

Demonstrating a Framework for KOS-based Recommendations Systems

Philipp Mayr, Thomas Lüke, Philipp Schaer

philipp.mayr@gesis.org

NKOS workshop @TPDL2013

2013-09-27

Background: Projects IRM I and IRM II

- DFG-funded (2009-2013)
- IRM = Information Retrieval Mehrwertdienste (value-added IR services)
- Goal: Implementation and evaluation of value-added IR services for digital library systems
- Main idea: Applying scholarly (science) models for IR
 - **Co-occurrence analysis** of controlled vocabularies (thesauri)
 - **Bibliometric analysis** of core journals (Bradford's law)
 - **Centrality** in author networks (betweenness)
- In IRM we concentrated on the basic evaluation
- In IRM2 we concentrate on the **implementation of reusable (web) services**

<http://www.gesis.org/en/research/external-funding-projects/archive/irm/>

Motivation

Query:

Total hits: 315726

1. [Socialism \(1987\)](#)
book (0-335-15388-7)
2. [For socialism \(1978\)](#)
book (0-914386-11-5)
3. [Socialism \(1895\)](#)
book
4. [Socialism \(1972\)](#)
book (0-8415-0141-6)
5. [Socialism \(1975\)](#)
book

social
social Advertising
social Demand Approach
social Sponsoring

Term Completion

Search-Term-
Recommendation

social
social Advertising
social Demand Approach
social Sponsoring
Alternative Search Terms:
Social Security
Social Work
Welfare State
Social Workers
Cultural Capital
History of Sociology

[Social Sciences](#) [Social Services](#) [Social Support](#)
[Social Work Education](#)
[Social Problems](#)
[Sociological Theory](#)
[Social Security](#) [Social Policy](#)
[Social Workers](#)
[Social Environment](#) [Social Factors](#) [Social Democracy](#) [Social Relations](#)
[Social Competence](#) [Social Development](#) [Social Order](#) [Social Change](#)
[Social Constructionism](#)
[Social Work](#)

see Hienert et al., 2011

Why custom KOS-based recommenders

- The more specific the dataset, the more specific the recommendations
- Customized for your specific information need (see Improving Retrieval Results with Discipline-specific Query Expansion, TPDFL 2012, Lüke et. Al, <http://arxiv.org/abs/1206.2126>)

Overview: recommendation in DL

Site	TS	QS	OAI-PMH
ACM Digital Library	no	no	no
Google Scholar	no	no	no
MIT Repository	no	no	yes
arXiv	no	no	yes
PubMed	yes	no	yes
MS Academic Search	yes	yes	no

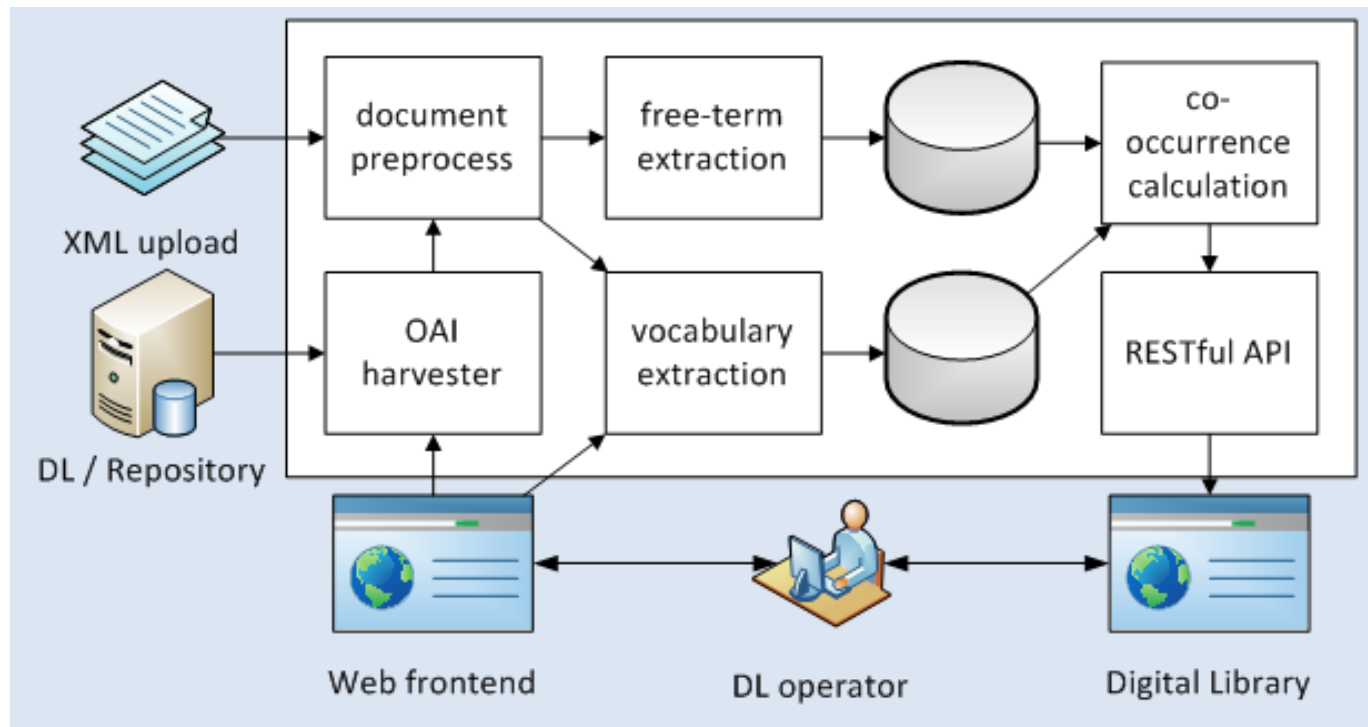
term suggestion (TS): try to add or replace single words or phrases

