

Plagiátorstvo a spôsoby jeho obmedzenia

Plagiarism and ways of its restriction

Ján GENČI - Milena MAŤAŠOVSKÁ - Ondrej PLOČICA - Henrieta TELEPOVSKÁ

Abstract

Plagiarismus is a phenomenon that is present in our society for decades. After a long period of quiet tolerance it has recently received increased attention. The contribution maps current state of the problem in Slovakia and at the Slovak universities and presents ways how to identify plagiarisms. It presents the basic structure of the system designated to detect plagiarisms in electronic documents.

Keywords

Plagiarism, detection, state of the art

Úvod

Plagiátorstvo (odpisovanie) je podľa Slovníka cudzích slov (akademický)¹ „tvorenie plagiátov“. Pod pojmom *plagiát* ten istý slovník špecifikuje „*nedovolené napodobenie, preberanie umeleckého al. vedeckého diela bez uvedenia vzoru al. autora; dielo takto vzniknuté*“; a Krátky slovník slovenského jazyka [2] zase uvádza „*napodobnenie al. doslovný odpis cudzieho diela bez udania predlohy; dielo, kt. takto vzniklo*“. Ešte dôslednejšie sa k tomuto problému vyjadruje prof. Meško, ktorý plagiátorstvo definuje ako „nedovolené používanie cudzích publikovaných i nepublikovaných myšlienok, formulácií, poznatkov, výsledkov bádania alebo iných výsledkov tvorivej práce, ako aj ilustrácií, tabuliek, fotografií a pod. bez referencie.“ [1].

Problematike plagiátorstva sa v posledných cca troch rokoch venuje pomerne hodne pozornosti aj v médiách. Pravdepodobne k tomu prispeli medializované kauzy, ktoré poukázali na niektoré prípady plagiátorstva. Z diskusií k článkom resp. z blogov (napr. [3]) je ale jasné, že problém bude zrejme ďaleko hlbší. Svedčí o tom aj článok neznámeho autora „Diplomovka: odstrihni a prilep“ uverejnený v denníku Sme 4.5.2007, zaoberajúci sa touto problematikou – v časti „Čo je to plagiátorstvo“ sú vety veľmi podobné vetám uverejneným v [1].

Už v roku 2004 sa v časopise TREND objavil článok Zuzany Horníkovej a Jána Záborského [4], ktorý odkazoval na prieskum agentúry FreshMinds. Výsledkom prieskumu bolo, že 25% vysokoškolákov vo Veľkej Británii priznalo, že sa dopúšťa (resp. dopustilo) plagiátorstva. Článok čiastočne mapoval aj situáciu na Slovensku – Vysoká škola manažmentu (VŠM) v Trenčíne (ako pobočka City University Bellevue, USA; v novších materiáloch prezentovaná ako pobočka City University of Seattle, USA) už v tomto období používala profesionálny systém na identifikáciu podozrenia z plagiarizmu, STU v Bratislave prezentovala existenciu systému na kontrolu zdrojových textov programátorských zadaní.

¹ On-line dostupný na <http://slovník.juls.savba.sk/>

Jozef Hvorecký, na svojom blogu v príspevku „Boj proti študentskému plagiátorstvu. Proti plagiátorstvu sa na vysokých školách bojuje. Len nie všade rovnako dôsledne.“ [5] prezentoval situáciu na VŠM. Azda za najzaujímavejšie z jeho príspevku možno označiť fakt, že sylaby všetkých predmetov poskytovaných VŠM obsahujú paragrafy týkajúce sa problematiky plagiátorstva. Verejne dostupné sylaby predmetov „Odborná písomná komunikácia a kritické myslenie“ [6] a „Kritické myslenie“ [7] sú toho svedectvom.

Ďalšiu zmienku riešenia problematiky plagiátorstva sme objavili na Katedre masmediálnej komunikácie a reklamy, UKF Nitra, kde sa v [8] uvádza: „Známkou “nedostatočne“ sa klasifikuje práca, ktorá nezodpovedá základným požiadavkám z obsahového i formálneho hľadiska, prípadne ktorá nedosahuje predpísanú odbornú úroveň, alebo je práca nesamostatným dielom (plagiát).“

V [9] sa uvádza, že Katedra politológie Filozofickej fakulty UK taktiež schválila pre študentov nariadenie k plagiátorstvu.

