

STW (Thesaurus for Economics) web service applied to library applications

Timo Borst
Joachim Neubert
IT Development
German National Library of Economics
Leibniz Centre for Economics

8th European Networked Knowledge Organization Systems
(NKOS) Workshop
at the 13th ECDL Conference, Corfu, Greece



Contents



- 1) ZBW, STW Thesaurus for Economics and motivation for STW web services
- 2) STW web services design and API
- 3) Applications consuming the services
 - a. Use case 1: indexing / author upload
 - b. Use case 2: retrieval support
 - c. Excuse: thesaurus hit rates
- 4) Technical infrastructure
- 5) Conclusion

ZBW and STW Thesaurus for Economics



- German National Library of Economics (ZBW) / Leibniz Information Centre for Economics is the world's largest economics library
- STW Thesaurus for Economics was developed in the nineties by four economics institutions as "Standard-Thesaurus Wirtschaft", funded by German government, for the purpose of unifying the use of subject headings in economics
- now maintained by ZBW as a controlled vocabulary of about 5,800 terms in German and English
- published as Linked Data in March 2009: SKOS, RDFa Web Pages, RDF/XML download, SPARQL endpoint
- licensed under Creative Commons (by-nc-sa)

- Home
- Alphabetical descriptor list

- ▶ A General descriptors
- ▶ B Business economics
- ▶ G Geographic names
- ▶ N Related subject areas
- ▶ P Commodities
- ▼ V Economics
 - V.00 Economics
 - ▶ V.01 Economic Theory
 - ▶ V.02 Microeconomics
 - ▶ V.03 Macroeconomics
 - ▶ V.04 Business Cycle, Growth, and Economic Structure
 - ▼ V.05 Money and Financial Markets
 - V.05.01 Monetary Theory
 - V.05.02 Monetary Policy
 - V.05.03 Inflation
 - V.05.04 Interest Rate
 - V.05.05 Credit
 - ▼ V.05.06 Financial Markets

Financial crisis

Finanzmarktkrise (german)

used for: Financial instability

Narrower Terms

- Banking crisis 
- Currency crisis 
- Stock market crash 

Broader Terms

- Economic crisis 

Related Terms

- International financial market 

Subject Categories

- V.05.06 Financial Markets
- V.07.07 International Financial Markets

Persistent Identifier (for bookmarking and linking)

- <http://zbw.eu/stw/descriptor/19664-4>

Motivation for STW web services



We wanted

- to make use of STW in our own library applications
- to provide a simple-to-use interface for "non-semantic-web-people"
- applicable to legacy and newly-build applications in a least-intrusive manner
- also for third party use

STW web services design



- Experimental public service at <http://zbw.eu/beta/stw-ws>
- Thin layer on top of SPARQL queries (hides complexity)
- RESTful architecture: simple, uniform interface through HTTP GET requests
- strongly use-case oriented
- delivers all information necessary in one round-trip
- does not try to be complete
 - in contrast to SKOS API and some other SOAP-based approaches
- users can run custom queries against the public SPARQL endpoint and offer the same interface to their application

STW web services example



/synonyms

takes any string, returns synonym terms, eg.

<http://zbw.eu/beta/stw-ws/synonyms?query=free+trade+zone>

STW Web Service Query Results

| synonym |
|------------------------------|
| "Freie Produktionszone" @de |
| "Free trade zone" @en |
| "Zollausschluss" @de |
| "Freilager" @de |
| "Freizone" @de |
| "Foreign trade zone" @en |
| "Export processing zone" @en |
| "Zollausschluß" @de |
| "Zollfreigebiet" @de |
| "Foreign-trade zone" @en |

STW web services API: Services



/concepts – Thesaurus concepts (by label search)

/narrower – Narrower concepts for a concept

/labels – All Labels for a concept

/synonyms – Alternate terms for a search term (from matching labels)

/suggest – Suggestions for descriptors (starting with a given string)

/combined1 – All of the above (and more) as needed for retrieval applications

STW web services API: Parameters



dataset (stw / jel)

query (searchstring)

concept (concept uri)

lang (en / de)

output (html / sparql-xml / sparql-json)

callback (name of a Javascript callback routine)



