PRACTICAL ISSUES FOR ACADEMICS USING THE TURNITIN PLAGIARISM DETECTION SOFTWARE

Dr. Karl O. Jones

Abstract: The Turnitin plagiarism detection system allows individual student assignments to be uploaded and matched for similarity with content on the web, all other assignments uploaded by institutions using the system and certain journals. An online report is produced for each submission identifying the sources of those similarities and the percentage match. There is a significant benefit in using Turnitin to identify possible cases of plagiarism. This paper highlights some issues that might be raised in employing Turnitin and suggests some approaches that academics might utilise to allow efficient use of Turnitin.

Keywords: Plagiarism, detection software, Turnitin.

INTRODUCTION

All forms of academic impropriety are of growing importance across Universities internationally, and there is growing evidence that University students do ‘cheat’, that is, engage in academic practices considered unacceptable under University regulations [1]. There is also considerable empirical evidence that such cheating is increasing [2]. Park [3] reports data from 6000 US students, indicating between 63% and 87% (depending on academic discipline) admitted having cheated during their college career. Over recent years, academics have had increased concern over the amount of academic impropriety undertaken by students, and in particular the number of plagiarism cases. This concern has coincided with the explosion in the use of the Internet as a tool for academic study and research. McCabe [4] noted “there is evidence that cheating has increased in the last few decades, and the Internet is likely to intensify the problem”. In practice, plagiarism is far easier with the Internet as a medium, rather than through printed material, after all operations such as ‘copy and paste’ are far easier from an electronic source than from a book.

Universities should be committed to helping their students acquire the skills necessary to support their progression and to achieve their academic potential. Students start university with a wide range of academic experiences and thus may not always be familiar with the formal protocols of academic writing and referencing. Student’s failure to follow these protocols may be considered as plagiarism. Most departments offer advice and information on appropriate referencing, recommend that if students are in doubt about what constitutes plagiarism, cheating or collusion, that they should discuss it with their tutors, further Universities employ a range of methods to support academic writing, as well as to identify and discourage plagiarism: including detection services such as Turnitin.

THE TURNITIN SERVICE

The Turnitin system [5] is probably the global leader in electronic plagiarism detection. It is recognised as a tried and trusted system in use around the world. In the UK over 80% of universities employ it, as well as a number of Further Education colleges and High schools. Developed by iParadigms LLC [6], Turnitin offers one approach to the growing problem of 'cut and paste' plagiarism.

Documents submitted to Turnitin, such as student assignments, are compared against billions of internet documents, an archived copy of the internet, a local database of submitted student papers, and a database of some periodicals, journals, and on-line publications, such as Emerald Publishing (Figure 1).
For each document submitted to Turnitin, an originality report is generated to estimate the percentage of matches between that document and the previous list of documents (Figure 2). Ignoring the commonly used words, the software looks for matches of strings of eight to ten words [7]. While this list seems exhaustive, work by Kaner and Fiedler [8] suggest that there are some major omissions such as the ACM [10] and IEEEXplore [9] databases of conference and journal papers, which is of particular concern for engineering academics.

Legal Issues related to using Turnitin

There have been some concerns voiced about the validity of using systems such as Turnitin with respect to student personal information and intellectual property rights. For example, depending on how Turnitin is used, student work will be submitted and may include ‘personal details’ such as name, email address and course details. In the UK there are limitations on when and how personal data can be ‘processed’, hence for any institution to validly process personal data, it must comply with certain conditions as defined under the Data Protection Act 1998 (DPA) [11], including “The data subject has given consent to the processing” and “The processing is necessary for the purposes of legitimate interests …… except where the processing is unwarranted in any particular case by reason of prejudice to the rights and freedoms or legitimate interests of the data subject”. Hence, UK Universities are registered as a data controller under the terms of the DPA, and when students sign their enrolment form, they give consent to the storage and processing of personal data by the University for legitimate purposes, such as student information systems, virtual learning environments, email and applications such as Turnitin. However, assignments can be submitted to Turnitin by an academic without this personal data included. In terms of Intellectual Property Rights, iParadigms has no interest in acquiring the intellectual property rights for any submitted document. The copyright will continue to reside with either the student or the University depending on individual University regulations.
PRACTICAL EXAMPLE OF USING TURNITIN

At class of 115 students were asked to submit work to Turnitin through the direct link on the University’s Virtual Learning Environment (Blackboard). In all 113 followed the instructions, with the remaining 2 students submitting the work on paper. This approach required the academic to print each assignment for marking and to manually request the class Originality Reports. Since this was very inefficient, the next class of 68 students were required to submit their work on paper and upload an electronic version to Turnitin (all students followed the instruction). This dual submission was far more efficient for marking, and was also improved with an automatic request for Originality Reports since a deadline for coursework submission can be set on Turnitin with all work submitted by the date then being processed for matching. Experience has shown that a Turnitin date of a couple of days after the requested deadline ensures that any late submissions are accounted for without any input from the academic.