Je však potrebné poznamenať, že sa nejedná o špecifický slovenský problém. Odpisuje sa zrejme na celom svete. V dôsledku toho západné univerzity aplikujú technologické riešenia buď v podobe komerčných produktov alebo situáciu riešia vývojom vlastných riešení. Zároveň sa problematike venujú aj také reprezentatívne fóra, ako je napr. IEEE Transaction on Education (2008, Vol. 51, Issue 2)

Je teda zrejmé, že problém plagiátorstva je komplexným problémom, ktorý pokrýva najmä otázky legislatívne, etické a pedagogické. Technologickým cieľom je kontrola a samotná detekcia plagiátov, je potrebné však myslieť aj na prevenciu, aby sa výskyt plagiátorstva v akademickom prostredí minimalizoval. Komplexné riešenie si preto vyžaduje najmä organizačné zázemie v rámci inštitúcie cez príslušné organizačné smernice a predpisy, počínajúc povinnosťou elektronického odovzdávania prác. Ide o tieto oblasti:

- pedagogika: preventívne pedagogické opatrenia, vysvetľovanie problematiky, inštrukcie ako správne citovať literatúru, zaradenie do riadnej výučby, správna voľba tém prác (neopakovať tie isté témy)),
- informačné technológie: používateľsky jednoduchý komplexný informačný systém, detekčné nástroje,
- organizačné opatrenia a interná legislatíva: jasne definované procesy, zrozumiteľné a jasné následné represívne opatrenia a postihy, ak sa plagiátorstvo naozaj potvrdí, zakomponované do príslušných interných predpisov (študijný poriadok, etický kódex).

Stav na slovenských vysokých školách

Problematika plagiátorstva na slovenských vysokých školách nie je vôbec nová. Nové v tejto oblasti je to, že po dlhom čase privierania očí sa o tejto problematike začalo otvorene hovoriť. Azda najviac k tomu prispeli medializované kauzy posledných rokov, spomenuté vyššie. Paradoxne, záujem médií neodštartovalo, ako by sme asi najviac očakávali, odpisovanie v radoch študentov ale samotných učiteľov.

Rozvoj a všeobecná dostupnosť Internetu spôsobili, že proces vzdelávania zasahuje enormná vlna plagiarizmu. Študenti považujú za úplne samozrejmé, ak si "požičajú" časti prác, ktoré sú verejne dostupné na internete, bez uvedenia zdroja (citácií) a prezentujú ich ako svoje dielo. V najmarkantnejších prípadoch kopírujú celé práce.

Tento proces spravidla začína už na základných školách a pokračuje na ďalšom stupni vzdelávania (viď napr. sídlo www.referaty.sk). S týmito návykmi študenti prichádzajú aj na vysoké školy. Situácia došla už tak ďaleko, že skupiny študentov prevádzkujú verejne dostupné portály, kde sa zadania nielen vymieňajú, ale sa s nimi dokonca obchoduje, vrátane „množstevných“ zliav (viď napr. portál www.tu-ke.com). Do procesu niekedy vstupujú aj absolventi, pre ktorých sa táto činnosť stáva doplnkovým príjmom.

Zahraničné skúsenosti ukazujú, že už len zverejnenie informácie o nasadení akýchkoľvek antiplagiátorských opatrení vedie k zníženiu plagiátorstva medzi študentmi. Mnohé univerzity prevádzkujú antiplagiátorské systémy - softvérové riešenia určené na odhaľovanie plagiátov, niektoré dokonca poskytujú službu detekcie plagiátov aj iným subjektom. Existujú komerčné firmy, ktoré poskytujú takéto služby za úplatu (viď napr. www.turnitin.com).

V našich podmienkach je situácia trochu komplikovanejšia, o čom svedčia naše osobné skúsenosti - napr. miera plagiarizmu v predmete spoluautora, kde bolo nasadené jednoúčelové riešenie na odhaľovanie plagiátov zdrojových kódov programov, dosahuje až 90% a to i napriek tomu, že študenti sú na detekciu podobnosti zdrojových kódov upozornení vopred. Takáto vysoká miera plagiarizmu zrejme súvisí s tým, že ešte v nedávnej minulosti sme nepozorovali ochotou manažmentov riešiť uvedený problém a študentov - plagiátorov postihovať.