- 1) ZBW, STW Thesaurus for Economics and motivation for STW web services
- 2) STW web services design and API
- 3) Applications consuming the services**
 - a. Use case indexing / author upload
 - b. Use case retrieval support
 - c. Excuse: thesaurus hit rates
- 4) Technical infrastructure
- 5) Further application areas and future prospects

Use Case 1: Indexing



- Common use case for librarians
- User groups: librarians + scientists (?) + library users (?)
- Operation: Entry of descriptive metadata, often from a controlled vocabulary
- Part of Open Access relevant activities like self-publishing, self-archiving and social tagging
- Common Open Source-repository software (DSpace, EPrints, Fedora...) yet does not support the integration of *externally* controlled metadata
(D. Salo: Name authority control in institutional repositories;
<http://minds.wisconsin.edu/handle/1793/31735>)
- Question: Can we support this operations with STW web service, thus generating ‚better‘ controlled metadata in terms of normalization and standardization?

Use Case 1: Indexing



Beta System: <http://87.106.250.18/beta/econstor/>

Geben Sie eine Zusammenfassung des Inhalts der Publikation ein.

Abstract

Impact of the financial crisis on the computer industry.

Geben Sie hier die STW-Begriffe ein.

STW-Begriffe

Computerindustrie

financia

Weitere hinzufügen

Klassifizierung der Publikation entsprechende Kategorie

JEL

Bitte wählen Sie die Sprache aus, in der

Sprache

N/A

STW Begriffe:

- Financial Engineering
- Financial Futures
- Financial administration → Finanzverwaltung
- Financial aid → Kapitalhilfe
- Financial arbitrage → Arbitragegeschäft
- Financial asset market → Finanzmarkt
- Financial assets → Betriebliches Finanzanlagevermögen
- Financial audit → Wirtschaftsprüfung
- Financial centre → Finanzplatz
- Financial crisis → Finanzmarktkrise

terature Classification System"

fügen

< Zurück

Weiter >

Abbi

trieben von der [Deutschen Zentralbibliothek für Wirtschaftswissenschaften](#) [Finanzmarktkrise](#) [Akt - Über Eco](#)

Use Case 1: Indexing



Beta System: <http://87.106.250.18/beta/econstor/>

Geben Sie eine Zusammenfassung des Inhalts der Publikation ein.

Abstract

Impact of the financial crisis on the computer industry.

Geben Sie hier die STW-Begriffe ein.

STW-Begriffe

Computerindustrie

Finanzmarktkrise

Weitere hinzufügen

Klassifizierung der Publikation entsprechende Kategorien aus dem "Journal of Economic Literature Classification System"

JEL

L6

Weitere hinzufügen

Bitte wählen
Sprache

JEL Klassifikation:

- L6 - Industry Studies: Manufacturing
- L60 - Industry Studies: Manufacturing: General
- L61 - Metals and Metal Products; Cement; Glass; Ceramics
- L62 - Automobiles; Other Transportation Equipment
- L63 - Microelectronics; Computers; Communications Equipment**
- L64 - Other Machinery; Business Equipment; Armaments
- L65 - Chemicals; Rubber; Drugs; Biotechnology

Bitte Publikation verfasst ist.

< Zurück

Weiter >

Abbi

trieben von der [De](#)

[Wirtschaftswissenschaften \(ZBW\) - Kontakt - Über Eco](#)

Use Case 2: Retrieval



- Most common use case in library applications
- User groups: librarians + end users
- Operation: Mostly simple fulltext search on metadata + fulltexts (if available and depending from configuration of search engine)
- Question: Can we support this operation with STW web service and can we get ,better' results?

Use Case 2: Retrieval



Beta System: <http://87.106.250.18/beta/econstor/>

Input: „Telework“

Hits: 9

EconStor.
DIGITAL ARCHIVE FOR ECONOMICS AND BUSINESS STUDIES

In EconStor suchen

[Erweiterte Suche](#)

- [Startseite](#)

Blättern in:

- [Institutionen & Reihen](#)
- [Titeln](#)
- [Autoren](#)
- [Stichwörtern](#)
- [Nach Datum](#)

Anmelden für:

- [Abonnements](#)
- [Mein EconStor](#)
Registrierte Benutzer
- [Profil bearbeiten](#)

[DSpace@ZBW >](#)

Suchergebnisse

Suchen:

nach

Telearbeit Telekommunikation Virtuelles Unternehmen

Ergebnisse 1-9 von 9.

