

IN-CONFERENCE WORKSHOPS

Introduction to Wikidata

The Wikipedia of Facts

By Clifford B. Anderson, Vanderbilt University

ABSTRACT This paper briefly summarizes an in-conference workshop about Wikidata and its potential uses for scholarly communications in religious studies and theology.

This in-conference workshop introduced Wikidata,¹ a project to collect and organize factual information from Wikipedia and other sources. Wikidata started in October 2012² and has grown to more than 50,000,000 items during its first six years.³ These facts do not relate to the English Wikipedia alone; increasingly, they connect the 300 language editions of Wikipedia.

Wikidata organizes data in a structured format called “statements.”⁴ These statements consist of three primary parts: items (“subject”), predicates (“property”), and objects (“values”). For instance, Wikidata records the founding date of the Christian Theological Seminary in Indianapolis as “Christian Theological Seminary” (item), “inception” (property), “1855” (value).⁵ Since there are different language editions of Wikipedia, and Wikidata cannot afford to privilege any single edition (like the English Wikipedia), Wikidata uses its own system of letters and numbers to refer to items, properties, and values. The internal representation of the statement above

is “Q5110194 P571 1855.” Thankfully, Wikidata provides robust tools to translate from codes into the language of your choice.

All data in Wikidata is queryable. The query language of Wikidata is SPARQL or the “SPARQL Protocol and RDF Query Language,” which the W3C developed to query linked data.⁶ These SPARQL queries underscore the power of Wikidata, allowing users to ask complex questions (“Which paintings of Mary include images of books?”) and to visualize the results.⁷ The possibilities for visualizations include charts, image galleries, maps, and timelines, among others. Visualizations may be embedded on websites, and data from queries may be downloaded and reused freely.

The types of information described in Wikidata vary broadly, from architectural objects to films to people. A large percentage of Wikidata entries (42.4%) are about scholarly articles. A movement called WikiCite⁸ is working toward adding bibliographic data about all scientific literature to Wikidata, including scholarly articles in the field of religious studies and theology. In “Scholia, Scientometrics and Wikidata,” Finn Årup Nielsen, Daniel Mietchen, and Egon Willighagen describe their work on Scholia,⁹ an open source tool to build scholarly and institutional profiles on the basis of bibliographic and related data in Wikidata.¹⁰

Since Wikidata is a crowdsourced resource like Wikipedia, it suffers from unevenness. Much more information exists about churches, for example, in New York City than in Indianapolis. The gaps in Wikidata’s knowledge base reflect the distribution and demographics of its volunteer editors. As Wikidata becomes increasingly central to the future of scholarly communications, theological librarians should take efforts to add and expand items about articles, books, denominations, institutions, and scholars in the field.

The full presentation, along with sample SPARQL queries, is available on Github.¹¹

WORKS CITED

- DuCharme, Bob. *Learning SPARQL: Querying and Updating with SPARQL 1.1*. 2nd ed. Sebastopol, CA: O’Reilly Media, 2013.
- Nielsen, Finn Årup, Daniel Mietchen, and Egon Willighagen. “Scholia, Scientometrics and Wikidata.” In *The Semantic Web: ESWC*

2017 Satellite Events, edited by Eva Blomqvist, Katja Hose, Heiko Paulheim, Agnieszka Ławrynowicz, Fabio Ciravegna, and Olaf Hartig, 237–59. *Lecture Notes in Computer Science*. Springer International Publishing, 2017. doi:https://doi.org/10.1007/978-3-319-70407-4_36.

Vrandečić, Denny, and Markus Krötzsch. “Wikidata: A Free Collaborative Knowledgebase.” *Communications of the ACM* 57, no. 10 (September 2014): 78–85. doi:10.1145/2629489.

NOTES

- 1 See <https://www.wikidata.org>.
- 2 Denny Vrandečić and Markus Krötzsch, “Wikidata: A Free Collaborative Knowledgebase,” *Communications of the ACM* 57, no. 10 (September 2014): 80, doi:10.1145/2629489.
- 3 See <https://www.wikidata.org/wiki/Wikidata:Statistics>.
- 4 See <https://www.wikidata.org/wiki/Wikidata:Introduction>.
- 5 See <https://www.wikidata.org/wiki/Q5110194>.
- 6 Bob DuCharme, *Learning SPARQL: Querying and Updating with SPARQL 1.1*, 2nd ed. (Sebastopol, CA: O’Reilly Media, 2013).
- 7 See <http://tinyurl.com/y7p67o24>.
- 8 See <https://meta.wikimedia.org/wiki/WikiCite>.
- 9 See <https://www.wikidata.org/wiki/Wikidata:Scholia>.
- 10 Finn Årup Nielsen, Daniel Mietchen, and Egon Willighagen, “Scholia, Scientometrics and Wikidata,” in *The Semantic Web: ESWC 2017 Satellite Events*, ed. Eva Blomqvist et al., *Lecture Notes in Computer Science* (Springer International Publishing, 2017), 237–59, doi:https://doi.org/10.1007/978-3-319-70407-4_36.
- 11 See <https://gist.github.com/CliffordAnderson/97116b3a3874c59f61c18a9cc2f05576>.