

CRZP/APS – podporný nástroj na zvýšenie kvality výstupov vedeckej činnosti na VŠ



Rozoberať dnes na tomto mieste zoširoka problematiku plagiátorstva – t. j. neoprávneného preberania, resp. napodobovania výsledkov cudzej duševnej činnosti – nie je cieľom príspevku. S prudkým rozvojom a dostupnosťou informačno-komunikačných technológií prakticky pre každého, ako i dramatickým rastom dostupného obsahu na internete sa plagiátorstvo rozmohlo nad únosné medze. Azda najvypuklejšie je to v oblasti školstva, kde časť študentov zneužíva možnosť získať aj celé vypracované texty rýchlo a bez vlastného príspevku do svojich prác. Odhaľovať a potlačovať tento nežiaduci fenomén je pomerne náročnou a vôbec nie jednoduchou spoločenskou úlohou.

Problém plagiátorstva záverečných a kvalifikačných prác (ZP) je asi mediálne najznámejším a reálne je jedným

z najškodlivejších druhov plagiátorstva. Autorom-plagiátorom prinášajú nezaslúžený profesijný, spoločenský, no nezriedka i finančný benefit. Na opačnej strane však v prípade odhalenia plagiátori kazia dobré meno inštitúcie, ktorá im umožnila takúto prácu obhájiť. Po málo dostatočným spôsobom, keď sa jednotlivé vysoké školy (pravdepodobne tie najserióznejšie) pokúšali s problémom vyrovnávať vlastnými, väčšinou menej efektívnymi riešeniami, sa v roku 2008 na podnet Ministerstva školstva SR začalo realizovať komplexné riešenie na celoslovenskej úrovni.

Ciele projektu

Primárnym cieľom bolo navrhnuť a realizovať centrálnu úložisko všetkých záverečných prác, ktoré vznikajú ako

výstupy štúdia na slovenských vysokých školách, s garanciou ich dlhodobého bezpečného uloženia. Na tento prvotný cieľ nadväzovala myšlienka využitia centrálného úložiska ako porovnávacieho korpusu pre plánovaný systém na kontrolu originality záverečných a kvalifikačných prác, ktorý by bol k dispozícii všetkým vysokým školám pôsobiacim na Slovensku podľa slovenského právneho poriadku. Zvýšenie kvality výstupov sa malo dosiahnuť aj nastavením jednotného kritéria na kontrolu originality záverečných a kvalifikačných prác. Zámerom bolo vybudovať vlastný antiplagiátorský systém (APS) pre potreby slovenských vysokých škôl, ktorý by spĺňal špecifikované požiadavky a dal by sa ďalej dynamicky rozvíjať podľa aktuálnych potrieb.

Budovanie systému

Samotné riešenie prebehlo vo viacerých etapách. Kľúčovou bola najmä úvodná analytická a koncepčno-organizačná etapa, ktorá sa realizovala pod gesciou UKF v Nitre a riešila centrálné úložisko všetkých záverečných prác. Ako externý dodávateľ systému bola vybraná firma SVOP, s. r. o. Počas tejto etapy sa v spolupráci s dodávateľskou firmou identifikoval východiskový stav v oblasti zberu záverečných a kvalifikačných prác a navrhla sa koncepcia riešenia, ktorá sa zaoberala aj legislatívou, problematikou autorských práv a licenčných zmlúv na diela, finančným a organizačným zabezpečením projektu v celej jeho šírke, identifikáciou logistiky a dosahov na akademické informačné systémy a pod. Výsledky mapovania východiskovej situácie ukázali napríklad aj to, že väčšina vysokých škôl už zbiera záverečné práce v elektronickej forme a ukladá ich vo svojich akademických informačných systémoch alebo knižniciach. Keďže problematika zberu záverečných (bakalárske, diplomové a dizertačné) a kvalifikačných (rigorózne a habilitačné) prác sa dotýka viacerých zákonov (zákon o vysokých školách, zákon o knižniciach, autorský zákon a i.), bolo už v tejto etape nutné v spolupráci s MŠ SR pripraviť vydanie Metodického usmernenia MŠ SR tak, aby po vybudovaní systému boli vysoké školy pripravené posilať záverečné a kvalifikačné práce do centrálného registra (kontrola originality je podmienkou pripustenia práce na obhajobu). Okrem toho sa riešila tiež problematika definovania požiadaviek na lokálne úložiská vysokých škôl voči CRZP a odhad nákladov na úpravu ich lokálnych IS, problematika uzavretia licenčných zmlúv s autormi, úprava zmluvy medzi ministerstvom a prevádzkovateľom registra, úprava *Metodického usmernenia 14/2009-R o náležitostiach záverečných prác* a príprava všeobecne záväzného právneho predpisu, ako aj koncepcia technického riešenia.

