Presented at the 10th DLSU Arts Congress
De La Salle University, Manila, Philippines
February 16, 2017

Turning it in: Experiences, Challenges and Recommendations for the Appropriate Use of Plagiarism Detection Software

Luis F. Razon[1]*, Raymond R. Tan[1], Michael Angelo B. Promentilla[1],
Kathleen B. Aviso[1] and Krista Danielle S. Yu[2]
[1] *Chemical Engineering Department, De La Salle University*
[2] *Economics Department, De La Salle University*
**Corresponding Author: luis.razon@dlsu.edu.ph*

Abstract: The global availability of vast amounts of information via the World Wide Web has led to significant progress in academic research. However, the copying of original material is now much easier and the tracking of the original source material of plagiarized submissions is more difficult. Many prominent personalities have been involved in sensational cases of plagiarism as either perpetrators or unwitting victims. The challenge of detecting plagiarism in academic work has led to the development of plagiarism detection software, one of which is Turnitin.com. Services like Turnitin automatically detect word clusters in a submitted document that are identical to material in its repositories or on the Web. Turnitin provides a "similarity report", which consists of a document with portions of the submitted text highlighted to indicate identical matches with documents in its repositories and a "similarity score" -- the percentage of the submitted document that matches other documents. The availability and ease-of-use of plagiarism detection software has made it very popular in academic institutions. Similar software is used by major academic journals as well. There are inappropriate uses of these tools, however, such as setting a hard numerical target for the similarity score and assuming that a paper with a low similarity score is not plagiarized. In this paper, we share our experiences with the use of plagiarism detection software and make recommendations for its appropriate use.

Key Words: Plagiarism, Turnitin, ethics, plagiarism detection software

1. INTRODUCTION

The Internet and the World Wide Web has made vast amounts of information to on a scale that would have been difficult to imagine even 10 years ago. The material available is not only limited to text sources but also images, music and videos. The digital form greatly facilitates the copying of materials and recasting in similar but not identical forms. The vast amounts of available materials also seem to lead students and authors to believe that it is humanly impossible to identify the original sources of plagiarized materials. Current pedagogical methods, such as transformative learning, also encourage instructors to assess student learning not via the traditional proctored classroom examinations but via submitted papers composed outside the classroom without close supervision. These factors have led to an explosion in incidences of "copy-and-paste" plagiarism (Heckler, Rice, & Hobson Bryan, 2013), not only among students but also in high-impact scientific journals. See (Andrade, Pérez, Sebastian, & Eapen, 2011a, 2011b), for example. The threat of increasing numbers of plagiarism has led to the development of software and Web-based tools that detect whether there are word clusters that are identical to other documents in the "depositories" of the software vendor (Bensal, Miraflores, & Tan, 2013). These depositories include available digital materials and previous submissions of student work to the vendor. Among the most prominent of these "originality checking services" or "plagiarism detection software" vendors is Turnitin which is the platform that has been selected by De La Salle University (Bensal et al., 2013).

Turnitin is organized according to classes and assignments that are set up by the instructor. Students are provided with a unique class number and a password. Students are expected to "enroll" in a class and upload their submissions into the appropriate assignment database. In setting up the assignment, the instructor has the option to set up parameters that will determine how submissions will be graded. The instructor has the option to: (1) set the number of words that are compared for matches; (2) exclude bibliographic materials from similarity checking; (3) exclude materials within quotation marks and (4) choose which document repositories will be included. Turnitin generates an "originality report" that includes an Adobe Acrobat file with highlights on the submitted text that was found to match text in Turnitin's depositories. For example, see Figure 1. Note that, Turnitin also color-codes the highlighting such that each individual source of matching text is assigned a color. Clicking on a highlighted text moves the cursor to the source. Another click on the links on the right side shows the context in which the matching text was used in the original source.

