

Roč. 10, č. 11 (2006)

2. 11. 2006

Odhalování plagiátů na Masarykově univerzitě

(software | plagiátorství | autorské právo | Informační technologie a elektronická komunikace)

English title:

Tracing and Fighting Plagiarism at Masaryk University

Abstrakt:

Development of information technologies and their use for teaching purposes and archiving students' theses calls for tools capable of tracking traces of plagiarism. Since there have recently been numerous requests for such tools on the part of Masaryk University teachers, the team of Masaryk University Information System (IS MU) developers decided to create software providing this service, which was seamlessly integrated in the system. Its tools are to ensure that students' theses (as well as assignments) submitted are of the required quality and do not violate the fundamental principles of copyright.

Obsah:

Odhalování plagiátů nebo ochrana autorských práv je letité a stále aktuální téma, které získalo na intenzitě a významu právě s rozvojem informačních technologií, zejména s rozvojem elektronické podpory výuky a ukládáním závěrečných prací do studijních systémů. Masarykova univerzita (MU) má nejen značný zájem na tom, aby se dbalo na kvalitu studentských prací, ale zároveň je průkopníkem ve využívání IT v celé šíři univerzity a v celé hloubce administrativy studia, výuky a komunikace. Proto vývojový tým Informačního systému Masarykovy univerzity (IS MU) reagoval na stesky učitelů v boji proti opisování a vyvinul unikátní softwarovou službu na odhalování plagiátů. Unikátnost služby spočívá v tom, jakým způsobem je integrována přímo do studijního informačního systému IS MU, v jejím neobvyklém rozsahu a funkcionalitě.

Princip služby pro odhalování plagiátů

Obecný princip vyhledávání podobných textů je v branži informačních technologií znám, přesný implementovaný algoritmus je však speciálním know-how vývojového týmu IS MU. Implementovaná služba je vysoce účinná zejména proto, že studijní informační systém v sobě integruje nejen administrativní část studia, ale i kompletní úložiště dokumentů (závěrečných prací, dokumentů, studijních materiálů) a e-

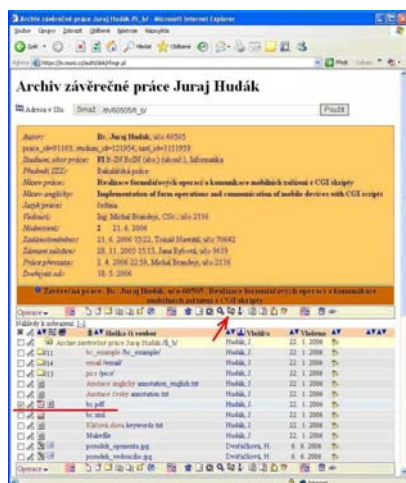
learningové agendy vyznačující se širokou nabídkou souborů ve formě předmětových studijních materiálů nebo odevzdáváním studentských prací. Aktuálně se nyní prohledává přes 250 000 uložených dokumentů (odevzdaných domácích úloh, esejí, závěrečných prací a dalších). Vyhledávání souborů online mezi tímto množstvím dat je účinné a rychlé.

Vícefázová analýza dokumentů

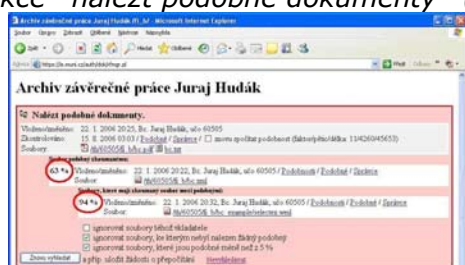
V první fázi procesu odhalování podobných dokumentů se každý text dokumentu vloženého do IS MU automaticky po jeho vložení analyzuje a srovnává se všemi texty ostatních souborů v systému bez ohledu na předměty, semestry, fakulty a agendy (Studijní materiály, Poskytovny, Dokumentový server, Závěrečné práce, Můj web, Úschovna, Odpovědníky - jde o aplikace IS MU). Předpokladem pro tuto analýzu je, aby byl každý soubor v textovém formátu. Proto mezi důležité vlastnosti datového úložiště souborů patří automatická konverze dokumentů do formátu TXT a PDF - stačí soubor umístit do příslušného datového úložiště a za určitou dobu "vedle něj" vzniknou ikony dvou dalších formátů: PDF (Portable Document Format) a holý text. Zpravidla do jedné hodiny po jeho zavedení do systému je pak dokument připraven pro uživatelské hledání. Analýza vloženého dokumentu probíhá opakovaně.

