Centralised National Corpus of Electronic Theses and Dissertations

Július Kravjar, Marta Dušková

Slovak Centre of Scientific and Technical Information

Abstract

ETDs are a significant source of grey literature and not only for the academic community. Slovakia has made a big step forward by implementation of the Centralised National Corpus of Electronic Theses and Dissertations in 2010. The national ETD corpus consists of bachelor's, master's, dissertation and habilitation theses. This implementation was coupled with the concurrent implementation of the National Plagiarism Detection System (aka the originality check or the anti-plagiarism system). Both systems have to be used by all higher education institutions operating under the Slovak legal order. The new theses and dissertations incoming into the national ETD corpus are compared to this corpus and to the selected internet resources. The higher education institutions pay no fee for the service, the system acquisition costs were covered by the Ministry of Education, Science, Research and Sport, and the operating costs are also paid by the Ministry. The formation of both systems and the first two years of its existence is analyzed.

The first signs of activities towards ETD in Slovakia were recorded on the threshold of the Millennium. March 2004 was to become a significant milestone: sixteen academic libraries of twelve Slovak universities decided to solve the ETD.SK project: "Building Digital Academic Libraries - Collecting and Providing Access to Full Texts of Slovak University Publications". The ETD.SK project marked the beginnings of cooperation on a national level in this area, with the effort to follow up international ETD activities. Unfortunately, the project was not sufficiently implemented due to the lack of financial and personnel resources, but mainly because of the lack of legislative support.

The ICT and internet penetration, low copyright awareness and the rapid growth in the number of higher education institutions and students in our country contributed to the ex-

pansion of plagiarism. There was also an inherent lack of systemic action that would act as a barrier for its future growth. The establishment of the nationwide electronic theses and dissertation repository and their originality check was considered as a perspective solution.

A significant step in this matter was made in 2008: the Ministry of Education decided to implement a comprehensive nationwide solution for the collection and processing of theses and dissertations produced at Slovak higher education institutions. The goal: creation of the national theses and dissertations repository, increase in the quality of theses by their originality check, copyright and intellectual rights protection. In 2009, the Higher Education Act was amended and the most relevant change was this: Before the defence of the thesis, the higher education institution forwards the thesis in the electronic form to the Central Repository of Theses and Dissertations (CRTD) and the originality check is performed.

During 2009, the CRTD was built and the whole system with the originality check became reality at the end of April 2010. The existence of such a system has a preventive effect, and not just in the student community. CRTD is now a publicly-available source of grey literature.

A yearly increase in CRTD is about 80 thousand items of bachelor's, master's, dissertation and habilitation theses.

Centralised National Corpus of Electronic Theses and Dissertations

Július Kravjar, Marta Dušková

Slovak Centre of Scientific and Technical Information

The paper analyzes the creation and two years' operation of the national corpus of bachelor, master, diploma, dissertation and habilitation theses of Slovak higher education institutions and the follow-up plagiarism detection system. The national corpus is called The Central Repository of Theses and Dissertations (CRTD). Each thesis has to be entered in CRTD before defence and it is then checked for plagiarism.

Creating Collections of Higher Education Theses: First Steps

"Creating digital collections of own academic production and making them available as full text digital documents, accessible in the computer network, has been gradually taken up by Slovak academic libraries in the last ten years. Projects aimed at the electronic processing of publications of university employees and at the electronic processing of theses and dissertations were the pilot projects in this area. Academic libraries implemented the first projects already at the beginning of the millennium.

