

*Pomáháme vám využívat
vaše informace*

Automatická indexace šedé literatury hesly Polytematického
strukturovaného hesláře

OBSAH

- NUŠL
- PSH
- Automatická indexace - co, k čemu a proč
- Použité principy
- Problémy
- Výsledky implementace
- Další možný rozvoj

NUŠL

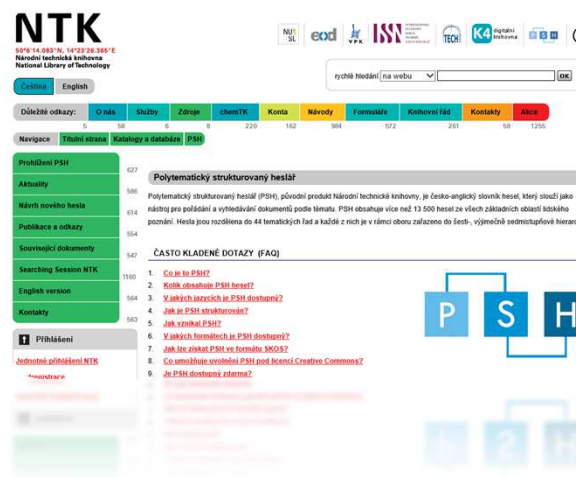
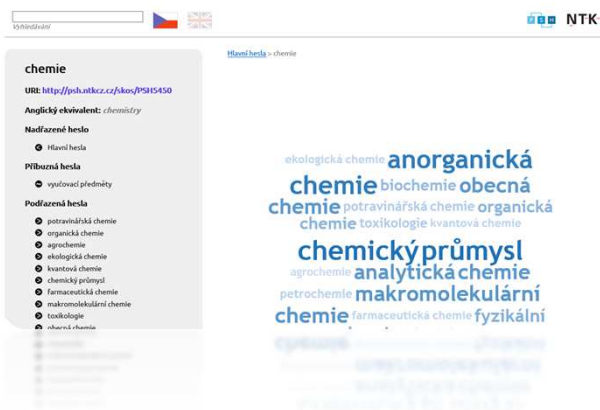
- Národní Úložiště Šedé Literatury
- Služba NTK
- Centralizovaný přístup k Šedé Literatuře založený na centrálním vyhledávání
- Oblast vědy, výzkumu a vzdělávání
- INVENIO, MS FASTSearch



PSH

- Polytematický Strukturovaný Heslář
 - Česko-Anglický slovník
 - Verze 3.0 = 13.971 hesel (21.10.2013)
 - ve 44 tematických řadách
 - 6-7 stupňů hierarchie

- Produkt NTK



AUTOMATICKÁ INDEXACE

- Přiřazení hesel PSH k záznamům v systému INVENIO na základě obsahu vybraných polí

(těm, která nemají hesla přiřazena intelektuálně/manuálně)

VÝCHOZÍ STAV A PŘEDCHOZÍ ZKUŠENOSTI

- Záznamy určené k indexaci v systému INVENIO
- Harvestované záznamy z různých institucí
- Odlišná popisná pole i úroveň popisu

- Předchozí zkušenosti – projekt LCC pro STK [NTK]

ANALÝZA

- Softwarové nástroje
 - Open source řešení
 - Využití Lucene SOLR
- Doporučení pro využití PSH
 - celý strom / několik úrovní
- Vytipování polí využitelných při indexování
- Další rozvoj systému
 - gramatický slovník pro PSH
 - stop words

PRINCIP PŘIŘAZOVÁNÍ HESEL

- Normalizace záznamu
- Porovnávání vůči PSH
 - Názvové údaje
 - Abstrakt
 - Klíčová slova
 - Konference
 - Instituce

