

Plagiarism from Tools to Techniques: A Review

Krishna Singh
UG student
DYP SOET
Maharashtra (INDIA)

Payal Yadav
UG student
DYP SOET
Maharashtra (INDIA)

Pallavi Dhormal
UG student
DYP SOET
Maharashtra (INDIA)

Arti Waghmare
Assistant professor
DYP SOET
Maharashtra (INDIA)

ABSTRACT

Plagiarism is an act of copying someone else's works or ideas and presenting it as their own work. It is a growing problem in various areas. It violates the intellectual property of people. This paper focuses on various techniques used by plagiarists, tools to detect plagiarism, impact of plagiarism in society, and landscape of various plagiarism detection algorithms.

General Terms

Plagiarism, repackaging, obfuscation, detection, impact, intentional

Keywords

Symbolic execution, Flow chart, structured programming, semantics of programming languages, similarity measures, algorithm, algorithm design and analysis, semantics.

1. INTRODUCTION

Plagiarism is an act of illegally stealing other's work or idea and manipulating it and claiming it to be their own. More appropriate words to describe it would be – literary theft, piracy, or cyber-cheating [1]. Usually it is deliberate but even an improper acknowledgement of sources in a re-published work could fall under plagiarism. Plagiarism is not only about cheating. It is one of the most serious intellectual crimes. It isn't only about morals and ethics; it could lead to reputations being destroyed and life's work turned into ashes [2].

It violates the terms of original license. Plagiarism can also be considered as a moral offence against the one who provides the plagiarist with a benefit in exchange for the original content. Plagiarism can be seen at various levels ranging from schools to big organizations. For example students copy their friend's answers in exam is also a kind of plagiarism. It can take several forms like it may include taking other's words from another source and presenting as your own or taking other's idea and presenting as your own without giving credit to the original author. People often do this to reduce their stress, effort or time and if plagiarists do not get caught, they get motivated to continue plagiarism.

Plagiarism has a major impact on the society. It becomes difficult for pursuers of knowledge to know what information is reliable and what source it actually came from; if one person is the origin of a particular information, who they are and his or her context can affect the value put on the data. Information based off plagiarism can negatively impact future research, publication or whatever is developed because this new information is unintentional and unknowingly based off of unrealistic origin. Imagine if everyone steals. Misinformation would become accepted as the truth and the whole world would eventually become misinformed, worst case scenario.

In academic plagiarism allegations can cause a student to be suspended or expelled. A professional business person or public figure may find that the damage from plagiarism follows them for their entire career. Many recent news reports and articles have exposed plagiarism by journalists, authors, public figures, and researchers [2].

Plagiarists use various techniques so that it becomes very difficult to differentiate between the plagiarized and the original content. For example obfuscation techniques [1] can be used by plagiarists to manipulate a code but preserving the semantics in such a way that it is difficult for humans to understand. This code can be used by plagiarists for reducing the cost of their software development. This copying of code could vary from open source code reuse to open source application repackaging.

So it is required to detect such plagiarism cases, for which various readily available tools can be used such as Doc Cop. Doc Cop is a collusion and plagiarism detection tool that allows to conduct scan from a file-upload and a web check. File Check can scan and compare a maximum of eight Microsoft Word or Adobe PDF files of up to 50,000 words each against one another. There are many other tools apart from Doc Cop which can help in detecting plagiarism like scopus, MOSS, JPlag etc. [8]

2. IMPACT OF PLAGIARISM ON SOCIETY

As discussed earlier plagiarism has major impact on society. This section will describe the impacts in detail. The impact of plagiarism can be personal, professional, ethical or legal. With readily available plagiarism detection software in use, plagiarists are being caught at an alarming rate. Once accused of plagiarism, a person will always be suspected as plagiarist. Plagiarists include academics, professionals, students, journalists, authors, and others.

2.1 Impact on student reputation

A student can be suspended or expelled due to plagiarism allegations. The student may be stopped from entering college or from high school because schools, colleges and universities take plagiarism very seriously. Many schools suspend students for first violation itself [3].

2.2 Impact on Professional Reputation

For a professional person the damage from plagiarism can follow them for their entire career. Not only will they be removed or asked to leave down from their present position, but they will also have to face problem of obtaining another good job. Depending upon the crime and the plagiarist's public status, plagiarism may ruin the plagiarist's name and make it impossible for him or her to make a meaningful career. [6]

2.3 Impact on academic reputation

The consequences of plagiarism are widely seen in world of academic .Plagiarism allegations may completely destroy persons academic's career. Publishing is an important part of a prestigious academic profession. If someone loses the ability to publish then it most likely means the end of an academic position for him or her and a destroyed reputation.[12]

2.4 Legal impact

The legal impacts of plagiarism can be quite serious. As the copyright laws are absolute ,an author has the right to sue a plagiarist .Some plagiarism may be considered as criminal offence ,in such cases the plagiarists may be sentenced to jail .Most susceptible to plagiarism cases are the people who write for a living such as journalists or authors.

