Added Value of Centralised Plagiarism Detection System on a National Level

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

Introduction

The use of plagiarism detection (strictly speaking text-matching or detection of text similarities) by means of information and communication technology became a standard for originality checks. Plagiarism detection software [3] is a good assistant that serves as a support for human decision making process in plagiarism matters. Software for detection of text similarities cannot determine plagiarism, it identifies similarities within the checked document with other sources and these similarities may represent plagiarism. The outputs of text similarities detection do not confirm whether the checked document is original or not, the final decision is made by authorities.

Text similarities detection contributes partially to plagiarism reduction, but text similarities detection alone is an insufficient measure for plagiarism removal or reduction. It is necessary to adopt academic integrity measures, including the academic integrity management system.

According to M. Bek (2018) "The main prerequisite for successful defense against plagiarism has always been and will be the quality work of supervisors with students working on their theses". J. Brandejsová (2018) said: "The essential responsibility lies with the supervisor, who is an expert in the field and is well versed in the relevant literature." Evering and Moorman (2012) consider that the most effective way of dealing with plagiarism is actively addressing issues through instruction and not by means of rules or codes. And they added:

> "The current emphasis on testing and grades has made educators and students alike lose track of the more important goals of schooling, such as lifelong learning and national and global citizenship. Refocusing on higher-order goals can persuade students that plagiarism and other forms of academic dishonesty are not in their long-term best interests."

Upbringing and education towards values in childhood should also continue during school years. Comprehensive curricula at all levels of education with an emphasis on values can significantly contribute to the shaping of the character of pupils and students and also to the culture of academic integrity and, consequently, to plagiarism reduction.

Objectives

[3] The term "plagiarism detection software" is widely used, but more exact term is "text-matching software" or "text similarities detection software".

The objective is to show the advantages of metadata collection in the case of a centralised text similarity detection system. In Slovakia, the Centralised Plagiarism Detection System is working closely with the Centralised Repository of Theses and Dissertations – both are in operation since April 2010. All Slovak higher education institutions (HEIs) are required to use this system according to Slovak law. The paper is focused on analytical possibilities of such a system based on uniform collection of theses and metadata.

Methodology

The cooperating systems, Centralised Repository of Theses and Dissertations and Centralised Plagiarism Detection System, are known under a single name SK ANTIPLAG. The Slovak Centre of Scientific and Technical Information (SCSTI) has operated both systems already for ten years. Five types of theses are collected. Today, a rich collection of theses (more than 0.6 million) and metadata are archived and they are used as a base for a wide spectrum of analytical insights useful for HEIs and the Ministry of Education. Several examples of simple and complex insights will be presented.

Uniform collection methodology (UCM) ensures consistent metadata from all Slovak HEIs using the XML format, which is mandatory for exporting electronic versions of theses and metadata (batch mode) into the SK ANTIPLAG system.

Text similarities detection at higher education institutions

The majority of HEIs are free to decide which text similarities detection system will be used in their academic environment. This also applies to Slovak HEIs with one exception: one designated system is used on an obligatory basis according to the Higher Education Act Amendment (2009). The implementation of SK ANTIPLAG is the first worldwide use of a centralised text similarities detection system, which cooperates with a centralised repository of theses and dissertations (both systems are developed in Slovakia). Before the launch of SK ANTIPLAG, only three HEIs used text similarities detection services. Within a year, all Slovak higher education institutions (public, private, state) started to use the SK ANTIPLAG system and it was a significant step forward. SCSTI is open to share its experience with the use of the SK ANTIPLAG system. The first delegation that wanted to know the Slovak experience with the system was a parliamentary and governmental delegation from Poland - they visited SCSTI already in 2011.

Since January 1st, 2019 Poland is the second country in the world that has implemented a centralised text similarity detection system named Jednolity System Antiplagiatowy (JSA), cooperating with the central repository Ogólnopolskie Repozytorium Pisemnych Prac Dyplomowych (ORPPD) – both systems were developed in Poland (jsa.org.pl). In Poland, text similarities detection was widely used already before JSA's operation.

The Slovak system checks the originality of five types of theses: bachelor's, master's,

rigorous, doctoral and habilitation theses and the access to the theses is open to the general public at www.crzp.sk (in Slovak language). The Polish system checks the originality of bachelor's, master's and doctoral theses, and the access to these theses is open for thesis supervisors, research promoters and for the teaching staff. Metadata related to theses are collected.