Druhou etapou, ktorá sa časovo čiastočne prekrývala s činnosťami prvej etapy, bola implementácia Centrálného registra záverečných a kvalifikačných prác (CRZP). Táto etapa zahŕňala vypracovanie technického návrhu riešenia, vývoj programového vybavenia, vybudovanie technickej infraštruktúry, inštalovanie potrebného softvérového vybavenia u prevádzkovateľa i na VŠ, testovacie práce a pod. Tieto práce realizoval v prevažnej miere externý dodávateľ. Počas tejto etapy rozhodlo MŠ SR o prevá-

dzkovateľovi, ktorým sa stala ministerstvom priamo riadená organizácia – CVTI SR so sídlom v Bratislave. Súčasťou tejto etapy, prebiehajúcej v roku 2009, bolo aj dobudovanie požadovanej infraštruktúry na strane vysokých škôl a testovanie funkcionality prenosových rozhraní medzi CRZP a AIS (akademický informačný systém). Podstatnou časťou činnosti v tejto etape bola aj príprava pokynov, vysvetlení a pod., súvisiacich s prijatou právnou normou, a vynútená úprava systémov na školách v súvislosti s exportom prác do centrálného registra.

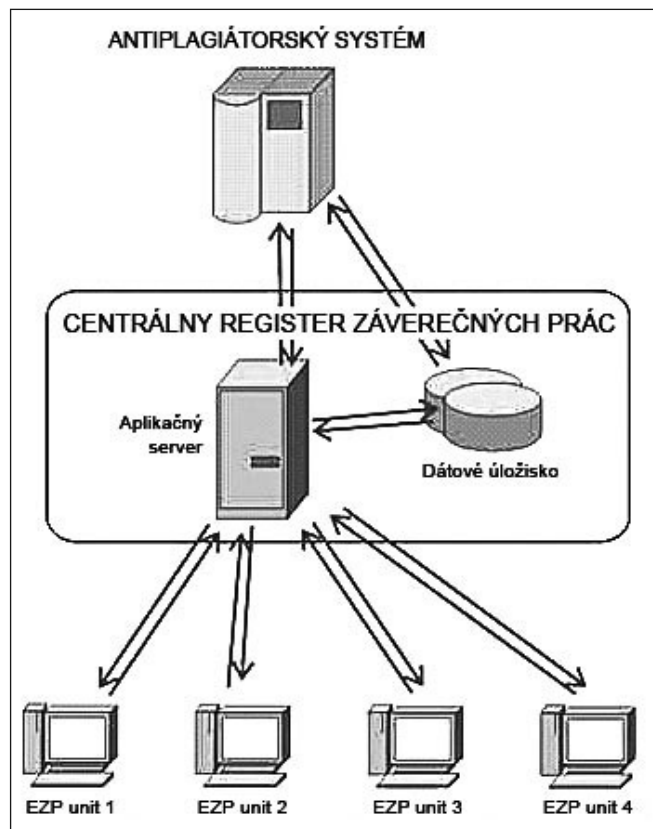
Prakticky od začiatku sa rávalo s tým, že budovaný CRZP bude slúžiť aj ako porovnávací korpus pre aplikáciu na kontrolu originality záverečných a kvalifikačných prác – tzv. antiplagiátorský systém. MŠ SR na tento účel, ktorý môžeme považovať za tretiu etapu riešenia, uvoľnilo finančné prostriedky a poverilo CVTI SR obstaraním antiplagiátorskej nadstavby nad CRZP. Po prieskume trhu a po definovaní požiadaviek na takýto sofistikovaný systém bolo vypísané verejné obstaranie. Súťažné podklady si vyzdvihlo 9 potenciálnych dodávateľov, ale vzhľadom na tvrdé podmienky (najmä z hľadiska časového, keďže bolo zámerom kontrolovať originalitu prác už za akademický rok 2009/2010, obmedzeného finančného limitu, ako aj požiadavka, aby systém fungoval v slovenčine) samotnú ponuku nepodal ani jeden zo záujemcov. V nasledujúcom priamom rokovacom konaní s tromi vybranými dodávateľmi (jedným z nich bola aj MU v Brne) sa CVTI SR napokon dohodla s firmou SVOP, s. r. o., že dodá register CRZP. Táto firma mala východiskovú situáciu uľahčenú najmä tým, že dokonale poznala systém, nad ktorým mal byť APS nadstavbou. CVTI SR ako obstarávateľ išlo do rizika najmä preto, že dodávateľ ponúkal svoje originálne, ale neoverené riešenie, ktoré predtým nikde v praxi nenasadil. Po podpise zmluvy bol naozaj termín dodávky dodržaný a systém na kontrolu originality záverečných a kvalifikačných prác bol koncom apríla 2010, t. j. na začiatku masívneho zberu záverečných prác akademického roku 2009/2010, uvedený do ostrej prevádzky.