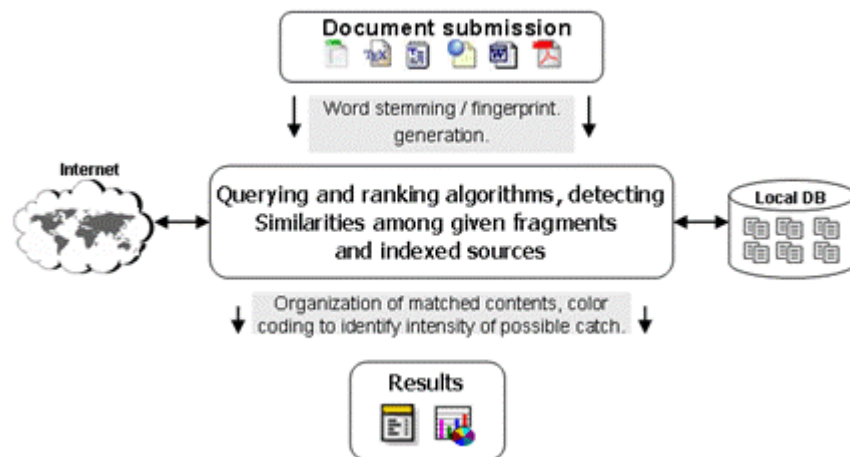
Spôsoby identifikácie plagiátorstva

Nástroje na detekciu plagiátorstva pracujú na rôznych princípoch, ktoré môžeme rozdeliť na tieto hlavné skupiny [10]:

- Nástroje **pracujúce s obsahom dokumentov** :
 - porovnávanie obsahu časti alebo celého dokumentu s množinou dokumentov (korpus).
 - § **Intrakorpálne** – porovnávanie dokumentu s množinou iných dokumentov sa vykonáva iba v rámci toho istého korpusu.
 - § **Extra-korpálne** – dokumenty v rámci korpusu sa neporovnávajú medzi sebou, ale v externých databázach, t.j. porovnávaný dokument je v rámci korpusu a jeho potencionálny zdroj je mimo korpusu.
 - § **Zmiešané** – používajú kombinovaný princíp intrakorpálnych a extrakorpálnych nástrojov.
 - **Intrinsic** – nástroje pracujúce iba na úrovni jedného dokumentu, odhaľujú plagiáty na základe štylometrickej analýzy samotného dokumentu [9]. Detekcia prebieha až po odovzdaní dokumentu (nie v čase tvorby alebo odovzdávania) na základe analýzy štýlu písania autora. Program odstráni z textu určité slová (napr. každé piate slovo) a autor je vyzvaný, aby chýbajúce slová doplnil. Sleduje sa počet správne doplnených slov, chyby, čas a iné údaje. Na ich základe je vypočítané skóre pravdepodobnosti plagiátorstva.
- **Nástroje nepracujúce s obsahom dokumentov** – pracujú s dátami, ktoré sú do dokumentov pridávané za účelom odhalenia plagiátov, ale zamedzujú vzniku plagiátu používajúc rôzne prístupy, nevykonávajú porovnávanie medzi dokumentmi.
 - Neviditeľné značkovanie – študent sa prihlási do systému pod svojim používateľským menom a heslom a následne odovzdá elektronickú verziu do

systému, pričom do dokumentu sú v procese elektronického odovzdania pridané skryté identifikačné údaje slúžiace na následné porovnávanie.

- o Používanie editora APE (AntiPlagiarism Editor) [12] – editor má obmedzenú funkciu kopírovania a vkladania (copy&paste) do dokumentu.



Obr. 1: Princíp detekcie plagiátov na princípe porovnávania dokumentov. Prevzaté z [11].

Obmedzenia detekčných systémov

Detekčné systémy majú výhody, ale zároveň aj obmedzenia, ktoré je potrebné si uvedomiť. Keďže jednotlivé systémy a softvéry na odhaľovanie plagiátorstva využívajú rôzne spôsoby spracovania dokumentov, analýzy textov a rôzne algoritmy detekcie možných plagiátov, výsledky detekcie plagiátorstva resp. originality toho istého dokumentu v rámci rôznych nástrojov sa môžu líšiť.

