2008

Deterring Digital Plagiarism, How Effective Is the Digital Detection Process?

Jayati Chaudhuri
University of Northern Colorado

Follow this and additional works at: http://digscholarship.unco.edu/libfacpub

Part of the Scholarly Communication Commons

| **Home** | **Table of Contents** | **Titles & Subject Index** | **Authors Index** |

Deterring digital plagiarism, how effective is the digital detection process?

Jayati Chaudhuri

M.Sc., M.L.I.S., Assistant Professor, University of Northern Colorado Libraries, Greeley, CO 80639, USA. Email: Jayati dot chaudhuri at unco dot edu

Abstract

Academic dishonesty or plagiarism is a growing problem in today's digital world. Use of plagiarism detection tools can assist faculty to combat this form of academic dishonesty. In this article, a special emphasis is given to text-matching software called SafeAssignment^{TM}. The advantages and disadvantages of using automated text matching software's are discussed and analyzed in detail. The advantages and disadvantages of using automated text matching software's are discussed and analyzed in detail.

Keywords

Introduction

While academic dishonesty is not a new phenomenon, there is no agreement about why plagiarism is so prevalent in the academic world. It is broadly acknowledged that online plagiarism is really high because of the easy availability of information (Mundava & Chaudhuri, 2007).

Plagiarism comes from the Latin word *plagiarius*, which means abducting or kidnapping (Hansen, 2003). According to Encyclopedia Britannica plagiarism is 'the act of taking the writings of another person and passing them off as one's own' (*Britannica Online Encyclopedia*). Plagiarism is unfair use of somebody else's work without giving credit for it. It is necessary to cite and acknowledge the sources even if those ideas are paraphrased and re-written with different words. Plagiarism is unethical and can hurt any academic institution's reputation. There is a difference between plagiarism and copyright infringement. Plagiarism is imitation of ideas or writings without any acknowledgement as opposed to copyright infringement which is extensive use of somebody's work without permission, with or without acknowledgment (Plagiarism

[Tutorial](#)).

Literature Review

A review of the literature reveals that plagiarism is widespread in any level of the society. There are several cases in which authors, historians, and even a university president faced accusation of plagiarism. This article, however, specifically focuses on students' plagiarism which is a rising problem not just in the U.S.A. but all over the world. The availability of the Internet entices students to do 'cut and paste' plagiarism. As Prof. Susan Bassnett from University of Warwick indicates "Across UK universities, we now have a cut and paste culture which is becoming difficult to detect" (Adenekan, 2003). The recent plagiarism case flashed in the media involved a Harvard University Sophomore student who was accused of using plagiarized material from multiple sources in her book 'How Opal Mehta Got Kissed, Got Wild and Got a Life' (Madray, 2007). A Wall Street Journal, August 2006 article reveals engineering students' plagiarism at Ohio University (Tomsho, 2006). Another news article reports how rampant plagiarism is at Oxford University, one of the premier academic institutions in the world (Smith, 2006). In a noteworthy article in 2006, Prof. Grafen mentioned that 'vigilance is required for the sake of the education our students receive, both in the substance of the subject and in the proper scholarly practice; and also in order not to create implicit understandings that plagiarism is acceptable in practice, despite preaching and signing of affidavits' (Smith, 2006).

Undoubtedly this issue is even more complicated among international students or students for whom English is the second language. Certainly, many international students with their high intelligence and academic achievements get admitted to U.S., U.K. or Australian universities. However, they frequently have very little knowledge or training to avoid plagiarism. Statistics shows that a very large numbers of international students come to western countries every year to pursue a graduate degree. According to a National Science Foundation Study 2008, in 2005, 59% of all doctoral degrees and 43% of all higher-education degrees in engineering and science had been awarded to temporary residents in USA (as cited in Broache, 2008). Obviously there is an enormous need to provide instruction for students grappling with this issue of avoiding unintentional plagiarism.

