ANTI-PLAGIARISM APPLICATION

Bhoomi Lad [1] | Foram Joshi [1] | Sheetal Patel [1]

[1] Department Of Computer Engineering, Vidyavardhini's College of Engineering & Technology, K.T. Marg, Vasai Road (West), Palghar-401202, Maharashtra

ABSTRACT

Plagiarism is snitching of other people's thoughts, words or ideas without proper acknowledgement of source. Plagiarism has seen widespread activities in recent times which affects the trustworthiness/ positiveness /reliability/integrity of organizations as well as its ability to ensure quality of its clients. Plagiarism detection of research papers deals with checking similarities with other research papers. Manual detection of plagiarism is not very easy and is time consuming due to vast amounts of data available and also the assigned reviewer may not have adequate knowledge in the research disciplines. They may have different way of thinking, views, creating errors and misinterpretations. Therefore, there was a need for an emphatic and feasible approach to check the submitted research papers with support of automated software.[1][2]

KEYWORDS: Plagiarism types, plagiarism detection, plagiarism techniques, plagiarism prevention, plagiarism algorithms.

I. INTRODUCTION

Plagiarism has become a world-wide problem and is increasing day by day. This problem is increasing further mainly because of the increase in the number of online publications. Using Plagiarism detection technique we can compare the given document with the target/original document.[9]

Plagiarism is borrowed from the Latin Language, meaning "to Kidnap". According to the *Merriam Webster Online Dictionary*, plagiarism means

i)　To snitch and pass off the words and ideas of another as one's own idea.

ii)　Use another's work without crediting the source.

iii)　To commit classical or scholarly theft.

iv)　Present as new or original idea or product derived from an existing source.[3]

Plagiarism-detection tools are currently used in a large number of institutions to perform an initial assessment of students work, for identification of plagiarized work.[8]

Our aim is to introduce plagiarism, its types, numerous websites devoted to this issue, to recite reasons for increase in plagiarism cases, how to avoid plagiarism, how to detect plagiarism, do's and dont's of plagiarism and finally prevention and punishment for plagiarism.

Objective is to impart the habit of respecting the academic integrity and discipline, to identify any act of dishonesty in academic work which constitutes academic misconduct.

Plagiarism may be due to: copying (using another person's language and/or ideas as if they are your own), Collusion (unauthorized collaboration).

Methods include:
i)　Turning in someone else's data or illustrations without clear indication as your own.

ii)　Manipulating the critical work of others without due acknowledgement – even if you change some words or the order of the words, this is still plagiarism if you are using someone else's original ideas and are not properly acknowledging it.

iii)　Failing to put a quotation in quotation marks.

iv)　Copying and pasting from the Internet to make a 'patchwork' or pastiche/mixture of online sources.

v)　Colluding with another person, including another candidate.

vi)　Submitting as part of your own report or someone else's work without identifying clearly who did the work (for example, where research has been contributed by others to a joint project).[14]

Plagiarism can occur in respect to all types of sources and all media which are as follows:

i)　Not just text, but also illustrations, musical quotations, experimentations, computer code etc.

ii)　Not just text published in books and journals, but also downloaded from websites or drawn from other media.

iii)　Not just published material but also unpublished works, including lecture handouts and the work of other students.

II. REVIEW OF LITERATURE

A) Domain explanation:
As described above, this project is all about detecting the extent of plagiarism in a particular document. One of the important steps while developing such a system is to examine all the research areas thoroughly. After introduction, it is now necessary to know the basic details regarding plagiarism i.e. why people plagiarize, what are its consequences. Also for designing such systems, existing plagiarism detection systems are studied. Thus based on literature, the methodology and programming tools for this system is justified.

B) Common Techniques used by Plagiarists:
It is always profitable to know what common plagiarism techniques are practiced so that it would be easier to detect them. Some common techniques which are used by plagiarists are mentioned below:

a)　Changing the word using synonyms.

b)　Altering the order of their(words) occurrence.

c)　Mixing/infusing the original and copied text.

d)　Incorrect references(changing reference name).

e)　Changing the variable names, function names, class names.

