národní
úložiště
šedé
literatury

Uživatelské rozhraní Národního úložiště šedé literatury 2012

Pavlová, Lucie
2012

Dostupný z http://www.nusl.cz/ntk/nusl-126840

Dílo je chráněno podle autorského zákona č. 121/2000 Sb.

Tento dokument byl stažen z Národního úložiště šedé literatury (NUŠL).

Datum stažení: 26.04.2024

Další dokumenty můžete najít prostřednictvím vyhledávacího rozhraní nusl.cz .

We help you to get the most from your information

USER INTERFACE – NATIONAL REPOSITORY OF GREY LITERATURE

[2012]

MS FAST SEARCH FOR INTERNET SITES

5th year of the Seminar on Providing Access to Grey Literature – October 2012

Mgr. Lucie Pavlová , INCAD

AIM OF NLGR UI

PROVIDE NATIONAL REPOSITORY OF GREY LITERATURE - **NRGL** - WITH STABLE **SINGLE POINT OF ACCESS INTERFACE FOR SEARCH, NAVIGATION AND INFORMATION SHARING**, WHICH INTEGRATES HETEROGENOUS DATA SOURCES AND DOCUMENT FORMATS

CHANGES SINCE 2011

UI

- Query completion
- Multi-level navigation: Data sources
- New navigation: Format
- New UI for mobile devices

SYSTEM

- Query processing optimization
- New fields indexation
- Analysis of duplications

DATA (October 2012)

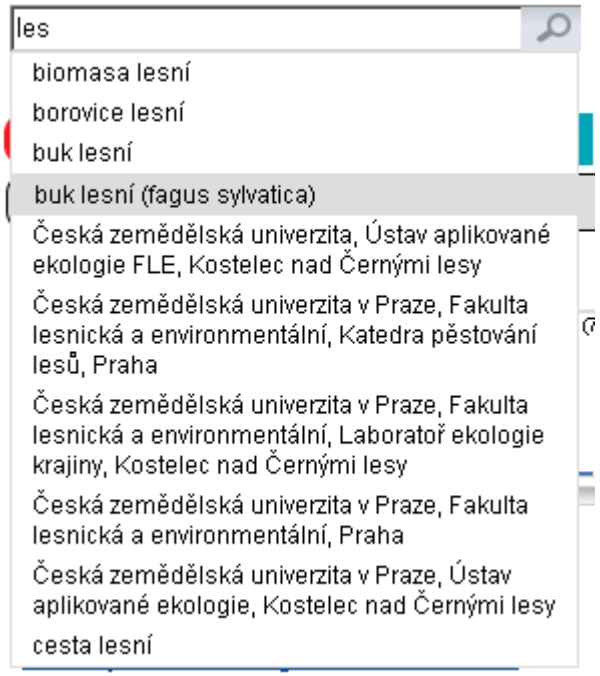
- Number of records: 203 200
- Number of data sources: ??

QUERY COMPLETION

Based on:

