Review Article

Indian Journal of Library and Information Science
Volume 15 Number 1, January - April 2021
DOI: http://dx.doi.org/10.21088/ijlis.0973.9548.15121.3

War Against Plagiarism and IPR Violation

B V Chalukya

How to cite this article:
B V Chalukya/War against Plagiarism and IPR Violation/Indian J Lib Inf Sci 2021; 15(1): 29–34.

Authors Affiliation:

Librarian, Shri Chhatrapati Shivaji College, Omerga, Osmanabad, Maharashtra 413606, India.

Address for Correspondence:

B.V. Chalukya, Librarian, Shri Chhatrapati Shivaji College, Omerga, Osmanabad, Maharashtra 413606, India.

E-mail: drbvcin@gmail.com

Abstract

For several institutions and organizations, plagiarism and intellectual property rights (IPR) abuses have become a serious concern. There are many opportunities for such violations to become much more common through the revolutionary growth of the Web. The possibility of adopting a 'culture of mediocrity' is generated by this situation. This paper explores these problems and suggests ways of solving the issues through the implementation of feasible technical solutions. New media, especially the Internet, are contributing to an explosion in violations of both plagiarism and IPR.

Key words: Plagiarism; Intellectual Property rights; Violations; Google; Detection tools; Plagiarism assertions; Stylometry; etc.

Introduction

At present, the Web is evolving so rapidly that deciding whether anything is truly new is becoming a challenge. At lightning speed, web content is created, shared, and transferred, making it incredibly hard to work out the degree of originality. Therefore, plagiarism and infringement of intellectual property rights (IPR) are issues that plague many institutions and organizations. For instance, by evaluating their academic or literary achievements, educational institutions need to determine the caliber of their students. The novelty of their own IPR inventions must be decided by organizations. And the publishing or creation of original works is rewarded in both environments.

Plagiarism is the unauthorized use or near imitation of the work of another author portrayed as an original work of one's own. Without proper attribution or acknowledgment of its source, plagiarism can therefore be seen as the stealing or 'borrowing' of published work. We classify plagiarism as the use of the content of others (text, photographs, movies, etc.) without precise source specification, whether the material is unchanged or in some sort of derivative form. On the other hand, IPR infringement requires the use or misuse of works that exceed the scope of its legal security. IPR infringement is the unauthorized use of content covered by IPR law in a way that violates the exclusive rights of the original copyright owner to reproduce or expand on the copyright work. Therefore, IPR infringements are content uses, in original or derivative form, which go beyond what is allowed under legal copyright exceptions (such as 'criticism and review'), whether the original source is cited or not.[1]

We consider them to be violations of any reasonable and ethical code of conduct since they are closely connected and one always leads to the other. Among the two, plagiarism is the more common, and in the academic field, it has thus been extensively deliberated. On the other hand, infringement of IPR appears to be taken even more harshly because it can have a direct effect on copyright owners' earnings, contributing to the loss of both revenue and power over the way the content is used. This paper highlights the negative effects of both, calling for an urgent solution to resolve what we see as potential risks and dangers.

Plagiarism and IPR violations are not new phenomena; however, both are causing an explosion in the new media, especially the Internet. Infringements in all types of digital forms, including volatile media such as SMS, chat and mail, will now occur outside print media. This makes the systematic treatment of plagiarism and IPR infringement even more complicated.[2]

The Web has made it possible to publish instantaneously, primarily by providing instant access to myriad information sources. Such a phenomenon has a huge effect on the quality of creative thought and writing, and thus on the quality of life. There are a range of instruments to tackle plagiarism and violations of IPR; we recommend ways to resolve the problem effectively, including an institutional approach, and through the use of viable technologies.