March 2004 was a significant milestone in the area of Slovak academic digital libraries when 12 Slovak Universities presented the central development IT project "Building Digital Academic Libraries – Collecting and Providing Access to Full Texts of Slovak University Publications (ETD.SK)". ETD is the internationally used abbreviation for the university graduation theses in the electronic form, i.e. Electronic Theses and Dissertations. ETD.SK Project started the cooperation in this area at the national level with the effort to continue in international activities in the ETD area. The Project involved the specification of the organisational, technical and technological requirements for ETD collection, digitalisation workplaces were created in individual libraries and two hardware storage sites were built (in the cities of Košice and Prešov). Unfortunately, the Project was not implemented sufficiently in all libraries,

mainly for insufficient financial and staff coverage; however, mainly for insufficient legislative support." (Haľko, 2011)

A positive effect was achieved as the universities started to make theses and dissertations accessible through online catalogues of academic libraries available on the Internet.

Plagiarism

Plagiarism existed in the past and will probably continue to exist in the future. In the Ancient Rome the word plagiarist (*plagiarius*) designated a kidnapper, mainly kidnapping children or slaves. Plagiarism action (*plagium, crimen plagii*) also meant offering refuge to an escaping strange slave.

The content of the word has been significantly modified and is kept only in a figurative meaning. Today "kidnapping" is understood as a theft of ideas, opinions or other intangible property that are illegally "appropriated" and presented to be original, authentic. Plagiarism can be defined as use of original ideas and creative formulations of another person with the aim to present them as one's own ideas or formulations. (Szattler, 2007)

In the absence of suitable instruments for the prevention and suppression of plagiarism, this phenomenon may overgrow to unacceptable dimensions. It would probably not be possible to fully eliminate all kinds of plagiarism in the future; however, it is necessary to build barriers where possible. We must not ignore or tolerate it.

The first higher education institution in Slovakia – the "early bird" – started using the system to reveal plagiarism in 2001 – it was a lonely runner during long seven years.

Background Situation

Slovakia with its 5.4 million inhabitants is confronted with plagiarism of higher education theses and dissertations like other countries. Plagiarism was growing at the higher education

scene before 2010. There were no systemic measures preventing its growth. Rapidly increasing number of higher education institutions and students, growing Internet penetration, low understanding of copyright and intellectual property rights – this helped the growth of plagiarism. If we look at 1989 (year of Velvet Revolution[1]) as a basis – and compare it with 2011, the number of schools increased three times to 39 and the number of students increased four times to 250 000. In 1989, Internet penetration was zero; it reached 74.3% in 2010 and 79.2% in 2011.

How does the Internet help plagiarism? In two ways. The Internet offers free papers, final theses, dissertations and also expert literature. Insensitive use of the "copy and paste" functions with no adequate quotations compound in the text signed by the "author" – this is a typical example of plagiarism. Offers related to the preparation "supporting materials" for various types of theses can be found on the Internet with little effort. There are web pages offering the preparation of a wide range of theses (seminar, graduate, bachelor, diploma, dissertation, MBA etc.) covering markets of more than ten countries, and there are also web pages aimed only at local markets or markets with similar language or history. It is necessary to realise that "publication of another person's work as one's own is plagiarism". (SME.SK, Adamová, 2012)

Order and payment for the prepared bachelor, diploma or dissertation theses is not only the Slovak speciality, this is a global problem. If it is found before the graduation that the thesis was ordered, such student may be dismissed from his studies prematurely. If this is found after the graduation, nothing happens. Our Higher Education Act does not define the removal of a university degree. The degree will remain in the hands of the owner (Hospodárske noviny, 2012). It seems that this will change in the near future. The removal of fraudulently acquired academic degree should be defined in the Criminal Code as informed by the daily news in August 2012 (TASR, 2012).

---

[1] The Velvet Revolution (Czech: *sametová revoluce*) or Gentle Revolution (Slovak: *nežná revolúcia*) was a non-violent revolution in Czechoslovakia that took place from November 17 to December 29, 1989. Dominated by student and other popular demonstrations against the one-party government of the Communist Party of Czechoslovakia, it saw to the collapse of the party's control of the country, and the subsequent conversion to capitalism (Wikipedia).

