



MASARYKOVA UNIVERZITA

Globálne vyhľadávanie podobných súborov, prezentácia a export výsledkov

Martin Stančík,
Masarykova univerzita, Fakulta informatiky

História

- 2004 – Archív záverečných prác IS MU
- 2006 – Zverejnenie záverečných prác
- 2006 - **System pre odhaľovanie plagiátov**
- 2008 – Národný register záverečných prác Theses.cz
- 2009/10 – Seminárne práce, šedá literatúra



Proces vyhľadavania podobných súborov

- Import a vloženie prác do centrálného úložiska
- Prevod súborov do textovej podoby (OCR)
- Vyhľadávanie podobností v jednotlivých prácach
- Vyhľadávanie podobností medzi zdrojmi Internetu
- Prezentácia výsledkov podobností
- Export výsledkov podobností do lokálnych systémov
- Dôsledná kontrola práv pri zobrazovaní podobných súborov

Technická realizácia



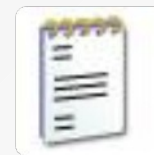
Režimy vkladania prác:

- Priame vkladanie študentmi.
- Hromadný import
 - OAI – PMH
 - Strojové
 - Webový formulár

- Stiahnutie všetkých súborov práce
- Prevod do textu (pdf, txt, OCR).

Podporované formáty

- Formáty MS Word (.doc, .docx.).
- OpenOffice (.odt).
- TeX (.tex).
- PDF (.pdf).
- TXT (.txt).
- Další formáty převoditelné na „holý text“ (.ppt, .xls, .csv, ...).



Vyhľadavanie plagiátov

- Priame vyhľadavanie
 - Vyhľadanie podobností ku konkrétnemu súboru
 - Príklad použitia: Užívateľ si chce overiť na koľko percent je daná práca podobná s niečim iným
- Globálne vyhľadavanie
 - Špeciálna aplikácia pre nájdenie všetkých súborov, ktoré majú podobnosť vyššiu ako zadaná hodnota a spĺňajú zadané parametre.
 - Príklad použitia: správca periodicky kontroluje nové súbory, ktoré majú vyššiu podobnosť ako 30 percent
- Zobrazenie podobností vo vzorovom dokumente

Vyhledávat podobné soubory k souborům

[Statistika vložených souborů v Národním registru](#)

Aplikaci si typicky jednou za čas spustí držitel práva t
podobné soubory.

Pracoviště: th1222 - Zemědělská fakulta - Jihočeská

Pozn. srovnává se vždy proti všem fakultám

zveřejněným po datu: 01 07 2008

s podobností větší než: 40 %

s textovou verzí větší než: 40000 znaků

- ignorovat soubory závěrečných prací na adresách majících shodné dvě nejvyšší složky
- ignorovat dvojice souborů, které byly ručně označeny, že jsou plagiáty
- ignorovat dvojice souborů, které byly ručně označeny, že nejsou plagiáty
- ignorovat všechny soubory podobné souboru označenému 'nikdy není plagiátem'
- omezit na soubory, které byly ručně označeny, že jsou plagiáty

Vyhledat

Lučka Šedláková: Založení pekařství NODY s r.o. v Písku (2008)

Název souboru: [W7_pedlakovsk.doc](#)

Vloženo/změněno: 28. 11. 2008 08:59, Ing. Petr Havel, učo 10079

Zkontrolováno: 28. 11. 2008 / [Podobné](#) / [Správce](#)

Soubory: [W7/7jpoau/W7_pedlakovsk.doc](#) [W7_pedlakovsk.txt](#)

soubor, ani jemu podobné nemohou být nikdy plagiátem (jen PRIVILEG)

Soubory podobné zkoumanému:

58 %

Lučka Šedláková: Vyhodnocení strategického záměru rozvoje mikroregionu

Název souboru: [W7_sedlak.doc](#)

Vloženo/změněno: 30. 11. 2008 10:26, Ing. Petr Havel, učo 10079 / [Podob](#)

Soubory: [W7/pdvbxt/W7_sedlak.doc](#) [W7_sedlak.txt](#)

tento není je plagiát; soubor, ani jemu podobné nemohou být nikdy p

56 %

Lučka Šedláková: Možnosti eliminace daňových úniků (2008)

Název souboru: [W7_sedlak.doc](#)

Vloženo/změněno: 30. 11. 2008 10:26, Ing. Petr Havel, učo 10079 / [Podob](#)

Soubory: [W7/sjedob/W7_sedlak.doc](#) [W7_sedlak.txt](#)

tento není je plagiát; soubor, ani jemu podobné nemohou být nikdy p

45 %

Lučka Šedláková: Vyhodnocení současné pozice společnosti na trhu a



VYSOKOŠKOLSKÉ KVALIFIKAČNÍ PRÁCE

Podobné dokumenty

Bc. Luboš Lunter (učo 10028)