Plagiarism and IPR Violation

A variety of formats such as university term papers, theses, and academic papers; essays and other written assignments in a school; and all sorts of other media such as project papers, news stories, and web material can be affected by plagiarism and IPR violation (blogs, wikis, etc.). Especially in an academic setting, plagiarism is a major concern, where it may impact both the reputation of institutions and their capacity to ensure the quality of their graduates. Plagiarism has been growing; the Internet and the World-Wide Web are largely responsible for the growth. According to officials from the University of California-Berkeley, from 1993 to 1997, cheating on campus increased by about 744 percent. As cited by Plagiarism.org, a national study released in Education Week highlighted that 54 percent of students acknowledged plagiarism from the Internet and 74 percent of students admitted having committed 'serious' cheating at least once. Many students prefer to take plagiarism lightly and deem it completely appropriate to have a degree of copying.[3]

In one severe instance, shortly before he was supposed to obtain his degree, a student was found plagiarizing. The student, who confessed to uploading Internet essays, justified himself by saying,' It would be fair enough if they had pulled me up with my first essay at the beginning and warned me of the problems and consequences. But with decent grades, all my essays were handed back and no one noticed it.' He then actually went on to sue his college for not catching him sooner! This clearly highlights a lack of accountability on the part of students who, without due regard to the legality of their acts, prefer to turn to the quickest means of getting work done. A significant number of 'paper mills' operate to make the situation worse, directly assisting students in the preparation of term papers. While there are reports about the ethical usage of their services on some of these pages, they make plagiarism way too simple for students to avoid. There are other sources of information that can be used by students in addition to these 'paper mills': web directories, Wikipedia, online bookstore book reviews, academic journals, and so on, all easily retrieved through search engines.Parents will spend $75,000 on high school tuition and $120,000 on a private college, as illustrated by Fox News, and then pay even more to ensure that their child does not learn anything. This article is an account of the life of a professional paper writer who, by producing their term papers for them, helped fully uninterested students to earn high marks.[3]

However, plagiarism and IPR infringements are not limited to students; they can even include professors, a college vice-president, or even a prime minister. Therefore, journals and conferences would have to take plagiarism seriously, as authors' submissions will well be essentially self-plagiarized (plagiarized from their own past works). A high degree of self-plagiarism suggests that a large part of the paper has been published before, and this may lead to a copyright violation inadvertently committed by a journal or conference.

Even government and commercial organizations, rather than plagiarism per se, are primarily concerned with IPR violations and excessive spending. It was reported that one government agency was more concerned about the risk of being sued for breach of copyright than about acknowledging the effects of plagiarized works. Likewise, the US federal government takes measures only against the plagiarism of the works it has sponsored, not against the plagiarism of the works of others. In other words, their laws mostly protect their own intellectual property and, more broadly, do not fight plagiarism.

The severity of overlooking plagiarism has been largely neglected; however, in research it may lead to increased dishonesty. The case of Frederick Cook vs. Robert Peary shows the challenge of resolving disagreements over potentially false research claims; plagiarism will make it impossible for researchers to prove their own discoveries with a prior argument. In the other hand, where none has happened, there are still instances of innocent persons being convicted of plagiarism. Furthermore, plagiarism may occur from genuine

incompetence in documenting and attributing a reference; an individual may accidentally omit a reference, for instance. The author was found not guilty of plagiarizing the Hertzberg piece in the case of the Star Tribune' Plagiarism Investigation '; he had failed to distinguish in a reported transcript between direct quotations and paraphrased ideas, and consequently did not identify the original author.[4]

Copy-Paste Syndrome by Google

The 'Google Copy-Paste Syndrome' (GCPS) is the name given to the widespread practice of copying, extracting and reusing passages from existing texts instantly, easily (and typically casually researched) by scientists and journalists alike. They actually carry out easy searches instead of gaining true insights through a systematic process of learning through scientific discovery; information from the Web is often used without even considering the integrity of its original source.[5]

A proliferation of plagiarism has resulted from GCPS. As the answer (or at least an answer) shows up conveniently, with minimal effort, it can potentially hinder the enquiry-driven science method. This syndrome, by de-emphasizing the need for intentional and insightful reasoning, has thus threatened original writing and thinking. As a consequence of the lack of careful thinking and comprehension, this new phenomenon promotes mediocrity in published works. A society without brains is developing as the 'global brain' takes form by supplying responses to all queries.Thus, the view of reality offered by the Web is considered to be a replacement for the hours that would otherwise be spent on initial inquiry and reflection. Weber aptly notes that 'by googling' we are in the process of creating reality.

