

DOI: [10.20472/TE.2019.7.1.001](https://doi.org/10.20472/TE.2019.7.1.001)

# **PAPER REVIEW: PROTECTING STUDENTS' INTELLECTUAL PROPERTY IN THE WEB PLAGIARISM DETECTION PROCESS**

*HASHEM ALMUSAWI*

## **Abstract:**

Students' from all over the world face an undisclosed infringement to their intellectual property. Mainly their written submissions in the form of essays, assignments and homework are archived in plagiarism detection services (PDS) without their consent. It is fair to say that these documents are never used for anything other than to report on similarity and cross matching newly submitted documents for plagiarism detection, or so it seems. However, it still is a form of infringement, especially when PDS providers are making profit out of the whole process. This paper reviews the potential application and implications of implementing the proposed architecture in the reviewed paper. The proposed architecture highlights a great resolution to the problem. This paper addresses an overview of the reviewed paper, as well as key notes to consider when adopting the proposed architecture.

## **Keywords:**

Plagiarism detection services, PDS, copyright infringement, college students.

**JEL Classification:** I20, I21, I29

## **Authors:**

HASHEM ALMUSAWI, Arabian Gulf University, Kingdom of Bahrain, Bahrain, Email: macaholiq8@gmail.com

## **Citation:**

HASHEM ALMUSAWI (2019). Paper review: Protecting students' intellectual property in the web plagiarism detection process. International Journal of Teaching and Education, Vol. VII(1), pp. 1-6., 10.20472/TE.2019.7.1.001

## Paper Summary

The paper discusses the potential violation and exploitation of students' copyrighted materials such as assignments, essays, research, dissertations, or any other form of written material submitted for review through third party Plagiarism Detection Services (PDS). It shows that universities usually tend to outsource plagiarism reports of students' work for review using commercial PDS such as Turnitin or SafeAssign, uploading them directly to the PDS servers, leaving a full copy of the material for future references and comparisons. The process seems to exploit copyrighted materials without prior consent of the copyright owners, whom in this case are the students. Universities must be aware whether PDS archive students' submissions for future comparisons, and whether those reports are discarded after sending them back to the university.

In that context, the authors propose a new technical architecture in order to reform the process of plagiarism detection and move it from third party PDS back to the university boundaries. This proposed architecture guarantees that any form of copyrighted material doesn't fall in the hands of the third party, rather it compiles a set of suspicious phrases and paragraphs that could potentially be plagiarized and send them to PDS for processing and reporting back on similarity. The main goal of the research is to answer two questions in regards to the proposed architecture which are; (1) How does the amount of information used in the search affect the search quality? (2) What is the minimal amount of information required to guarantee detection of different levels of plagiarism? The results of the experiment showed reliable detection of multi-level plagiarism, giving unorganized portions of the submitted document (between 5-50% of the actual document), thus making it impossible for third party PDS to recompile the original document.

The paper is well framed and the topic is covered thoroughly with deep technical aspects and a well oriented plan to tackle the research problem. It discusses related works and existing solutions that were already available at the time of the study. However, none of the existing solutions seem to outline a practical solution that insures third party PDS don't exploit students' intellectual property, hence the authors proposed the new architecture. One interesting fact regarding the matter in hand is that there were several lawsuits accusing PDS, Turnitin to be specific, of copyright violation. Even though Turnitin had not been found guilty, it generated damaging publicity for both the service and universities involved. The authors also note that if these lawsuits were in Europe instead of the US the results would probably be different since Europe intellectual property protection is less tolerant to violation and infringement than the US.

It is worth noting that the process of protecting students' intellectual property isn't solely based on the students alone. It must be part of universities duty to make sure their students' works are protected from any potential exploitation and misuse when implementing systems such as automated PDS (Brinkman, 2013). The proposed architecture insures that universities have control over their students' submissions and doesn't allow third parties to recover the full document.