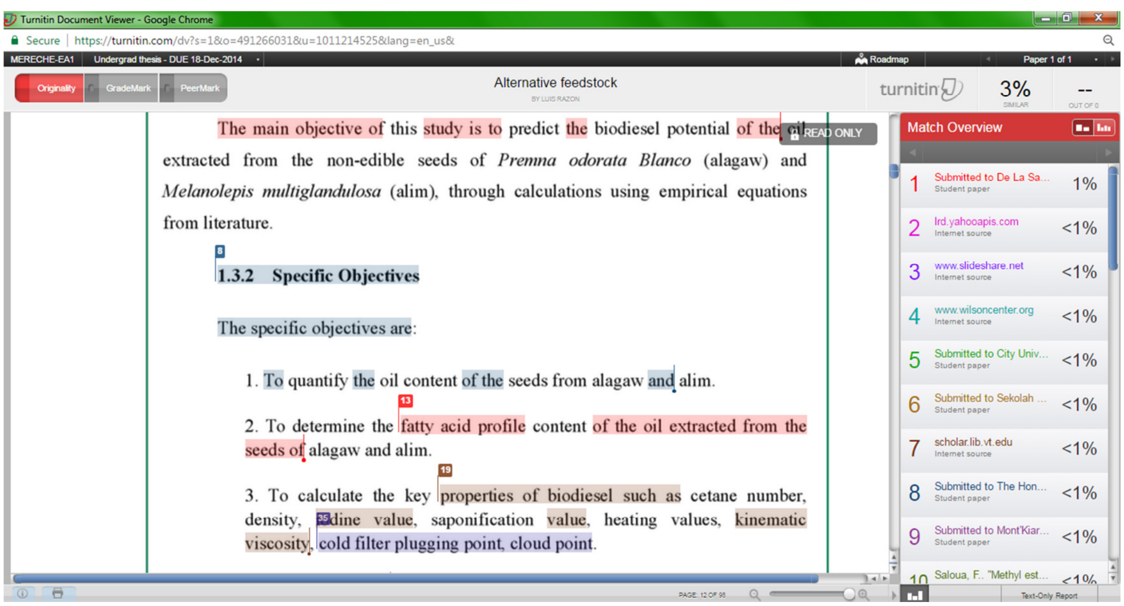

Fig. 1. Example of an originality report.

In addition to the originality report, Turnitin generates a "similarity score" or the percentage of the text in the submission that was found to match those in other documents. The similarity scores are classified into five categories: blue (0% matching), green (greater than zero and less than 25% matching), yellow (25% to 49% matching), orange (50% to 75% matching) and red (greater than 75%) (IParadigms LLC, 2011).

Services like Turnitin have been very popular and the response has range from enthusiastic to more cautionary. Some authors have found that student foreknowledge of the use of plagiarism detection software was a good deterrent to copy-and-paste plagiarism (Heckler et al., 2013; Martin, 2005). An informal survey by one of us (LFR) among chemical engineering undergraduate students also showed that most of the respondents believe that using Turnitin is a good deterrent. However, there have been many ways found to bypass plagiarism detection software. Patel et al. (2011) list 10 methods, some of them truly amazing for their creativity, for defeating plagiarism detection software. It is also well known that Turnitin cannot detect the copying of figures and equations. This is actually a more serious concern because the unauthorized use of figures is a violation of copyright law, whereas plagiarism is more of an ethical issue.

Turnitin takes great pains to specify in its website that Turnitin cannot by itself decide whether the submission was plagiarized. This warning appears in one of its documents (IParadigms LLC, 2011):

> "Warning: These indices do not reflect Turnitin's assessment of whether a paper has or has not been plagiarized. Originality Reports are simply a tool to help an instructor find sources that contain text similar to submitted papers. The decision to deem any work plagiarized must be made carefully, and only after in depth examination of both the submitted paper and suspect sources in accordance with the standards of the class and institution where the paper was submitted."

Unfortunately, it is very tempting to use the similarity score as an indicator of the presence or absence of plagiarism because it is so easy to do. Student anecdotes have indicated that several faculty members within De La Salle University set a hard numerical target, or a number below which a work is judged to be not plagiarized and above which is judged to be plagiarized, prior to accepting student submissions. Student submissions with a score above a number set by the instructor (say 15%, for example) are rejected outright or even threatened with failure in the course. Furthermore, even though similarity scores are low, students may have plagiarized a certain part of their work and claim that they did not, given that it is well below the number set by the instructor.

Academic journals have also started using plagiarism detection software (Butler, 2010). While this should be not be a problem *per se*, there are indications that hard numerical targets are being set by these journals (Editage Team, 2013). This has also resulted in some consternation among some authors, as we will see in the cases we discuss in this paper.

2. METHODOLOGY

In this paper, we describe individual cases where the results of Turnitin may lead to false judgements of plagiarism or its absence. These cases are taken from our own experiences and those of students. We classify the cases as either Type I errors or Type II errors. Type I errors ("false positives") are cases where the similarity was flagged by Turnitin whereas there was not any actual plagiarism that occurred or was intended. On the other hand, a Type II error ("false negative") is an instance wherein plagiarism did occur but was not flagged by Turnitin.

By analyzing these examples, we illustrate the cases wherein false judgments about the presence or absence of plagiarism and we make recommendations for faculty and journal policies that reflect the appropriate use of plagiarism detection software.