Proces zberu a overovania prác

Podľa prieskumu vykonaného v rámci analytických prác v roku 2009 asi 75 % vysokých škôl už vykonávalo zber elektronických verzií prác a väčšina ostatných sa na to pripravovala. Pri dodržaní platnej legislatívy bolo treba nastaviť podmienky a pravidlá tak, aby sa práce zo škôl dostali v požadovanej forme do CRZP, či už majú práce uložené vo svojom lokálnom systéme evidencie záverečných prác (EZP) alebo v inej samostatnej softvérovej aplikácii, alebo prostredníctvom špeciálneho webového rozhrania. Toto lokálne úložisko musí byť vybavené kompatibilným rozhraním s CRZP, prostredníctvom ktorého sa potom vykonáva prenos dát medzi CRZP a EZP.

Samotnému uloženiu záverečných prác do CRZP predchádza proces zberu na VŠ. Počas tohto procesu sa vytvorená ZP prevedie do požadovaného formátu, opatrí sa definovanými metadátami a definujú sa parametre licenčnej zmluvy na prístup k plným textom. Takto pripravené práce sú uložené do EZP školy a usporiadané do dávky čakajú na pokyn na kopírovanie do CRZP. Po po-

žiadavke/pokyne sa cez dátové rozhranie prenesú meta-dáta popisujúce ZP do CRZP, kde sa z nich extrahujú informácie o umiestnení záverečných a kvalifikačných prác. Potom na ich základe sa automaticky stiahnu súbory záverečných prác z EZP do dátového úložiska CRZP.



Obr. 1 Miesto CRZP v procese zberu záverečných prác

Následne sa na základe požiadavky z VŠ spustí kontrola originality prijatej dávky záverečných a kvalifikačných prác voči porovnávaciemu korpusu vytvoreného z prác zaslaných a uložených v predošlom období a voči internetovým zdrojom. Výsledky kontroly originality sú potom automaticky posielané späť príslušným vysokým školám. Samotné texty prác sa začlenia do porovnávacieho korpusu antiplagiátorského systému.

Protokoly o kontrole originality (obr. 3) sú *podkladom* na posúdenie a hodnotenie práce príslušnou komisiou a nie sú potvrdením, že práca je originálom alebo plagiátom. Posúdenie, či ide alebo nie o plagiátorstvo, patrí do kompetencie príslušnej komisie.

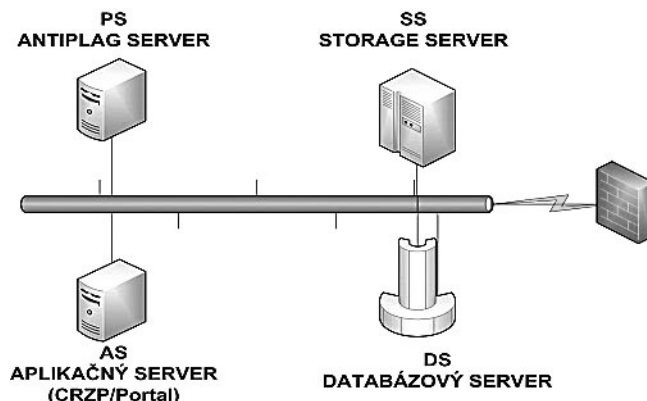
Stručný popis technického riešenia

Technicky sú systémy CRZP a APS konštruované ako systém spolupracujúcich serverov.

- *Aplikačný server* plní funkciu komunikačného servera viditeľného v prostredí internetu, poskytuje portálové služby a zabezpečuje sťahovanie plných textov.
- *Storage server* je univerzálne úložisko pre originálne

súbory i plain texty, výstupné protokoly, logy a pod.