## Paper Methodology

The methodology of the research isn't concisely clear as it is blended throughout the explanation of the proposed architecture. However, the authors clearly state the steps and measures taken into consideration when implementing the proposed architecture. The research includes both qualitative and quantitative data. Qualitative data represented in the form of existing processes

of PDS outsourcing, types of documents used in the implementation of the proposed architecture, and an algorithm equation that handles queries for the purpose of processing suspicious documents. Quantitative data is also presented in the form of a set of documents (500 documents, each have 3,000 words). These documents were recompiled from Wikipedia.com to represent different fields which are; countries, objects from specialized domains, social concepts and actions, and general objects.

The algorithm in the proposed architecture randomly allocates and selects six-word sentences with a minimum of 150 words and a maximum of 900 words for each document. This process assures that the document is not entirely shared with commercial PDS. This proposed architecture algorithm is effective for intrinsic detection within the university database as well. Vani & Gupta (2016) claim that it's better for extrinsic plagiarism detection algorithms to classify queries to one of the following; string based, vector space model, semantic based, syntax based, structural based, or citation based. That classification is not presented in the reviewed paper, meaning all documents in the experiment were handled as a single class for extrinsic detection.

One can simply check for plagiarism by searching for suspicious text through convenient search engines. It is however a hectic and time-consuming process. It also sends out the document without encryption making it susceptible for exploitation. Therefore, the reason for using PDS, whether internally or externally, is to assist manual detection (Dhir, Arora, & Arora, 2008). The proposed architecture surpasses manual detection and moves documents directly to automatic detection.

## Paper Results

The main purpose of the paper is to propose a new architecture for PDS to preserve students' intellectual property rights, specifically copyright, and eliminate third party PDS chances of making profit from students' works and prevent recovering their full documents. As far as the initial experiment; the proposed architecture succeeded in detecting plagiarized documents while preserving students' copyrights. Results show that even with 5% plagiarism, the architecture had 17% level of confidence for detection when 50% of document queries were used. That percentage rises significantly when using documents with 50% plagiarism where the level of detection confidence reaches 100% for as little as 5% of search queries. In order for the proposed architecture to detect insignificant amounts of plagiarism it needs to send out as much as 50% queries for significant detection. On the contrary, the architecture only needs 5% of queries for to detect significant amounts of plagiarism. In other words, the higher the plagiarism, the less queries needed to detect it, and vice versa.

Some inconsistency with the results was caused by connection timeouts and inaccessibility of certain sources to compare suspicious documents. To overcome this obstacle there needs to be better connection to the internet and an overall robust IT infrastructure with consistent computational power to sustain the search process. The authors proposed a solution for such cases in the form of manually preselecting particular segments from the suspicious document to be sent out for checkup but that conflicts with the general idea of using an automated PDS architecture such as the one being proposed.

## Paper Contribution to the Field

Plagiarism is a heated topic among academia, with the growing availability of digital content over the internet, it rises so many questions regarding academic integrity and there are several efforts to prevent such malpractice (Hariharan, 2012). It has been more evident with the spread of the internet yet it's been easier to detect with new technologies such as PDS (Green, 2002). Even though the discussed paper tackles only one part of intellectual property rights, which is copyright, it still delivers solutions in regards to many other aspects. One of which is universities IT infrastructure. Universities can invest in overhauling their IT infrastructure to handle PDS internally, while third party PDS would only handle exhaustive search queries that were undetectable by the internal PDS. The latter part helps in reducing the workload on the university IT department. Another aspect is the probability of interconnecting universities without the involvement of commercial PDS, since no party is disclosing the original document rather they share queries to find similarity in other documents elsewhere.

The experiment in the proposed architecture only resolves the issue of copyright in the form of written documents within the educational field. Those documents must incorporate Natural Language Processing "NLP" for detection (Paul & Jamal, 2015). Other fields could benefit from this method if it integrated other forms of intellectual property such as video, audio, or even mathematical equations for that matter, in its search queries.