2.5 Monetary impact

In recent years there have been many reports which have exposed plagiarism by journalists ,authors and researchers .If in case an author take legal action against a plagiarists ,the author may receive monetary compensation .If the plagiarist is a journalist working for some magazine or even a student in school ,he or she could have to pay monetary compensation.

2.6 Plagiarized research

A very atrocious form of plagiarism is plagiarized research .Particularly if the field is medical in nature the effect of plagiarism could result loss of people's lives .The consequences of plagiarism are far-reachable .Neither stature nor ignorance can be a excuse for someone to commit plagiarism.

2.7 Looser Morals

Plagiarism is correlated to other forms of cheating are .A student who cheats on a paper is likely to cheat on a test. This may become a habit and also can spread to other areas in life. An accountant who cheated in college is likely to cheat his clients .So , it indicates that plagiarism indirectly leads to looser morals.

3. TYPES OF PLAGIARISM

Plagiarism can be of two type .It may intentional or unintentional. The next section describes these two types in detail in the following section.

3.1 Intentional plagiarism

Intentional plagiarism is claiming lone origination of a slog that is known to have been written by someone else. Examples of Intentional Plagiarism. Purchasing a pre-written paper .Letting someone else write part or all of a paper for you.

3.1.1 Paraphrasing poorly

In this type of plagiarism, the plagiarist changes a few words without changing the sentence structure of the original, or changes the sentence structure but not the words.

3.1.2 Quoting poorly:

In this type, the plagiarist puts quotation marks around some part of a quotation but not around all of it, or puts quotation marks around an extract that is partly rephrased and partly quoted.

3.1.3 Citing poorly

In this type, the plagiarist may exclude an occasional excerpt or citing inaccurately.

3.2 Unintentional Plagiarism

Unintentional plagiarism is plagiarism that results from the ignorance for proper pedagogic steps. Examples of Unintentional Plagiarism: Failure to reference source that is not common . Failure to "quote" or block quote author's same words, even if it is documented.

3.2.1 Copy

This type of plagiarism includes copying something from the web page ,net source or database without giving credit to original owner.

3.2.2 Cut and Paste

This type includes cutting and pasting from more than one source and creating a paper without quoting.

3.2.3 Borrow

This type of plagiarism includes borrowing ideas from other students or sources without giving them proper credit.

