

Awareness of Plagiarism: Marching Approaches and Renovations-A Survey

C.I.Arthi,

Assistant Professor, Dept of CMPN
Vivekananda Education Society's Institute of Technology
Mumbai, India

Dimple Bohra,

Assistant Professor, Dept of CMPN
Vivekananda Education Society's Institute of Technology
Mumbai, India

Abstract—Plagiarism is an act of stealing or reflecting others ideas or thoughts without the permission of authority or not mentioning that the ideas are the reflection of the authority. Plagiarism of others paper works is a crime, has been caught by plenty of intelligently coded plagiarism detection systems. The reason for this is lack of awareness about plagiarism. Some people are aware about plagiarism detection system's mechanism and they follow various ways to escape from the code, finally they do, because machine is a machine after all. In order to avoid this crime, there are plenty of papers are available insisting “knowledge theft” should be stopped by presenting wonderful detection techniques. But still we are in need of a concrete plagiarism detection system that can be able to track plagiarism by shuffling or manipulating synonyms of words and sentences in paper work. In this paper we discuss the evolution of plagiarism detection techniques along with the pros and cons of those methods.

key words: Intelligently coded plagiarism detection system, lack of awareness, knowledge theft, shuffling or manipulating synonyms of words

I. INTRODUCTION

Plagiarism in layman's term can be defined as “Copy – Paste”. However, in technical terms it can also be defined in terms of hidden crime called “knowledge theft”, ie stealing ideas of an authority and narrating that as their own paper work. Knowledge theft in turn can further defined as stealing Rational resource of human mind. Rational resource of a human includes the findings and original creativeness. And that should be legally protected. Even though there exist intellectual property rights such as copyrights, and lot of research has been going in valuable Algorithm and interpretations of plagiarism detection techniques to match and detect the true positives of plagiarism using data mining on material resources available internationally there is a lack of detecting thesaurus oriented plagiarism. Before knowing about the plagiarism techniques understanding its types is very important.

II. TYPE S OF PLAGIARISM

If a person is plagiarizing unintentionally the only solution to come out of plagiarism crime case is avoiding plagiarism by identifying the types of plagiarism. There are two major categories of plagiarism.

- I) Intentional plagiarism
- II) Unintentional plagiarism

I) Intentional plagiarism:

It is also called as direct or self plagiarism.

- a) Obtaining others creation in paper works as one's own creation.
- b) Collecting and copying data from internet without legal permission of the authority. It includes copyright data too.
- c) Clubbing two or three research works and making a new work without legal intimation from authority.
- d) Creating a new sentence from already existing sentence by slightly changing the synonyms of one or two words in the sentence.

This type of crime is considered as a perfect crime and penalty such as black listing and blocking the author from the service given.

II) Unintentional plagiarism:

It is also called accidental plagiarism.

- a) If a person unintentionally paraphrases the data by same group of words in sequence or similar form of sequence or similar form of structure without attribution, then it comes under unintentional plagiarism.
- b) Manipulating the position of comma, quotation marks and changing the pattern of content.
- c) Combining data from various citations and showing as a fresh original data of him.

In case of unintentional plagiarism there is a chance of false positives. Even though plagiarizing was done without intention still it is considered as a crime and the laws for it cannot be ignored. However In such cases minor penalty is conferred. Like simply dispensing warning and pausing the attempts of the author could be granted.

III. CRISIS OF PLAGIARISM: SAMPLES

1. Plagiarism suppresses the concept of assignment given to students. It is like substituting someone else's intelligence and somebody else submitting that as their own work. Therefore, plagiarism also affects goal of teaching students to learn and write, not just copy.
2. Plagiarism is considered to be a form of crime, because the professor is expecting students to learn and write, not someone else's already done work. Plagiarism destroys the relationship that should exist between teacher and student.
3. Plagiarism affects the purpose of scholarship. As it is Unacceptable by all scholars. The motive of scholarship is to innovate, understand and implement. That purpose is

affected when old knowledge is fraudulently presented as a student's new work. Hence plagiarism is not allowed in nonacademic professional fields such as journalism and creative article writing.

There some times in industry professions where originality and authorship are not important and it is sometimes appropriate to take other people's ideas without citing them and that too with authors' intimation. However, even if a student is training for such a profession, the student should learn and implement, not to plagiarize.