- *Databázový server* slúži na prevádzkovanie MS SQL databáz.
- *Antiplag server* obsahuje indexačné a prehľadávacie jadro algoritmu na odhaľovanie plagiátov (počet týchto serverov sa bude s rastom počtu prác zväčšovať).



Obr. 2 Schéma prepojenia jednotlivých častí systému

Samotný antiplagiátorský systém je súborom viacerých aplikácií a originálnych algoritmov, ktoré vytvorili vo firme SVOP, s. r. o. Sú navrhnuté ako agentový systém a sú spúšťané dynamicky na základe podnetov.

Komunikační agenti

- *Input* – sťahuje súbory z lokálneho úložiska (metadáta a plné texty) a vykonáva konverziu do plain textov.
- *Output* – transformuje výsledky porovnania na plagiátorstvo do tvaru PDF, generuje výsledky vo forme metadát, uploaduje súbory pre agentov CRZP.

Indexační agenti

- Rozkladajú (fragmentujú) text, detegujú jazyk dokumentu, identifikujú texty umelo pripravené na obídenie systému.

Detekční agenti

- Prostredníctvom špeciálnych algoritmov detegujú zhody v textoch.

Čísla a štatistiky

Do systému CRZP/APS posielala v súčasnosti záverečné a kvalifikačné práce na kontrolu originality 33 slovenských vysokých škôl.

V porovnávacom korpuse sa za prvé dva akademické roky prevádzkovania zhromaždilo asi 150 000 záverečných a kvalifikačných prác. Predpokladá sa, že každoročne do neho pribudne približne 80 000 prác. Kvôli kontrole originality bolo z internetu stiahnutých asi 3,4 milióna dokumentov v objeme asi 1,4 TB dát.

Porovnanie jednej práce oproti dnešnému korpusu vrátane spracovania metadát a vygenerovania PDF protokolu trvá v priemere 10 sekúnd. Systém bez problémov zvlá-

Protokol o kontrole originality

Kontrolovaná práca

Citácia	Percento*
BIBLIOGRAFICKÁ IDENTIFIKÁCIA TESTOVANÉHO DOKUMENTU plagID: 765	13,61% << GP

*Číslo vyjadruje percentuálny podiel textu, ktorý má prekryv s indexom prác korpusu CRZP.

Práce s nadprahovou hodnotou podobnosti

Dok.	Citácia	Percento*
1	BIBLIO.IDENTIFIKÁCIA NÁJDENÉHO DOKUMENTU č.1	2,84% << P1
2	BIBLIO.IDENTIFIKÁCIA NÁJDENÉHO DOKUMENTU č.2	13,08% << P2

*Číslo vyjadruje percentuálny prekryv testovaného dokumentu len s dokumentom uvedenom v príslušnom riadku.

Dátum vygenerovania protokolu (protokol1.bmp) Identifikátor protokolu a dokumentu v CRZP (aj vo forme čiarového kódu)

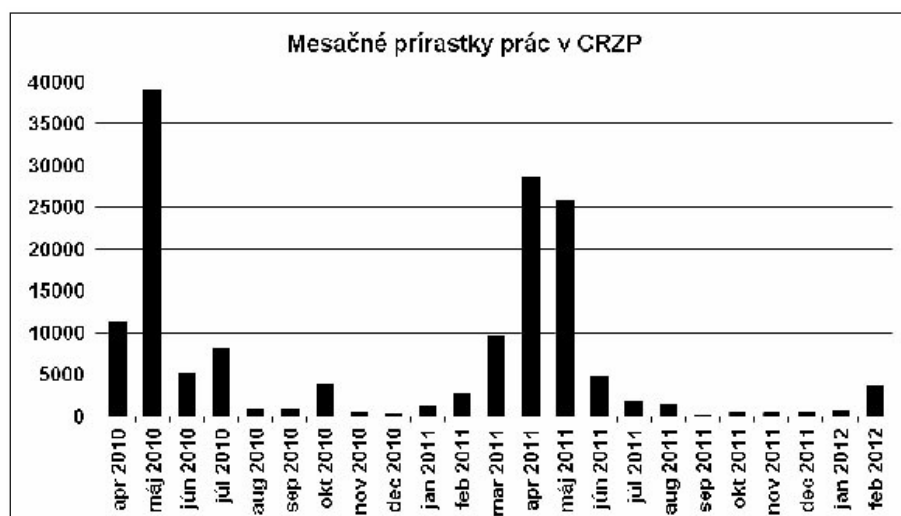
25.04.2010 2391A9FE66FD4429A227FA21BF0B6BC2 - 1 - www.crzp.sk

