



MASARYKOVA UNIVERZITA



# Theses.cz

**7. října 2008, Systémy pro zpřístupňování eVŠKP 2008**

Miroslav Křipač

Michal Brandejs, Jitka Brandejsová, Jan Kasprzak, Martin Stančík

Masarykova univerzita

Národní registr VŠKP a systém na odhalování plagiátů

## Obsah

- ⇒ Mezinárodní rozměr
- ⇒ Design
- ⇒ Implementace aplikací pro vkládání dat přímo studenty
- ⇒ Vyhledávání
- ⇒ Podpora formátu
- ⇒ Nový hardware
- ⇒ Výstupy vyhledávání plagiátů
- ⇒ Statistiky

# Co je nového?

## 1. Projekt přesáhl rozměry národního projektu, aktuální počet škol není konečný:

---

- **Akademie múzických umění v Praze**
  - **Česká zemědělská univerzita v Praze**
  - **Janáčkova akademie múzických umění v Brně**
  - **Jihočeská univerzita v Českých Budějovicích**
  - **Masarykova univerzita**
  - **Ostravská univerzita v Ostravě**
  - **Slezská univerzita v Opavě**
  - **Univerzita Hradec Králové**
  - **Univerzita Jana Evangelisty Purkyně v Ústí nad Labem**
  - **Univerzita Palackého v Olomouci**
  - **Univerzita Tomáše Bati ve Zlíně**
  - **Vysoká škola báňská Technická univerzita Ostrava**
  - **Vysoká škola ekonomická v Praze**
  - **Vysoká škola polytechnická Jihlava**
  - **Vysoká škola technická a ekonomická v Českých Budějovicích**
  - **Vysoká škola uměleckoprůmyslová v Praze**
  - **Západočeská univerzita v Plzni**
- 
- **Vysoká škola finanční a správní**
  - **Brno International Business School**
- 

➤ **Ekonomická univerzita v Bratislave**

# Co je nového?

## 2. Theses.cz se převlékl:

Vysokoškolské kvalifikační práce - Microsoft Internet Explorer

Soubor Úpravy Zobrazit Oblíbené Nástroje Nápověda

Zpět Hledat Oblíbené

Adresa http://theses.cz/ Přejít Odkazy

Google mediální obraz Go Bookmarks 9 blocked Check AutoLink AutoFill Send to mediální obraz Settings

VYSOKOŠKOLSKÉ KVALIFIKAČNÍ PRÁCE

## Vysokoškolské kvalifikační práce

- Theses

Theses.cz je systém pro odhalování plagiátů mezi závěrečnými pracemi. Slouží vysokým školám a univerzitám (nejen v ČR) jako národní registr závěrečných prací (informací o pracích – název, autor, ...) a jako úložiště prací pro vyhledávání plagiátů. Systém umožňuje zástupcům zapojených škol vkládat práce a vyhledávat mezi nimi plagiáty.

Veřejnosti jsou zpřístupňovány záznamy o práci, příp. plné texty (dle rozhodnutí školy), a vyhledávání mezi nimi. Systém nabízí další služby, funkce a aplikace a bude rozvíjen dle potřeby.

### Vyhledávání ve veřejné části registru


Toto není Google! Čtěte, jak zadat hledaný text.

- Tematické vyhledávání

Přihlásit se do neveřejné části

Kontakní e-mail: [theses@fm.uni.cz](mailto:theses@fm.uni.cz)

Úschovna



# Co je nového?

## 2. Theses.cz se převlékl:

Vysokoškolské kvalifikační práce - Microsoft Internet Explorer

Soubor Úpravy Zobrazit Oblíbené Nástroje Nápověda

Zpět Hledat Oblíbené

Adresa http://www.theses.cz/ Přejít Odkazy >>

Google 5P marketingový mix Go Bookmarks 11 blocked Check AutoLink AutoFill Send to 5P marketingový mix Settings

**VYSOKOŠKOLSKÉ KVALIFIKAČNÍ PRÁCE**  
**Vysokoškolské kvalifikační práce**

- [Theses](#)

Theses.cz je systém pro odhalování plagiátů mezi závěrečnými pracemi. Slouží vysokým školám a univerzitám (nejen v ČR) jako národní registr závěrečných prací (informací o pracích – název, autor, ...) a jako úložiště prací pro vyhledávání plagiátů. Systém umožňuje zástupcům zapojených škol vkládat práce a vyhledávat mezi nimi plagiáty.