Another benefit from the paper could be derived as an awareness message for the field. Scholars could positively foresee their expectations from such automated systems for plagiarism detection when they are aware that their students' copyrights are protected (Jocoy & DiBiase, 2006). Furthermore, higher education institutions (HEI) can manage and minimize ethical risks associated with students' copyright infringement such as access, consent, and confidentiality where there could be security breaches by PDS malpractice in the form of spreading awareness among students and scholars (Veuger et al., 2018; Glogovac, Krasic, Vulovic, Rakecivic, & Horvat, 2017).

## Paper Assessment

The paper is well written and well presented with a clear goal and a practical solution to the research problem that had been experimented and proven to be effective on a small scale. There are however some limitations to the proposed architecture. The major one is that it cannot detect plagiarized documents that are from different universities or could not be accessed online. For example, if a document from another university has been the source of plagiarism and that paper was never uploaded to the internet, it is impossible for the proposed system to detect it. Even though this draws a major flaw in the proposed system, it still doesn't count for much since up to 90% of plagiarism comes from internet resources.

Furthermore, there are factors to be considered for adopting the proposed architecture on a larger scale. These factors are;

- Increase in computational power and storage.
- Internal database management system (DBMS).
- Integrating conventional search engines such as Google or Bing to look for indexed text within documents that are available online.

- Decreasing subscription costs of third party PDS by investing more on internal IT infrastructure.

Therefore, according to the paper, for a sustainable IT infrastructure to handle the proposed architecture on a larger scale, universities need to have a Learning Management System (LMS), which according to Al-Bataineh, Hallatt, Huss, and Unsbee (2015) is also known as Course Management System (CMS) or Virtual Learning Environment (VLE). As well as internal PDS, external PDS, and conventional search engines.

Another limitation of the proposed architecture is the randomness of queries sent out for detection. Documents with very few plagiarized contents have lower probability for detection since random queries don't necessarily cover the plagiarized portions of the document for checkup. As previously mentioned in paper results part, the authors have suggested a solution to overcome this setback but it seems to be inconvenient since the whole idea of the proposed architecture is to provide automated detection. Manually preselecting parts of the document to be examined can be frustrating and time consuming.

Trying to reflect on the paper we see that commercial PDS are dependent on two methods for reporting on similarity; using web search to compare with online documents and comparing submitted works with formerly archived submissions. The latter is an obvious attempt to use copyrighted material for commercial use. Even if the similarity report doesn't actually show the content of the original document it is still considered copyright infringement under derivative work beside the fact that they are making profit from that particular service.

## Conclusion

After reviewing the paper and reading related topics we can conclude that the process of protecting students' intellectual property, namely copyright of their written materials, is an overwhelming and generally overlooked task. It involves continuous precautions when trying to implement commercial PDS as they're always striving to obtain and archive as much information from submitted documents as they can. Hence, the proposed architecture tries to constrain third party PDS from taking advantage of students' copyrights. In order to implement the new architecture, some measures need to be taken into consideration. These measures include providing awareness of PDS exploits of copyrights, overhauling IT infrastructure to process plagiarism detection internally, interconnecting universities with the proposed architecture, submitting only small and random parts of the documents to commercial PDS, and integrating the proposed architecture to handle different kinds of copyrighted materials.

Recommendations for future research could be extracted from the following questions;

- What obstacles face implementing the proposed architecture on a larger scale?
- How to adapt the proposed architecture to handle different file formats (video, audio, images) beside written documents?
- How to raise awareness among both scholars and students about potential copyright exploitation by commercial PDS?
- Would encrypting documents resolve the issue of connecting with commercial PDS?