However, there are numerous evidences of native speaker of English accused of plagiarizing. According to Duke University's Center for Academic Integrity 2005 study "40% of all U.S. college students admit to having woven unattributed material from the Internet into their written work" (as cited in Tomsho, 2006). In most cases, student misunderstanding regarding plagiarism is a major reason behind it. The penalties can be severe; they can vary from just a warning to expulsion from the university. Librarians are offering instruction to spread the awareness against plagiarism. There are many academic institutions all over the world which are using plagiarism detection tools to detect intentional Internet plagiarism.

Plagiarism and Detection

In light of these increasing plagiarism incidents, like many academic institutions across the world, at the University of Northern Colorado, we started using SafeAssignment[TM] since fall 2005 to assist faculty members in detecting plagiarism. SafeAssignment[TM] can be used as standalone software or can be integrated with Blackboard Learning

Environment (www.blackboard.com). The MyDropBox Suite of services is the choice of hundreds of schools and institutions around the world (MyDropBox.com: Get the Facts). Note that *SafeAssignment* is bought by Blackboard Inc and now called *Safe Assign* (Keuskamp & Sliuzas, 2007). It is now offered as part of Blackboard service.

SafeAssignment^TM sample report

The SafeAssignment^TM text matching feature compares submitted papers for any plagiarism sign and provides a sample report. The SafeAssignment^TM sample report has the following features:

a. provides a 0-100% matching text index, in most cases scores above 40% need to be reviewed for any sign of plagiarism,
b. presents the lists of suspected sources from where the particular paper may have been plagiarized,
c. different color coded sentences on the manuscript text section indicate different plagiarized sources, this clickable color coded sentences open a source comparison window to compare between the original line and copied line, and
d. instructors evaluating the paper have option to save, print or e-mail the document.

Methodology

In an attempt to examine the effectiveness of this particular text matching software, 50 plagiarized papers were submitted and analyzed to evaluate the effectiveness of SafeAssignment^TM plagiarism detection system. These 50 papers were gathered from the following sources (see Table 1).

Table 1. No. of papers submitted to SafeAssignment^TM from the following sources

No. of Papers Submitted	Sources
i. 10 completely plagiarized papers	From freely available PubMed interface and ProQuest databases
ii. 5 completely plagiarized re-submitted papers	Re-submitted papers
iii. 15 completely plagiarized papers	From 3 subscription databases
iv. 5 completely plagiarized papers	From open access journals
v. 5 partially plagiarized papers	From open sources
vi. 10 completely plagiarized papers	From 5 different search engines
Total Papers = 50	

Different types of file formats (doc, pdf, rtf, and html) were used to examine the compatibility factors of this text matching tool.

Results

Inconsistency in results

10 completely plagiarized papers from ProQuest Newspapers database and PubMed database were submitted to analyze the result patterns generated from this detection

tool. The matching scores were 100% for three papers, 20% for one paper, 0% for one paper from ProQuest Newspapers database (see Table 2). The percentages of matching scores varied widely depending on the copied source.

Similarly, the author received 100% matching report for three articles and 0% for two other articles from PubMed database, while all of them were completely plagiarized papers. The text matching results show inconsistencies in the SafeAssignment^{TM} plagiarism detection. Hence, the plagiarism detection was somewhat effective when copied texts were from ProQuest Newspapers database and from freely available PubMed interface.

SafeAssignment^{TM} plagiarism detection tool saves all submitted papers to one single institutional database. Supposedly, it can detect plagiarism from the papers that have been copied and submitted second time to SafeAssignment^{TM} plagiarism detection. Five papers were resubmitted to analyze the matching results. This tool was able to detect the plagiarized texts from four out of five resubmitted papers. Therefore, same discrepancy in results is observed in text matching scores with resubmitted papers.