Veřejnosti jsou zpřístupňovány záznamy o práci, příp. plné texty (dle rozhodnutí školy), a vyhledávání mezi nimi. Systém nabízí další služby, funkce a aplikace a bude rozvíjen dle potřeby.

**Vyhledávání ve veřejné části registru**

Toto není Google! Čtěte, jak zadat hledaný text.

- [Tematické vyhledávání](#)

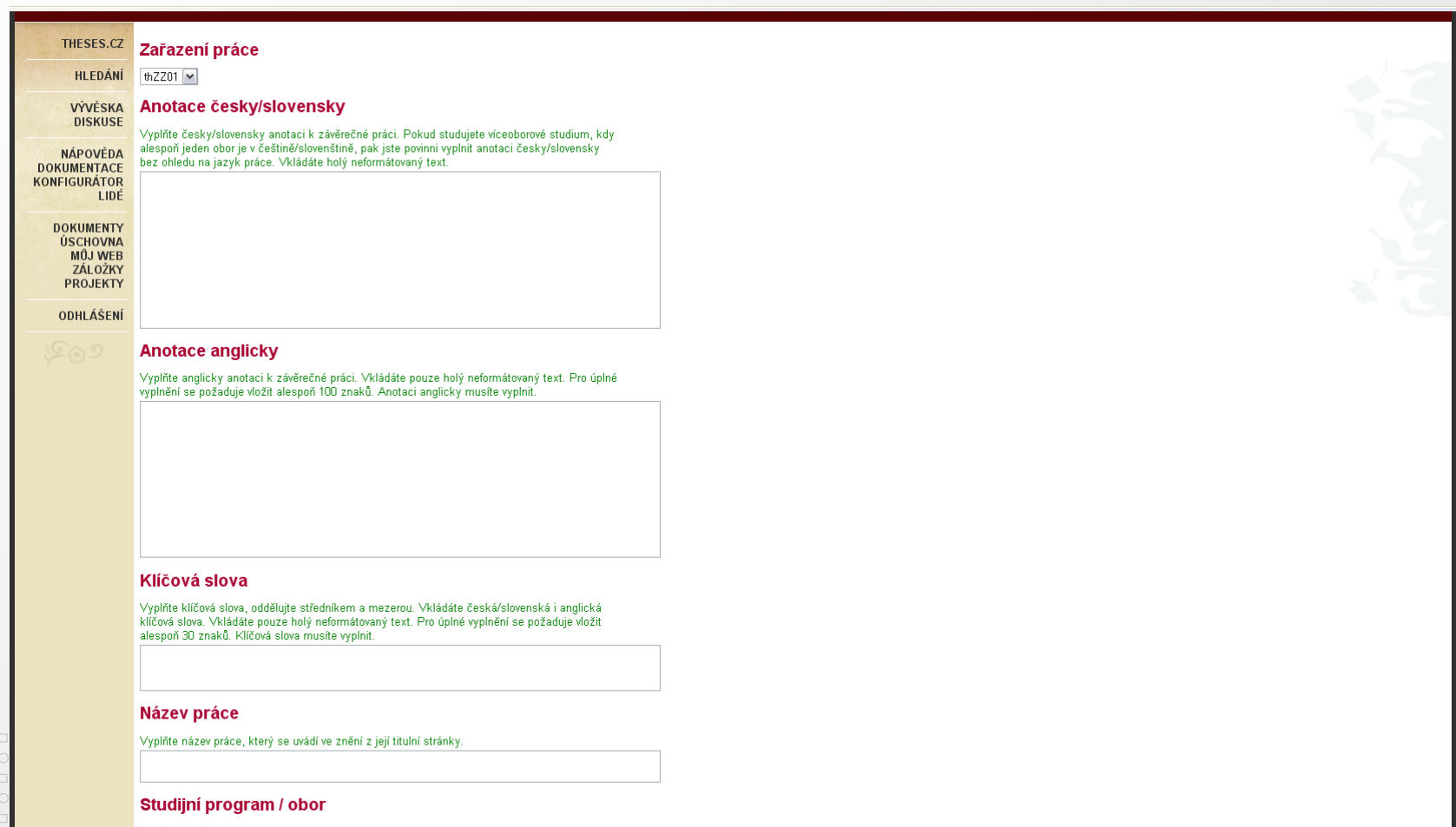
[Přihlásit se do neveřejné části](#)