- Keywords
- People

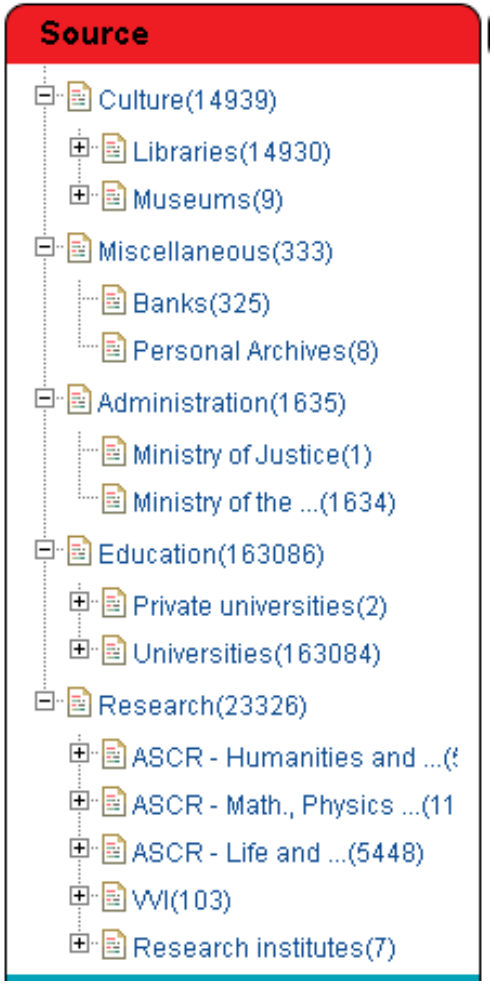
Min.3 letters
Anywhere in field text



NAVIGATION

Adapted navigation: DATA SOURCES

- More than 90 data sources
- 3 level tree

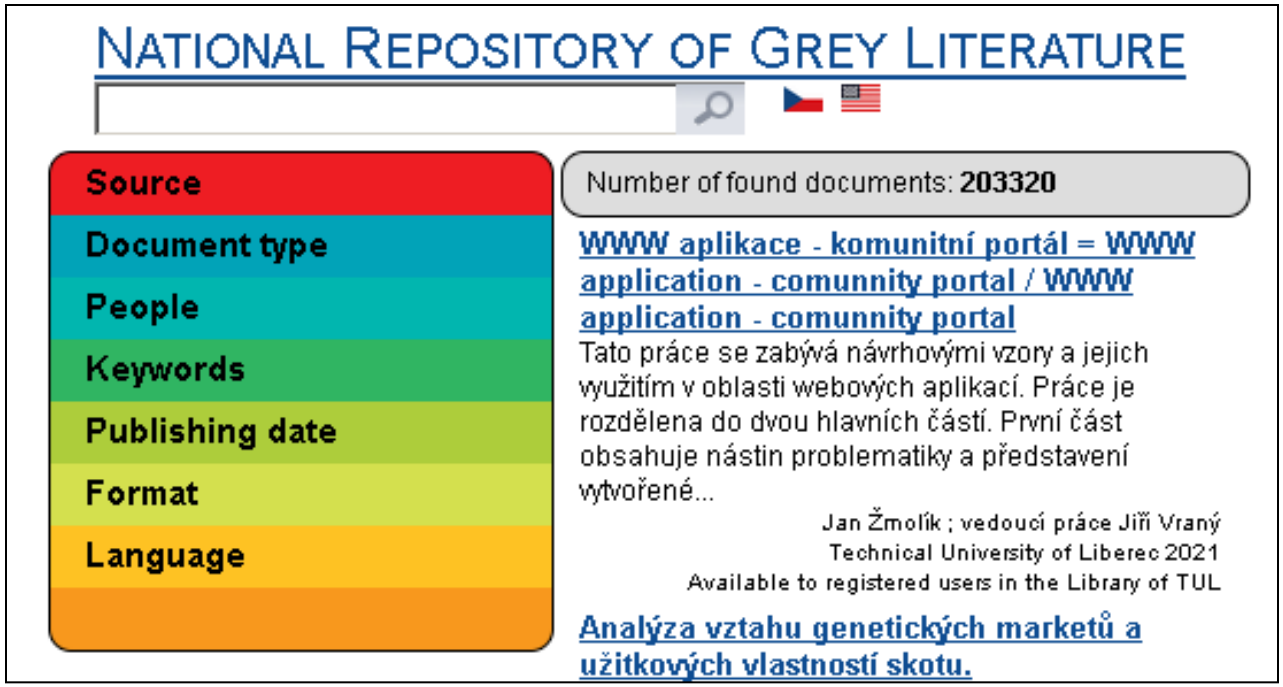


New navigation: FORMAT

- Text
- Video
- Image

UI FOR MOBILE DEVIDES

- UI simplification (no logo,…)
- Closed navigation
- One column of results



QUERY PROCESSING OPTIMIZATION

QUERY

- Elimination of diacritic
- Stemming

RANGING OF RESULTS

- Based on the location and frequency of words
- Document parts by weight
    1. Main title
    2. Person
    3. Keywords and other titles
    4. Document body

NEW FIELDS IN INDEX

New fields selected

Conference name
Project number
Project supplier
ISBN
ISSN

Modification of conversion tables

Data ssources
Document type (*Program to detect unmapped types of document*)

DUPLICATE DETECTOR

- Intended for the analysis of duplications across sources
  x duplications are mainly in one source

- **1. Duplications within the source**
  - List of duplicate records
  - Correction of data within the source
  - Reindexing

- **2. Duplications across sources**
  - grouping into one search result

DUPLICATE DETECTOR

Duplicate estimation based on:

1. All metadata outside of unique identifiers

2. Subset of metadata

3. With "optimized" fields (length restriction and removed diacritics)

   AUTOR_20,
   TITLE_50,
   TITLE2_50,
   TITLE3_50,
   DATA_20,
   TITLE4_50,
   ZDROJORIG_50
   SHA

DUPLICATE DETECTOR - Parameters

What will be compared? In what sources?