III. PROPOSED SYSTEM

First we update the database until the papers submitted till date. The soft copies present in the database are loaded and then comparison begins. If an article/paper is submitted contains article which crosses 20% criteria of copied stuffs then the article will be given a "Plagiarized stamp" and it will give you a description and details from where it has been copied. If less than 20% of the article/paper is found with similarity it will be given a "Non-Plagiarized Stamp".
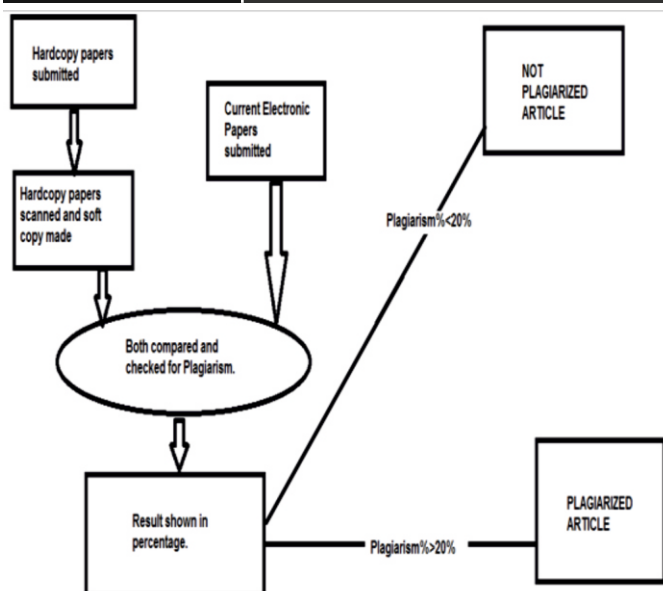
Visualization plays an important role while displaying large amount of information or data because it is very essential to compress such large amount of data into small area showing comparisons. Our system consists of 3 screens:

1) **Main Screen:** It will provide user to upload his/her paper/article.

2) **Browsing the files to be checked:** Here the files to be uploaded are browsed in order for checking the plagiarised content.

3) **Result:** Screen on which result will be displayed.
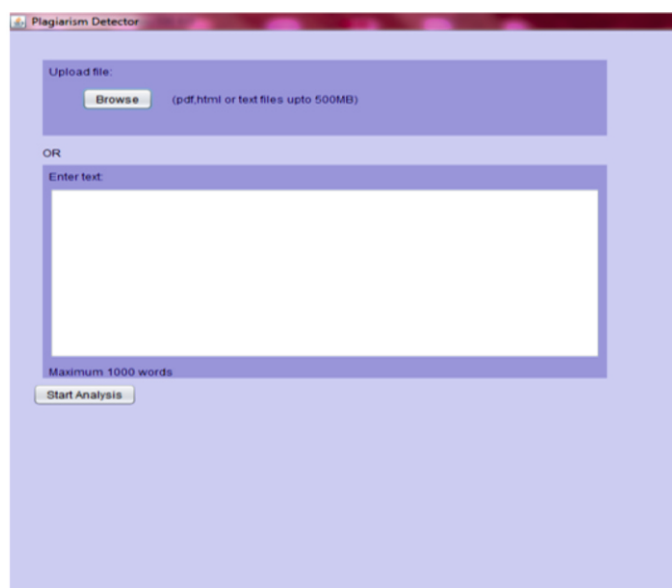
IV. MATERIALS AND METHODS

Methodology is analysis of the tasks to be done in order to obtain the desired output. Here, for this system, a number of methodologies were considered and the most efficient ones are used. This doesn't mean that only one particular method is used. According to the system, the most appropriate ones are used in combination. The methodology once decided is changed during the project if there arise any circumstances where the design emerged any flaws. Thus based on the situations appropriate methodologies/ techniques are implemented.[9][12] The algorithm used here is "Rabin Karp" algorithm which is as follows:

[1] Take the Text String that is to be matched for a pattern.

[2] Convert the alphabetic string to be matched to the alphabet values for [A,B, C.................Z] = [0,1,2.....................25].

[3] Take the Pattern String to be matched.