Table 2 – SafeAssignment^{TM} matching scores from freely available PubMed, ProQuest databases and resubmitted papers

Databases	File Format	Matching %
1. ProQuest Article 1	rtf	100%
2. ProQuest Article 2	doc	100%
3. ProQuest Article 3	doc	100%
4. ProQuest Article 4	doc	0%
5. ProQuest Article 5	rtf	20%
6. PubMed Article 1	doc	0%
7. PubMed Article 2	doc	0%
8. PubMed Article 3	doc	100%
9. PubMed Article 4	doc	100%
10. PubMed Article 5	doc	100%
11. Resubmitted Paper 1	doc	0%
12. Resubmitted Paper 2	doc	100%
13. Resubmitted Paper 3	doc	100%
14. Resubmitted Paper 4	pdf	62%
15. Resubmitted Paper 5	pdf	100%

Effectiveness with commercial/subscribed databases

One of the major issues with SafeAssignment^{TM} is that it cannot detect plagiarism from any subscribed databases. To verify, 15 completely plagiarized papers were submitted from the subscribed databases, namely Academic Search Premier, Academic One File and JSTOR. In each case five articles were submitted to the SafeAssignment^{TM} plagiarism detection.

The matching results for five Academic Search Premier articles were below 5%. The matching scores for five Academic One File articles were 0%. The results for the four

JSTOR articles were below 10% and the matching result for one article was 27% (see Table 3). Overall, in all cases, matching scores were below 27% which means there was no sign of plagiarism from these papers. This establishes that the plagiarism detection is least effective with commercial or subscribed databases. As a result, an enormous range of library resources cannot be checked by SafeAssignment^{TM} plagiarism detection.

Table 3 - SafeAssignment^{TM} matching scores from subscribed/commercial databases

Databases	File Format	Matching %
1. Academic Search Premier (EBSCO) Article 1	pdf	0%
2. Academic Search Premier (EBSCO) Article 2	pdf	0%
3. Academic Search Premier (EBSCO) Article 3	pdf	2%
4. Academic Search Premier (EBSCO) Article 4	word	3%
5. Academic Search Premier (EBSCO) Article 5	pdf	4%
6. Academic One File (Gale/Thompson) Article 1	pdf	0%
7. Academic One File (Gale/Thompson) Article 2	pdf	0%
8. Academic One File (Gale/Thompson) Article 3	pdf	0%
9. Academic One File (Gale/Thompson) Article 4	word	0%
10. Academic One File (Gale/Thompson) Article 5	pdf	0%
11. JSTOR Article 1	pdf	0%
12. JSTOR Article 2	pdf	7%
13. JSTOR Article 3	pdf	4%
14. JSTOR Article 4	pdf	27%
15. JSTOR Article 5	pdf	1%

False Positive Result

Several times, this detection tool generated false matching scores from strings of words cited from somewhere else. The author has also received false positive results while doing self-checking since this tool cannot distinguish between citation information and text within quotation marks that has been properly cited. This typical stricture reconfirms that this tool is not perfect yet and the matching scores need to be verified by the evaluator.

Only Text matching

Detecting plagiarism only through text matching may work well with text based subject areas like humanities and social sciences (Talab, 2004). However, it requires more sophisticated technology to detect plagiarism from scientific subject areas. For example: MOSS (Measure of Software Similarity), a free software is used for determining similarities among different computer programs such as Java, C, C++, Paschal, Ada, Lisp, or Scheme programs (Plagiarism). Different types of detection tools can be useful for different subject areas.

Compatibility Issue

Despite the fact that SafeAssignment^{TM} accepts diverse set of applications or file formats like .zip, .doc, .txt, .pdf, .rtf and .html files, however, there are still certain

technical limitations with SafeAssignment^{TM}. It is not compatible with all types of file formats; for example one cannot upload Microsoft Office 2007 file or word 07 file (docx). Similarly, WCopyfind, one free plagiarism detection desktop software, was not able search on the web document earlier. Now it is able to search web document for any sign of plagiarism (Plagiarism Resource Site Windows Software Page). This common limitation ascertains that every automated detection tool has some technical restrictions with it.

