

Online Plagiarism Checker

Neha Joshte¹, Sushmitha Kadam², Sakshi Tembhurne³, Prof. Prajakta Gotarne⁴

^{1,2,3} SNTD Women's University, BTech in Computer Science & Technology Usha Mittal Institute of Technology Mumbai, India

⁴ Professor, SNTD Women's University, BTech in Computer Science & Technology Usha Mittal Institute of Technology Mumbai, India

Abstract—This paper is solution to find plagiarised content in file. Plagiarism is a bad practice that everyone should avoid to reduced bad consequences in academics front. Plagiarism is detected through plagiarism tool that compares two strings which also being consider as a long string. It tell how much content is copied or used in other files. This will help to reduce human effort to check line by line similarity.

Index Terms—Plagiarism, plagiarism tool, string matching algorithm, karp-rabin, ocr

I. INTRODUCTION

Plagiarism is using someone else's intellectual property (texts, ideas, or results), thereby implying as their own.

Two components of plagiarism:

- 1) To appropriate the work of someone else
- 2) Passing it off as one's own by not giving proper credit.

II. HOW IT CAN BE DETECTED?

The way that plagiarism detection software works is to identify content similarity matches. That is, the software scans a database of crawled content and identifies the text components and then compares it to the components, or content, of other work. Three general methods to detect plagiarism are:

- 1) Document source comparison
- 2) Manual search of characteristic phrases proper credit.
- 3) Stylometry

III. PROBLEM STATEMENT

According to one research it was found that in early days every third paper was containing plagiarized content. The need for checking papers came into existence for plagiarised contents. However practically it became impossible to check manually. Thereafter, many software was introduced for our convenience. Online plagiarism Checker will be one of our effort in this field to make plagiarism detectable with mere modifications than existing systems available today.

IV. LITERATURE SURVEY

In plagiarism research, Gert Helgesson [7] mentioned that plagiarism reprehensible is that it involves an unfair acquisition of scientific credit. In addition, intentional plagiarism involves dishonesty. In plagiarism of data or results, fabrication is also implied. While studying plagiarism detection tools, Alberto Bugarian and Maria. J. Carreira [9] studies found that for better results one can use Turnitin over Jplag. They mentioned the comparative use of both the softwares in educational fields. Checking survey related to same in that Hermann A Maurer, Frank Kappe and Bilal Zaka [8] described three general methods of detecting plagiarism. In journal related to computer application, Kiran Sonawane and Prabhudeva S [6] mentioned the system

to be of 85 percent of precision and 10 percent minimization of failed detection. Then Andysah Putera Utama Siahaan, Solly Aryza, Eko Hariyanto, Rusiadi, Andre Hasudungan Lubis, Ali Ikhwan, Phak Len Eh Kan made a combination of Rabin-Karp and Levenshtein algorithms makes a useful contribution. The level of similarity can be optimized so that documents can be assessed correctly according to the content of the document. This combination can also improve the accuracy of reviewers based on some parameter, N-Gram, Base, and Modulo. This parameter can be adjusted according to the desired sensitivity level of the algorithm. The higher the N-Gram value, the lower the sensitivity level. The lower Base and Modulo values, the lower the test sensitivity level.

Jithin S Kuruvila, Midhun Lal V L, Rejin Roy, Tomin Baby, SangeethaJamal Sherly K K [5] compared the flowcharts and by comparing it with both the shape orientation. Since this approach creates graph from the flowchart, it is capable to detect the plagiarism with same shaped objects even though the orientation of the graph is different. R. Mittal and A. Garg

[1] stated that OCR software can be used to convert a physical paper document, or an image into an accessible electronic version with text.