In both cases the standard Originality Reports indicated matched text levels from 0% through to 100%. Figure 3 provides a histogram of the levels of matched text for the 113 assignments. Closer analysis of the percentage matched text revealed (in this case) that the students in the range 90-100% were primarily copying work from each other (possible collusion) rather than plagiarism from other sources. This is still academic impropriety, and would have been rather difficult to identify through just reading 113 essays.

Within Turnitin a filtering system can exclude low percentage reports allowing the academic to concentrate on checking documents with a higher similarity score. The percentage reported cannot be used as a measure of plagiarism, since Turnitin only indicates the level of matched text. Two obvious issues with this purely matching ability are reference lists used in the work and correctly quoted material, although the software does provide links to exclude these items. Figure 4 illustrates an example of a document that had a reported level of matched text of 9%, which after selecting the ‘exclude bibliography’ reduces to 2% and eventually becomes a level of 1% when quoted material is excluded. A significant drawback for efficient use is that these actions have to be performed by the user and cannot be automated.
It would appear that there are some problems with the text matching algorithm. The results obtained thus far from the use of Turnitin have revealed occasions where text highlighted can be considered as innocent. The first instance relates to a number of students within a class including the question within their answer (Figure 5). This is clearly not plagiarism and is also not something that the software can identify at the moment – the ability to include the question as part of the comparison documents but for exclusion purposes would alleviate this problem. The next problem is the identification of phrases that might be called ‘common language’ with respect to the topic of the assignment, such as those shown in Table 1 that have come from recent student work that Turnitin has processed. For example, an essay on Evolutionary Computation is highly likely to have statement 1, although most academics would not expect such a phrase to be referenced.

**Table 1 Illustrations of ‘Common Language Matching’**

| 1) “based on Darwin’s Theory of Evolution” |
| 2) “require some form of mathematical model of the process” |
| 3) “Technology has the potential to cause harm in the wrong hands” |

Another area of concern is the matching of parts of phrases. Any section of text matched by Turnitin is displayed in a box with the similar words represented in bold (lower part of Figure 5), while words within the phrase that are not an identical match are left as normal font (upper part of Figure 5). Table 2 provides a selection of such phrases found in student assignments. The matched text increases the percentage level reported, however the relationship between the amount of bold text versus the non-bold text is cause for concern since the small groups of words highlighted can hardly be considered plagiarism. It is suggested that this is more a case of ‘noise’ interfering with the matching process.

**Table 2 Illustrations of ‘Noise matching’**

| 1) “The potential benefits from AI technology to society are unlimited as the utilization of” |
| 2) “find the shorted path between their nest and a food source” |
| 3) “accuracy of the model parameters, the DE algorithm determines” |

**CLOSING COMMENTS**

Sunderland-Smith and Carr [12] found that academics expected that “the software would highlight passages of text and identify those passages as plagiarised”. It must be made clear that Turnitin should not really be considered a plagiarism detection system, it is merely a text matching system. Furthermore, Sunderland-Smith and Carr state that academics “expected they would need very little input in identification of plagiarism”. It is still the role of the academic to review the reports to determine if plagiarism has actually
occurred, because Turnitin does not differentiate between correctly cited references and unacknowledged copying. Also one must be aware that there is a difference between incorrect referencing (for example through student inexperience) and intentional plagiarism, this judgement can only really be made by a human and not by software. Across the University academics have reported broad support for the use of Turnitin although they have expressed reservations about its capacity and its complexity of use.

In the early stages of use, it is recommended that academics concentrate their time and efforts on the extremely high levels of matched material. By doing this one can be certain of identifying the major instances of similar text. It is important to remember that just because Turnitin identifies student work as having a certain percentage of matching text, it is not by itself evidence of plagiarism. It is up to the academic to carefully examine the material and view sources of matched text, and make a judgment on the applicability of the reported matching. Academics must also be aware that if students submit both paper and electronic copies of their work, then a sample of the electronic submissions should be compared against the paper versions to ensure that students are not attempting to present innocuous documents for use by Turnitin.

Overall Turnitin is a very valuable tool in the fight against academic impropriety; however it is not quite the automatic tool it first seems. To correctly identify instances of plagiarism, academics must spend a little time to ascertain whether matched text falls into any of the categories mentioned above. Only after this manual intervention can the true level of plagiarism be determined.

REFERENCES


ABOUT THE AUTHOR

Dr. Karl O. Jones is a Principal Lecturer in the School of Engineering at Liverpool John Moores University. He has a particular interest in the levels of student engagement in academic impropriety.

Phone: +44 151 231 2199, E-mail: k.o.jones@ljmu.ac.uk.