[4] Convert the Pattern String to the alphabets value same as step 2

[5] Choose a Prime number.

[6] Calculate the mod of the Pattern String against the Prime number chosen in step 5.

[7] Divide the Text String length to the every possible length of the Pattern String(if the length of the pattern string is 5 then divide the Text String to the every possible length of 5).

[8] Find out Mod of the every possible length of the Text String to the prime number.

[9] Match the every possible mod value calculated in the previous step to the mod value of the Pattern String.

[10] Do the manual matching for all the matched mod results for the Spurious Hits and Valid Hits against the actual alphabetic Pattern String that you want to find out in the text String.

V. RESULT AND DISCUSSION

This project aimed to develop a plagiarism detection system that detects the extent of plagiarism in a particular document uploaded by the client. Subsequent numbers of literatures were reviewed before starting the project. Design considerations were then carefully undertaken and implemented. The result obtained by implementing different algorithms and methods are within the desired framework. "Rabin Karp" algorithm and different methods are used and the result is shown as desired. The developed system is also compared with the existing plagiarism detection system. Though the system needs some improvements and the future enhancement is also a challenging task, the overall outcome of the project is as expected in its design consideration.

VI. CONCLUSION

The main reason why students and researchers plagiarize is because they do not understand what constitutes plagiarism. At present, there are no foolproof tools/techniques available to detect plagiarism but sincere efforts are being made in this direction and here professional can play a vital role. It is true, that no one can prevent plagiarism but sincere efforts can be made to reduce plagiarism.

REFERENCES

1. Mundava, M., & Chaudhuri, J. (2007, March). Understanding plagiarism, the role of librarians at the University of Tennessee in assisting students to practice fair use of information.College & Research Libraries News, 68(3), pp.170-173.

2. The Random House Dictionary of the English Language – unabridged

3. Merriam-Webster Online Dictionary

4. United States, the Office of Research Integrity (ORI) http://en.wikipedia.org/wiki/ Office_of_Research_Integrity (Accessed on 21/12/2008)

5. Society for Scientific Values. http://www.scientificvalues.org/cases.html (Accessed on 20/12/2008)

6. Devoss, Danielle and Rosati, Annette C. It wasn't me, was it? Plagiarism and the Web. Computers and Composition, 2002, 19, pp. 197.

7. Smith, Brian C. Fighting Cyber plagiarism. Library Journal Net Connect, summer 2003, pp.22.

8. Mcmrtry, Kim. E-cheating: Combating an 21st Century Challenge. T.H.E. Journal, November 2000, pp.37-38.

9. Wiebe, Glenn. Are You Legal? Copyright & Plagiarism in the Classroom. http://www.slideshare.net/glennw98/are-you-legalcopyright-plagiarism-in-the-classroom-202177 (Accessed on 24/12/2008)

10. Society for Scientific Values. http://www.scientificvalues.org/cases.html (Accessed on 22/12/2008)

11. Now, Pune scientists face allegations of plagiarism. Yahoo India (March 10 2007)

12. R. Karp and M. Rabin, "Efficient Randomized Pattern-Matching Algorithms," *IBM Journal of Research and Development*, 2007.

13. S. Grier, "A tool that detects plagiarism in Pascal programs," *ACM SIGCSE Bulletin*, 2001.

14. J. A. Faidhi and S. K. Robison, "An empirical approach for detecting program similarity within a university programming environment," *Computers and Education*, 2008.

15. K. L. Verco and M. J. Wise, "a comparison of automated systems for detecting suspected plagiarism," *The Computer Journal*, 2005.

16. M. J. Wise, "YAP3: improved detection of similarities in computer program and other texts," *Proc. of SIGCSE'96 Technical Symposium*, 2006.

17. L. Prechelt, G. Malpohl, and M. Philippsen, "Finding plagiarisms among a set of programs with Jplag," *Journal of Universal Computer Science*, 2008.

18. Professor Guide: Mr. Sunil Katkar

   Department Of Computer Engineering, Vidyavardhini's College of Engineering & Technology, K.T. Marg, Vasai Road (West), Palghar-401202, Maharashtra