V. FIGURES

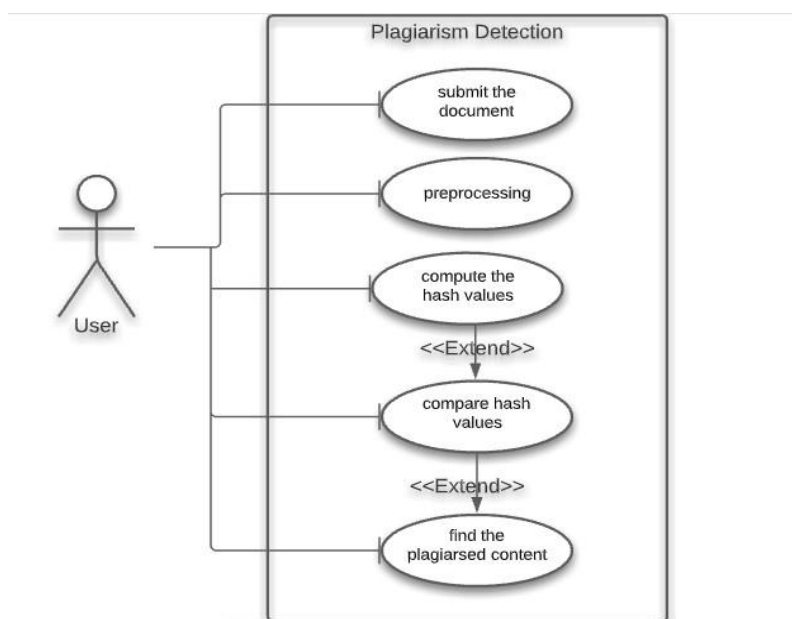


Fig. 1. Project Use Case

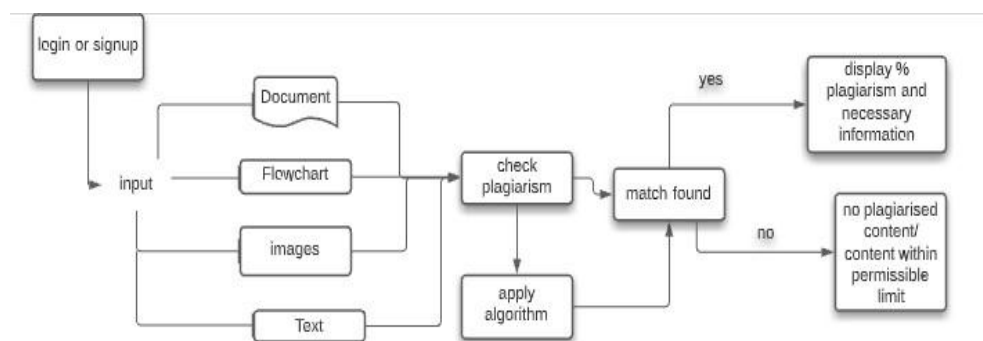


Fig. 2. Project Block Diagram

VI. ALOGRITHMS

A. *SequeceMatcher*

For searching the repeating pattern between two strings there is need to check the similarity between the same. For that different algorithm suggest token based edit based sequence based algorithms is used. Here, difficult to choose between three algorithms, sequence based according to convenience is used. SequenceMatcher checks the similarity based on long common string within two string that is to be checkcked for similarity.

B. *Advantages*

This algorithm mainly calculates string similarity build on the length of the longest common sub-sequence and recursive lengths of common characters in different portion of the string.

C. *Disadvantages*

Disadvantage of this matching system is, it generates multi- ple combinations and sometimes the shorter natural match is proximate.

VII. IMPLEMENTATION

We have created an web application to compare two docu- ments against each other to detect plagiarism. It also has functionality to ask user to upload the file, that will compare file from database / internet. We have used Karp-Rabin algorithm.

VIII. CONCLUSION

Detection of plagiarism by plagiarism tool makes life easy to find originality of work of student / professional / researcher. Best mean to find plagiarism is to use string matching algo- rithm like karp-rabin.

ACKNOWLEDGMENT

We would like to express our special thanks of gratitude to our guide Prof. Prajakta Gotarne madam as well as our principal Dr. Sheeka Neema madam who gave us the golden opportunity to do this wonderful project on the topic Online Plagiarism Checker, which also helped us in doing a lot of Research and we came to know about so many new ways to solve the problem that help to build good idea to create application to resolve an issue.

REFERENCES

- [1] R. Mittal and A. Garg. "Text extraction using ocr: A sys- tematic review, second international conference on inventive re- search in computing applications (icirca), 2020, pp. 357-362, doi: 10.1109/icirca48905.2020.9183326. ", 2020.
- [2] Krunal binekar Shubham juvekar, Bushan Bhopatrao and Prachi Gad- hire.r. "Textmage: Plagiarism checker, volume: 06 issue" , 05 May 2019.
- [3] J. L. Ganesha. "Plagiarism detection using levenshtein distance with dynamic programming", January 2018.
- [4] E. Hariyanto A. H. Lubis A. Ikhwan P. L. E. Kan. A.P.U.Siahaan, S. Aryza. "Combination of levenshtein distance and rabin- karp to improve the accuracy of document equivalence level", 2018.
- [5] Rejin Roy Tomin Baby Sangeetha Jamal Sherly K K. Jithin S Kuruvila, Midhun Lal V L. " Flowchart plagiarism detection system: An image processing approach", 24 August 2017.
- [6] S Prabhudeva. Kiran Sonawane. " International journal of computer applications(0975-8887) volume 116 -no.23" , April 2015.

- [7] Stefan Eriksson. Gert Helgesson. Plagiarism in research., July 2014
- [8] Bilal Zaka. Hermann A. Maurer, Frank Kappe. " Plagiarism - a survey", January 2006.
- [9] Xose Manuel Pardo. Alberto Bugarin, Maria J. Carreira Manuel Lama. "Plagiarism detection using software tools: A study in a computer science", July 2005.