Treffer in Publikationen:

| Datum | Titel | Autoren |
|-------|--|--------------------------------------|
| 2001 | Why Do Firms Recruit Internationally? | Winkelmann, Rainer |
| 2007 | Flexicurity : Reconciling Social Security with Flexibility ; Empirical Findings for Europe | Seifert, Hartmut / Tangian, Andranik |
| 2007 | Is flexible work precarious? : a study based on the 4th European survey of working conditions 2005 | Tangian, Andranik S. |

Use Case 2: Retrieval



Mouse-over „Telearbeit“

[DSpace@ZBW](#) >

Suchergebnisse

Suchen:
nach

Telearbeit Telekommunikation Virtuelles Unternehmen

-9 von 9.

Suchterme:

- Telework
- Telearbeit
- Teleworking
- Telecommuting
- Distance working
- Teleheimarbeit
- Elektronische Heimarbeit
- EDV-Heimarbeit

Relevanz: 6
Treffer (match) der Suchanfrage:
- Telework

| Datum | Titel | Quelle |
|-------|--|---------------|
| 20 | Internationally? | Wink Raine |
| 20 | Social Security Findings for | Seife Tang |

Use Case 2: Retrieval



Hits: 20 (!) with expansion through synonyms

[DSpace@ZBW](#) >

Suchergebnisse

Suchen:
EconStor gesamt

nach

Telearbeit Telekommunikation Virtuelles Unternehmen

Ergebnisse 1-10 von 20.

Treffer in Publikationen:

| Datum | Titel | Au |
|-------|--|-----------------------------|
| 2001 | Qualifikationsanforderungen und Arbeitsformen in der Neuen Ökonomie | Christenser |
| 2007 | Flexicurity : Reconciling Social Security with Flexibility ; Empirical Findings for Europe | Seifert, Har Tangian, Ar |

Excuse: Query logs



- Question: How well does STW retrieval support work *in reality*?
- Large scale evaluation: 1.3 million "simple" queries from econis.eu log files (2 years)
- Queries hardly repeat (49.2 % of all queries not even once in two years)
- Only 4.4 % occurred more than a hundred times
- Most queries use 1-3 words

| | |
|-------|------------|
| 28.8% | 1 |
| 37.7% | 2 |
| 16.8% | 3 |
| 16.7% | 4 and more |

Excuse: Thesaurus hits in queries



Three strategies for searchstring ./ thesaurus concept matching:

- 8.0% searchstring ./ thes label exact matches
(often more than one concept involved!)
- 43.3% split searchstring in separate words and word groups, match exactly against thes labels
- 62.7% searchstring full-text matches against thes labels
(not yet optimized: no bilingual stemming, no score thresholds applied)

