

Testing of Support Tools for Plagiarism Detection

Tomáš Foltýnek, Mendel University in Brno, Czechia

Dita Dlabolová, Mendel University in Brno, Czechia

Alla Anohina-Naumeca, Riga Technical University, Latvia

Salim Razi, Canakkale Onsekiz Mart University, Turkey;

Julius Kravjar, Slovak Centre of Scientific and Technical Information, Slovakia

Laima Kamzola, Riga Technical University, Latvia

Jean Guerrero-Dib, Universidad de Monterrey, Mexico

Ozgur Cerlik, Balikesir University, Turkey

Debora Webber-Wulff, HTW Berlin, Germany

Keywords: plagiarism detection, text-matching software, software testing, plagiarism

Acknowledgement: We aim to present the findings from our study in the form of poster. At the time of the conference, our results will be already published as a preprint and submitted to a journal. This contribution does not bring any new results. Its purpose is to allow the conference participant to learn more details about our research.

Abstract

There is a general expectation that software must be able to easily do things that humans find difficult. Since identifying plagiarism in a text and finding its sources is not easy, there is a wide-spread expectation that must be simple for software or some sort of technology to determine if a text is plagiarized or not.

According to Fishman (2009, p. 5): “Plagiarism occurs when someone uses words, ideas, or work products, attributable to another identifiable person or source, without attributing the work to the source from which it was obtained, in a situation in which there is a legitimate expectation of original authorship, in order to obtain some benefit, credit, or gain which need not be monetary.”

Foltýnek, Meuschke, and Gipp (2019) identify three layers of addressing plagiarism:

1. Plagiarism detection methods, given a suspicious document, are expected to identify possible source document(s) in a (large) repository;
2. Plagiarism detection systems maintain a database of potential sources, employing various detection methods and providing an interface to users;
3. Plagiarism policies define institutional rules and processes to prevent and deter plagiarism.

Software focuses on the second layer but it cannot determine plagiarism. It can

Software focuses on the second layer but it cannot determine plagiarism. It can work only as a support tool for identifying some text similarity that may constitute plagiarism. The question is: How well do the various systems work? How effective they are to find such similarities?

Testing similarity detection tools is not a new endeavor, however, previous studies generally have fallen short in providing satisfactory results. The research was designed based in the analysis of previous efforts and tries to overcome their problems and shortcomings. It is the output of an intensive collaboration and systematic effort conducted by researchers from seven countries.

The study compares fifteen test for web-based similarity detection tools (Akademia, Copyscape, Dodoloc, DPV, Dupli Checker, intihal.net, PlagAware, Plagiarism Software, PlagiarismCheck.org, PlagScan, StrikePlagiarism, Turnitin, Unicheck, Urkund and Viper) using two main criteria (coverage and usability), analyzing testing documents (single-source and multi-source) in eight languages (Czech, English, German, Italian, Latvian, Slovak, Spanish, and Turkish), compiled from several sources (Wikipedia, online articles, open access papers, student theses available online) and using various disguising techniques (copy & paste, synonym replacement, paraphrase, translation, white characters, homoglyphs and image).

Using a rigorous methodology, we classified all systems into four categories (useful systems, partially useful systems, marginally useful systems and unsuited for academic institutions) for each criterion. For the coverage criteria, we tested how much of the known plagiarism was found and found 5 partially useful systems, 7 marginally useful systems, and 3 systems unsuited for academic institutions. For the usability criteria, we assess the testing process, the understandability of report, and other usability aspects we found 5 useful systems, 5 partially useful systems and 5 marginally useful systems.

The most important findings of the study are:

1. Some systems work better for a language or languages family. Coverage of sources written in major languages (English, German, Spanish) is in general much larger than coverage of minor-language sources (Czech or Slovak).
2. The systems performance varies according to the source of the plagiarized text (for instance, most systems are good at finding similarity in Wikipedia texts but not as good for open-access papers, thesis and online articles).
3. The performance of systems is also different depending on the disguising technique used. The performance is partially satisfactory in synonym replacement and quite unsatisfactory especially for paraphrased and translated texts.
4. The systems perform better at catching similarity in multi-source documents than single-source ones. Considering that patch-writing is a common technique used
5. As for the usability perspective, this study clearly portrays how useful the similarity reports and how user-friendly the testing process of the systems are. The users can see

which features are supported by the systems and which are not. Also, vendors can benchmark their features with other systems.