Whatever ways of studies and education available in our real world, plagiarism should be avoided and prevented by the people this environment.

Thus plagiarism is described as a form of stealing or copyright infringement. However, it is always unacceptable to plagiarize, even if the author of the work says you can use it.

IV. STATEMENT OF PROBLEM

The plagiarism problem can be defined as taking the writings, inventions or works of another person as one's own without proper intimation of author. In general, plagiarism could be done with copying textual or a pattern oriented documents. Examples include essays, paper reports, and codes of some programming

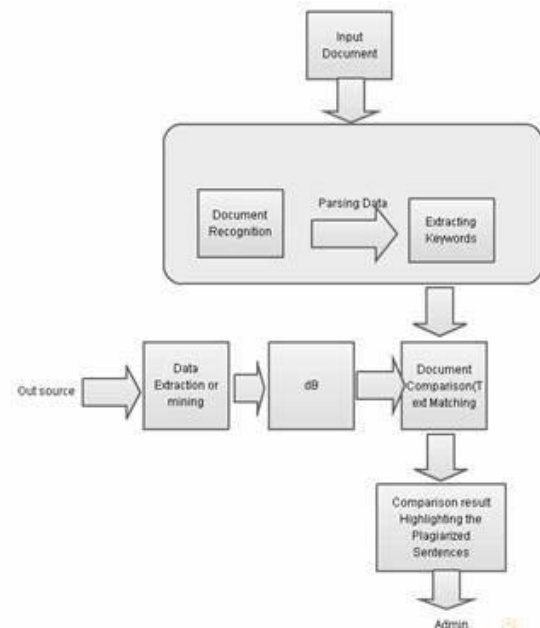
Languages. Several prior works has been done for detecting textual plagiarism [2][4] and programming code plagiarism [1]. In addition, as one may notice, the structural information which is implicitly contained in programming codes is taken as additional clues for plagiarism detection in corresponding works.

A special case of plagiarism occurs on the Internet that is "duplicate" or "near-duplicate" web pages. News, technical reports, specifications, articles are typical examples in which the same information is distributed in different web sites.

V. COMMON PDS FRAMEWORK

[2] The degree of similarity between patterns depends on the number of citations included in the pattern, and the extent to which their order and/or the range they cover is alike. Thus, literally matching

Sub-sequences of citations in two documents are a strong indicator for semantic similarity. The same is true for texts containing patterns that span over similar ranges, even if the order of citations in the pattern does not necessarily correspond towards each other. The width of the covered range can be expressed with regard to sequential positions of citations in the pattern, textual ranges or combinations of both. Measuring range width in units reflecting some semantics, e.g. paragraphs or sentences, is assumed beneficial compared to considering purely syntactic character or citation counts. For example, documents containing several matching citations, one of them within a single section, the other distributed over several chapters are less likely to be similar. However, if both share identical citations e.g. within a paragraph, then their potential similarity is respectively higher. Alternatively, e.g. the document tree may be used to identify semantic clusters in the form of chapters etc.



VI. DETECTION SYSTEM EVALUATION

The interest in automatic plagiarism detection is on the rise and increasing number of systems are being developed and implemented every year in order to satisfy the demand. The demand for new and more efficient systems is coming from both the publishing industry and the academia. The publishing industry is being wholeheartedly supported by governments that are enacting new, more restrictive and punitive intellectual property legislation. The general sentiment in academia is that the problem of plagiarism is getting more severe and educational institutions are making steps to procure PDS and implement them in their work flow. It is reasonable to assume that this trend of enforcing stricter regulations on the Internet will continue for some time regardless of the opposition and that it will encourage further research in this area so the number of available solutions will experience constant growth. In this light, an efficient methodology for evaluating available systems is needed.

A. Online plagiarism detection system

In this paper [5], the author Yi-Ting Liu proposed an online detection system that reduces the search engines misapplication. Collaboration of plagiarism detection system and search engines are used for extracting and verifying the documents. Specially with this proposed design, different priorities are given to extracted text segments, when linking them to search engines as the ascertainment of plagiarized texts.

This idea helps to reduce useless and unnecessary works when performing plagiarism detection comparisons.

B. Multilingual plagiarism detection

[6] In this paper a system for generating multilingual corpora to determine performance of plagiarism detection systems is used. A parallel language corpus is used in the implemented method and it can be applied to any language because of its scalability. The Authors have collected data from five parallel corpora and enabled corpus generation

for Croatian, French, German, Spanish and Italian language in order to do comparisons.