Vybrané všeobecné limitné faktory systémov na automatizovanú detekciu plagiátov:

- Nástroje pre automatizované odhaľovanie plagiátorstva neodhaľujú plagiáty, ale len podozrivé znaky, ktoré je potrebné ďalej vyhodnotiť.
- Žiaden dostupný nástroj v súčasnosti nevie odhaliť plagiát, ktorý vznikol prekladom dokumentu z cudzieho jazyka, aj keby bol preklad doslovný.
- Automatizované nástroje nemôžu odhaliť plagiát zo zdroja, ktorý im nie je známy (dostupný), t.j. :
 - o Nie je možné zistiť plagiátorstvo zo zdrojov, ktoré sú tlačené a neexistujú v elektronickej forme.
 - o Nástroje pracujúce v intrakorpálnom režime nemôžu odhaliť plagiáty mimo svojho korpusu.
 - o Nástroje pracujúce v extrakorpálnom režime nemôžu odhaliť plagiáty mimo tých (vymenovaných) externých zdrojov, s ktorými pracujú (napr. licencované databázy komerčných producentov).
- Dokumenty v rámci intrakorpálnych nástrojov sú si rovné z hľadiska originality autorstva, pri porovnávaní sú považované aj za možný zdroj aj za možný plagiát.
- Detekovaná podobnosť dvojice dokumentov v rámci toho istého korpusu nemusí znamenať, že sa jedná o intrakorpálne plagiátorstvo, keďže zdroj oboch dokumentov sa môže nachádzať mimo korpusu.
- Intrinsic nástroje vedia odhaliť plagiát len vtedy, ak sa v dokumente vyskytujú časti od rôznych autorov (viac pisateľských štýlov). V prípade podozrenia na plagiátorstvo

nedávajú presný dôkaz o tom, odkiaľ bolo dielo (časť) prevzaté, nelokalizujú potenciálny zdroj (ako je to v reportoch u iných nástroj – podozrivý text je zvýraznený a obsahuje hyperodkazy k pravdepodobným zdrojom).

- Väčšina nástrojov nevie odhaliť parafrázovaný text, alebo pozmenený text (využívajúci synonymá, tezaury).
- Mnohé nástroje nedokážu rozlíšiť medzi textom, ktorý bol správne citovaný a tým, ktorý nebol citovaný (plagiáty nachádzajú aj v bibliografických odkazoch), čím sa skresľuje skóre zhody.
- Nástroje/webové služby nefungujú, ak je problém so sieťou.
- Nástroje/služby, ktoré si vyžadujú upload dokumentu do internej databázy poskytovateľa si vyžadujú analýzu z hľadiska ochrany autorských práv tvorcov dokumentov (študenti) v súlade s platnou národnou autorsko-právnou legislatívou.
- Existujú metódy, ako sa dá každý detekčný nástroj “oklamať“, niektoré návody sú aj na internete.

Detekcia plagiátov

Problematiku detekcie plagiátov je možné rozdeliť do niekoľkých fáz:

1. zber elektronických dokumentov
2. konverzia dokumentov do spoločného formátu
3. výber algoritmov vhodných pre porovnávanie podobnosti
4. určenie podobnosti dokumentov, resp. ich častí.
5. prezentácia výsledkov

Problematiku zberu dokumentov v minulosti riešil projekt ETD, výsledkom ktorého na Technickej univerzite v Košiciach je databáza textov diplomových prác za posledné 4 roky.

Konverzia dokumentov je technicky problém, ktorý je pre pomerne veľký počet formátov dokumentov dobre zvládnutý. Posledné verzie niektorých programov už zvládajú aj prevod diakritických znakov v slovenských textoch, čo ešte nedávno bolo problémom. Taktiež je možné nájsť programy, ktoré umožňujú pracovať v dávkovom režime, ktorý je ako jediný vhodný pre prevod veľkého množstva dokumentov v takomto systéme.

Dobrý prehľad algoritmov na detekciu plagiátov možno nájsť v [13].