The study of Králiková "Introduction of Rules of Academic Ethics at Slovak Higher Education Institutions" summarises the state of academic ethics at Slovak higher education institutions and shows that the majority of inquired pedagogical workers directly experienced fraud of students. The most discussed topic in the media was the plagiarism of students, mainly the ways in which the students cheat and the instruments used by higher education institutions to eliminate plagiarism. Other themes included plagiarism of pedagogical workers or persons in high public positions. The absence of a wide discussion on academic ethics has also other consequences. One of them is that the academic environment members and the general public do not understand the importance of academic ethics and they are also less sensitive to a breach of it." (Králiková, 2009)

Issues related to the collection and processing of higher education institutions theses in electronic form and issues of plagiarism were often repeated in discussions in the academic area; however, with no significant progress. Seeds of future changes were seeded at the meeting of the Slovak Rectors' Conference in September 2006 (Slovak Rectors' Conference, 2006) when two documents on academic ethics were approved. These nationwide documents deal with ethics of pedagogical workers and students. However, the measures proposed to prevent plagiarism did not come to life (Králiková, 2009). In February 2008, the Slovak Rectors' Conference returned to the problems of plagiarism and asked The Ministry of Education, Science, Research and Sport of the Slovak Republic to coordinate activities connected with the acquisition of the plagiarism detection system. Higher education institutions were recommended to punish plagiarism and to create electronic archives of theses (Slovak Rectors Conference, 2009).

Although the plagiarism and the need to fight against it were being discussed a lot, the final decision was adopted by The Ministry of Education, Science, Research and Sport of the Slovak Republic in 2008 (in 2008, only two higher education institutions used the plagiarism detection system) for all higher education institutions operating under the Slovak legal order:

- To establish The Central Repository of Theses and Dissertations (CRTD), higher education institutions have to send any thesis or dissertation to CRTD;

- Any thesis and dissertation with the name and surname of the author and higher education institutions will be stored in CRTD for a determined period of time from the day of registration;
- Theses and dissertations must be checked for originality before defence, as proved by the originality check protocol;
- The originality check protocol is the result of the comparison with all the theses and dissertations in CRTD and to Internet sources and other available electronic sources;
- Higher education institutions' local repositories of theses and dissertations are the CRTD's communication partners.

The defined tasks were aimed to increase the quality of higher education institutions studies and also:

- Copyright and intellectual property rights protection, its better understanding;
- Theses and dissertations quality improved due to the originality check;
- The creation of The Central Repository of Theses and Dissertations;
- The creation of barriers to growing plagiarism.

Establishment of the Central Repository of Theses and Dissertations

In 2009, the Higher Education Act was amended and the most important modifications were:

1. The Ministry of Education, Science, Research and Sport of the Slovak Republic will administer the Central Repository of Theses and Dissertations
2. Before any person is allowed to defend his/her thesis/dissertation, the higher education institution sends his/her thesis/dissertation to the Central Repository of Theses and Dissertations in electronic form and undergoes the originality check.
3. Thesis/dissertation will be stored in the Central Repository of Theses and Dissertations with the name and surname of the author and the name of higher education institution for the period of 70 years from the day of registration. (Zbierka zákonov (Collection of Acts), 2009)

Theses and dissertations registered in the Central Repository of Theses and Dissertations after 31 August 2011 are publicly available – this was defined in the amendment to the Higher Education Act. (Zbierka zákonov, 2011)

The Central Repository of Theses and Dissertations of the Slovak higher education institutions and the Plagiarism Detection system became real at the end of April 2010. Higher education institutions operating under the Slovak legal order are obliged to use both systems. Approximately 80 000 theses are registered in the Central Repository of Theses and Dissertations yearly. As of 31 October 2012, there were 223,757 theses registered.

What did the implementation of the Central Repository of Theses and Dissertations and the plagiarism detection system (the originality check) bring?