PRINCIP PŘIŘAZOVÁNÍ HESEL

- Hodnocení = váha pro jednotlivé parametry
 - Pole, ve kterém se dané slovo nachází ve zdrojovém dokumentu
 - Pole, ve kterém je dané slovo nalezeno v PSH
 - Příslušnost k oboru (proximity)
 - Plná shoda dotazu (exact)
 - Více slov dotazu (multiple)
 - Úroveň hesla ve stromové struktuře PSH (level)
 - Heslo přidělené automaticky vyhledávacím systémem

UKÁZKA PROSTŘEDÍ - PŘIŘAZENÍ

PSH Automatické přidělování PSH hesel

prachu statistickými metodami

num docs: 19739

Metodika identifikace zdrojů pevných částic a pouličního prachu statistickými metodami
 Methodology for identification of sources of aerosol particles and street dust using statistical methods

Morfologie, chemická a toxikologická charakteristika pouličního prachu a suspendovaných prachových částic, včetně určení zdrojů původu
 Morphology, chemical and toxicological characterisation of road dust and particulate matter including sources apportionment
 Závěrečná zpráva - souhrn

Identifikace zdrojů resuspendovaných pevných částic statistickými metodami
 Identification of sources of resuspended aerosol particles using statistic methods

Morfologie, chemická a toxikologická charakteristika pouličního prachu a suspendovaných prachových částic, včetně určení zdrojů původu

Inventarizace emisí - vědecké a technické aspekty vstupu do Evropské

Metodika identifikace zdrojů pevných částic a pouličního prachu s
methods
 Metodika popisuje postup pro kvantifikaci podílů jednotlivých zdrojů na znečištění statistických metod a receptorového modelování. Methodology describes a process of measurement of chemical composition of particulate matter and street
subjects: (pollution sources) (particulate matter) (receptor modeling)

prachové částice (7455.5) ?
 /geofyzika/aerokosmické metody/průzkum okolí Země/prachové částice
prach in csAltLabel_hl
dust particles in enPrefLabel_hl
 prachové částice in csPrefLabel_hl
 znečištění vzduchu (7439.1) ?
 /biologie/životní prostředí/znečištění životního prostředí/znečištění vzduchu
znečištění ovzduší in csAltLabel_hl
air pollution in enPrefLabel_hl
znečištění vzduchu in csPrefLabel_hl
 statistické metody (7381.4) ?
 /věda a technika/výzkum/statistické metody
statistical methods in enPrefLabel_hl
 statistické metody in csPrefLabel_hl
 hmota (4194.3) ?
 /fyzika/obecná fyzika/fyzikální veličiny/hmota
matter in enPrefLabel_hl
 znečištění (4146.8) ?
 /obecnosti/znečištění
pollution in enPrefLabel_hl
 znečištění in csPrefLabel_hl

UKÁZKA PROSTŘEDÍ - PŘIŘAZENÍ

Automatické přidělování PSH hesel

pském srovnání – léto 2013

num docs: 18041

Hodnocení členství v EU a postoj k euru v středoevropském srovnání – léto 2013 Evaluation of EU Membership and Attitude to Euro in Central European Comparison - Summer 2013

Hodnocení ekonomické situace a materiálních životních podmínek v středoevropském srovnání – léto 2013 Evaluation of Economic Situation and Material Living Conditions in Central European Comparison - Summer 2013

Hodnocení dopadů členství v EU – duben 2013 Impacts of EU Membership on the Czech Republic - April 2013

Hodnocení ekonomické situace a materiálních životních podmínek v středoevropském srovnání – léto 2012 Evaluation of Economic Situation and Material Living Conditions in Central European Comparison - Summer 2012

Hodnocení ekonomické situace a materiálních životních podmínek v středoevropském srovnání – léto 2010 Evaluation of Economic Situation and Material Living Conditions in Central European Comparison - Summer 2010

Hodnocení ekonomické situace a materiálních životních podmínek v středoevropském srovnání – léto 2011 Evaluation of Economic Situation and Material Living Conditions in Central European Comparison - Summer 2011

Hodnocení členství v EU a postoj k euru v středoevropském srovnání – léto 2013 Evaluation of EU Membership and Attitude to Euro in Central European Comparison - Summer 2013

