

Odhalovat plagiáty se daří díky spolupráci a odpovědnosti vysokých škol

Cooperation Among Conscientious Universities Reveals Plagiarism

Michal BRANDEJS – Jitka BRANDEJSOVÁ – Martin STANČÍK - Ľuboš LUNTER

Abstract

The issue of plagiarism has recently been intensively discussed in academic and scientific circles. The availability of learning materials, theses and other literature on the Internet fosters plagiarism and tempts students, and sometimes even teachers, to pass someone else's work off as theirs. This situation calls for tools capable of revealing instances of plagiarism, which would, at the same time, function as a deterrent to the would-be copyright violators.

In 2006, Masaryk University provided its students and staff with a unique service by means of which they can search the University's archive of theses for similarities across the individual works it hosts. At that time, demand for a similar service arose at other schools as well. This demand represented an impetus for the joint project of Theses.cz – National Registry of Theses and Plagiarism-Tracing System (<http://theses.cz/>) starting in 2008 and involving Masaryk University as well as 16 other state-owned universities. Later on, some other schools joined the project currently totaling 22 (20 Czech and 2 Slovak).

At the present time, the schools are cooperating in revealing plagiarism not only in theses, but also in other works such as seminar papers, essays, etc. – an effort which gave rise to the project titled Searching for Traces of Plagiarism via Odevzdej.cz (<http://odevzdej.cz/>). Its coordinators, Masaryk University and its development team at the Faculty of Informatics, put the new system into a testing phase in June, 2009. The Theses.cz and Odevzdej.cz databases are interconnected so that similarity searches can be done across all the files they contain.

To make the searches yet more effective, the development team plan to include the files available on the Internet in the search operations, too. The plagiarism detection has recently become an object of interest among other research institutions as well since these would like to apply it to scientific publications and thus protect their and their staff's intellectual property against possible misuse.

Keywords

plagiarism tracing, theses, IS MU, Theses.cz, Odevzdej.cz

Historie

Národní registr vysokoškolských kvalifikačních prací a systém na odhalování plagiátů <http://theses.cz/> vznikl v roce 2008 jako projekt 16 veřejných vysokých škol z ČR.

Myšlenka projektu Národního registru VŠKP a systému na odhalování plagiátů vznikla jako přímý důsledek tržní poptávky, zejména ze strany vysokých škol, po službě odhalování plagiátů mezi vysokoškolskými pracemi. Již v roce 2006 dala Masarykova univerzita svým studentům, vyučujícím a zaměstnancům k dispozici velmi kvalitní nástroj pro odhalování podobných dokumentů v Informačním systému Masarykovy univerzity (IS MU). Učinila tak, jak je jejím zvykem, na základě poptávky vlastní akademické obce. MU je univerzitou otevřenou a maximálně podporuje zpřístupňování elektronických studijních materiálů ve výuce stejně jako elektronického archivu závěrečných prací aj. Na druhé straně si je vědoma,

že každá otvorenosť nese riziko potenciálneho zneužití. Nový technologicky úspešný nástroj poskytuje účinné prostriedky pro technické odhalení podobných dokumentů, což je první důležitý krok v procesu správného posouzení, zda se jedná o plagiát. [1]

V současnosti je do systému Theses zapojených 22 vysokých škol nejen z ČR, ale i ze Slovenska a zájem o zapojení do systému projevili další veřejné i soukromé vysoké školy.

Cíl systému

Hlavním cílem systému Theses je umožnit zapojeným školám vyhledávat v systému ke svým pracím potenciální plagiáty. Ukázalo se, že vedoucí závěrečných prací nutně potřebují efektivní nástroje na zkoumání původnosti předkládaných závěrečných prací. Přístupnost elektronických zdrojů totiž láká studenty vydávat cizí myšlenky za své, a proto učitelé poptávali systém, který je nejen upozorní na opsané texty, ale bude mít zároveň preventivní účinky. Systém slouží současně jako archiv závěrečných prací, s možností vkládat práce do systému individuálně studenty, nebo automatizovaně napojením na existující lokální úložiště školy. Systém Theses nabízí mnoho dalších užitečných služeb jako je např. tematické vyhledávání v pracích, automatický převod prací do formátů PDF a TXT, komunikační nástroje (diskusní fóra a informační vývěsky pro vzájemné konzultace v této oblasti, ...) aj. Hromadný import prací z lokálních úložišť zapojených škol probíhá pomocí metadatového záznamu XML. Na stejném principu mohou školy přebírat výsledky vyhledaných podobností (podezřelých sborů, plagiátů) zpět.

