

## STW Thesaurus for Economics Web Service applied to library applications

Timo Borst, Joachim Neubert

German National Library of Economics (ZBW) – Leibniz Information Centre for Economics

After having published the STW Thesaurus for Economics as Linked Data (<http://zbw.eu/stw>), the ZBW German National Library of Economics wants to make use of it in library applications. Frequently, these applications are not thesaurus-driven (let alone Semantic Web enabled), so thesaurus support has to be added in an unobtrusive way “from the outside”. For this purpose we provide access to the thesaurus concepts and relations in a web service (<http://zbw.eu/beta/stw-ws>). The web service works on the SKOS-representation of the STW, offering a simple, REST-oriented API for resources like “concepts”, “narrower” or “synonyms”. It can be applied to enhance and facilitate both retrieval and indexing in library applications (prototypically implemented in a repository application).

*In the field of retrieval* the thesaurus web service can

- 1) support (even cross-lingual) retrieval by an optional query expansion with synonyms (in environments where only full-text retrieval is available), and
- 2) support retrieval in holdings indexed using a controlled vocabulary by providing the user with links to searches for narrower, related etc. terms
- 3) provide users with links as mentioned in 2), expanded with synonyms as mentioned in 1), to support full-text queries with narrower and related terms

STW with its 5,800 bilingual concepts, 17,000 synonyms and 25,000 relations is well suited for this.

For a large-scale advance simulation of the effects of query enrichment facing “real life” queries (and for the evaluation of the actual thesaurus coverage) we analyzed more than a million “simple” search queries mined from the logs of a representative bibliographic database in economics (<http://econis.eu>). As we found out, half of the (already normalized) queries only occurred once in two years (only less than 1% showed up 10 or more times). In the case of queries with multiple occurrences we looked up the thesaurus and found less than 15 % direct matches. The main reason for this rather low rate was that users often combine multiple concepts within one query. We therefore will present different approaches to deal with this, especially a) syntactical splitting of search strings (which increases the hit rate to more than 40%), and b) an application of full-text matching techniques to the lookup itself. For the identified thesaurus concepts the synonyms, broader, narrower and related terms are evaluated statistically.

*In the field of indexing*, the thesaurus web service can assist users with indexing or “tagging”. In our DSpace repository for scientific articles in economics, indexing is prospectively done by authorized end users submitting and uploading their articles. These users, e.g. scientific staff from an economics faculty, generally have no experience in applying controlled vocabularies. To support them and to foster the submission of electronic resources we will offer an autosuggest service for thesaurus terms very similar to the one already in use on the STW web site (<http://zbw.eu/stw>). Finally, we will extend this service to another well-known terminology, the “Journal of Economic Literature” (JEL) classification.

The web service builds on Semantic Web technology. A thin PHP wrapper translates the resource requests to SPARQL queries. A Joseki server operating directly on a memory model of the RDF/XML file and alternatively a SPARQLite server with a Jena TDB triple store and LARQ (Lucene free text indexing for SPARQL) are used to execute the queries.