query suggestion (QS): often based on query log analysis (complete query strings)

IRSA

- Information Retrieval Service Assessment (IRSA) component based on **OAI-PMH harvested metadata**
- Calculating search term suggestions based on co-occurrence analysis.

IRSA: Workflow



Analysis

Title terms (free)

Classification / subject terms (controlled)

Abstract terms (free)

Co-occurrence analysis of free and controlled terms (using Jaccard, NWD, Dice, Cosine etc.)

```

header:
  identifier : oai:gesis.izsoz.de:19389
  datestamp : 2011-01-10T13:46:00Z
  setSpec : SSOAR

metadata:
  dc:
    identifier: http://nbn-resolving.org/urn:nbn:de:0168-ssoar-193894
    title: How can international donors promote transboundary water management?
    creator: Mostert, Erik
    creator: Deutsches Institut für Entwicklungspolitik gGmbH
    subject: Political science (320)
    subject: Life sciences, biology (570)
    subject: International Relations, International Politics, Development Policy (10505)
    subject: Ecology, Environment (20900)
    subject: Management; Afrika; Entwicklung; Entwicklungsland; Akteur; Wasser
    source: Born
    source: DIER Discussion Paper (1860-0441) 8/2005
    description: "This paper discusses how international donors can promote the development of transboundary water management. It assumes, first, that cooperation will take place whenever the major stakeholders consider cooperation to be a better option than non-cooperation. The perceptions and motivations of the stakeholders are therefore crucial. Secondly, this paper assumes that the major stakeholders are not 'states', but specific groups and individuals: individual politicians, sectoral government bureaucracies, regional and local governments, farmers, electricity companies, etc. Some of these may be involved in the international negotiations themselves, others may be needed to get international agreements ratified or implemented, and still others may be affected by transboundary water management but lack the means to exert any influence." (author's abstract)
    language: English
    rights: Deposit Licence - No Redistribution, No Modifications
    contributor: SSOAR - Social Science Open Access Repository
    date: 10.01.2011 13:46
  
```


Output

```
<searchtermrecommendations entryterm="money" limit="10" count="260" sort="confidence" proj
  <term name="fiscal policy " confidence="0.60726434" type="subject" vendor="test"/>
  <term name="economic policy " confidence="0.5725895" type="subject" vendor="test"/>
  <term name="monetary theory " confidence="0.55314887" type="subject" vendor="test"/>
  <term name="capitalism" confidence="0.54612887" type="subject" vendor="test"/>
  <term name="economy" confidence="0.5368112" type="subject" vendor="test"/>
  <term name="economic development" confidence="0.52294195" type="subject" vendor="test"
  <term name="income" confidence="0.5048042" type="subject" vendor="test"/>
  <term name="Federal Republic of Germany " confidence="0.5019167" type="subject" vendor
  <term name="power" confidence="0.5017163" type="subject" vendor="test"/>
  <term name="wage policy " confidence="0.4999472" type="subject" vendor="test"/>
</searchtermrecommendations>
```

Integration

gesis [▲]
sowiport

Home Search Topics Service

Overview About Sowiport Partners Imprint

SOWIPORT at a glance:

- > 7.444.163 Literature references
- > 55.518 Research projects
- > 9.053 Institutions
- > 6.300 Events
- > 5.094 Studies
- > 29 Topic feeds
- > 41 Research Special

Searching for social science information? You want to have it all in one place; fast and easily available?

