

Linking knowledge organization systems via Wikidata

Presentation

Joachim Neubert
ZBW – Leibniz Information
Centre for Economics,
Germany
j.neubert@zbw.eu

Keywords: Wikidata; STW; KOS; LOD; linked open data; alignment; matching tool

Abstract

Wikidata is a large collaboratively curated knowledge base, which connects all of the roughly 300 Wikipedia projects in different languages and provides common data for them. Its items also link to more than 1500 different sources of authority information. Wikidata can therefore serve as a linking hub for the authorities and knowledge organization systems represented by these “external identifiers”. In the past, this approach has been applied successfully to rather straight-forward cases such as personal name authorities¹. Knowledge organization systems with abstract concepts are more challenging due to, e.g., partial overlaps in meaning and different granularities of concepts.

Our work is based on the ongoing mapping effort of “STW Thesaurus for Economics” to Wikidata². Contrary to other vocabularies, and just like Wikipedia, Wikidata can be extended by everybody and for every domain of human know knowledge. I will discuss the pros and cons of such extensions with regard to Wikidata itself and also to the indirect linking to Wikipedia pages, which is often one of the goals of mapping approaches to Wikidata. As an alternative to creating new Wikidata items, the newly introduced “mapping relation type” qualifier³ – which comprises the SKOS mapping relations – allows for in-exact mappings of Wikidata items to external identifiers.

During mapping creation, in particular Wikidata’s “Mix’n’match” tool and tailored SPARQL queries have proved useful. Existing mappings of other KOS to Wikidata can be exploited for deriving indirect mappings to these vocabularies, but also for generating mapping suggestions where a direct mapping already exists (as evaluated for a sample set of STW/GND mappings).

The extensibility of Wikidata by everybody also raises maintenance issues, as there is no single ownership and responsibility for a mapping. Wikidata’s SPARQL query service allows tracing inconsistencies with according reports, taking into account the mapping relation types⁴. Coverage of newly introduced Wikidata and external concepts over time and possibly differing practices and interests of contributing parties pose additional challenges for the long-term maintenance of collaboratively used and curated vocabulary mappings. Besides technical challenges, this requires an understanding of Wikidata’s policies and the communication within its community.

¹ Joachim Neubert: Wikidata as authority linking hub: Connecting RePEc and GND researcher identifiers, ZBW Labs 2017-11-30, <http://zbw.eu/labs/en/blog/wikidata-as-authority-linking-hub-connecting-repec-and-gnd-researcher-identifiers>

² <https://github.com/zbw/stw-mappings>

³ <https://www.wikidata.org/wiki/Property:P4390>

⁴ https://www.wikidata.org/wiki/Property_talk:P3911#Reports_for_the_maintenance_of_the_STW_ID_.2F_Wikidata_mapping