Záver

Príspevok prezentuje postupy na identifikáciu plagiátov. Avšak, ako zvyčajne, aj v tejto oblasti môže byť veľmi účinnou prevencia, zvlášť v situácii, kedy sú nasadené technologické riešenia na detekciu plagiátov. Univerzity, ktoré takéto riešenia zasadili, zvyčajne prezentujú okamžité zníženie úrovne plagiarizmu. Veríme, že nasadenie podobného systému na slovenských VŠ bude otázkou veľmi krátkeho času.

Abstrakt

Plagiarizmus je fenomén, ktorý je v našej spoločnosti prítomný už desaťročia. Po dlhom období tichej tolerancie sa mu v poslednom období venuje zvýšená pozornosť. Príspevok mapuje aktuálny stav problematiky na Slovensku a na slovenských vysokých školách a prezentuje spôsoby identifikácie plagiátov. Taktiež prezentuje základnú štruktúru systému na detekciu plagiátov na báze elektronických dokumentov.

Kľúčové slová

Plagiarizmus, detekcia, stav problematiky

Literatúra

- [1] Meško D.: Plagiátorstvo. <http://www.etd.sk/doc/plagiatorstvo.doc>
- [2] <http://slovník.juls.savba.sk/>
- [3] <http://sudor.blog.sme.sk/c/70990/Plagiatori-a-psie-hovienka.html>
- [4] Zuzana Horníková, Ján Záborský: Na plagiátorov nasadili softvér zatiaľ len dve slovenské univerzity. Odpisovanie z cudzích zdrojov nie je cudzie iba študentom, ale aj niektorým pedagógom. <http://www.etrend.sk/technologie/pocitace-a-softver/na-plagiatorov-nasadili-softver-zatial-len-dve-slovenske-univerzity/37379.html> (2004)
- [5] Jozef Hvorecký: Boj proti študentskému plagiátorstvu. Proti plagiátorstvu sa na vysokých školách bojuje. Len nie všade rovnako dôsledne. 6.1.2008
<http://blogy.etrend.sk/jozef-hvorecky/boj-proti-studentskemu-plagiatorstvu/122914.html>
- [6] <http://www.vsm.sk/files/syllabus/07fall/classroom/BABC200s.doc>
- [7] <http://www.cutn.sk/bsba/svk/BABc301sdl.doc>
- [8] Diplomová práca – Návod. Katedra masmediálnej komunikácie a reklamy.
<http://www.kmkr.ff.ukf.sk/oznamy/Navod.pdf>
- [9] Diplomovka: odstrihni a prilep. *Sme*, 4. 5. 2007.
<http://www.sme.sk/c/3278980/Diplomovka-odstrihni-a-prilep.html>
- [10] Lancaster, T. – Culwin, F.: Classifications of Plagiarism Detection Engines. In: *Italics*. Vol. 4 (2), 2005. Internet: <http://www.ics.heacademy.ac.uk/italics/Vol4-2/Plagiarism%20-%20revised%20paper.pdf>
- [11] Maurer Hermann, Kappe Frank, Zaka Bilal: Plagiarism : A Survey. In: *Journal of Universal Computer Science*. Vol. 12 (8), 2006. [citované 7. 6. 2008]. Internet: http://www.jucs.org/jucs_12_8/plagiarism_a_survey/jucs_12_08_1050_1084_maurer.html
- [12] Vamplew, P. and Dermoudy, J. An anti-plagiarism editor for software development courses. In: *Proceedings of the 7th Australasian Conference on Computing Education*. Volume 42 (Newcastle, New South Wales, Australia). A. Young and D. Tolhurst, Eds. ACM International Conference Proceeding Series, vol. 106. Australian Computer Society, Darlinghurst, Australia, 83-90. 2005.
- [13] Mozgovoy M.: Enhancing Computer-Aided Plagiarism Detection. Academic Dissertation. University Of Joensuu 2007.

Kontakt

Ing. Ján GENČI, PhD. Technická univerzita v Košiciach, Ústav výpočtovej techniky a Katedra počítačov a informatiky

E-mail adresa: genci@tuke.sk

Recenzent: Mgr. Ján Skalka, PhD., UKF Nitra