Ontological Modelling in Wikidata

Markus Kroetzsch[1]

[1]Technical University of Dresden, Germany

Abstract. Wikidata, the knowledge base of Wikimedia, has been extremely successful in building and sustaining new communities of editors and users. Since its inception in 2012, it has developed from an experimental "data wiki" into a well-organised reference knowledge base with an amazing array of applications. Developing an ontological schema for such an open and rapidly expanding project is a huge undertaking, and difficult challenges arise on many levels. The community has directed significant efforts towards vocabulary development, many guidelines and rules have been created, and tools are used for helping editors to avoid and correct modelling errors. Nevertheless, the distributed nature of Wikidata editing often means that ontology design, too, is distributed, and a coherent global view is only being worked on once significant amounts of data have been added. The result is a knowledge graph with a widely varying modelling quality across different sub-domains.

The big question for researchers is how their insights and methods can help here. The Wikidata community is widely aware of semantic web activities and existing standards and academic publications play a role in many discussions. Yet, there seems to be only little direct exchange between the communities. In this talk, I will review the current state of Wikidata and its connection to semantic web standards such as RDF and SPARQL. I will try to raise awareness of the particular requirements of Wikidata, and argue that these are of general interest for the data-driven curation of knowledge graphs.