3. RESULTS AND DISCUSSION

1.1 False positives

The first example involves a paper by one of us (LFR) that was rejected by a high-impact journal for "too much similarity". Figure 2 shows examples of text that were highlighted as being similar to previous publications. The highlights are from Turnitin while the journal used CrossCheck. Because the originality report was not fed back to the authors, we are forced to speculate that the highlighted text shown in Figure 2 are the same as those seen by the journal editors. Figure 2(a) shows a standard set of material used by the principal author in all of his papers. Figure 2(b) shows a series of citations that has a series of "et al." and years that happened to coincide with each other. Note that the similarity checkers also seem to tag similar phrases even when there are words in between. This is probably done to counter attempts to mask actual plagiarism. Unfortunately, this also leads to the comical situation shown in Figure 2(b). Finally, Figure 2(c) shows a series of stock phrases used to describe the methods for the chemical analysis of biodiesel. The original sources are by the authors themselves, so it can be argued that this is a case of self-plagiarism. It is difficult, however, to see how these instances can be construed to be. What is to be gained by rephrasing standard methods of analysis so that these are more original? Is it even possible to rephrase addresses and disclaimers?

a) National Center for Agricultural Utilization Research, Agricultural Research Service, U.S. Department of Agriculture; Peoria, IL 61604, U.S.A.

b) Department of Chemical Engineering, De La Salle University; 2401 Taft Avenue, 1004 Manila, The Philippines

c) Department of Biology, De La Salle University; 2401 Taft Avenue, 0922 Manila, The Philippines

Corresponding author: Dr. Gerhard Knothe; USDA/ARS/NCAUR; 1815 N. University St.; Peoria, IL 61604; U.S.A.; Phone: +1-309-681-6112; Fax: +1-309-681-6524; E-mail: gerhard.knothe@ars.usda.gov

Disclaimer: Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture. USDA is an equal opportunity provider and employer.

(a)

[Moser and Vaughn, 2010b], cuphea [Knothe et al., 2009], Indian milkweed [Phoo et al., 2014], jatropha [Foidl et al., 1996; Shah et al., 2004], moringa [Rashid et al., 2008], mustard species [Jham et al., 2009], okra [Anwar et al., 2010], pennycress [Moser et al., 2009], pongamia (karanja) [Scott et al., 2008], rubber seed [Ramadhas et al., 2005], thespesia [Rashid et al., 2011]

(b)

an equation given in the literature [Knothe, 2014]. Oxidative stability measurements were carried out with a Rancimat (Metrohm; Herisau, Switzerland) equipped with software for statistical evaluation using the European standard method EN14112. Lubricity was investigated utilizing a high frequency reciprocating rig (HFRR) lubricity tester following the ASTM D6079 method as used in previous work [Knothe and Steidley, 2005b]. Cloud and pour point determinations were conducted with a phase technology (Richmond, BC, Canada) cloud, pour, and freeze point analyzer. Density was determined with an Anton Paar (Ashland, VA, USA) DMA 4500M density meter. Water content was determined by Karl Fischer titration with a Metrohm (Herisau, Switzerland) 831 KF coulometer. Acid values were determined with the AOCS (American Oil Chemists' Society) method Cd3d-63. Analysis of Na, K, P, S, Ca, and Mg were performed with a Perkin-Elmer (Waltham, MA, USA) Optima 7000 DV inductively coupled plasma – optical emission spectrometer (ICP-OES).

(c)

Fig. 2. Text highlighted as similar in a submission to a high-impact journal

The second example is from the experience of a PhD student in chemical engineering. PhD students in chemical engineering are required to present two papers in international conferences and have at least one paper accepted in a Scopus-indexed journal prior to submission of the final dissertation and defense. After these requirements were fulfilled, the student uploaded his dissertation into Turnitin and received very high similarity scores. This is because his dissertation was found to be similar to his two conference proceedings and his journal article. Should this be again considered a case of self-plagiarism? Surely not, unless there is an expectation that the dissertation is completely different from conference proceedings and journal article submissions.

1.2 False negatives

Our first example of a false negative concerns a student submission to a journal. Prior examination of the draft via Turnitin showed only a 10% similarity score. However, further examination of the paper after the first round of reviews showed that there was an entire paragraph taken from an article by another author. This paragraph accounted for 6% of the total 10% similarity. This illustrates how even a low similarity score can lead one to conclude the absence of plagiarism whereas there actually is some plagiarism.

The second example is from a student submission for coursework. The submission obtained a low similarity score but some of the writing seemed to be too good to be true and the instructor spent a considerable amount of time locating the old textbook from which the text had been copied. Since the source was old, its contents were not available online.

4. CONCLUSIONS

The examples show the limitations of Turnitin and the dangers of some of the practices in its implementation.

Turnitin and plagiarism detection software are only as good as the algorithms that they implement. The current generation of plagiarism software detect only the copying of word clusters whereas the more serious problem of stealing ideas without appropriate attribution is actually a more serious problem, Science articles, since they employ a considerable amount of stock phrases or what others would call jargon or clichés are particularly vulnerable to receiving high similarity scores. Sun (2013)compared the similarity scores from Turnitin from a large number of randomly selected papers from science, technology, engineering and mathematics (STEM) and social sciences. She found that the STEM papers had higher similarity scores than those from the social sciences. The implementation of more stringent standards where strings of text that have intervening words or numbers may only drive up similarity scores even higher even when there was no actual plagiarism.