## References

- Al-Bataineh, A., Hallatt, D., Huss, M., & Unsbee, C. (2015). *Homework Completion: Perceptions and Comparisons Of 6th-12th Grade Students Using Traditional and Digital Submission*. Paper presented at the Teaching & Education Conference, Amsterdam, The Netherlands. doi: 10.20472/TEC.2015.001.001 Retrieved February 8, 2019 from <http://www.iises.net/proceedings/teaching-education-conference-amsterdam/table-of-content/detail?article=homework-completion-perceptions-and-comparisons-of-6th-12th-grade-students-using-traditional-and-digital-submission>
- Brinkman, B. (2013). An analysis of student privacy rights in the use of plagiarism detection systems. *Sci Eng Ethics*, 19(3), 1255-1266. doi:10.1007/s11948-012-9370-y
- Butakov, S., Dyagilev, V., & Tskhay, A. (2012). Protecting students' intellectual property in the web plagiarism detection process. *The International Review of Research in Open and Distributed Learning*, 13(5). doi:10.19173/irrodl.v13i5.1239 Retrieved April 29, 2018 from <http://www.irrodl.org/index.php/irrodl/article/download/1239/2391>
- Dhir, A., Arora, G., & Arora, A. (2008). Architectural Designing and Analysis of Natural Language Plagiarism Detection Mechanism. *Journal of Theoretical and Applied Information Technology*, 4(12), 1150-1170. Retrieved April 29, 2018 from <http://www.jatit.org/volumes/research-papers/Vol4No12/3Vol4No12.pdf>
- Glogovac, M., Krasic, M., Vulovic, A., Rakicevic, Z., & Horvat, A. (2017). *Importance of Specific Risks in Higher Education* Paper presented at the 4th Teaching & Education Conference, Venice, Italy. doi: 10.20472/TEC.2017.004.005 Retrieved February 8, 2019 from <http://iises.net/proceedings/4th-teaching-education-conference-venice/table-of-content/detail?article=importance-of-specific-risks-in-higher-education>
- Green, S. P. (2002). Plagiarism, Norms, and the Limits of Theft Law: Some Observations on the Use of Criminal Sanctions in Enforcing Intellectual Property Rights. *SSRN Electronic Journal*. doi:10.2139/ssrn.315562
- Hariharan, S. (2012). Automatic Plagiarism Detection Using Similarity Analysis. *The International Arab Journal of Information Technology*, 9(4), 322-326. Retrieved April 29, 2018 from <http://www.iajit.org/PDF/vol.9,no.4/2796-4.pdf>
- Jocoy, C. L., & DiBiase, D. (2006). Plagiarism by Adult Learners Online: A case study in detection and remediation. *The International Review of Research in Open and Distributed Learning*, 7(1). doi:10.19173/irrodl.v7i1.242 Retrieved April 29, 2018 from <http://www.irrodl.org/index.php/irrodl/article/download/242/495>
- Paul, M., & Jamal, S. (2015). An Improved SRL Based Plagiarism Detection Technique Using Sentence Ranking. *Procedia Computer Science*, 46, 223-230. doi:10.1016/j.procs.2015.02.015
- Vani, K., & Gupta, D. (2016). Study on Extrinsic Text Plagiarism Detection Techniques and Tools. *Journal of Engineering Science and Technology Review*, 9(4), 150-164. Retrieved April 29, 2018 from <http://www.jestr.org/downloads/Volume9Issue4/fulltext23942016.pdf>. <https://doi.org/10.25103/jestr.094.23>
- Veuger, S., Gallagher, S., Creighton, H., Robson, I., Cookson, L., & Ridley, M. (2018). *Uncovering Expectations and Perceptions of The Final Year UG Dissertation - Supporting the Student Research Journey and Enhancing Their Graduate Attributes*. Paper presented at the 5th Teaching & Education Conference, Amsterdam, The Netherlands. doi: 10.20472/TEC.2018.005.009 Retrieved February 8, 2019 from <http://iises.net/proceedings/5th-teaching-education-conference-amsterdam/table-of-content/detail?article=uncovering-expectations-and-perceptions-of-the-final-year-ug-dissertation-supporting-the-student-research-journey-and-enhancing-their-graduate-attributes>