Duplicates analyzer

Select fields for duplicate detection	Select datasource
☐ all	invenio.nusl.cz
☐ AUTOR	narodohospodarsky_ustav
☐ AUTOR_20	orientalni_ustav
☐ CISLOPROJEKTU	parazitologicky_ustav
☐ DATA	prague_college
☐ DATA_20	psychologicky_ustav
☐ DOCDATETIME	slovansky_ustav
☐ ISBN	sociologicky_ustav
☐ ISSN	univerzita_karlova_v_praze
☐ NAZEVKONFERENCE	ustav_analyticke_chemie
☐ POSKYTOVATELPROJEKTU	ustav_anorganicke_chemie
☑ SHA	ustav_archeologicke_pamatkove_pece_severozapadnich
☐ TITLE	ustav_biologie_obratlovcu
☐ TITLE2_50	ustav_chemickych_procesu
☐ TITLE3_50	ustav_dejin_umeni
☐ TITLE4_50	ustav_experimentalni_botaniky
☐ TITLE_50	ustav_experimentalni_mediciny
☐ ZDROJORIG_50	ustav_fotoniky_a_elektroniky
	ustav_fyzikalni_chemie_j_heyrovskeho
	ustav_fyziky_atmosfery
	ustav_fyziky_materialu

Minimal 2 occurrence to recognize duplicate

Detect duplicates

DUPLICATE DETECTOR – Query results

Suggested duplicates

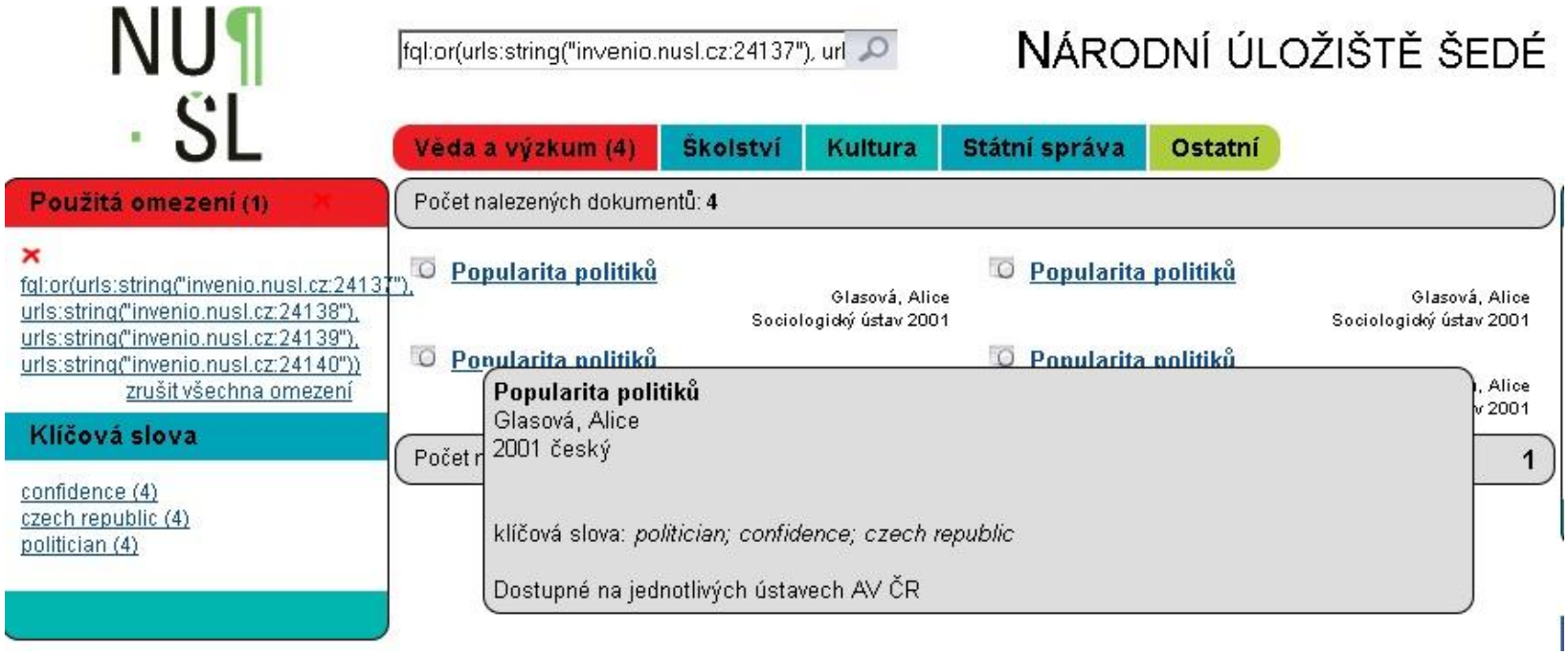
Total records in duplicate DB: 1502

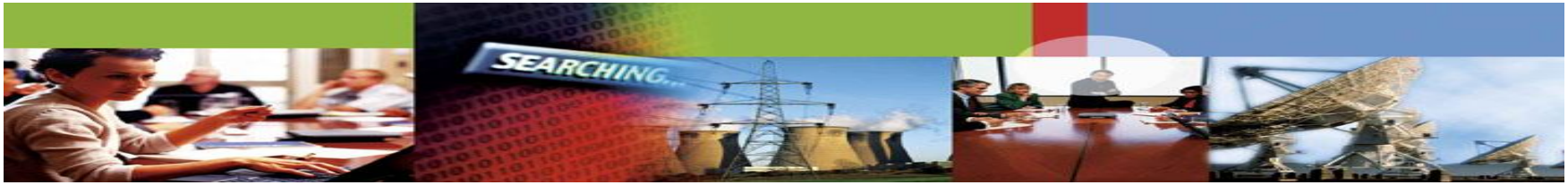
Using fields: SHA
Datasource: where url like '%invenio.nusl.cz%'
Source: and zdrojorig_50 = 'sociologicky_ustav'
Minimale occurrence to be duplicate: 2

Count	Date	Source	Title	URLs
