

A Survey: Content Verification and Duplication Avoidance System for UG Project Title

Shubham Podhade¹, Saily Deore¹, Mohini Pandey¹, Padmanabhi Saoji¹, Prof M. L. Bangare²
Student, Department of Information Technology Engineering, Smt. Kashibai Navale College of Engineering,
Pune, India¹

Assistant Professor, Department of Information Technology Engineering, Smt. Kashibai Navale College of
Engineering, Pune, India²

ABSTRACT: Title, Content Verification and duplication avoidance has become an important issue in Colleges. In existingsystem everything done manually. It's too slow process and takes too much time for checking. Normally, Duplication happens in traditional work. This paper focuses on preventing repetition of project and students implement new con-cepts. Proposed system store well organized repository of the previous projects.

KEYWORDS: Title verification, content extraction, Data mining.

I. INTRODUCTION

Data mining is used for finding the useful information from the large amount of data. Data mining techniques are used to implement and solve different types of re-search problems. This system wants to verify project title using data mining and Text Extraction techniques.

Text Extraction Techniques:

Extracting Text is one of the most important tasks when working with text. Readers benefit from keywords be-cause they can judge more quickly whether the text is worth reading. Website creators benefit from keywords because they can group similar content by its topics. Algorithm programmers benefit from keywords because they reduce the dimensionality of text to the most important features. Multiple Methods and Algorithms for Text extraction and String Matching. But we use Aho-corasick algorithm for Text Attractions.

Algorithm 1. Pattern matching machine.

Input. A text string $x = a_1a_2 \dots a_n$ where each a_i is an input symbol and a pattern matching machine M with go to function g , failure

Function f , and output function $output$, as described above.

Output. Locations at which keywords occur in x .

International Journal of Innovative Research in Science, Engineering and Technology

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

Visit: www.ijirset.com

Vol. 7, Issue 1, January 2018

Method.

Step 1: begin

Step 2: state \leftarrow 0

Step 3: for i \leftarrow 1 until n do

Step 4: begin

Step 5: while g (state, a_i) = fail do state \leftarrow f (state)

Step 6: state \leftarrow g (state, a_i)

Step 7: if output (state) \neq empty then

Step 8: begin

Step 9: print i

Step 10: print output (state)

Step 11: end

Step 12: end

Step 13 end

Background:

1. No such System
2. Everything done manually
3. Time consuming process
4. No repository Available
5. Domain wise analysis is time consuming

Motivation:

1. Project Work Duplication
2. Centralized Portal
3. Avoid Manual Process
4. Technical Justification
5. Domain-wise selection

International Journal of Innovative Research in Science, Engineering and Technology

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

Visit: www.ijirset.com

Vol. 7, Issue 1, January 2018

II. EXISTING SYSTEM

Hemlatha A.M, M.Subha: The current plagiarism de-tECTION 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 paraphrased semantically. The big chal-lenge is to provide plagiarism checking with appropriate algorithm in order to improve the percentage of finding result and time checking. The important question for the plagiarism detection problem in this study is whether it is possible to apply new techniques such as Semantic Role Labeling to handle plagiarism problems for text documents many documents are available on the internet and are easy to access. Due to this availability, users can easily create a new document by copying and pasting from these resources. Sometimes users can reword the plagiarized part by replacing the word with their syno-nyms. Motivation of the paper is to find the most plagia-rism content that should be copied from anywhere iden-tified in the efficient manner. Further it helps to as pla-giarism detection process in applications to user or indi-vidual publish their journals.

Kamalakar Pallela, Sneha Talari: In this paper, a study has been attempted to evaluate possible approach-es to introduce a methodology in Indian education sys-tem whereby students going abroad or working in scien-tific research are protected from the academic code of conduct or copyright issues. A critical assessment was carried out on various imparting factors of education system. A decent procedure and adoptable ethical waysof teaching and training methodologies are being sum-marized in this paper.

Joshi O D, Pudale A H, and Ghorpadde R D: Detect-ing the plagiarized data is of most concern to research organizations, industries and educational institutions on the computational systems for such a task has become important. Existing methods for detection of plagiarized document compute the similarity of document-to-document basis. We proposed system that will do text mining, exploring the use of keywords and semantic analysis of statements as a feature for analyzing a docu-ment. The main goal is to analyze the statements and keywords, looking for segments of the document that is written by other authors. This is considered as a seman-tic analysis using keyword based statement analysis where paragraphs with unique in style. This approach does work on the use of keywords and semantic analysis of statements, so it is do not require any language speci-fications. We feel that this feature shows improvement in this area. It will achieve tremendous results compared to existing models.

