



The significance of metadata at theses collection

Július KRAVJAR

Slovak Centre of Scientific and Technical Information, Slovakia

Prevailing model at plagiarism detection is that the theses are collected to be processed by plagiarism detection software. No metadata are collected. This results in the fact that the only output is the originality protocol and there is no space for deeper analytical outputs.

If theses collection at plagiarism detection is accompanied by metadata collection then there are possibilities to analyse detected similarities at more detailed levels. The use of uniform collection methodology of theses and metadata in Slovakia allows to exploit deeper analytical views for MinEdu, higher education institutions, faculties. The spectrum of analytical outputs will be demonstrated on real data from the nationwide SK ANTIPLAG (Slovak Central Repository of Theses and Dissertations and Plagiarism Detection System) which is mandatory for all Slovak higher education institutions operating under the Slovak legal order since April 2010. Today there are about 600 thousands of theses and dissertations with metadata stored in the SK ANTIPLAG. Different analytical views on theses and dissertation are available, for example by key words, by thesis types, by faculties, by higher education institutions, by study fields, by language, by year etc. Some of them point to non-standard academic behaviour.

Keywords: theses, collection, metadata, plagiarism, detection, analyses, higher education, academic misconduct, behaviour.