Before the implementation of the project, it was necessary to consider and address:
- The centralised national corpus of electronic theses and dissertations;
- The growing plagiarism of theses and dissertations;
- The occasional efforts of higher education institutions to check originality of theses and dissertations;
- The absence of system instruments to fight against plagiarism at the national level;
- The public access to theses and dissertations from one spot;
- The increased understanding of copyright and intellectual property rights.

The published information on the prepared centralised national corpus of electronic theses and dissertations and the originality check brought preventive effect even before the implementation. Other effects of the implemented project were the following:
- The principal breakthrough in the fight against plagiarism in Slovakia, obligatory use of the Central Repository of Theses and Dissertations and the originality check (with the Central Repository of Theses and Dissertations, Internet sources and other electronic sources).
- The existence and real operation of the instrument for the protection of copyright and suppression of plagiarism.

- Unified methods for collecting theses and dissertations (creating centralised repository for all higher education institutions) and unified system for the originality check (Plagiarism Detection System).

- Automated collection of theses and dissertations, originality checks and distribution of originality check protocols.

- The existence of CRTD (as one of the sources of grey literature) and the system to detect plagiarism is preventive in the community of students and others. It increases the understanding of copyright and intellectual property rights at least in the academic area; it improves how students work with literature, Internet, quotations, and improves the quality of theses and dissertations.

- All higher education institutions in Slovakia operating under the legal order of the Slovak Republic are obliged to use the complex of the CRTD with the originality check at the national level. This applies to 35 higher education institutions of 39.

- All theses and dissertations are collected in the CRTD where they are stored for 70 years.

- The public may verify the suspected plagiarism on the web page where theses and dissertations are published: http://www.crzp.sk/crzpopac?fn=*searchform.

- The originality check protocol does not confirm that a thesis/dissertation is original or a plagiarism. The protocol is the basis for the examination committee's decision, it helps – it informs about documents a supervisor or an opponent might have overlooked. The originality check protocol identifies the parts of the text of the presented thesis/dissertation identical with the parts of texts deposited in CRTD and with Internet sources. The originality check protocol is available to the examination committee for evaluation and it is a part of the final (national) exam records.

- Higher education institutions do not pay to use these systems; the procurement costs were covered by the Ministry of Education, Science, Research and Sport of the Slovak Republic and the operation costs are also paid by the Ministry.

In Conclusion

There are large reserves in the fight against plagiarism in the education of the young generation. Plagiarism should be prevented and the process of education to the non-cheating cul-

ture must start gradually and adequately from the earliest age, from the pre-school education level. (Skalka, et al., 2009).

A correctly oriented and properly timed educational process and the implementation of advanced technologies have high potential in the fight against plagiarism. Technologies, however, are not a panacea. Very important and irreplaceable is the role of education – from the beginning of the educational process – in close connection with the prevention and plagiarism uncovering with clearly defined rules and penalties, and with mutual interaction of all these parts. (Kravjar, 2011).

The implementation of the CRTD together with the originality check at the national level in everyday practice is a unique solution in Europe and probably in the world. The system has high potential to be implemented in many areas. The originality check can be done where a written work is the outcome. The following may be considered:

- Seminar and other works at higher education institutions;
- Research reports;
- Applications for projects, grants, their outcomes;
- Final works for increasing qualification of pedagogical and other professions;
- Secondary school works within the curriculum;
- Secondary school works beyond the curriculum;
- General publication activities (the establishment of relationships with publishers);
- The repositories of grey literature.

The system is still being developed. Last year, the system supplier (a company called SVOP, Ltd, http://svop.sk/en/Default.aspx) won the first prize at the international competition of anti-plagiarism systems in Amsterdam "PAN 2011 Lab Uncovering Plagiarism, Authorship, and Social Software Misuse" held in conjunction with the "CLEF 2011 Conference on Multilingual and Multimodal Information Access Evaluation", where their algorithm detecting plagiarism was the best in all four indicators (plagiarism detection, recall, precision, granularity). It is necessary to remind that the competition corpus did not include Slovak texts, only English, German and Spanish texts, and to point out that the system was also able to uncover the so-

called "translated plagiarism", and that the detection is invariant against a change of word order, against the occurrence of changed words, against omission or additions of words in the text in a suspicious document.