In both countries, HEIs do not pay for using centralised text similarities detection, or for licences, implementation, technical support and updates. All collected theses and metadata are stored in one centralised repository. The use of centralised systems is obligatory in Slovakia and in Poland due to amendments to the Higher Education Act.

In Slovenia, all major HEIs use the same system and in the near future it is expected that all HEIs will use this system, which was developed in Slovenia (Ojsteršek, 2018). In Czechia, there is a system used by about 50% of all Czech HEIs; the system was developed in Czechia (www.theses.cz). In Slovenia and Czechia, HEIs use text similarities detection systems on a voluntary basis. In the literature, there have been several declarations that all HEIs in the country use the text similarities detection system. However, a deeper analysis showed that it was not true (Kravjar, 2015).

The role of metadata

Metadata is the key and gate to analyses. If the theses originality check is not accompanied by metadata collection, an opportunity for deeper insights is missed. To name a few of metadata: author, type of thesis, study field, thesis title, thesis subtitle, unique thesis identifier, language, abstract, key words, number of pages, year, supervisor, opponents, department, faculty, HEI, thesis downloadability, date and time of thesis registration in the central repository, similarity percentage, originality protocol creation date and time, date of thesis publication at www.crzp.sk.

The palette of analytical insights will be demonstrated on data from the nationwide SK ANTIPLAG system, which is mandatory for all Slovak higher education institutions operating under the Slovak law since April 2010. Many different analytical views on theses and dissertation are available, for example by supervisor, by thesis type, by faculty, by higher education institution, by type of higher education institution, by study field, their combination, etc. There are some analytical views that show a violation of academic integrity by academic staff. One may say that SK ANTIPLAG is not only a detector of text similarities but to some extent a detector of academic misconduct.

Conclusion

If text similarities detection systems collect metadata, then their indisputable advantage are analytical insights. The absence of metadata collection means the absence of analytical insights. Metadata collection means more work that is rewarded by spectra of analytical

possibilities. On the other hand, centralised systems are relatively new and their comparative corpora are not as rich as those of the systems existing for twenty or more years. The ability of text similarities detection of SK ANTIPLAG in the local language is very good, but the ability of text similarities detection in other than the local language is worse.

SK ANTIPLAG collects theses and metadata according to the uniform collection methodology and it provides analytical insights that have common and comparable data base. This feature is out of reach for an academic environment where text similarities detection systems collect theses only. The systems for the detection of text similarities are not a panacea, they have inherent limitations. One of them is the comparative corpus, which is the base for the comparison. No comparative corpus is all-embracing. These systems are only an element of the whole mosaic that helps to reduce plagiarism and increase the level of academic integrity.

References

Bek, M. (2018, August 1). In: Masarykova univerzita řeší, jak zdokonalit boj s plagiáty, Zprávy z MUNI, https://www.em.muni.cz/udalosti/10816-masarykova-univerzita-resi-jak-zdokonalit-boj-s-plagiaty, Accessed: Nov 5, 2019

Brandejsová, J. (2018, August 1). In: Masarykova univerzita řeší, jak zdokonalit boj s plagiáty, Zprávy z MUNI, https://www.em.muni.cz/udalosti/10816-masarykova-univerzita-resi-jak-zdokonalit-boj-s-plagiaty, Accessed: Nov 5, 2019

Evering, L. C. & Moorman, G. (2012). Rethinking Plagiarism in the Digital Age. Journal of Adolescent & Adult Literacy, pp. 35-44

Kozłowski M. (2019). Personal communication

Kravjar, J. (2015). SK ANTIPLAG is bearing fruit, Plagiarism across Europe and Beyond 2015, pp. 147-163, https://plagiarism.pefka.mendelu.cz/files/proceedings_15.pdf

M. Ojsteršek, personal communication

Unknown (2019). Centrálny register záverečných a kvalifikačných prác, www.crzp.sk, Acessed Nov 7, 2019

Unknown (2019). Jednolity System Antiplagiatowy, https://jsa.opi.org.pl, Accessed: Nov 19, 2019

Unknown (2019). Theses.cz – Vysokoškolské kvalifikační práce, www.theses.cz, Accessed: Nov 11, 2019