ukážka textu [môže byť krátený] s vyznačením zhodných plnovýznamových slov

Details - zistené podobnosti

Odsek	Citácia
	BIBLIOGRAFICKÁ IDENTIFIKÁCIA NÁJDENÉHO DOKUMENTU plagID: 290
1 80%	Činu podvodu sa vyžaduje, aby bolo preukázané, že páchateľ v čase, keď sa zmluvný záväzok uzatváral, konal s úmyslom, že záväzok v dohodnutom čase nespĺní, alebo ho nebude môcť splniť a tým veriteľa uvádza do omylu, aby sa obohatil na jeho škodu. Majetkom sa rozumie všetky majetkové hodnoty, ktoré tvoria nielen veci, ale aj pohľadávky, iné práva, peniaze a iné oeceniteľné hodnoty. Cudzím majetkom sa rozumie majetok, ktorý nepatrí páchateľovi alebo nepatrí výlučne len jemu. Škodu na cudzom majetku sa rozumie len újma majetkovej povahy. Obohatením seba alebo iného sa rozumie neoprávnené rozmožnenie majetku páchateľa alebo nejakej inej osoby, či už rozšírením majetku, alebo ušetrovaním nákladov, ktoré by inak boli z majetku, páchateľa alebo inej osoby vynaložené. Ide o neoprávnené obohatenie, obohatenie sa nemusí rovnať spôsobenej škode. Uvedenie do omylu je konanie, ktorým páchateľ predstiera okolnosti, ktoré nie sú v súlade so skutočným stavom vecí. Uvedenie do omylu môže byť spôsobené nielen konaním, ale aj opomenutím konania, alebo aj konkludentným činom. Využitie omylu spočíva v tom, že páchateľ sám k vyvolaniu omylu neprispel, ale po zistení omylu iného a v príčinom vzťahu k nemu koná tak, aby na škodu cudzieho majetku seba alebo iného obohatil. <small>Kvalifikácia</small>

Obr. 3 Príklad protokolu o kontrole originality



Obr. 4 Grafické znázornenie mesačných prírastkov prác v CRZP

dol aj maximálnu dennú dávku dokumentov, ktorá mala 4 900 súborov, čo trvalo asi 10 hodín (obr. 4).

Ohlasy, prínosy, hodnotenia a ocenenia

Téma plagiátorstva je mediálne veľmi príťažlivá, preto sme po uvedení systému do prevádzky s napätím očakávali zvýšený záujem médií i reakcie akademickej obce. Počiatkový mediálny záujem však rýchlo ustal, keďže sa nevyskytli žiadne kauzy. Podobne zo strany vysokých škôl prichádzali väčšinou pozitívne reakcie napriek tomu, že koncepcia riešenia je postavená na direktívnom vyžadovaní porovnania práce na zhody ešte pred jej obhajobou.

Samotné technické riešenie sa ukázalo ako veľmi stabilné a po počiatkovej osвете voči používateľom aj jeho využívanie ako vcelku bezproblémové. Systém spĺňa požiadavky, ktoré boli pri jeho výbere na kladené. Zatiaľ sa javí, že porovnávací algoritmy sú dostatočne rýchle a systém stíha produkovať výsledky s časovou rezervou, čo je veľmi pozitívne. Treba brať napr. do úvahy, že času medzi odovzdaním práce a obhajobou je zväčša málo (asi 2 týždne) a je nutné stihnúť kontrolu originality, posúdenie jej výsledkov komisiou aj napísanie samotných posudkov.

Po takmer dvoch rokoch používania je možné čiastočne vyhodnotiť, či systém napomohol k dosiahnutiu cieľa – obmedzeniu plagiátorstva v záverečných prácach. Faktom je, že niekoľko desiatok prác bolo zaslaných na prepracovanie kvôli nadlimitnému percentu zistených zhody. Podľa vyjadrení napr. prezidenta Slovenskej rektorskej konferencie Libora Vozára, rektora Ekonomickej univerzity v Bratislave Rudolfa Siváka alebo predsedu Rady vysokých škôl Viktora Smieška môžeme usudzovať, že systém CRZP/APS momentálne priniesol najmä preventívny účinok. Dlhšie používanie systému nám v budúcnosti prinesie možnosti objektívnejšie vyhodnotiť jeho prínos.