Databáze porovnávaných dokumentů

Systém v současnosti porovnává nejen mezi závěrečnými pracemi 22 zapojených vysokých škol (16 veřejných, 3 soukromé, 1 státní a 2 zahraniční VŠ), ale také mezi pracemi vloženými do systému na odhalování plagiátů v seminárních pracích Odevzdej.cz, dokumenty v dalších systémech provozovaných MU a některými vybranými zdroji na Internetu (např. Wikipedia). Referenční databázi tvoří téměř **1,5 mil.** dokumentů, mezi nimiž se prohledávají podobnosti. Počty vložených dokumentů mají vzrůstající tendenci, protože do systému jsou neustále vkládány další a další práce a systém se rozvíjí o další projekty. Vývojový tým Fakulty informatiky MU v současnosti testuje vyhledávání podobností s celým Internetem, které plánuje do systému implementovat v roce 2010.

Tým odborníků z Fakulty informatiky Masarykovy univerzity řeší v roce 2009 spolu s 9 veřejnými vysokými školami z ČR projekt Odhalování plagiátů v seminárních pracích (Odevzdej.cz – <http://odevzdej.cz/>). Ukázalo se, že problematiku plagiátorství je potřeba řešit nejen u prací závěrečných, a proto je cílem systému Odevzdej nabídnout vysokým školám možnost vyhledávat plagiáty i mezi seminárními pracemi, referáty, esejemi, úlohami, projekty, laboratorními cvičeními, protokoly, zprávami, slohovými cvičeními aj.

Odevzdej.cz umožňuje učitelům zřizovat pro účely odevzdání seminárních prací tzv. odevzdáárny, kam mohou studenti učitelům vkládat práce podle pokynů pedagoga. Jde o e-learningový model pro odevzdávání prací, který přináší pedagogům i mnohé jiné výhody než jen vyhledávání podobných souborů (podezřelých plagiátů). Například učitel neztrácí čas osobním sbíráním prací, nemusí si nechávat práce zasílat e-mailem jako přílohy, neboť stahování velkých souborů, je-li nízká přenosová rychlost připojení nebo omezená kapacita poštovní schránky, může způsobovat problémy. Učitel není nucen vytvářet si vlastní systém adresářových složek na svém lokálním počítači tak, aby se v narůstajícím počtu souborů

vyznan. Systém nahradí učiteľa práci s archiváciou či zálohovaním pro prípad, že by se poškodil lokálny počítač. Učiteľ odpadá sledovanie v kalendári, do kedy ktorým študentom zadal termín odevzdania prác, a nastaví presné datum pro obdobie odevzdania práce (od – do). Práca lze také učiteľom ohodnotiť a študenti si môžu hodnotenie prečítať. Každý učiteľ si môže založiť hneď niekoľko odevzdávaní podľa počtu svojich predmetů alebo skupín študentů a sdílet je podľa potreby s ďalšími svojimi pedagogickými kolegy. Po zpracovaní vložené práce môže vyučujúci nechať práci kedykoľvek zkontrolovať systémom na odhalovanie podobností, ktorý ji porovná s ostatnými vloženými prácami včetně záverečných prác uložených v Theses.cz (<http://theses.cz/>) a s ďalšími evidovanými zdrojmi dát (napr. Wikipédie). [2]

Import prác do systému Theses

Vysoký počet zapojených škôl kladie netriviálne nároky na otvorenosť systému pro různé typy jejich požadavků. Mají napr. různé politiky zpřístupňování a zveřejňování závěrečných prací, některé školy mají vlastní úložiště pro sběr prací apod. Práce lze do systému vkládat nejen studenty samotnými, ale i dalšími centralizovanými, automatizovanými způsoby, napr. prostřednictvím XML souboru. V takovém případě se vytvoří soubor XML, který obsahuje metadata o pracích určených pro import do Theses.cz. XML soubor se vytváří dle potřeb školy buď pravidelně, nebo napr. týden před obhajobou. Samotný import prací se následně uskuteční různými způsoby (import pomocí webového formuláře, automatizovaný strojový import, OAI - PMH).

Jednoduchý import prací probíhá pomocí webového formuláře. Správce vyplní formulář, do kterého vloží metadatový soubor a práce naimportuje.

Další možností je importovat práce hromadně pomocí nástroje curl. K této možnosti je potřeba naprogramovat jednoduchou aplikaci v lokálním systému, která zabezpečí automatický import prací. Správce lokálního systému si vytvoří OAI-PMH server a oznámí správcům Theses metadatový prefix a adresu báze OAI-PMH serveru.

V metadatovém záznamu se nachází také informace o cestě k plnému textu práce. Na základě této informace systém stáhne požadované soubory do svého úložiště a přiřadí soubory k danému archivu závěrečné práce. V systému Theses.cz je implementován systém OCR, díky němuž se rozpozná text i z „netextového“ souboru PDF.

Školy, které nemají vlastní archiv závěrečných prací, mohou pro tyto účely využít systému Theses. Všechny práce vložené do systému procházejí antivirovou kontrolou a jsou zálohovány.