Dr. S.Vijayarani Ms. R.Janani: Information retrieval is the method of retrieving the knowledge relevant to an issue of curiosity. It locates the relevant documents, on the premise of user's question which includes keywords or example files. Probably the most acquainted applica-tion of information retrieval system is search engine corresponding to Web search, Desktop search, Federated search, Mobile search, Enterprise search and Social search. This research work mainly focused on the desk-top search. Desktop search is the specified variant of enterprise search, where the information foundations are the files stored on a personal computer, together with email and websites established on content analysis. Con-tent Analysis is a group of manual or computer based approaches for contextualized explanations of docu-ments. To analyze the content the different text pattern matching algorithms are used and it is used to discover all the existences of a limited set of patterns inside an input document. Commonly these algorithms are used in several applications that include information security bio-informatics, plagiarism detection, text mining and document matching. String matching is essential for finding text patterns that are in online and offline. String matching algorithm is used to matches the pattern pre-cisely or about in the input document. The main objec-tive of this research work is to analyze the performance of existing string matching algorithms. For this compari-son there are four algorithms are used namely, two way algorithm, Colussi algorithm, optimal mismatch algo-rithm and Maximal shift algorithm. From this analysis it is observed that the Colussi string matching algorithm gives the better result.

Dr. (Ms). Ananthi Sheshasayee1 Ms. G. Thailambal2:

Text Mining is an emerging area of research where the necessary information of user needs to be provided from large amount of information. The user wants to find a text P in the search box from the group of text infor-mation T. A match

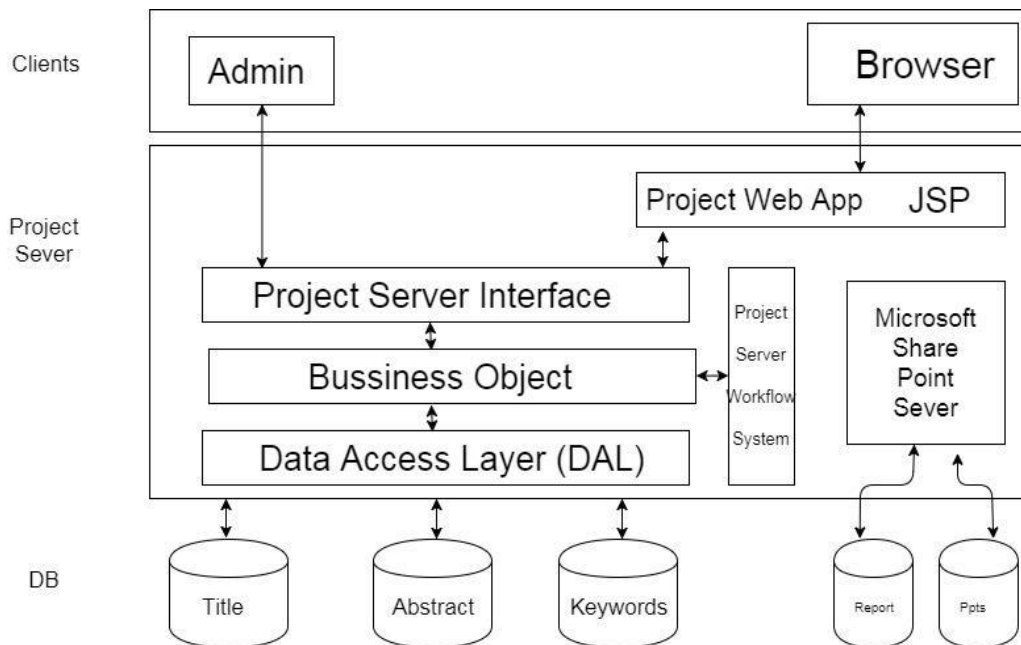
International Journal of Innovative Research in Science, Engineering and Technology

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

Visit: www.ijirset.com

Vol. 7, Issue 1, January 2018

needs to be found in the information then only the search is successful. Many String matching algorithms available for this search. This paper discusses three algorithms in unique pattern searching in which only one occurrence of the pattern is searched. Knuth Morris Pratt, Naive and Boyer Moore algorithms implemented in Python and compared their execution time for different Text length and Pattern length. This paper also gives you a brief idea about time Complexity, Characteristics given by other authors. The paper is concluded with the best algorithm for increase in text length and pattern length.



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 Staff Module:

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

Admin Module

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

International Journal of Innovative Research in Science, Engineering and Technology

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

Visit: www.ijirset.com

Vol. 7, Issue 1, January 2018

III. CONCLUSION

The proposed system is capable enough to handle the reduplication avoidance in projects and hence keep the reservoir of the same.

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 Oc-tober 2016
3. 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 Jour-nal of Computer science engineering(IJCSE) Mach-2016 ISSN 2319-7323
4. Matching Algorithm", , Int. J. Comp. Tech. Appl., Vol 2 (6), 1877-1883. ISSN: 2229-6093, December 2011.
5. 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
6. KamalakarPallela, SnehaTalari, "Plagiarism : A serious ethical issue for indian students", 2016 IEEE International Symposium on technology and society (ISTAS) 20-22 October 2016
7. 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
8. Dr.S.VijayaraniMs.R.Janani: String Matching Algorithms for Retrieving Information from Desktop – Comparative Analysis.
9. Dr. (Ms). Ananthi SheshasayeeI Ms. G. Thailambal2: A COMPARITIVE ANALYSIS OF SINGLE PATTERN MATCHING ALGORITHMS IN TEXT MINING