The plagiarism detection system received a significant award at the international conference ITAPA 2011 in Bratislava (Information Technology and Public Administration). In the category New Services, the nationwide plagiarism detection system and CRTD ranked second.

This paper is a reworked and updated version of the conference paper Barrier to thriving plagiarism (Kravjar, 2012).

References

HAĽKO, P. (2011). *Elektronický zber záverečných prác na Prešovskej univerzite v Prešove*. [ONLINE] Available at: http://www.pulib.sk/elpub2/PU/Uninfos2011/data/ P11_Halko.pdf. [Accessed 26.6.2012].

KRÁLIKOVÁ, R. (2009). *Zavádzanie pravidiel akademickej etiky na slovenských vysokých školách.* [ONLINE] Available at: http://www.governance.sk/assets/files/publikacie/akademicka_etika .pdf. [Accessed 21 March 2012].

KRAVJAR, J. (2012). *Barrier to thriving plagiarism*. [ONLINE] Available at: http://www. plagia-ris madvice.org/documents/conference2012/finalpapers/Kravjar_fullpaper.pdf. [Last Accessed 22.8.2012].

KRAVJAR, J. (2011). *Antiplagiátorský systém.* [ONLINE] Available at: http://www.inforum.cz/pdf/2011/kravjar-julius.pdf. [Accessed 23 March 2012].

SKALKA, J. et al. (2009). *Prevencia a odhaľovanie plagiátorstva.* [ONLINE] Available at: http://www.crzp.sk/dokumenty/prevencia_odhalovanie_plagiatorstva.pdf. [Accessed 21 March 2012].

SZATTLER, Eduard. *Právne a morálne aspekty plagiátorstva. Duševné vlastníctvo : Revue pre teóriu a prax v oblasti duševného vlastníctva* [online]. 2007, 1, [cit. 2011-04-06]. Dostupné z WWW: <http://www.solain.org/clanky/20070325192403.pdf>.

SME.SK (e.g. 2011). *Veľký biznis: Diplomovka na kľúč v akcii za pár stoviek eur* . [ONLINE] Available at: http://www.sme.sk/c/6329494/velky-biznis-diplomovka-na-kluc-v-akcii-za-par-stoviek-eur.html. [Accessed 28.6.2012].

HOSPODÁRSKE NOVINY (2012). *Napíšem vám diplomovku. Stačí zaplatiť.* [ONLINE] Available at: http://style.hnonline.sk/vikend/c1-55146090-napisem-vam-diplomovku-staci-zaplatit. [Accessed 23 March 2012]

SCHÖPFEL, J. (2010). *Towards a Prague Definition of Grey Literature*. [ONLINE] Available at: http://archivesic.ccsd.cnrs.fr/docs/00/58/15/70/PDF/GL_12_Schopfel_v5.2.pdf. [Accessed 22.8.2012].

TASR.SK (2012). *Podvodne získaný titul bude trestným činom* . [ONLINE] Available at: http://www.sme.sk/c/6511466/podvodne-ziskany-titul-bude-trestnym-cinom.html. [Accessed 30.8.2012].

ZBIERKA ZÁKONOV (2009). *Zbierka zákonov č. 496/2009.* [ONLINE] Available at: http://www.zbierka.sk/sk/predpisy/496-2009-z-z.p-33272.pdf. [Accessed 20 March 2012].

ZBIERKA ZÁKONOV (2011). *Zbierka zákonov č. 6/2011.* [ONLINE] Available at: http://www.zbierka.sk/sk/predpisy/6-2011-z-z.p-33957.pdf. [Accessed 20 March 2012].