Open Access Journals

Nowadays, thousands of open access journals are available freely on the Internet. Five totally plagiarized open access journal articles were submitted to this detection tool. This text matching software was able to find plagiarized texts from the three open access journal articles. However, this tool could not detect plagiarized texts from other two open access journals, *Biomedical Research* and *Library Philosophy and Practice* (see Table 4). Therefore, disparities in results were found even from the open access journal articles.

Table 4 - SafeAssignment^{TM} matching scores from open access journals

Open Access Journals	Matching %
Biomedical Research	0%
Journal of Social Sciences	100%
Library Philosophy & Practice	29%
Molecules	94%
The Industrial Geographer	69%

Translated Papers

Few detection tools such as, CopyCatch and SafeAssignment provide multilingual support and are compatible with quite a few European languages in detecting plagiarism (CopyCatch Front Catch Screen). However, plagiarism detection software's are capable of detecting plagiarism only from the same language. So far, it is not possible for any detection tool to detect plagiarism from translated papers, for example, papers originally written in Chinese or German, but translated and plagiarized later into English.

Product volatility

Another concern with plagiarism detection software is its volatile nature. A good number of detection software is in the market for few years; some of them no longer exist. McKeever informed about multiple automated detection service, out of which: Howoriginal, Integriguard, Plagiserve and Edutie, no longer exist (McKeever, 2006). Moreover, SafeAssignment^{TM} is now bought by Blackboard Inc., and called Safe Assign^{TM} (Keuskamp & Sliuzas, 2007). Even though it carries more or less same characteristics and functionality, product instability can have profound influence on the purchasing decision of the academic institutions.

Copyright infringements

Some of this text matching software's (for instance: TurnItIn) save the submitted student's paper to one single database, which can be an infringement of student

copyright. SafeAssignment^{TM} attempts to avoid the copyright infringement by having a different database per institution, instead of one single database such as the one maintained by TurnItIn.

Advantages of automated detection

Regardless of all potential negative aspects, one cannot deny the advantages of using plagiarism detection software. There are different types of plagiarism detection tools available to combat especially, intentional plagiarism.

Explosion of information

With the advent of Web 2.0 technology there have been tremendous changes on how information is being processed, organized, disseminated and used in academic and research world. There is exponential growth in scholarly publication, making it harder for faculty and instructor to detect the plagiarized sources or even making assumption about the suspected sources. Nowadays students are faced with thousands of virtual choices for their research and assignments. Use of plagiarism detection tools can equip faculty members to fight back against this form of academic dishonesty.

Discrepancy in search engines results

Many argue in favor of using search engines especially Google instead of plagiarism detection tools. Google is probably the largest and undoubtedly the most popular search engine on the Internet. However, it is possible to use Google as a plagiarism detection tool. As most of our students nowadays start their research from Google and many of them are satisfied with the result they get from the Google. Data reveals that "77% start their research through the Internet, not the library's resources (electronic or otherwise)" (Waldman, 2003). Presumably, it is possible to detect plagiarism in the same way as students find information from Google by simply copying the suspected line and pasting it into a Google search. It is little difficult to search for an entire paper in Google. Moreover, there are many search engines Yahoo, AltaVista, MSN, Lycos, AOL, etc. are a few top names among many. One important thing to remember is that Google and other search engines can give different search results.

Many contend that faculty can use search engines as one of the many measures to deter plagiarism and to foster academic integrity. However, using search engines for locating and cross checking original text for plagiarism sign is a time consuming process. Since our faculties are already overburdened with their assigned duties, a time consuming process can discourage them from using any of these search engines to detect plagiarism.

Effectiveness with patch work plagiarism

To better evaluate the effectiveness of this particular text matching software, five partially plagiarized articles were examined. The author submitted parts of her own 'yet not published' paper and copied some information from the Internet and submitted to this detection tool. In all cases matching score was above 56%. For two articles matching scores were above 80% and for one article matching score was above 94% (see Table 5). This text matching software was able to detect the plagiarized text from the Internet sources. Thus the result shows the efficacy of this software with patchwork plagiarism.