Based on our results, we offer the following recommendations to the vendors for the improvement of their systems:

1. Consider employing semantic analysis techniques to successfully detect all types of plagiarism, particularly those coming from translation and/or paraphrase.
2. Identify the original sources of plagiarism when a text has been found similar to different sources, for example Wikipedia and a blog article that has copied/used text from Wikipedia, the system should show both as possible sources of plagiarism, prioritizing showing Wikipedia first because it is more likely to be the real source of plagiarism.
3. Avoid asking users to enter metadata (for example, author, subject, among others) in the system along with the text or file as mandatory information. It is good to have this feature, but it should not be mandatory.
4. Design useful reports and documentation. They must be readable and understandable in any format. Special care must be taken with printed (static) forms. It must show users the compromised text (plagiarism suspected) along with the possible sources of plagiarism highlighting the text that seems similar.
5. Distinguish false positives from real plagiarism. Many of these false positives occur due to commonly used phrases within the context or language employed.

We also provide valuable insights to academia, policy makers and users. Particularly, a number of important points for educators need to be emphasized:

1. Despite the systems being able to find a good bit of text overlap, they do not determine plagiarism. There is a prevalent misconception about these tools. In the literature, most of the studies use the term ‘plagiarism detection tools’. However, plagiarism and similarity are very different concepts. What these tools promise is to find overlapping texts in the suspicious document. Overlapping texts do not indicate plagiarism all the time. Therefore, the similarity reports of these tools should be inspected by a human being such as a teacher or an academic to decide whether the similarity causes plagiarism or not.
2. Translation plagiarism can be found by a match in references.
3. Another problem related to these tools is the risk of their any possible cooperation with essay mills; this is because technically they can store uploaded documents and share them with third parties. In the ‘Terms and Conditions’ sections of some tools, this notion is clearly stated. Uploading documents to such websites can cause violation of ethics and laws, and teachers may end up with legal consequences. Thus, users are expected to be suspicious about the credibility of the free online tools before uploading any documents to retrieve a similarity report.
4. Detecting plagiarism happens too late in the academic writing process. It is necessary to establish institution-wide efforts to prevent academic misconduct and to develop a culture of excellence and academic integrity. This encourages genuine learning and shows how things can be done right, instead of focusing on policing and sanctioning.

Considering both the number of participating systems, and the number of testing documents and language variety, this paper describes the largest testing which has ever been conducted. We hope the results will be useful both for educators and for policymakers who decide which system to use at their institution. We plan to repeat the test in three years to learn whether the vendors accepted our recommendations.

References

Fishman, T. (2009) “We know it when we see it” is not good enough: toward a standard definition of plagiarism that transcends theft, fraud, and copyright. Proceedings of the Fourth Asia Pacific Conference on Educational Integrity (4APCEI), 28–30 September 2009, University of Wollongong NSW Australia. Retrieved from <https://www.bmartin.cc/pubs/09-4apcei/4apcei-Fishman.pdf>

Foltýnek, T., Meuschke, N., & Gipp, B. (2019). Academic Plagiarism Detection: A Systematic Literature Review. *ACM Computing Surveys*, 52(6), Article 112. doi: 10.1145/3345317

Appendix

Contact URLs for the systems used

Tool	URL
Akademia	paneli.akademia.al
Copyscape	www.copyscape.com
Docol©c	www.docoloc.de
DPV	dpv.openscience.si
Duplichecker	duplichecker.com
Intihal	intihal.net
PlagAware	plagaware.com
Plagiarism Software	www.plagiarismsoftware.net
PlagiarismCheck	plagiarismcheck.org
PlagScan	www.plagscan.com
StrikePlagiarism	panel.strikeplagiarism.com
Turnitin	www.turnitin.com
Unicheck	unicheck.com
Urkund	www.urbund.com
Viper	www.scanmyessay.com