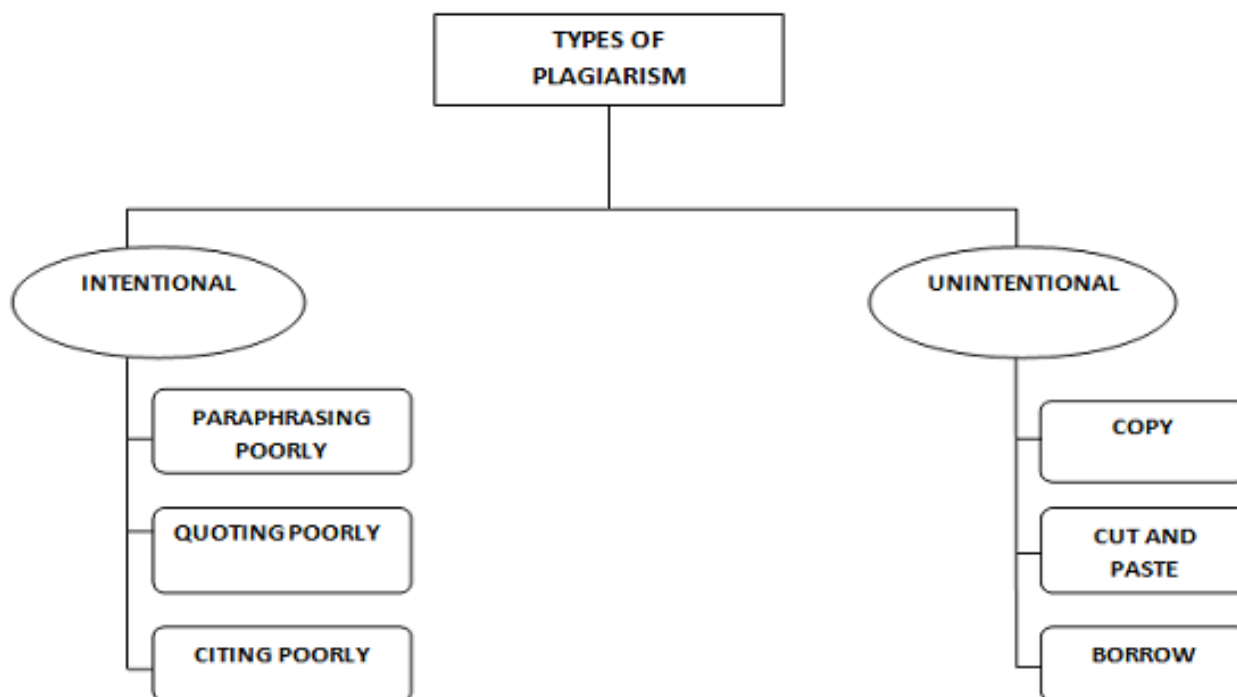


Fig 1: Types of plagiarism

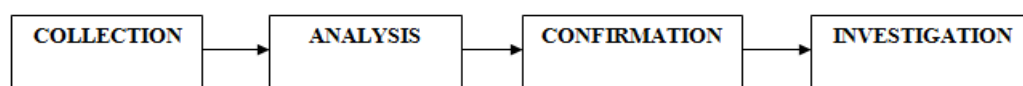


Fig 2 : Text based plagiarism detection

4. PLAGIARISM DETECTION

Plagiarism detection is the process of scanning a document and analyzing its content in order to disclose portions of that document that have been plagiarized or taken from other sources without giving appropriate credit .It may include bringing up similar source of documents together, if available[3].

Plagiarism can occur in any type of document, not only in academic works, it is also found in the Press, in the program code, in art designs, etc. The extensive use of computers and the rise of the Internet has made it easier to plagiarize the work of others. Even if it is being a practice which breach morality, plagiarism is occurring far more frequently than people can imagine. The main reason behind the growth in plagiarism is the ease with which people can achieve access to online content .Detection of plagiarism can be done either manually or it may be software-assisted. Manual detection of plagiarism requires significant effort and excellent memory, and in cases where large documents must be compared, or original documents are not available for comparison it becomes impractical. Due to the inability of humans to

process large documents and to recover all suspected portions and original sources ,the need for computerized plagiarism detection systems becomes feasible. [5]

4.1 Plagiarism Detection Techniques

There are many plagiarism detection techniques available like such textual based plagiarism, citation based plagiarism, and

shape based plagiarism for flowchart. .The following sections gives a brief introduction of text based and citation based plagiarism detection techniques.[7]

4.1.1 Text based plagiarism detection

This type of plagiarism detection technique uses vector space model to detect similarities among documents .It enumerate and tally the repetition of words in the document and then the fingerprints are used for each document to match it with the fingerprints of other documents in order to find the similarity. This method works well for non fractional plagiarism but for partially plagiarized documents it does not provide good results. This may include copy and paste or modification of word of original document from online sources. [4]

A) Stages in text based plagiarism detection

1) Collection.

This is the first stage. it requires the student or researcher to upload their documents to the web engine. This web engine performs as user interface.

2) Analyses

The second stage is analysis. N this stage are the collected documents are scanned through a similarity engine to find out similarity between documents.

3) Confirmation

This stage determines if the pertinent text that has been plagiarized from some other text.

4) Investigation

This is the final stage and it requires human intervention. In this step a human expert determines if the system performed correctly and also determines if the result is actually plagiarized or it is simply cited.[9]

4.1.2 Citation based plagiarism detection

Citation based plagiarism detection (CbPd) uses references and citations to determine similarities between documents in order to identify plagiarism. References and citations contain valuable information about the content of a document and related works in the academic environment. [3]

A) Identification of citation pattern

The core idea of CbPd is to find similar pattern in the citations used within two texts. C1 and C2 are citation tuples of two documents 1 and 2 respectively that consist of partially shared references by which it can be concluded that they are similar to each other.

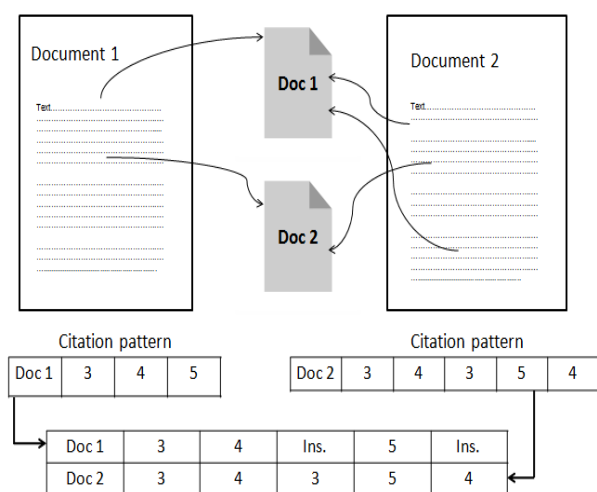


Fig 3: citation based plagiarism detection

The extent of similarity between texts depends upon the number of citations included in the pattern or the degree to which their order or range matches. Thus matching subsequences of citations in two documents strongly indicates semantic similarity. [10]

B) Process of CbPd

- 1) The documents are anatomized and a series of investigation is applied to process the citations.
- 2) This step includes matching of citations according to their entries in their bibliography.
- 3) Then, the similarity among documents is calculated on the basis of citations.

