

14th European Networked Knowledge Organization Systems (NKOS) Workshop

To be held during the TPDL 2015
Conference in Poznań, Poland

Friday 18th September 2015

Presentation Proposal

The missing link: a vocabulary mapping effort in economics

Andreas Oskar Kempf (a.kempf@zbw.eu) and Joachim Neubert (j.neubert@zbw.eu)
ZBW – Leibniz Information Centre for Economics, 20354 Hamburg, Germany

In economics there exists an internationally established classification system. Research literature is usually classified according to the JEL classification codes¹, a classification system originated by the Journal of Economic Literature and published by the American Economic Association (AEA). Complementary to keywords which are usually assigned freely, economists widely use the JEL codes when classifying their publications. In cooperation with KU Leuven, ZBW – Leibniz Information Centre for Economics has published an unofficial multilingual version of JEL in SKOS format.²

In addition to this, exists the STW Thesaurus for Economics³ a bilingual domain-specific controlled vocabulary maintained by the German National Library of Economics (ZBW). Developed in the mid-1990s and since then constantly updated according to the current terminology usage in the latest international research literature in economics it covers all sub-fields both in the economics as well as in business economics and business practice containing subject headings which are clearly delimited from each other. It has been published on the web as Linked Open Data in the year 2009.

In our presentation we would like to report about the effort to map both subject terminologies onto each other. To be more specific, we would like to present results of a vocabulary mapping effort to establish relations from the JEL classification codes to the subject categories of the STW Thesaurus for Economics. This mapping project serves a specific use case. Building on the fact that economists who publish their own work as working papers or journal articles are usually quite familiar with the JEL classification codes we, this way, hope to animate economists to provide a more fine-grained content description of their publications with a standardized controlled vocabulary. Starting with the assignment of JEL classification codes subject headings might be suggested that would fit to the content of a publication. Considering the huge increase of publication numbers this vocabulary mapping effort could be part of an overall indexing strategy to win authors themselves for indexing their research publications by using a controlled vocabulary.

Cross-concordances between two terminologies or classification systems stand for a specific approach to treat semantic heterogeneity. They are directed, and oftentimes relevance rated

¹ <https://www.aeaweb.org/econlit/jelCodes.php>

² http://zbw.eu/beta/external_identifiers/jel

³ <http://zbw.eu/stw>

relations between the terms or classes of the terminologies or classification systems applied. It could be distinguished between three different types of relation. According to the SKOS mapping relations⁴ it could be distinguished between `skos:exactMatch`, `skos:closeMatch`, `skos:broadMatch/skos:narrowMatch`, and `skos:relatedMatch` relations. Usually established to support a distributed search over several information systems in a digital library, vocabularies are mostly related bilaterally; that is, a cross-concordance relating terms from vocabulary A to vocabulary B as well as a cross-concordance relating terms from vocabulary B to vocabulary A. This is not to say, that they are necessarily symmetrical. Having the specific use case mentioned above in mind we, so far, are particularly interested in a cross-concordance from the JEL classification codes to notations of the STW subject categories.

The two knowledge organization systems applied are characterized by a number of differences. Not only does the one KOS represent a classification while the other is a subject category system. Both terminologies are also of different size and scope. The JEL classification consists of nearly 1,000 classification codes on three different hierarchical levels representing a specific Anglo-American understanding of economics. The subject categories of the STW Thesaurus for Economics, by contrast, only contain nearly 500 subject categories. They range between one and three hierarchical levels covering a broader scope of economics which includes business economics and business practice representing a more European understanding of economics.

Building this mapping the AMsterdam Alignment GenerAtion MEtatool (AMALGAME) will be tested. It is a web-based interactive platform for creating, analyzing and evaluating vocabulary alignments. The tool supports a multi-step mapping process of automated creation and intellectual evaluation of mapping suggestions. The evaluation will be done intellectually by a domain expert. On the basis of two mapping iterations within AMALGAME we seek to determine whether a reasonable vocabulary mapping is possible. While for the first iteration the bare labels of the STW subject categories and JEL classes were used, for the second iteration both KOS were enriched with additional word material. The labels of the JEL classes were enriched with additional keywords taken from the online JEL guide. The labels of the STW subject categories were enriched with the descriptors assigned to a subject category, their synonyms, and with all the exact equivalent relations and their synonyms from the other vocabularies the STW has already been mapped to in the past. Results could motivate the development of a web service. On the basis of JEL classification codes this web service could suggest a reasonable selection of STW subject categories economists could go through to choose STW subject headings to index their own research papers in economics.

References

- Mayr, P. & Petras, V. (2008). Building a terminology network for search: the KoMoHe project. In: *Proceedings of the International Conference on Dublin Core and Metadata Applications*, Berlin, 177-182.
- Patel, M., Koch, T., Doerr, M., & Tsinaraki, C. (2005). *Semantic Interoperability in Digital Library Systems*. Other. UKOLN, University of Bath.

⁴ See: <http://www.w3.org/TR/2009/REC-skos-reference-20090818/>

Zeng, M. L., & Chan, L. M. (2004). Trends and Issues in Establishing Interoperability Among Knowledge Organization Systems. *Journal of the American Society for Information Science and Technology*, 55(3), 377-395.