International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 06 Issue: 05 | May 2019 www.irjet.net p-ISSN: 2395-0072

PLAGIARISM CHECKER

Shubham Juvekar[1], Bhushan Bhopatrao[2], Krunal Binekar[3], Prachi Gadhire[4]

[1,2,3]Student, Computer Department, DRIEMS, Maharashtra, India.
[4] Professor, Computer Department, DRIEMS, Maharashtra, India.

Abstract - *Plagiarism is any identical or lightly-altered use of one's own or someone else's work. Text plagiarism detection systems are widely available. Students achieve the best results in learning by writing and doing exercises. This mandates a large number of written exercises. However, limited resources and distribution of assessment work lead to problems when students' answers need to be checked for plagiarism. Plagiarism or copy pasting is difficult to notice in a large volume of documents. The demonstrated project focuses on computer-assisted plagiarism detection in medium to large volumes of text-based submissions. Moreover, the project supports automated search for web sources.*

***Key Words*: Cheating, plagiarism, computer assisted assessment, text documents, web based**

1. INTRODUCTION

Nowadays, plagiarism is a really serious issue within the professional environment, or even within the education system. Since Internet is accessible to everyone, it is easy to use Internet as a source of information. However, copying documents from Internet can be considered as plagiarism: what can be found on Internet can come from a book, a research document or an article. It can even result to some legal problems, such as copyright infringement. Plagiarism detection of first case is easy to detect while the later requires complex analysis and algorithms.

Plagiarism detection of first case is easy to detect while the later requires complex analysis and algorithms. Finding plagiarized parts of a text document is very slow labor for teachers of copy and paste.

Finding plagiarized parts of a text document is very slow labor for teachers. Even with a limited number of texts it relies on the teacher's ability to read and remember every submission. This is slow and ineffective. The Internet makes the problem even worse, since the teacher would be required to read and remember related material on the Internet to find the possible sources of copy and paste. Search engines can certainly help to find sources that have been copied, but picking suitable search terms and manually searching is tedious and repetitive. Plagiarism eats resources that would be better used in real work. It is wasting everyone's time without any gain on the course material.

2. IMPLEMENTED SYSTEM

The plagiarism checkers that academics use today compare a document against a wide variety of databases. Instead of having to use different software for each database that you want to search, one will be enough to search all of those available. The majority of teachers use a standard type that anybody has access to; just choose which one seems most reliable and start uploading your documents.

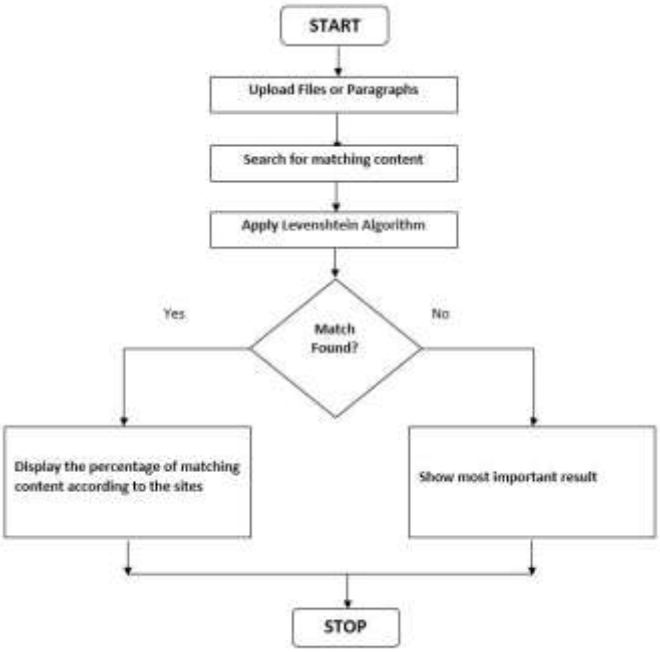

Fig 1: Flowchart Diagram

The main goal of this project is to create a plagiarism system that is primarily based on Levenshtein Algorithm. This system primarily focuses on paragraph detection. The database contains paragraphs stored in a single file. The system will retrieve and rank this database based on a given query by the user.