telework

free trade zone brazil

emerging free trade zones in china

Telework

Free trade zone

Brazil

Free trade

Free trade area

Free trade zone

China

Excuse: Effects of STW support in retrieval

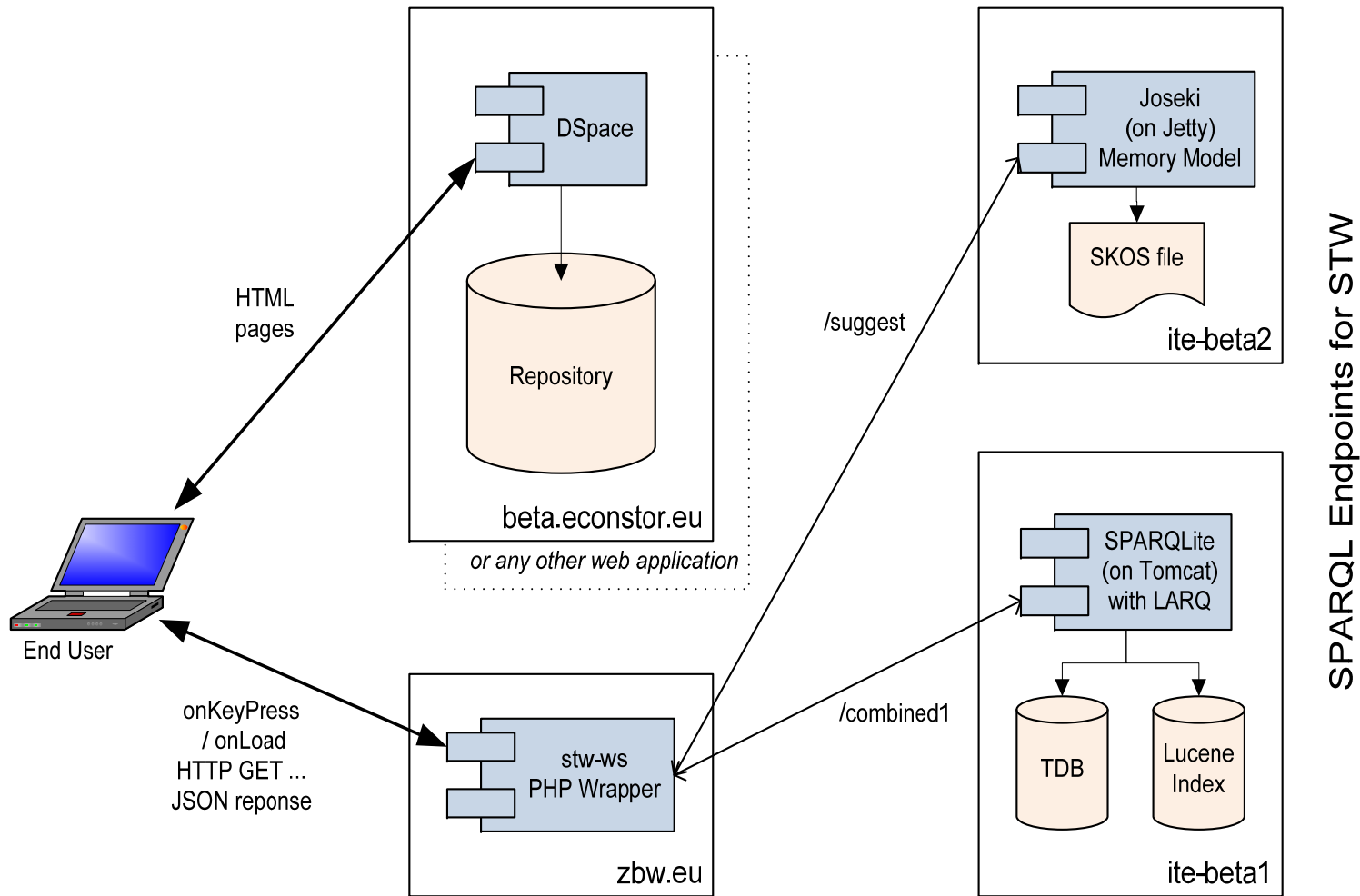


For each thesaurus hit, STW offers (on average)
4.7 narrower and related concepts, and for each concept
4.2 synonyms

STW can be applied effectively to support real-life queries

- offering „near-by“ alternate concepts (serendipity)
- gaining better recall through synonyms

Infrastructure



Applications building on thesaurus web services



Lightweight enhancements to existing "legacy" applications

- minimal intrusive
- easy-to-adapt to different systems
- don't require data model changes (which are hard!)

Types of applications:

- Repositories (often open source, plug-in architecture)
- Catalogs (often proprietary and monolithic)
- Portal solutions (heterogenous content)

It's no substitute for real thesaurus-aware software
which also starts coming up - for DSpace e.g.:

- Functional requirements Thesaurus Add-on (from FAO work)
<http://code.google.com/p/dspace-agrisap/wiki/ThesaurusAddOn>
- Discussion about Authority control on Dspace-tech and Dspace-devel lists

let alone a real linked data application

... but can be done right now!



Thank you!

Questions?

Links:

<http://87.106.250.18/beta/econstor/>

<http://zbw.eu/beta/stw-ws>

<http://zbw.eu/beta/sparql>

<http://zbw.eu/stw>

On Publishing STW as SKOS and Linked Data:

http://events.linkeddata.org/ldow2009/papers/ldow2009_paper7.pdf

Timo Borst

Joachim Neubert

German National Library of
Economics (ZBW) /

Leibniz Centre for Economics
{t.borst,j.neubert}@zbw.eu