Search in Sowiport

money

- money **in circulation**
- money **laundering**
- money **wage**

Alternative Search Terms:

- Economics
- Finances**
- Sports (marketing)
- Professional sports
- Sports
- Associations

Services of sowiport partners see ► Service

term cloud ▼

Kaufkraft Geldtransfer
Währungsunion Geldwirtschaft
Wirtschaftspolitik Tausch
Finanzierung EWWU
Geldpolitik Europäische
Zentralbank Verschuldung
Deutsche Bundesbank
Geldtheorie Währung
Ware Globalisierung
Währungspolitik Budget
Finanzpolitik

Demo

IRSA Toolkit

[Home](#) [Prototype](#) [Repositories](#)

Hello test ([show user details](#) | [logout](#))

Repository List

Name	Processing Error	Harvesting Successful	STRs Created	Author Centrality ready
testr	None	false	false	false
SSOAR	None	true	true	false
Solis	None	false	true	false
new repo	None	false	false	false
Pedocs	None	true	true	true
prototypeSTR	None	false	true	false

- Add a new repository

Add a new Repository

Name



<http://multiweb.gesis.org/irsa/>

Demo

IRSA Toolkit

[Home](#) [Prototype](#) [Repositories](#)

Hello test ([show user details](#) | [logout](#))

Repository SSOAR

- Add OAI address of the repository
- Add date restrictions


General Configuration

Name *

Repository Size 0 Byte in 18 files.

OAI Configuration

[Help](#)

OAI-PMH URL 

From Date (optional, Format: YYYY-MM-DD)

Until Date (optional, Format: YYYY-MM-DD)

Set (optional)

Harvesting Error

Demo

Processing

[Help](#)

Metrics:

Cosine	<input checked="" type="checkbox"/>	
Jaccard-LOG	<input checked="" type="checkbox"/>	
Normalized Web Distance	<input type="checkbox"/>	
Jaccard	<input type="checkbox"/>	
Dice	<input type="checkbox"/>	

DublinCore Metadata Fields:

Term	<input checked="" type="checkbox"/>	
Subject	<input checked="" type="checkbox"/>	
Source	<input type="checkbox"/>	
Creator	<input type="checkbox"/>	

- Select different recommender
- Define co-word analysis entities

Demo

IRSA Toolkit

Home Prototype Repositories Assessments Administration

Hello admin ([show user details](#) | [logout](#))

Administration

User Management

Job Management

Journal Management

Repositories

List of available
Controllers

Scheduled Jobs

Date Created	Owner	Repository	Type	Processing Error	Delete
25.09.2013 09:38:51 MESZ	test	SSOAR (54413)	str		running

- Status of the repository

Benchmark:
 SSOAR ~ 26k docs
 It took ~ 1h to harvest all docs
 It took ~ 20min to compute the recommenders

Limitations

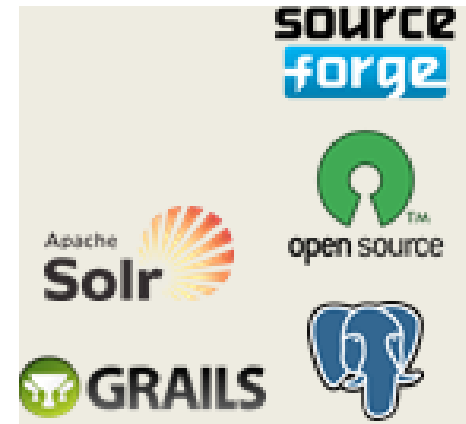
- Issues with OAI-harvested metadata
 - Wrong terms, typos and other ambiguous information (due to the Open-Access self-archiving policies of many repositories)
 - Mixed up classifications and subject terms in dc:subject
 - Disambiguation issues, abbreviations, etc.
 - No clear separation of subsets in OAI
- Huge datasets

Using IRSA

Check out and get an API key from

- <http://multiweb.gesis.org/irsa/IRMPrototype/>
- <https://sourceforge.net/projects/irsa/>

- Open source framework with build-in support for
 - Search term recommendation,
 - OAI harvesting, and Solr integration



References

- Lüke, T., Schaer, P., & Mayr, P. (2013). A framework for specific term recommendation systems. In Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval - SIGIR '13 (p. 1093). New York, New York, USA: ACM Press. doi:10.1145/2484028.2484207
- Mutschke, P., Mayr, P., Schaer, P., & Sure, Y. (2011). Science models as value-added services for scholarly information systems. *Scientometrics*, 89(1), 349–364. doi:10.1007/s11192-011-0430-x
- Lüke, T., Hoek, W. van, Schaer, P., & Mayr, P. (2012). Creation of custom KOS-based recommendation systems. In *NKOS Workshop 2012*. Paphos, Cyprus. Retrieved from <https://www.comp.glam.ac.uk/pages/research/hypermedia/nkos/nkos2012/abstracts/Luke.pdf>
- Lüke, T., Schaer, P., & Mayr, P. (2012). Improving Retrieval Results with discipline-specific Query Expansion. In *International Conference on Theory and Practice of Digital Libraries (TPDL 2012)* (pp. 408–413). Paphos, Cyprus: Springer Berlin Heidelberg. doi:10.1007/978-3-642-33290-6_44
- Hienert, D., Schaer, P., Schaible, J., & Mayr, P. (2011). A Novel Combined Term Suggestion Service for Domain-Specific Digital Libraries. In S. Gradmann, F. Borri, C. Meghini, & H. Schuldt (Eds.), *International Conference on Theory and Practice of Digital Libraries (TPDL)* (pp. 192–203). Berlin: Springer. doi:10.1007/978-3-642-24469-8_21