The main goal of this project is to create a paragraph plagiarism system that is primarily based on Levenshtein Algorithm. This system primarily focuses on paragraph detection. The database contains paragraphs stored in a single folder. The system will retrieve and rank this database based on a given query by the user.

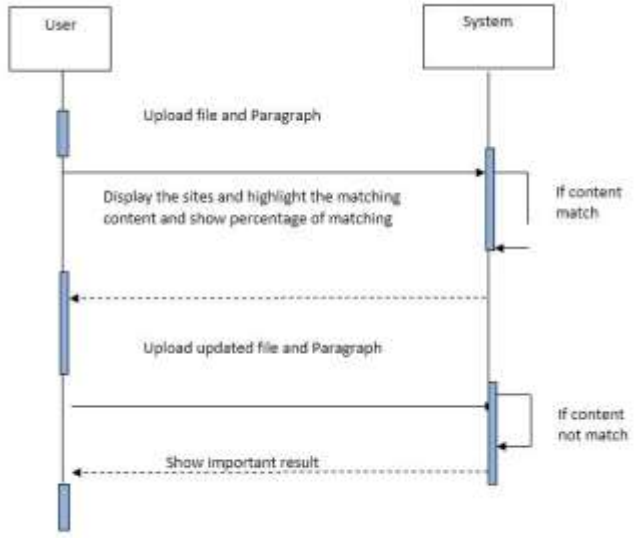

Fig 2: Sequence Diagram

This project will describe how the software works without going into details. Importing one single file for online plagiarism check. Concerning this part, the user is able to import paragraph to check for plagiarism. Then, the user can proceed to the configuration of the plagiarism detection through the configuration window. Then, the file analysis can start and system will complete some steps.

Levenshtein distance Algorithm:

In information theory, linguistics and computer science, the Levenshtein distance is a string metric for measuring the difference between two sequences. Informally, the Levenshtein distance between two words is the minimum number of single-character edits (insertions, deletions or substitutions) required to change one word into the other. It is named after the Soviet mathematician Vladimir Levenshtein, who considered this distance in 1965.

Levenshtein distance may also be referred to as edit distance, although that term may also denote a larger family of distance metrics. It is closely related to pairwise string alignments.

Mathematically, the Levenshtein distance between two strings (of length and respectively) is given by where is the indicator function equal to 0 when and equal to 1.

$$\operatorname{lev}_{a,b}(i,j) = \begin{cases} \max(i,j) & \text{if } \min(i,j) = 0, \\ \min \begin{cases} \operatorname{lev}_{a,b}(i-1,j) + 1 \\ \operatorname{lev}_{a,b}(i,j-1) + 1 \\ \operatorname{lev}_{a,b}(i-1,j-1) + 1_{(a_i \neq b_j)} \end{cases} & \text{otherwise.} \end{cases}$$

3. FEATURES OF THE APPLICATION

- Simple
- Efficient
- Reliable
- Reduces human effort

4. CONCLUSION

The experimental results shows that the proposed approach is capable to detect plagiarism in paragraph. . Line detection algorithms such as levenshtein algorithm etc. are used for this purpose which makes it more efficient from normal checkers which made us to develop this software.

REFERENCES

[1] H. Maurer, et al., "Plagiarism-a survey," Journal of Universal Computer Science, vol. 12, pp. 1050-1084, 2006.

[2] L. McKeever. Online plagiarism detection services-savior or scourge. Assessment & Evaluation in Higher Edu 2006; 31(2): 155-165.

[3] B. G. Vasudevan, et al. Flowchart knowledge extraction on image processing. In proceedings of IEEE International Joint Conference on Neural Networks, 2008; 4075-4082.

[4] K. S. Candan, M. L. Sapino. Data management for multimedia retrieval. London Cambridge University Press 2010.

[5] Flowchart Plagiarism Detection System: An Image Processing Approach Jithin S Kuruvila, Midhun Lal V L, Rejin Roy,Tomin Baby, Sangeetha Jamal, Sherly K K* Rajagiri School of Engineering and Technology, Kakkanad, Kochi 682039, India