C. Two-Phase Plagiarism Detection Method

In this paper[7] the author, Du Zou have implemented a two-phase plagiarism detection method, which author have used in the learning management system of SCUT to detect plagiarism in the courses of network engineering. Also by this method external plagiarism in the PAN-10 competition can also be detected. The method has been divided into two steps: the first step, called pre selecting, is to use the successive same fingerprint to increase the scope of detection; the second step, called locating, ways here between two documents using cluster method find and merge all fragments.

D. Source Code Plagiarism Detection

[8] The arrival of open source code search tools have made the source code of open source software (OSS) readily accessible, hence increasing legitimate reuse, it has also opened up the possibility of unaware employees plagiarizing code from OSS repositories. Plagiarism in software codes would not only lead to costly lawsuits, but also impairs the organizations credibility. Hence detecting plagiarism in proprietary software is a basic requirement. Eventhough there exist a number of techniques available for detecting plagiarism in student project assignments, they do not scale well in the case of large proprietary software codes. Especially when code snippets are plagiarized open source software which are available. In this paper they propose a method that applies Mining Software Repositories (MSR) based techniques to perform plagiarism detection in great extend. Author have created a programming style profile for each maintenance engineer by mining the version history and use that to detect source code that has been plagiarized from open softwares. Those plagiarism suspected code fragments can be analyzed and detected using the existing plagiarism detection techniques to confirm the plagiarism and to prevent it.

VII.TABLE 1 – FACTORS OF PLAGIARISM COMPARISON

S.NO	FACTORS OF PLAGIARISM	A	B	C	D
1	Text Matching	✓	✓	✓	✓
2	Pattern Matching			✓	✓
3	Side by Side Matching	✓	✓	✓	✓
4	Links to Outsource	✓	✓	✓	✓
5	Finding Synonyms of words	X	X	X	X
6	Concept Checking	X	X	X	X
7	Thesaurus checking	X	X	X	X

CONCLUSION

Dealing with plagiarism, seems endless problem for academics and paper of citations. Also here we demonstrated about the awareness of plagiarism and some of the detection techniques. The actions against academic dishonesty plagiarism cannot be taken by people. But the efficiency of the plagiarism detection system could be improved intelligently coded softwares. However we are lack in such efficient software that works like a human and traces all kind of plagiarized works. Here we discussed about the pros and cons of previous approaches to fight against plagiarism. Still it has been found that we are backward in finding the synonyms of words in sentences and checking the concept of the work. Hence even-though there occurs the chances of false positives in case of comparison of documents, by thesaurus oriented methods and concept checking method, it should be considered for getting 100% results and to adore a plagiarism free environment in education.

REFERENCES

- [1] B. Belkhouche, A. Nix, and J. Hassell, "Plagiarism Detection in Software Designs," Proceedings of the 42nd Annual Southeast Regional Conference, pages 207-211, April 2004.
- [2] citation Pattern Matching Algorithms for Citation-based Plagiarism Detection: Greedy Citation Tiling, Citation Chunking and Longest Common Citation Sequence" by Bela Gipp and Norman Meuschke
- [3] plagiarism Detection in Large Sets of Press Agency News Articles," Proceedings of the 17th International Conference on Database and Expert Systems Applications, pages 181- 188, September 2006
- [4]. Nicole, "Title of paper with only first word capitalized," J. Name Stand. Abbrev., in press.
- [5] Extending Web Search for Online Plagiarism Detection Yi-Ting Liu, Heng-Rui Zhang, Tai-Wei Chen, and Wei-Guang Teng Department of Engineering Science National Cheng Kung University Tainan, Taiwan
- [6] Multilingual plagiarism detection corpus Vedran Juričić*, Vanja Štefanec**, Siniša Bosanac* Department of Information and Communication Sciences Department of Linguistics Faculty of Humanities and Social Sciences
- [7] A Two-Phase Plagiarism Detection Method Du Zou School of Computer Science & Engineering South China University of Technology GuangZhou,China duzou@scut.edu.cn Wei-Jiang Long.
- [8] An Approach to Source-Code Plagiarism Detection and Investigation Using Latent Semantic Analysis by Georgina Cosma