Závěr

Skandály spojené s plagiátorstvím se poslední dobou objevují často. Dokonce se ukazuje, že tyto podvody se netýkají pouze studentů, ale také akademiků. Ojedinelá spolupráce vysokých škol v projektech na odhalování plagiátů se ukazuje jako efektivní nástroj boje proti plagiátorství. V současnosti se ozývají hlasy z řad vědců a výzkumníků, které poptávají podobné nástroje, jenž by napomohly k odhalování plagiátů i v odborných publikacích a jiných vědeckých pracích, a byly navíc propojeny se systémy Theses a Odevzdej, tedy s pracemi závěrečnými a seminárními.

Abstrakt

V odborných a akademických kruzích se diskutuje problém plagiátorství už niekoľik let. Dostupnosť študijných materiálov, záverečných prác a literatúry na Internetu uľahčuje plagiátorom jejich zneužití a existujú prípady, kdy študenti (a někdy bohužel i učitelé) vydávají cizí díla za svá vlastní. Proto se ukázalo jako potřebné hledat řešení, které nejen

odhalí prípady zneužití cizích děl, ale má i preventivní význam ve smyslu „vím, že moje práce bude kontrolována, proto nebudu opisovat“. Masarykova univerzita v rámci svého vlastního úložiště závěrečných prací, implementovaného v Informačním systému Masarykovy univerzity (IS MU), poskytla svým studentům, učitelům a administrativním pracovníkům už v roce 2006 unikátní službu pro odhalování podobností v souborech. V té době vznikla poptávka po podobné službě i na dalších školách. Proto v roce 2008 realizovala Masarykova univerzita spolu s dalšími 16 veřejnými vysokými školami projekt Theses.cz (<http://theses.cz/>) Národní registr VŠKP a systém na odhalování plagiátů. Postupně se do systému zapojují i další vysoké školy a dnes je v systému Theses.cz zapojených 20 vysokých škol z České republiky a 2 vysoké školy ze Slovenska. V roce 2009 pokračuje spolupráce v odhalování plagiátorství nadále, a to nejen u prací závěrečných. Učitelé především humanitních oborů poptávali podobný nástroj k vyhledávání podobností v souborech i pro práce seminární a jiné školní práce, protože i tyto jsou často opisovány, stahovány z různých serverů a vydávány za vlastní díla. Proto podalo 10 veřejných vysokých škol projekt Systém na odhalování plagiátů v seminárních pracích Odevzdej.cz (<http://odevzdej.cz/>). Jeho hlavním řešitelem je Masarykova univerzita a její tým z Fakulty informatiky, který systém implementoval a v červnu 2009 spustil do testovacího provozu. Databáze souborů Theses.cz a Odevzdej.cz jsou navzájem propojeny, takže vyhledávání podobností je umožněno napříč všemi dostupnými soubory. Aby bylo odhalování plagiátů ještě efektivnější, připravuje vývojový tým v blízké budoucnosti rozšíření vyhledávání podobností i o soubory na Internetu. O problematiku odhalování plagiátů se v současnosti zajímají i další vědecko-výzkumné instituce, které mají zájem odhalovat plagiáty mezi odbornými publikacemi a tím zvyšovat autorskoprávní ochranu svých zaměstnanců.

Klíčová slova

odhalování plagiátů, závěrečné práce, kvalifikační práce, IS MU, Theses.cz, Odevzdej.cz, Internet

Literatura

- [1] Brandejs, Michal - Brandejsová, Jitka - Křipač, Miroslav - Kasprzak, Jan. Národní registr VŠKP a systém na odhalování plagiátů. INFORUM 2008: 14. ročník konference o profesionálních informačních zdrojích, Praha : Albertina icome, 1, 1, od s. 1-1, 4 s. ISSN 1801-2213. 2008. ISSN 1801-2213.
- [2] Brandejs, Michal; Brandejsová, Jitka; Krhutová, Růžena; Mikulášová, Zuzana; Pekárková, Lucie; Stančík, Martin. Kontrola plagiátů v seminárních pracích prostřednictvím Odevzdej.cz. Ikaros [online]. 2009, roč. 13, č. 8. [cit. 2009-08-11]. Dostupný na World Wide Web: <<http://ikaros.cz/node/5641>>. URN-NBN:cz-ik5641. ISSN 1212-5075.
- [3] Vysokoškolské kvalifikační práce. Oficiální stránky Theses.cz [online]. 2009. Dostupný z www: <http://theses.cz/>. [cit. 2009-10-30]

Kontakt

Mgr. Martin Stančík, Fakulta informatiky, Masarykova univerzita, Botanická 68a, 602 00 Brno,

č.t.: +420 549 496 285,

E-mail adresa: stancik@fi.muni.cz

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