Testing of support tools for plagiarism detection

Tomáš FOLTÝNEK, Dita DLABOLOVÁ, Jan MUDRA
Mendel University in Brno, Czech Republic

Debora WEBER-WULFF
HTW Berlin University of Applied Sciences, Germany

Alla ANOHINA-NAUMECA, Laima KAMZOLA
Riga Technical University, Latvia

Styliani KLEANTHOUS
University of Cyprus, Cyprus

Salim RAZI
Canakkale Onsekiz Mart University, Turkey

Július KRAVJAR
Slovak Centre of Scientific and Technical Information, Slovakia

Jean GUERRERO DIB
University of Monterrey, Mexico

Introduction

Plagiarism detection tools, also known as text-matching software, are expected to use state-of-the-art methods to detect plagiarism. The expected output is a suspicious document, where the plagiarized parts are highlighted (showing WHAT was reused), together with links to potential sources (showing WHERE these parts came from). To reveal HOW the source was changed and whether a particular case constitutes plagiarism, human examination is needed.

Current detection systems are quite good at finding copy-paste plagiarism, but unfortunately (and surprisingly) fail in finding obfuscated plagiarism, such as translation, paraphrase and summary (Vani & Gupta, 2016). There are a number of systems available for free or as a paid service, some are available online, while others can be downloaded and used locally. Academics around the globe are naturally interested in the question: How far can these systems reach in detecting plagiarism and to what extent are they successful?

In this study, we will look at state of the art plagiarism detection software and provide a comparison based on specific criteria by following a systematic methodology.

Related Work

There have been several initiatives to test the current tools that claim to detect plagiarism. The most methodologically sound comparison was conducted by Debora Weber-Wulff and her team between 2004 and 2013. In their last testing experiment in 2013, the researchers

compared 15 tools which were selected based on previous comparisons (Weber-Wulff et al, 2013). The testing set contained documents mostly in English and German.

After that time, there have been several more attempts at testing. Luparenko (2014) published useful comparative tables of the tools. An overview of detection systems summarizing their most important characteristics was also provided by Pertile (2015). Chowdhury and Bhattacharyya (2016) provided an exhaustive list of tools, but each was presented briefly and authors did not make comparisons. Vani and Gupta (2016) compared 4 publicly available plagiarism detection tools; however, the selection criteria they used are unclear.

It should be noted that none of these attempts was as systematic and methodologically sound as that carried out by Weber-Wulff and her team. The international team formed under the European Network for Academic Integrity decided to fill this knowledge gap by conducting a broad test of such tools using a clear methodology with the ultimate aim of the results being generally accepted. Both the initiative and the group are called "Testing of Support Tools for Plagiarism Detection" (TeSToP).

Methodology

In order to test the systems, a large collection of intentionally plagiarized documents in 9 different languages was prepared: Czech, English, German, Greek, Italian, Latvian, Slovak, Spanish, and Turkish. The documents use various sources, various plagiarism techniques (cut and paste, paraphrase, translation) and various disguising techniques. The testing set also contains original documents to check for possible false positives. All testing documents were prepared by TeSToP team members or their collaborators.

The documents are available in PDF, DOCX, and TXT form. The amount of plagiarism in each document and the sources used are determined according to TeSToP document set methodology, which is unavailable to vendors before testing. Each document was numbered according to an internal document-numbering scheme. Each document may belong to one or more specific subsets according to the language, type of plagiarism and disguising technique.

Approximately 20 vendors have agreed to participate in the testing. In the next stage, the documents will be submitted to the systems by authorized TeSToP members at a time unknown to the vendor. To ensure comparability, the time difference between submission to different systems will be as small as possible.

The reports will be reviewed by participating TeSToP authorized members and judged qualitatively. System default parameters will be used at all times; if values such as minimum word run are discernable, they will be recorded. The evaluators will judge the following aspects:

- Coverage (How much of the known plagiarism was found? How much plagiarism was reported?)
- Usability (Understandability of report; Usefulness)
- Price (where available)

We are aware of the fact that percentages of similarity do not carry any information on the

actual extent of plagiarism and may even be misleading. Therefore we decided to evaluate the coverage by awarding 0-5 points for each test case, as per the following criteria for the amount of plagiarism detected (all (5), major portion (4), more than half (3), half or less (2), very minor portion (1), none (0)). For false positives, the scale is reversed.

The testers keep the right to make modifications in the evaluation criteria if the testing results will reveal new aspects that should be taken into account. However, the testers will notify vendors about possible changes before introducing them.

The final report will be made available to the vendors prior to publication. They can send a response to the assessment of their system. If any mistake is revealed, it will be corrected. Other responses will be included in the final report. The test report will be available online as an open access document. The authors of the study may publish academic papers on the test, but they are obligated to always include a reference to the open access document.

Expected Results

Based on the above mentioned points, usefulness and expert judgement of the testing team, the systems will be classified into four groups using a qualitative assessment:

1. Useful for academic institution
2. Partially useful for academic institutions
3. Marginally useful for academic institutions
4. Unsuited for academic institutions

At the time of submitting this abstract, the documents are being uploaded to the systems. The results will be fully evaluated by the spring of 2019. However, the conference presentation will endeavour to cover the most important findings of this evaluation.

Keywords: plagiarism detection, text matching software, testing.

References

Chowdhury, H. A., & Bhattacharyya, D. K. (2016). Plagiarism: Taxonomy, Tools and Detection Techniques. *In 19th National Convention on Knowledge, Library and Information Networking (NACLIN 2016).* Retrieved from http://arxiv.org/abs/1801.06323

Luparenko, L.A. (2014). Plagiarism Detection Tools Comparative Tables. Retrived from http://lib.iitta.gov.ua/4066/1/Comparative_Tables.pdf

Pertile, S. de L., Moreira, V. P., & Rosso, P. (2015). Comparing and combining Content- and Citation-based approaches for plagiarism detection. *Journal of the Association for Information Science and Technology, 67(10),* 2511–2526. https://doi.org/10.1002/asi.23593

Vani, K., & Gupta, D. (2016). Study on Extrinsic Text Plagiarism Detection Techniques and Tools. *Journal of Engineering Science & Technology Review, 9(5).* Retrieved from http://jestr.org/downloads/Volume9Issue5/fulltext2952016.pdf

Weber-Wulff, D., Möller, C., Touras, J., & Zincke, E. (2013). Plagiarism detection software test 2013. Retrieved from http://plagiat.htw-berlin.de/wp-content/uploads/Testbericht-2013-color.pdf