This declaration emphasizes the strong reliance on the content indexed by search engines such as Google and in content warehouses such as Wikipedia by many of us, particularly the younger generation.As defined by Kulathuramaiyer and Balke, search engines appear to limit or distort the view of users intentionally or unintentionally. Furthermore, search results are ranked using algorithms that create a bias against famous sites (i.e. often linked to) and can thus not provide an entirely authentic recording of historical events.[6]

Detection of plagiarism and IPR violation

In the identification of plagiarism and IPR infringement, there are a variety of factors to be taken into account, since they do not necessarily consist of simple verbatim copying of text parts. In addition, the text may be modified to a degree that makes it incredibly difficult to detect-copied text may be paraphrased or converted into another language. Plagiarism can often include copying smaller bits of material. If you fail to quote the original work, paraphrasing - condensing the work of another author or putting the words of the author into your own words-can be called plagiarism. As demonstrated by the guidance of Thomson Publishing to authors, paraphrased texts can also be seen as violations of IPRs.[7]

To be excluded from copyright, if you paraphrase content, it has to be radically different from the source. If, without difficulty, a reader finds similarities between the paraphrased text and the original edition, then permission must be requested. When paraphrasing content, there is no simple way to reliably quantify the need for permission, so it is usually advisable to ask the copyright holder for further advice.[7]

Both detection of plagiarism and detection of IPR violations depend on the ability to recognize similarities between documents. This includes calculating the degree of similarity between a (original) source document and a (potentially copied) target document. Thus, identification of document similarities includes a large database of documents and texts. Of course, the original source records may not always be available in digital form or may be inaccessible behind barriers to access; however, libraries and major search engine companies are now carrying out mass digitization initiatives to enable easy access to such sources of documents.

Furthermore, several publishers now allow search engines to index their subscription-based publications; search engines now have access to publications deposited on local servers by authors. This explains why it performs substantially better than the leading commercial plagiarism detection systems for a document similarity detection method using Google's search engine. However, coping with the deep Web (also known as the invisible or secret Web), World Wide Web material that cannot be viewed explicitly by search engines, is the greatest challenge faced in document similarity detection.. The DeepWeb is much larger than the user-familiar surface web; it comprises approximately 550 billion individual records, compared to 1 billion on the surface web. Less than 5 percent of the deep web is accounted for by subscription-access sites; others are database-driven outlets that generate items in response to queries. Plagiarism and the

identification of IPR violations will therefore remain a problem.[8]

Photos, as well as the information they contain, also need to be secured, but it is very difficult to identify similarities in all images and any text they can incorporate. For the extraction of text from images, OCR-based techniques are usually used. A large proportion of text in photographs can be digitized without much trouble, as OCR technology is reliable. However, as far as the images themselves are concerned, advanced techniques for image processing will be too computer-intensive to apply to massive image datasets.In order to extend identification to non-textual resources, novel methods are therefore required; Google, for example, uses collective image labeling as a basis for clustering similar images, rather than relying on image processing techniques.[9]

For plagiarism detection systems, there are other considerations. The work of a large team of researchers can result in publications. While a long list of authors may be popular in some fields, such as medicine, in others, such as computer science, authorship may be credited to only the key contributors to specific ideas in a report, instead of listing the names of all members of the research team. This growing result in the publishing of similar content by the various members of the original community independently of each other.[10]

Plagiarism allegations can have a significant effect on the credibility of the complainant, even if the allegation is found not to be justified. In certain circumstances, as it would be difficult to prove, writers whose works have been plagiarized can prefer to take no action. Software for detecting plagiarism may also help to substantiate the assertion of an author to be the original author of a published book.But to generate incontrovertible evidence, plagiarism detection systems can never be relied on; all they can do is suggest that plagiarism may have occurred! Therefore, a manual search is often sufficient to determine whether plagiarism actually occurs. Complete dependency on an automated detection system for plagiarism would eventually yield false positives, which could be catastrophic.[11]

Somefree plagiarism detection tools for e-learning professionals: Digital technology and the growth of the Internet have given us access to loads of information from anywhere on the globe, whenever we desire it. Initial concepts tend to get rarer and rarer. It seems like everybody is reproducing the thoughts of other people and posing them as their own. Although this activity is not something new, many people would argue that it has reached its height nowadays. I'll list the top 10 free plagiarism detection tools in this article that will allow eLearning professionals to tackle the nightmare of plagiarism.[12]