Table 5 - SafeAssignment^{TM} matching scores from partially plagiarized papers

Papers	Matching %
1. Article 1	67%
2. Article 2	82%
3. Article 3	82%
4. Article 4	94%
5. Article 5	56%

Excellent for Internet or web plagiarism

In order to evaluate the effectiveness of this particular text matching software with Internet resources, the author submitted 10 completely plagiarized papers gathered from the following five search engines (see Table 6).

Table 6 - SafeAssignment^{TM} matching scores from Internet resources

Search Engines	Matching %
1. Google 1	100%
2. Google 2	100%
3. Yahoo 1	100%
4. Yahoo 2	100%
5. MSN 1	100%
6. MSN 2	100%
7. AltaVista 1	100%
8. AltaVista	100%
9. AOL 1	100%
10. AOL 2	100%

The result shows that plagiarism detection was 100% effective in this case. Above Internet sources include online books, encyclopedias, *Wikipedia*, government pages, and papers from Internet paper mills. In all cases, SafeAssignment^{TM} plagiarism detection was able to find plagiarized sources. Undertaking this analysis proves usefulness for this type of text matching software in today's digital world.

Further, Internet paper mills, with their ever-increasing numbers, offer paper to students sometimes for a fee, many times for free. This detection software works really well to identify plagiarized text from Internet Paper mills. Several times this tool found similar text matching not from the particular paper mill site the author has used, but from another paper mill site, reconfirming the fact how widely papers from these term paper mills overlap. SafeAssignment^{TM} works extremely effectively to identify plagiarized paper from paper mill sites.

Affordability

There are few digital detection technology available free of cost. Some of these free text matching tools and web sites are AntiPlagiarist 1.8, scanmyessay.com, WCopyfind and MOSS: A System for Detecting Software Plagiarism. Obviously, it becomes more expensive for more sophisticated tool. Few fee based tools such as, SafeAssignment^{TM}

and others, are still affordable for many academic institutions.

Quick turnaround time, ease of use

In general, plagiarism detection tools are user friendly tool for both faculty and students. The turnaround time for SafeAssignment^{TM} plagiarism detection is really fast. Usually it takes 2-3 minutes to show the originality report on the Blackboard Learning environment. Faculty can view the matching report directly from Blackboards' grade book which is a timesaving process, since faculty can decide to change students' grade after evaluating the plagiarism matching report.

Not only detection, plagiarism prevention

A small number of digital tools (TurnItIn and SafeAssignment^{TM}) offer students an option to submit their assignment or paper as a draft assignment to avoid inadvertent plagiarism. If in doubt, students can check their paper through plagiarism detection process. In all likelihood, students will attempt to reduce their matching scores and improve their paraphrasing skills to avoid plagiarism. However, faculties need to give permission to students to submit their assignment as a draft assignment.

Conclusion

Frequently, the chief reason students plagiarize is because they do not understand what constitutes plagiarism. Thus, researcher Rebecca Moore has argued that 'teaching, not software' is the key to preventing plagiarism (Hansen, 2003). Contrarily, John Barrie, President of TurnItIn believes that 'digital plagiarisms is a digital problem and demand a digital solution' (Hansen, 2003). In an eloquent article, Lucy McKeever described 'plagiarism detection has existed for as long as plagiarism itself, only the automatic detection process has merged more recently' (McKeever, 2006). If, however, a student deliberately, intentionally, negligently, violates the academic honesty; the use of digital tool can be helpful in order to support scholarly creativity and academic integrity.

More than detection or catching the students, a plagiarism detection tool can be used as a beneficial educational tool and a preventive measure for both faculty and students. The threat that an instructor is using a plagiarism detection tool is more than enough to deter students from attempting to plagiarize. Since, for various reasons, there are few limitations to plagiarism detection software and this tool is not a perfect tool yet. It is more likely that there will be more sophisticated detection tools available in the future and more and more academic institutions will be using them if not to detect plagiarism at least to prevent plagiarism.