Kontakní e-mail: [theses@fi.muni.cz](mailto:theses@fi.muni.cz)

**Úschovna**

# Co je nového?

## 3. Vkládání prací samotnými studenty implementováno.



THESES.CZ

**Zařazení práce**

HLEDÁNÍ

VÝVĚSKA DISKUSE

**Anotace česky/slovensky**

Vypíšte česky/slovensky anotaci k závěrečné práci. Pokud studujete víceoborové studium, kdy alespoň jeden obor je v češtině/slovenštině, pak jste povinni vyplnit anotaci česky/slovensky bez ohledu na jazyk práce. Vkládáte holý neformátovaný text.

NÁPOVĚDA DOKUMENTACE KONFIGURÁTOR LIDÉ

DOKUMENTY ÚSCHOVNA MŮJ WEB ZÁLOŽKY PROJEKTY

ODHLÁŠENÍ

**Anotace anglicky**

Vypíšte anglicky anotaci k závěrečné práci. Vkládáte pouze holý neformátovaný text. Pro úplné vyplnění se požaduje vložit alespoň 100 znaků. Anotaci anglicky musíte vyplnit.

**Klíčová slova**

Vypíšte klíčová slova, odděluje středníkem a mezerou. Vkládáte česká/slovenská i anglická klíčová slova. Vkládáte pouze holý neformátovaný text. Pro úplné vyplnění se požaduje vložit alespoň 30 znaků. Klíčová slova musíte vyplnit.

**Název práce**

Vypíšte název práce, který se uvádí ve znění z její titulní stránky.

**Studijní program / obor**

# Co je nového?

## 4. Nové vyhledávání

- **Nový způsob jádra vyhledávacího algoritmu:**
  - **Náhrada za původní Oracle Text.**
  - **Plně distribuované prostředí.**
  
- **Vyhledávání s ohledem na přístupová práva**
  - **Nejedná se o internetové vyhledávání.**
  
- **Nový způsob zadávání dotazů:**
  - **Jednoduché textové pole (ve stylu „Google“).**
  - **Pokročilé vyhledávání (formulář omezujících podmínek)**
  
- **Právě ve vývoji...**

## Co je nového?

### 5. Nově podporovaný formát pro import metadat do Theses.cz.

Aktuálně podporované formáty metadat:

#### **Formát theses.cz, verze 1.0**

**Formát vyvinutý a podporovaný v rámci projektu theses.cz.**

#### **☒ Formát evskp.cz, verze 1.1**

**Formát vyvinutý Pracovní skupinou Odborné komise pro otázky elektronického zpřístupňování vysokoškolských kvalifikačních prací.**

Aktuálně podporované protokoly předávání metadat:

#### **☒ Ruční HTTPs (webový formulář).**

#### **☒ Automatické HTTPs (metadata vkládá přímo systém školy).**

#### **☒ OAI-PMH (theses.cz stahuje přímo ze systému školy).**

**Plné texty stahovány automaticky z lokálního uložště v libovolném formátu.**



## Co je nového?

### 6. Nový hardware:

**Pro projekt Theses.cz byl zvolen výkonný a spolehlivý hardware:**

#### ☒ **Servery SGI® Altix® XE250:**

- ☒ **Databázové servery.**

- ☒ **Server pro koordinaci distribuovaných algoritmů hledání podobností.**

#### ☒ **Disková pole SGI® InfiniteStorage 220**

- ☒ **Uložení metadat i plných textů prací.**

- ☒ **Geograficky oddělené zálohování dat.**

#### ☒ **Servery O&G**

- ☒ **Cluster pro distribuované výpočty.**

- ☒ **Cluster pro distribuované uložení dat.**

**Předpoklad využití hardware i pro další rozvoj a projekt Odevzdej.cz**

## Co je nového?

### 7. Školy zkoumají výsledky vyhledávání plagiátů:

Víme, že:

- ☒ **Byly nalezeny podobné soubory mezi školami (většinou podobnými s pracemi z MU a zdaleka ne jen se ZP). Nešlo o plagiáty, v jednom případě ukazovaly výsledky na kompilát.**

Nevíme, že:

- ☒ **Byly nalezeny podobné soubory mezi školami, a to ani mezi školami a MU. Jsme správci a není naším cílem sledovat konkrétní kauzy, nenapíše-li nám někdo.**

Cílem projektu není:

- ☒ **„Uspokojit“ média a veřejnost nálezy a exemplárními případy, tj. ani statistikami.**

Cílem projektu naopak je:

- ☒ **Případy opisování odhalit dříve než veřejnost, zamezit ostudě, ušetřit si následnou práci spojenou s veřejnou kauzou a prevence.**

# Způsoby vyhledávání plagiátů

## ➤ Přímé vyhledání

- **Volba funkce pomocí ikony u zvolených konkrétních souborů.**
- **Příklad použití: učitel má podezření na konkrétní plagiát.**

## ➤ Globální vyhledání

- **Zvláštní aplikace pro výpis všech nalezených podobností.**
- **Příklad použití: správce periodicky kontroluje nové soubory na plagiáty**

## Nastala doba vkládání ostrých dat - statistiky

Škola	Počet
Masarykova univerzita	12234
Slezská univerzita v Opavě	591
Vysoká škola ekonomická v Praze	6817
Vysoká škola finanční a správní, o.p.s.	2338
Vysoká škola polytechnická Jihlava	112
<b>Celkem prací:</b>	<b>22092</b>

Stav k 28. září 2008.

## Zveřejňování

## Plné texty

plné texty předávané Theses  
zveřejňují autentizovaným  
uživatelům Theses : 3 VŠ

plné texty předávané Theses  
se nezveřejňují: 4 VŠ

plné texty předávané Theses  
zveřejňuje světu: 2 VŠ

plné texty předávané  
Theses zveřejňují  
svým  
zaměstnancům: 4 VŠ

plné texty předávané Theses  
zveřejňují svým zaměstnancům  
a studentům: 5 VŠ

metadata  
předávaná  
Theses  
zveřejňuje  
autentizovaný  
m uživatelům  
Theses: 1 VŠ

## Metadata

metadata předávaná Theses  
zveřejňuje svým zaměstnancům: 1 VŠ

metadata předávaná Theses zveřejňuje světu: 16 VŠ

**Stav k 1. 10. 2008.**

## Unikátnost?

- ❏ Ve světě se systematicky nesbírají závěrečné práce z více škol proto, aby se v nich vyhledávaly plagiáty.
- ❏ Nikdy se nespojilo tolik vysokých škol v podobném projektu.
- ❏ Google nemůže vyhledávat text. řetězce proti konkrétní práci, protože nemá přístup k pracím (kromě z několika škol).
- ❏ Systém pracuje přesně, nikoli s heuristikou.
- ❏ Systém zvládá velké množství dat s krátkými odezvami.
- ❏ Systém má implementované rozhraní (což je z pohledu časové náročnosti na vývoj mnohem náročnější než implementace bez uživatelského rozhraní).

# Prostor pro Vaše otázky

Děkuji Vám za pozornost!

[theses@fi.muni.cz](mailto:theses@fi.muni.cz)

<http://theses.cz/>

<http://is.muni.cz/clanky/>