• [IS Odevzdej](#) > [Autentizovaná titulní strana](#)

ODEVZDEJ.CZ
NÁPOVĚDA
LIDÉ
MŮJ PROFIL
OS. STRÁNKA
KLÍČE

Název souboru: 147.251.48.246
 Vloženo/změněno: 23. 6. 2009 10:48, Neautentizovaný Uživatel, učo 4
 Zkontrolováno: Podobnost tohoto dokumentu byla zkontrolována. / [Podobné](#)
 Soubory: [/doc/gltele/h4wuzm/jan_novak_seminarni_prace.doc](#) [jan_novak_seminarni_prace.txt](#) [jan_novak_seminarni](#)

Obsah zkoumaného souboru je z X % podobný souborům níže:

98 %

Agenda: Dokumentový server
 Změněno: 27. 4. 2009 14:42:36
 Soubor pochází z jiného systému is.muni.cz / [Podobnosti](#) / [Informace o souboru](#)

13 %

Agenda: Dokumentový server
 Změněno: 27. 4. 2009 14:42:36
 Soubor pochází z jiného systému is.muni.cz / [Podobnosti](#) / [Informace o souboru](#)

Obsah souborů níže je z X % podobný zkoumanému souboru:

98 %

Agenda: Dokumentový server
 Změněno: 27. 4. 2009 14:42:36
 Soubor pochází z jiného systému is.muni.cz / [Podobnosti](#) / [Informace o souboru](#)

78 %

Agenda: Dokumentový server
 Změněno: 27. 4. 2009 14:42:36
 Soubor pochází z jiného systému is.muni.cz / [Podobnosti](#) / [Informace o souboru](#)

Odevzdej.cz

RELEVANTNÍ ODKAZY

Nájdienie podobného textu v inom dokumente

Tato aplikace porovnává dva dokumenty. Zobrazuje text prvního dokumentu (vzoru), a pokud se tentýž text nachází i ve druhém dokumentu, je odlišen červeně.

[Jen podobnosti](#)

Ve vzorovém dokumentu byly nalezeny tyto shodné části (s dokumentem druhým), které jsou označeny červeně:

INTELEKTUÁLNÍ DRIL: STUDENTI MĚNĚ OPAKUJÍ A VÍCE SI PAMATUJÍ
SPACED-REPETITION ALGORITHM IN UNIVERSITY CONTEXT:
DRILLING LESS BUT REMEMBERING MORE

Michal Brandejs, Jitka Brandejsová, Miroslava Misáková,

Jan Kasprzak,

Daniel Keder,

Ľuboš Lunter

brandejs@fi.muni.cz, brandejsova@fi.muni.cz, imladris@fi.muni.cz, kas@fi.muni.cz,

keder@fi.muni.cz,

lunter@fi.muni.cz

ANOTACE

V Informačním systému Masarykovy univerzity (IS MU) mají učitelé k dispozici e-learningovou podporu už od roku

2004. Kromě interaktivních osnov, studijních materiálů, odevzdávacích, diskusních fór a dalších objektů mohou vytvářet přímo v systému interaktivní testy (Odpovědníky) a zpřístupňovat je studentům v různých režimech - od dobrovolného procvičování až po testy určené k ostrému zkoušení u PC. Novinkou je nástroj pro efektivní zapamatování velkého množství slovíček nebo pojmů Dril.

KLÍČOVÁ SLOVA

dril, učení s prodlevami, křivky zapominání

SUMMARY

Information System of Masaryk University has been assisting teachers with its e-learning tools since 2004. Apart from using the applications such as Interactive Syllabi, Teaching Materials, Homework Vaults, Discussion Groups and other utilities, the teachers can create their own interactive material (ROPOT applications) and make it accessible to their students in various forms ranging from plain exercises to PC-based tests. The application called Drill, a comparatively novel tool, helps the students to learn large amounts of vocabulary items.

KEYWORDS

dril, spaced repetition, forgetting curves

1. Úvod

E-learningové aplikace Informačního systému Masarykovy univerzity (IS MU) již delší dobu obsahují nástroj Odpovědník, který umožňuje učitelům formulovat otázky a studentům na ně odpovídat. To, v jakém režimu student odpovídá (kolikrát, kdy, ...), rozhoduje učitel. Motivací pro vytvoření nové agendy byl záměr vytvořit prostředí, které by učební strategii optimalizovalo samo pro každého studenta tak, aby mohl učební položky procvičovat efektivně (nemusel se jim věnovat příliš často a přitom si hodně zapamatovat). Tvůrci měli na mysli učení se jednoduchých drilovatelných položek, např. slovní zásoby v jazycích.

2. Co je to Dril a proč je unikátní?

Dril simuluje

tzv. kartičkovou metodu s využitím principu „spaced repetition“,

tzv. opakování

s vhodně zvolenými prodlevami. Uživatel

dá systému zpětnou vazbu, jak kterou kartičku

ovládá a systém spočítá,

kdy se má kartička

nabídnout znovu k zopakování.