It is critical that all students of the university community stay informed about their rights and responsibilities when using scholarly or research works. The digital tool attempts to balance the conflicting interests between the academic honesty and students' academic misconduct/behavior to assure academic integrity.

Unquestionably, plagiarism detection tool has the potential to provide support to teaching, and can play a key role to bridge the gap between students and plagiarism, and a combined effort of both instruction and detection can ensure the academic honesty in today's digital world.

Acknowledgement

I am thankful to my supportive colleague Mark Anderson for reviewing the manuscript.

References

- Adenekan, S (2003, November). Students using net to cheat . Retrieved May 5, 2008, from http://news.bbc.co.uk/2/hi/uk_news/education/3265143.stm
- Broache, A (2008, March 12). Bill Gates to congress: let us hire more foreigners. Retrieved April 1, 2008, from http://www.news.com /8301-10784_3-9892046-7.html
- CopyCatch Front Catch Screen. Retrieved June, 18, 2008 from http://www.copycatchgold.com/CCGFrontCatchScreen.htm
- Hansen, B. (2003, September 19). Combating plagiarism. *CQ Researcher*, 13, 773-796. Retrieved March 10, 2008, from CQ Researcher Online, http://0-library.cqpress.com.source.unco.edu:80/cqresearcher/cqresrre2003091900
- Keuskamp, D., & Sliuzas, R (2007). Plagiarism prevention or detection? The contribution of text-matching software to education about academic integrity. *Journal of Academic Language and Learning*, 1(1), A91-A99. Retrieved March 3, 2008, from http://www.aall.org.au/journal/ojs/index.php/jall/article/viewFile /29/41
- Madray, A. (2007) Developing students' awareness of plagiarism: crisis and opportunities. *Library Philosophy and Practice*. Retrieved March 10, 2008, from http://www.webpages.uidaho.edu/~mbolin/madray.htm
- McKeever, L. (2006, April). Online plagiarism detection services—saviour or scourge? *Assessment and Evaluation in Higher Education*, 31(2), 155-165.
- 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), 170-173.
- MyDropBox.com: Get the Facts. Accessed April 2, 2008, from http://www.mydropbox.com/services/facts.php
- Plagiarism. Accessed April 2, 2008, from http://www.123helpme.com /plagiarism.jsp
- Plagiarism (2008). *Britannica Online Encyclopedia*. Accessed June 4, 2008 from http://www.britannica.com/eb/article?eu=61807
- Plagiarism Resource Site Windows Software Page. Retrieved May, 5, 2008, from http://www.plagiarism.phys.virginia.edu/Wsoftware.html
- Plagiarism Tutorial. Accessed June 4, 2008, from http://www.lib.ncsu.edu /scc/tutorial/plagiarism/versus.html
- Smith, A. (2006, March). Plagiarism 'rife' at Oxford. Retrieved March 2, 2008 from http://education.guardian.co.uk/higher/news/story/0,,1731423,00.html
- Talab, R. (2004). A student online plagiarism guide: Detection and prevention resources (and copyright implications!). *TechTrends: Linking Research and Practice to Improve Learning*, 48(6), 15-18.
- Tomsho, R (2006, August 15). Familiar words: student plagiarism stirs controversy at Ohio University; some foreign engineers' grasp of research fundamentals becomes a concern to peers; 'not really intending to copy'. *Wall Street Journal* (Eastern Edition), p. A.1. Retrieved March 2, 2008, from Wall Street Journal database. (Document ID: 1094699861).
- TurnItIn. Retrieved June, 20, 2008 from http://turnitin.com/static/home.html
- Waldman, M. (2003). Freshmen's use of library electronic resources and self-efficacy. *Information Research*, 8 (2), paper 150. Retrieved March 17, 2008, from http://informationr.net/ir/8-2/paper150.html

Bibliographic information of this paper for citing:

Chaudhuri, Jayati (2008).   "Deterring digital plagiarism, how effective is the digital detection process?"   *Webology*, **5**(1), Article 50. Available at: http://www.webology.ir /2008/v5n1/a50.html

Alert us when: New articles cite this article