V první polovině září 2013 CVVM zjišťovalo, jak občané hodnotí členství ČR v Evropské unii a jeho prospěšnost a jak se dívají na přijetí eura jako měny v průběhu června a července realizovány též v Polsku, Maďarsku a na Slovensku. In the first half of September 2013 Public Opinion Research Centre investigated the attitude of citizens towards EU membership and its benefits and how they look on the adoption of the euro as the currency of the country.

subjects: (European Union) (euro)

[měnový systém evropský](#) (14872.2) ?
/ekonomické vědy/ekonomie/makroekonomie/ekonomická integrace/měnový systém evropský
evropský měnový systém in csAltLabel_hl
měnová jednotka **evropská** in csAltLabel_hl
European Currency Unit in enAltLabel_hl
euro in enAltLabel_hl
European monetary system in enPrefLabel_hl
měnový systém **evropský** in csPrefLabel_hl
 [Evropská unie](#) (13584.3) ?
/právo/mezinárodní právo/evropská integrace/Evropská unie
EU in csAltLabel_hl
European union in enPrefLabel_hl
Evropská unie in csPrefLabel_hl
 [hospodářské unie](#) (9954.6) ?
/ekonomické vědy/ekonomie/makroekonomie/ekonomická integrace/hospodářské unie
European Union in enAltLabel_hl
hospodářské **unie** in csPrefLabel_hl
 [veřejné mínění](#) (2491.7) ?
/politologie/politická kultura/veřejnost/veřejné mínění
public opinion in enPrefLabel_hl
 [měna](#) (2162.5) ?
/ekonomické vědy/ekonomie/makroekonomie/měna
currency in enPrefLabel_hl
měna in csPrefLabel_hl
 [postoj](#) (2078.2) ?
/psychologie/psychologie osobnosti/vlastnosti osobnosti/postoj
attitude in enPrefLabel_hl
postoj in csPrefLabel_hl
 [hodnocení](#) (2068.2) ?
/věda a technika/metodologie vědy/hodnocení
evaluation in enPrefLabel_hl
hodnocení in csPrefLabel_hl

UKÁZKA PROSTŘEDÍ – NASTAVENÍ

Automatické přidělování PSH hesel nastavení procesy

fyzika hledat metod

Nastavení

přidělování

title	anotace	subject	konference	csPref	enPref	csAlt	enAlt	level	proximity	exact
4.3	1.0	8.0	1.1	6.4	6.3	0.1	0.1	10.0	8.0	40.0

max keys

stop words + -

/fyzika/jaderna/fyzika/atomove_jadro/jadern_e_reakce
jaderné reakce in csPrefLabel_hl
 [jaderné inženýrství \(1175.4\) ?](#)

PROBLÉMY

- Přesnost přiřazování hesel
- Hesla PSH jsou v hesláři pouze jednou > přidělení hesel z jiných větví (obsahově správných)
- Kvalita dat > různá úroveň záznamů neumožňuje nastavit prioritu jednotlivých hodnocených polí

VÝSLEDKY K 18.10.2013

- INVENIO
 - 107.469 záznamů
 - 24.587 ručně zadaných
 - 82.882 neindexovaných

- AUTOMATICKÁ INDEXACE
 - 107.286 automatická
 - 183 nezpracováno
 - Instituční heslo max 1455



DALŠÍ MOŽNÁ VYUŽITÍ ...

- Webová služba pro přiřazování hesel PSH
- Možnost využít proces i pro další hesláře
- Možnost vložení komponenty pro doporučení hesel do web formuláře jiných aplikací

PROJEKTOVÝ TÝM

- NTK
 - Bohdana Frantíková
 - Jana Slouková
 - Petra Pejšová

- INCAD
 - Alberto Hernandez
 - Tomáš Dočkal

- <http://code.google.com/p/pshassign/>



DOTAZY ?

- Děkuji za pozornost

INCAD

NU
ŠL