

Thesaurus Mapping: A Challenge for Ontology Alignment?

Dominique Ritze and Kai Eckert

Mannheim University Library

`dominique.ritze,eckert@bib.uni-mannheim.de`

Thesauri are hierarchical knowledge organization systems commonly used in libraries to categorize and index publications. While sometimes referred to as so-called lightweight ontologies [4], they actually fundamentally differ from ontologies in several aspects. Nevertheless, as thesauri are actively used, constantly maintained and improved, they offer an interesting background knowledge for semantic applications. This year, we reinstated the OAEI library track¹, i.e., we provide the ontology matching community with the challenge to create alignments between thesauri. First, we aim for interesting insights into the differences between ontologies and thesauri. Second, we try to further integrate existing thesauri by means of new alignments which leads to better search experiences within library systems. From 2007 to 2009, there has already been a library track in the OAEI [1]. They focused on matching thesauri describing the same topic but at a different level of granularity. For our track, we selected two very comparable thesauri with topical overlaps. To make sure that the created alignments are indeed used, we work closely together with the maintaining institutions. We apply the following two thesauri:

STW: The Thesaurus for Economics (STW) provides vocabulary on any economic subject: more than 6,000 standardized subject headings (in English and German) and 19,000 additional keywords. The entries are richly interconnected by 16,000 broader/narrower and 10,000 related relations. The vocabulary is maintained on a regular basis by ZBW German National Library of Economics – Leibniz Centre for Economics². The thesaurus is available in SKOS [3].

TheSoz: Similar to the STW, the Thesaurus for the Social Sciences (TheSoz) serves as a crucial instrument for indexing documents and research information in the social sciences. Overall, it contains about 12,000 keywords, from which 8,000 are standardized subject headings (in English, German and French) and 4,000 additional ones. The thesaurus is owned and maintained by GESIS - Leibniz Institute for the Social Sciences³. The thesaurus is available in SKOS [5].

The matching results are evaluated by means of a reference alignment which has been manually created by domain experts in 2006 [2]. It has not been adapted or further developed after its initial creation. Hence, it does not cover changes of the thesauri. Within the reference alignment, concepts are aligned to more than one concept

¹ <http://web.informatik.uni-mannheim.de/oaei-library/2012/>

² <http://zbw.eu/index-e.html>

³ <http://www.gesis.org/en/home/>

($n:m$ mapping). All in all, the alignment contains 2,839 exact matches and 1,450 sub-suggestions. Other generated correspondences will be evaluated by domain experts as well. It is planned to extend the reference alignment on the basis of manually evaluated matching results, if the quality is sufficient to justify the effort.

The participating matchers in OAEI are currently developed for (OWL) ontology matching. As a starting point for them, we provide an OWL version of the thesauri. Therefore, the SKOS predicates are mapped to RDF/OWL as follows:

| SKOS | RDF/OWL |
|--|--------------------------------|
| <code>skos:concept</code> | <code>owl:class</code> |
| <code>skos:prefLabel</code> , <code>skos:altLabel</code> | <code>rdfs:label</code> |
| <code>skos:scopeNote</code> , <code>skos:notation</code> | <code>rdfs:comment</code> |
| <code>skos:related</code> | <code>rdfs:seeAlso</code> |
| <code>skos:narrower</code> | <code>rdfs:superClassOf</code> |
| <code>skos:broader</code> | <code>rdfs:subClassOf</code> |

There are several issues with such a mapping: First and foremost, a `skos:concept` is not a class. Concepts sometimes represent classes, like `COMMODITIES`, but there are other concepts that clearly represent instances, like `GERMANY`. The mapping of the broader/narrower relationships is likewise problematic. In the STW, the narrower path `COMMODITIES` \rightarrow `METALS` \rightarrow `METAL PRODUCTS` \rightarrow `RAZOR` is found. All metals are commodities too, but metal products like a razor only consist of metal, but are no metal. And last, the expressiveness of SKOS regarding different types of labels, additional descriptive notes and general concept relations are lost in RDF/OWL.

Thus, the question arises to which degree the current matching systems are hampered by these oversimplifications and semantic inconsistencies. We indeed hope that specialized SKOS matchers will join the challenge and that they outperform the generic ontology matchers. This way, the library track can contribute to the integration of thesauri in real world applications. As a side-effect, we would like to raise the discussion, how thesauri relate to ontologies and which role they might play in the Semantic Web.

References

1. Antoine Isaac, Lourens van der Meij, Shenghui Wang, and Henk Mattheizing. Results of the OAEI 2007 Library Thesaurus Mapping Track. Technical report, VU Amsterdam, 2007.
2. Philipp Mayr and Vivien Petras. Building a Terminology Network for Search: The KoMoHe Project. In *Proc. of the Int. Conference on Dublin Core and Metadata Applications*, pages 177 – 182, 2008.
3. Joachim Neubert. Bringing the “Thesaurus for Economics” on to the Web of Linked Data. In *Proc. of the WWW Workshop on Linked Data on the Web (LDOW)*, 2009.
4. Michael Uschold and Michael Gruninger. Ontologies and semantics for seamless connectivity. *SIGMOD Rec.*, 33(4):58–64, 2004.
5. Benjamin Zopilko, Johann Schaible, Philipp Mayr, and Brigitte Mathiak. TheSoz: A SKOS Representation of the Thesaurus for the Social Sciences. *Semantic Web – Interoperability, Usability, Applicability*. accepted.