4	01.01.2001	sociologicky_ustav	Popularita politiků...	show on www.nusl.cz, invenio.nusl.cz:24137, invenio.nusl.cz:24138, invenio.nusl.cz:24139, invenio.nusl.cz:24140
4	01.01.2002	sociologicky_ustav	Spokojenost s politickou situací...	show on www.nusl.cz, invenio.nusl.cz:24109, invenio.nusl.cz:24111, invenio.nusl.cz:24112, invenio.nusl.cz:24113
4	01.01.2001	sociologicky_ustav	Spokojenost s celkovou politickou situací...	show on www.nusl.cz, invenio.nusl.cz:24132, invenio.nusl.cz:24134, invenio.nusl.cz:24135, invenio.nusl.cz:24136
2	01.01.2003	sociologicky_ustav	Hodnocení činnosti vlády a parlamentní opozice v Č...	show on www.nusl.cz, invenio.nusl.cz:24245, invenio.nusl.cz:24246
2	01.01.2001	sociologicky_ustav	Důvěra ústavním institucím...	show on www.nusl.cz, invenio.nusl.cz:24123, invenio.nusl.cz:24124
2	01.01.2002	sociologicky_ustav	Důvěra ústavním institucím...	show on www.nusl.cz, invenio.nusl.cz:24025, invenio.nusl.cz:24117
2	01.01.2001	sociologicky_ustav	Občané k ekologii...	show on www.nusl.cz, invenio.nusl.cz:24089, invenio.nusl.cz:24090
2	01.01.2003	sociologicky_ustav	Spokojenost s politickou situací...	show on www.nusl.cz, invenio.nusl.cz:24281, invenio.nusl.cz:24284

DUPLICATE DETECTOR – Reviewing records

- Duplicate
- Failure of metadata
- Change the search parameters



…

Thank you for your attention