- *Dupli Checker:* This is one of the Internet's most powerful free plagiarism detection methods. It definitely gets the job done well, although it doesn't have a fancy design.
- *Copyleaks:* This cloud-based authenti -cation framework helps you to control how content from eLearning is used all over the internet.
- *Paper Rater:* A free plagiarism detection multi-purpose tool that is used in over 140 countries.
- *Plagiarisma:* Basic and easy-to-use, multi-purpose plagiarism detection tool that is used by students, teachers, writers, as well as various members of the literary industry.
- *Plagiarism Checker:* User-friendly, entirely free plagiarism detection tool to check whether content is plagiarized.
- *Plagium:* Basic but fully functional free plagiarism detection tool with different levels of search.
- *PlagScan:* Tool for detecting plagiarism for both individuals and organizations, which also tests texts against online material, scientific journals and user documents.
- *PlagTracker:* Fast free plagiarism detection tool that searches both websites and academic databases by copying and pasting text, or file uploading.
- *Quetext:* Basic layout and functional interface that checks against the Internet, as well as various databases.
- *Plagiarismhunt:* Online plagiarism checker, which tests with one click on 5 different plagiarism software systems.

Usual approach

A document is split into a (large) collection of 'fingerprints' using the standard method to detect plagiarism. In order to distinguish similar documents, a fingerprint consists of one or more sentences which are then applied as query strings to search the Web or a specific database. Most software packages currently available use this technique; they differ only in the method used to select fingerprints, the type of the fingerprints, and the search engines used. The benefit of this approach

is that if the order of the text is rearranged, it is not invalidated. However, it will not detect synonyms and translations.[13]

Manual Detection

This method involves selecting a sentence, or one or more sentences, manually, describing a unique concept contained in a text. This selected text is then used by one or more search engines as a query. This method can be replicated, concentrating and refining the question phrase in the process a number of times.Although this technique is basic, its ability to discover plagiarism can be impressive. In formulating meaningful questions, the success of this method depends primarily on the domain specific expertise of a human expert. It is also likely that such an expert may know the probable source of a piece of text in advance, which would help narrow down the quest. This method may, instead, be partially automated or implemented in combination with other approaches.[13]

Checking Plagiarism Assertions

To disprove alleged plagiarism, unique instruments are needed. A Cloze technique may be used when a conflict occurs to judge the probable original author of published works. Cloze operates in a regular pattern by concealing individual words in a text. It is then important for the author to fill in the blanks with the words he or she considers acceptable. It has been found that a document's original author is more likely to choose the right terms than a plagiarizer.[14]

Stylometry

Stylometry is a methodology focused on similar trends that analyzes writing styles. A specific text, based on his or her past works, may be contrasted with an individual's traditional writing style. Alternatively, it is possible to equate the text in a single paragraph with the overall writing style as found elsewhere in a document. Stylometry is capable of detecting plagiarism without the need for an external corpus of documents, unlike the other methods mentioned. Within texts, such as syntactic forms and text structure, as well as the use of key words, it can identify stylistic trends.[15]

Creation of application using Google

Designed on top of Google's search engine, a home-grown plagiarism detection method has unexpectedly provided better results than some of the leading commercial software packages, such as Turnitin and Mydropbox. This is primarily because many more websites are indexed by Google. While most papers are not publicly accessible on the Internet, Google and other search engines are encouraged by most publishers to crawl and index the full text. Moreover, many writers publish a pre-print version of their own publications in institutional repositories or on personal websites, which can then be indexed by search engines (e.g. Google Scholar indexes institutional repositories). However, free access by Google to their search engine sets a cap of 1,000 queries a day.[15]

Advanced Identification of Plagiarism

Although current instruments of plagiarism seem sufficient, they cannot cope with a serious 'trained' plagiarist. Instead of only phrases, natural language processing methods may be used to identify plagiarism at the level of ideas; an essay-grading method has been suggested in a different context that produces a proprietary model of information representation-model responses written by teachers are compared to student responses, to evaluate grade assignment, and this approach has been able to balance the approach so far.In order to detect plagiarism in software development programs, the use of graphs has also been suggested to explain deeper trends such as dependencies between program flow and written code; such a method is unaffected by rearranging the order of the document. Other approaches to the determination of similarity at the level of ideas use concept maps to reflect the domain expertise of an expert.[16]