Nejedná se přitom o klasickou formu e-learningu

„shora“,

Prepojenie viacerých IS a iných zdrojov

Odevzdej.cz
Semestrální a školní práce

Porovnat dokument na shodu
Ověře si svůj textový dokument, zda není podobný dalším zobrazeným textům v systémech zápisů z Internetu.

Soubor z PC:
Nebo URL:
E-mail:

Po kontrole bude výsledek odeslán na zadaný e-mail

Mám kód: Píde typy volného kódu můžete v dalším kroku odeslat práci nebo si založit účet.

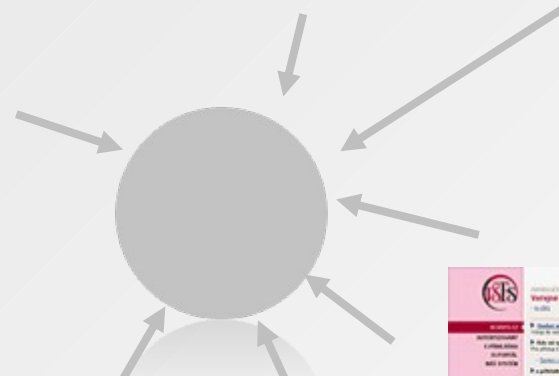
Skoly zapojene do projektu Odevzdej.cz
V roce 2009 byl vznik systému finančně podpojen z centralizovaného rozvojového projektu MSM1 C132009 „System na odhalování plagiatů v semestrálních pracích“

Hlávková: Další projekty: Služba Odevzdej.cz přinášíme Vysokým tým Informačního systému Masarykovy univerzity.
• [Theses.cz](#)
• [Informační systém MU](#)

© [Mazury](#) | Aktuální datum a čas: 19. 6. 2009 13:19:25 (čtyř) týden
Kontakt: plagiat@plagiat.cz

Hlavní strana
Vítejte ve Wikipeidii!
Wikipeidie je online encyklopedie, kterou mohou editovat všichni lidé. Wikipeidie je v současnosti největší encyklopedie na světě. Wikipeidie je v současnosti největší encyklopedie na světě. Wikipeidie je v současnosti největší encyklopedie na světě.

Internet



Vnitřní služby informačního systému
Vnitřní služby informačního systému. Vnitřní služby informačního systému. Vnitřní služby informačního systému. Vnitřní služby informačního systému.

Vysokoškolské kvalifikační práce
Theses.cz je systém pro odhalování plagiatů mezi závěrečnými pracemi. Službu vysokým školám a univerzitám (např. v ČR) jako národní registr závěrečných prací (informací o pracích – název, autor, ...) a jako úložný prostor pro vyhledávané plagáty. Systém umožňuje zápisům závěrečných škol kláde práce a vyhledávané mezi nimi plagáty. Všechny jsou zřizovatelským záznamy o práci, přílohu, před texty (sle rozhodnutí školy), a vyhledávané mezi nimi. Systém nabízí další služby, funkce a aplikace a bude rozvíjen dle potřeby.

Vyhledávání ve veřejné části registru
Toto není Google! Číslo, jak začít hledání text.

• [Tematické vyhledávání](#)

• [Přihlásit se do nového účtu](#)

Kontakt e-mail: theses@plagiat.cz

Úschevna
• [Vyhledání souboje do Úschevny osobit](#)

Národní registr je financován z Centralizovaného rozvojového projektu MSM1 C132009 "Národní registr vOIP a systém na odhalování plagiatů a je realizován pod vedením Masarykovy univerzity.

Vnitřní služby informačního systému
Vnitřní služby informačního systému. Vnitřní služby informačního systému. Vnitřní služby informačního systému.

Nedostatky súčasného stavu

- Nedokonalosť prevodu súborov na text
- Jazyková analýza textu
- Jedno číslo pre celkovú podobnosť daného dokumentu
- Heuristika štýlu daného textu
- Konkrétne vyznačení podobností medzi viacerými dokumentmi
- Globálne vyhľadávanie je závislé na viacerých súborových DB
- Korektnosť komunikácie medzi systémami
 - Informácia o nedostupnosti súborov pre stiahnutie
 - Chybný prevod na textový dokument
 - Poskytnutie výslednej informácie o podobnosti dokumentu
 - Informácia o zmene percenta podobnosti dokumentu

Antiplagiátorský systém dnes

- 36 zapojených škol do Theses.cz
- 30 zapojených inštitúcií do Odevzdej.cz
- 15 inštitúcií v Repozitar.cz
- Viac ako 250 tis. záverečných prác a 2,6 miliónov dokumentov

Ďakujem za vašu pozornosť!

stancik@fi.muni.cz