Podobný problém s plagiátorstvom ako na Slovensku začínajú riešiť aj

v Poľsku. Preto na pozvanie CVTI SR prišli na pracovnú návštevu dvaja vysokopostavení hostia – prof. Jan Kaźmierczak, poslanec Sejmu, predseda parlamentného výboru pre inovácie a informatizáciu, a prof. Zbigniew Marciniak, námestník ministra z Ministerstva vedy a vyššieho školstva v Poľsku. Bolo im odprezentované naše riešenie, ktoré hodnotili vysoko pozitívne a deklarovali jeho inšpiratívnosť a príťažlivosť pre ich plánovaný prístup.

Dodávateľ systému, slovenská firma SVOP, s. r. o., získala na celosvetovej súťaži antiplagiátorských systémov PAN v roku 2011 v Amsterdame so svojím algoritmom prvé miesto vo všetkých parametroch.

Podobne na prestížnom kongrese ITAPA 2011 získal systém na odhaľovanie plagiátov 2. miesto v kategórii „Nové služby“.

Perspektíva ďalšieho rozvoja

Jedna inovácia vlastne už bola realizovaná implementovaním novely vysokoškolského zákona, ktorá hovorí o sprístupnení záverečných prác, uložených v CRZP, verejnosti.

Perspektívy ďalšieho rozvoja systému sú pomerne široké, a to vo viacerých rovinách:

- Skvalitnenie porovnávacieho korpusu tým, že vysoké školy zašlú staršie práce, ktoré majú uložené vo svojich AIS alebo knižniciach. Zatiaľ bola snaha v tomto smere neúspešná. Plánujeme tiež vytypovať ďalšie internetové zdroje, ktoré by sa dali integrovať do porovnávacieho korpusu.
- Skvalitnenie samotného algoritmu porovnávania. Postupne budujeme do algoritmu morfológický a synonymický slovník (od JÚLŠ SAV) a slovníka koreňových morfém (od Filozofickej fakulty Prešovskej univerzity).
- V pláne je optimalizácia výstupných protokolov aj možnosť overovania voči primárnym zdrojom (zákony, normy, vestníky, predpisy...).
- Plánuje sa rozšírenie funkcionality vylepšením používateľského komfortu a širšou škálou možností nastavenia systému pre administrátorov a prevádzkovateľov systému (napr. možnosť nastavenia rôznej hodnoty tolerovaného percenta zhody pre rôzne študijné smery).

Možnosti rozvoja a využitia systému odhaľovania plagiátov sú široké a presahujú hranice porovnávania záverečných prác na vysokých školách. Systém je využiteľný aj v iných oblastiach, kde je potrebná ochrana autorských práv a potlačovanie plagiátorstva.

Záver

CRZP/APS funguje ako dlhodobé centrálné úložisko, plní funkciu efektívneho nástroja na odhaľovanie plagiátorstva. Ide o mimoriadne ekonomicky výhodné riešenie v porovnaní s outsourcingovým variantom pre všetky slovenské vysoké školy.

Fungovanie CRZP/APS prispieva k zvýšeniu uvedomenia autorov, o dodržiavaní autorskej etiky a zákonov. Veríme, že spustenie systému najmä svojimi preventívnymi účinkami prispelo a v budúcnosti iste ešte prispeje k skvalitneniu vedeckovýskumnej a vzdelávacej činnosti na vysokých školách. A možno nielen tam.

Link na stránku CRZP/APS: www.crzp.sk.

Použitá literatúra:

1. NOGE, Juraj: Central register of theses and dissertations in Slovakia and document originality verification as a centrally provided service. In *ProInflow, časopis pro informační vedy* [online], 2011, č. 2 [cit. 2012-03-02] <http://pro.inflow.cz/central-register-theses-and-dissertations-slovakia-and-document-originality-verification-centrally-pr>
2. RAVAS, Rudolf; GRMAN, Ján: *Technical aspects of plagiarism detection system*. Prezentácia v CVTI SR, 2011.
3. SKALKA, Ján a kol.: *Prevenia a odhaľovanie plagiátorstva : zber prác za účelom obmedzenia porušovania autorských práv v kvalifikačných prácach na vysokých školách*. Nitra : UKF, 2009. 126 s. ISBN 978-80-8094-612-8. http://www.crzp.sk/dokumenty/prevenia_odhalovanie_plagiatorstva.pdf [Online] [Dátum: 8. 1. 2011].
4. GRMAN, Ján: *Systémový popis APS (antiplagiátorský systém)*. Technická dokumentácia k APS, 2010.

Juraj Noge
(juraj.noge@cvtisr.sk)

Július Kravjar
(julius.kravjar@cvtisr.sk)