What are we capable of doing

A system that can assess the degree of similarity between works lies at the heart of the identification of plagiarism or IPR infringement. This skill is extremely valuable because it is possible to apply the same technology to other fields, such as checking the originality of patent applications or even providing answers in natural language to questions asked by users! Kulathuramaiyer and Maurer have identified a proposed center for the creation and maintenance of a 'federated' text database. This center will then hire suitable software to identify textual similarities, offering a range of intellectual property security services. In order to improve the transparency and expertise of information staff, it will also be responsible for developing training modules to encourage academic and scholarly best practices.

Conclusion

The significant consequences of plagiarism and IPR

infringement have been highlighted in this article. If the problem is not solved, sales will be lost and, worse, scientific culture will be degraded, and indeed culture in general. Therefore, technology for plagiarism and IPR violation detection is absolutely necessary. A detection center for plagiarism and IPR violation will help to pool resources in both textual and non-textual resources to establish universal instruments for the detection of plagiarism and IPR violation.

Reference

1. Wikipedia, Copyright infringement, http://en.wikipedia.org/wiki/Copyright_infringement.
2. Smithers, R. 2005. Crackdown urged on web examplagiarism. Education Guardian, 22 November 2005.Available at http://education.guardian.co.uk/gcses/story/0,,1648106,00.html.
3. Curtis, P. 2004. Quarter of students 'plagiarise essays'.Education Guardian, 30 June 2004. Available athttp://education.guardian.co.uk/students/work/story/0,,1250786,00.html.
4. Smith, A. 2006. Plagiarism 'rife' at Oxford. EducationGuardian, 15 March 2006. Available at http://education.guardian.co.uk/higher/news/story/0,,1731423,00.html.
5. Fredericks, M.A. 2002. Cheating, the copy-and-pasteway. The Star Online Exclusive Report. Available athttp://202.186.86.35/special/online/plagiarism/mike_cutpaste.html.
6. Jacobs, J. 2004. History without history, spelling withoutspelling. Fox News, 4 June 2004. Available athttp://www.foxnews.com/story/0,2933,121840,00.html.
7. Plagiarising student sues university for negligence. Guardian Unlimited, 27 May 2004. Available at http://education.guardian.co.uk/higher/news/story/0,1226148,00.html.
8. Collberg, C. and Kobourov, S. 2005. Self-plagiarism incomputer science. Communications of the ACM, 48:88–94.http://doi.acm.org/10.1145/1053291.1053293.
9. Maurer, H. and Zaka, B. Plagiarism – a problem andhow to fight it. In Proceedings of Ed-Media, AACE, USA, 2007, pp. 4451–8. Available at http://www.iicm.tugraz.at/iicm_papers/plagiarism_ED-MEDIA.doc.
10. Henderson, B. True North. Peary, Cook, and the Race tothe North Pole. New York, W.W. Norton, 2005.
11. Star Trib plagiarism probe clears writer. Fox News, 16December 2006. Available at http://www.foxnews.com/wires/2006Dec16/0,4670,StarTribunePlagiarismProbe,00.htm.
12. https://elearningindustry.com/top-10-free-plagiarism-detection-tools-for-teachers
13. Weber, S. Das Google-Copy-Paste-Syndrom: WieNetzplagiateAusbildung und Wissengefährden[TheGoogle copy–paste syndrome: how Net plagiariststhreaten education and knowledge]. Hanover, Heise,2006.
14. How to Recognize plagiarism. Indiana Universitywebsite. Available at http://www.indiana.edu/~istd/example1paraphrasing.html
15. Guide for Authors, Thomson Learning, http://hed.thomsonlning.co.uk/authors/10copyright.pdf.
16. Bergman, K. 2001. The 'Deep' Web: surfacing hidden value. Journal of Electronic Publishing, 7, July.Available at http://www.press.umich.edu/jep/07–01/bergman.html.