V druhé fázi jsou vyhledávány podobné dokumenty. Uživatel nejprve označí dokument nebo více dokumentů (lze vybrat jeden soubor, více souborů, celou složku, či celý strom složek), k nimž hledá podobné, a spustí operaci vyhledávání (označenou 'nalézt podobné dokumenty') "on demand" (pozn. "na požádání"). Operaci 'nalézt podobné dokumenty' provádí ve Správci souborů (jde o aplikaci "filemanager" v IS MU) pod ikonou "podobné jako vejce vejci". Samotný proces vyhledání podobných dokumentů probíhá velmi rychle, výsledky se zobrazí do několika sekund. IS MU vypíše u nalezených podobných souborů míru shody v procentech. V případě potřeby zobrazit shodující se text, vybere uživatel následující funkcionalitu 'Podobnosti' a systém mu vypíše také shodný text.

Vyhledání podobných dokumentů ke konkrétní závěrečné práci je ilustrováno na následujících dvou obrázcích.



Obr. 1 – Závěrečná práce, k níž se budou hledat podobné dokumenty spuštěním funkce 'nalezt podobné dokumenty' (ikona "podobné jako vejce vejci")



Obr. 2 – Nález podobných souborů s procentuálním určením shody

Vyhledávání v jazykových mutacích

Služba funguje neoptimálněji pro češtinu, slovenštinu a angličtinu, tedy pro jazyky, které jsou nejčastěji užívanými jazyky ve výuce na českých univerzitách.

Spuštění služby

Služba byla spuštěna ve středu 16. srpna 2006. Nejen učitelé, ale i studenti mají možnost dohledat podobné dokumenty. Každý uživatel však vždy pouze mezi dokumenty, které jsou mu přístupné - služba zohledňuje složitý systém individuálně nastavených práv. Další přínos pro uživatele ve vyhledávání podobných dokumentů spočívá v možnosti dohledat vlastní (i cizí) dokumenty například za účelem jejich odstranění, změny nebo aktualizace.

O vývoji a systému

Celouniverzitní studijní informační systém IS MU vyvíjí a provozuje tým inženýrů z Fakulty informatiky MU již od roku 1998.

Kvalitu Informačního systému Masarykovy univerzity potvrzují i mezinárodní uznání, například prestižní cena EUNIS Elite Award 2005, kterou Masarykova univerzita získala v loňském roce za implementaci svého informačního systému jako první, a dosud jediná, z univerzit tzv. východního bloku. Systém má neobvykle rychlý rozvoj, jako první v České republice implementoval komunikační nástroje do administrativního systému, první implementoval e-přihlášky (v loňském roce 97 % podaných e přihlášek), první implementoval e-learning do administrativy studia, první implementoval službu pro vyhledávání plagiátů, první dal k dispozici závěrečné práce v plném znění v souladu s novelou zákona o vysokých školách aj.

Literatura:

1. Masarykova univerzita. Masarykova univerzita vyvinula software proti plagiátům (Tisková zpráva). IS MU: Informační systém Masarykovy univerzity [online]. 2006 [cit. 2006-08-16]. Dostupný z www: <http://is.muni.cz/clanky/plagiatorstvi.pl>.
2. Informační systém Masarykovy univerzity. Masarykova univerzita vyvinula software proti plagiátům (článek). IS MU: Informační systém Masarykovy univerzity [online]. 2006 [cit. 2006-08-16]. Dostupný z www: http://is.muni.cz/info/nov_20060817_plagiat.pl.

Brandejsová, Jitka. Odhalování plagiátů na Masarykově univerzitě. *Ikaros* [online]. 2006, roč. 10, č. 11 [cit. 2006-11-24]. Dostupný na World Wide Web: <<http://www.ikaros.cz/node/3676>>. URN-NBN:cz-ik3676. ISSN 1212-5075.