the redundancy of the projects.

It also provides a central repository which will be useful for the students and also guide them. The development of web application with eclipse IDE and apache tomcat server is available free of cost. Also teachers as well as students can be able to access the software from anywhere and at any time.

4.2 APPLICATIONS

This software will be useful for students as well as faculty members of the colleges. This software can also be used by various colleges as well as by University to fully reduce the redundancy of the project. Teachers or Head of the department used this software to keep the thorough information about the teachers available in the department, and also about student's project. Through this software guide can also communicate with the students and keep track of the project time to time.

4.3 FUTURE SCOPE: The idea can be extended to the university level.

4.4 LIMITATIONS: Software is limited to college level now

ACKNOWLEDGEMENT

Authors would like to thank Prof. V.S.Khandekar & Prof. Y.V.Wagh (Project Co-ordinator) for their help. Also like to thank Prof. R.H. Borhade (HOD IT) who helped in all possible ways. Authors extend sincere thanks to SKNOCE Principal Prof. Dr. A.V. Deshpande and Vice Principal Prof. Dr.K.R.Borole. Also like to thank all staff members for their valuable guidance. Lastly, Authors also would like to thank their friends and family for their love and support.

REFERENCES

1. Hemlatha A.M, M.Subha, "A study on plagiarism checking with appropriate algorithm in data mining", International Journal of Research In Computer Application and Robotics, Nov 2014, ISSN 2320-7345
2. KamalakarPallela, SnehaTalari, "Plagiarism : A serious ethical issue for indian students", 2016 IEEE International Symposium on technology and society (ISTAS) 20-22 October 2016
3. Ababneh Mohammad, OqeiliSaleh and Rawan A Abdeen, Occurrences Algorithm for String Searching Based onBrute-Force Algorithm, Journal of Computer Science,2(1): 82-85, 2006.
4. Joshi O D, Pudale A H, Ghorpadde R D, "Text extraction technique applied to plagiarism detector : the semantic analysis for analyzing the writing style" International Journal of Computer science engineering(IJCSE) Mach-2016 ISSN 2319-7323
5. Dr. (Ms). AnanthiSheshasayee , Ms. G. Thailambal, "A Comparative Analysis of Single Pattern Matching Algorithm in Text Mining" Quaid-e-Milleth College for Women University, Kancheepuram, Chennai, India India
6. Zhou Tong1 and Haiyi Zhang2 , "A document exploring system on lda topic for Wikipedia articles" The International Journal of Multimedia & Its Applications (IJMA) Vol.8, No.3/4, August 2016
7. Raymond J. Mooney and Un Yong Nahm, "Text Mining with Information Extraction" Multilingualism and Electronic Language Management: Proceedings of the 4th International MIDP Colloquium, September 2003, Bloemfontein, South Africa, March 2017
8. AmrutaPatil, Nikhil Bomanwar, " A Survey on Different Plagiarism Detection Tools and Software's" (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 7 (5) , 2016,2191-2193 April 2017
9. NingZhong, YuefengLi and Sheng Tang Wu, "Effective pattern discovery for text mining" IEEE transactions on knowledge and data engineering, vol. 24, no. 1, January 2015