It is dangerous to use of a hard numerical target to decide whether a work is plagiarized or not. As was seen in the first example of a false positive and in the first example of a false negative, the use of a hard numerical target is very dangerous because a work with no actual plagiarism may be rejected while a work with actual plagiarism may be accepted. Furthermore, emphasizing numerical targets may mislead students into thinking that this is the only form of plagiarism. Hard numerical targets should be discouraged within the University. It is also prudent for authors to be aware that some journal editors are (foolishly) using hard numerical targets. In this case, there may be no choice for the meantime other than to run paper drafts through a plagiarism detection service and reword similar text when flagged, inconvenient though it may be. Journals should be more transparent and provide copies of originality reports, especially when a paper is rejected for too much similarity.

Turnitin and other plagiarism detection software are also as good as the contents of their repositories. In the second example of a false negative, a low similarity score was obtained when in fact there was egregious word-for-word copying that had occurred. The second example of a false positive is a little trickier because the student did reuse his own work for the dissertation but this is a commonly accepted practice. In fact, the current trend indicates that most dissertations in STEM are now compilations of papers rather than single unified documents (Gould, 2016). Does this constitute self-plagiarism? These discussions beg the question: "What is plagiarism?" There have been numerous attempts to define plagiarism. The authors prefer to paraphrase the definition of US Supreme Court Justice, Potter Stewart, of pornography (Jacobellis v. Ohio 1964) and declare, "We know plagiarism when we see it".

Plagiarism detection software, like Turnitin, have proven to be invaluable tools in the prevention and detection of plagiarism, given the large volume of student submissions and the large volume of potential source material. These tools, and it is important to remember that they are mere tools, should be used judiciously and should not be construed to be replacements for human judgement and common sense.

5. REFERENCES

Andrade, J. E., Pérez, A., Sebastian, P. J., & Eapen, D. (2011a). A review of bio-diesel production processes. *Biomass and Bioenergy*, *35*(3), 1008–1020. http://doi.org/10.1016/j.biombioe.2010.12.037

Andrade, J. E., Pérez, A., Sebastian, P. J., & Eapen, D. (2011b). Retraction notice to "A Review of Biodiesel Production Processes" [Biomass Bioenergy 35 (2011) 1008-1020]. *Biomass and Bioenergy*, *35*(7), 3284. http://doi.org/10.1016/j.biombioe.2011.05.019

Bensal, E. R., Miraflores, E. S., & Tan, N. C. (2013). Plagiarism: Shall we turn to Turnitin? *Call-Ej*, *14*(2), 2–22. Retrieved from http://www.scopus.com/inward/record.url?eid=2-s2.0-84886920048&partnerID=40&md5=132c3d5f48578d472a8f549335c7d246

Butler, D. (2010). Journals step up plagiarism policing. *Nature*, *466*(7303), 167–167. http://doi.org/10.1038/466167a

Editage Team. (2013). How much do journal editors rely on plagiarism detection software? | Editage Insights. Retrieved January 15, 2017, from http://www.editage.com/insights/how-much-do-journal-editors-rely-on-plagiarism-detection-software

Gould, J. (2016). What's the point of the PhD thesis? *Nature*, *535*(7610), 26–28. http://doi.org/10.1038/535026a

Heckler, N. C., Rice, M., & Hobson Bryan, C. (2013). Turnitin Systems. *Journal of Research on Technology in Education*, *45*(3), 229–248. http://doi.org/10.1080/15391523.2013.10782604

IParadigms LLC. (2011). Turnitin Instructor User Manual Chapter 2 : OriginalityCheck (pp. 48–60). IParadigms, LLC.

Jacobellis v. Ohio, 378 U.S. 184 (1964) Retrieved January 21, 2017, from https://www.law.cornell.edu/supremecourt/text/378/184

Martin, D. F. (2005). Plagiarism and Technology: A Tool for Coping With Plagiarism. *Journal of Education for Business*, *80*(3), 149–152. http://doi.org/10.3200/JOEB.80.3.149-152

Patel, A., Bakhtiyari, K., & Taghavi, M. (2011). Evaluation of cheating detection methods in academic writings. *Library Hi Tech*, *29*(4), 623–640. http://doi.org/10.1108/07378831111189732

Sun, Y. C. (2013). Do journal authors plagiarize? Using plagiarism detection software to uncover matching text across disciplines. *Journal of English for Academic Purposes*, *12*(4), 264–272. http://doi.org/10.1016/j.jeap.2013.07.002