4.2 Plagiarism detection tools

Many types of tools are available that help in reliably determining if a submitted document is plagiarized or not. For example Plagiarism Checker, PlagScan, iThenticate etc.

A) PlagScan

PlagScan is used to check document plagiarism. It is often used by school and provides different types of account with different features. It uses complex algorithms to check and inspect input document for plagiarism detection, based on up-to-date linguistic research. Unique signature extracted from

the document's structure that is then compared with PlagScan database and millions of online documents.[11]

B) iThenticate

iThenticate is one of the application or services designed especially for the researchers, authors' publisher and other. It is designed to be used by institutions rather than personal, but lastly they provided a limit service for single plagiarism detection user like master and doctoral students and this allows them to check a single document of up to 25,000 words. So they can use this service to insure or to check their draft thesis whether containing correct citation and content originality.[11]

C) PlagiarismDetection.org

PlagiarismDetection.org is an online service that provides high level of accuracy in result of plagiarism detection. Mainly designed to help the teachers and students to maintain and prevent and detect plagiarism against their academic documents. It provides quickly detect plagiarism with high level of accuracy.[11]

5. CONCLUSION

Plagiarism is an act of illegally stealing other's work or idea and manipulating it and claiming it to be their own. It is a serious problem in the society and it violates the intellectual property of people. This paper presents a review on plagiarism. It describes about what is plagiarism and what are the different areas where plagiarism can have its impact. This paper also focuses on the types of plagiarism such as intentional and unintentional plagiarism along with various techniques used by plagiarist to avoid getting caught. Few types of plagiarism detection methods and tools are also discussed. Future work can be, exploring more types of tools and techniques to detect plagiarism. This study can also be made more specific by focussing on a particular type of plagiarism and its detection techniques. Review about new and advanced plagiarism detection tools can also be added.

6. REFERENCES

- [1] Tripathi, Tiwari, P.Nithyanandam, "Avoiding plagiarism in research through free online plagiarism tools" 2015 IEEE 4th International symposium on emerging trends and technologies in information services.
- [2] Fangfang Zhang, Dinghao Wu, "Program logic based software plagiarism detection" 2014 IEEE 25th International Symposium on Software Reliability Engineering.
- [3] K. Žáková, J. Pištej and P. Bisták, "Online tool for student's source code plagiarism detection" 11th IEEE International Conference on Emerging eLearning Technologies and Applications • October 24-25, 2013.
- [4] S. Arabyarmohamady, H. Moradi, M. Asadpour "A Coding Style-based Plagiarism Detection", 2012 IEEE International Conference on Interactive Mobile and Computer Aided Learning (IMCL), November 6-8, 2012.
- [5] D. Chuda, P. Navrat, B. Kovacova, P. Humay, "The Issue of (Software) Plagiarism: A Student View", IEEE Transactions on Education, Vol. 55, No. 1, February 2012.
- [6] S. Mann and Z. Frew, "Similarity and originality in code: plagiarism and normal variation in student assignments," Proceedings of the 8th Australian conference on Computing education - Volume 52, 2006.

- [7] Asim M., El Tahir Ali, Hussam M. Dahwa Abdulla, "Overview and Comparison of Plagiarism Detection Tools" Department of Computer Science, Germany & UC Berkeley JCDL 2011.
- [8] L. Prechelt, G. Malpohl and M. Philippsen , "Finding plagiarism among set of programs with JPlag " J. Univ.Comput.
- [9] Ahmed Hamza Osman, Naomie Salim¹, and Albaraa Abuobieda, "Survey of Text Plagiarism Detection"
- [10] International University of Africa, Faculty of Computer Studies, Khartoum, Sudan Vol. 1, No. 1, June 2012Bela
- Gipp, Norman Meuschke, Joeran Beel" Comparative Evaluation of Text- and Citation-based Plagiarism Detection Approaches using GUTENPLAG".
- [11] S.A.Hiremath, M.S.Otari "Plagiarism Detection- Different Methods and Their Analysis: Review" International Journal of Innovative Research in Advanced Engineering (IJIRAE) ISSN: 2349-2163 Volume 1 Issue 7 (August 2014).
- [12] D. McCabe, "Cheating: Why students do it and how we can help them stop," in Guiding Students from Cheating And Plagiarism to Honesty And Integrity: Strategies for Change. Libraries Unlimited, 2005, pp. 237–246.