chem-bla-ics

Adding citations between existing articles in Wikidata

Published September 7, 2024

Citation

Willighagen, E. (2024, September 7). Adding citations between existing articles in Wikidata. *Chem-bla-ics*. https://doi.org/10.59350/7hjzg-ngr66

Keywords

Wikidata, Bioclipse, Opencitations

Copyright

chem-bla-ics

Scholarly articles provide context to the factualness of statements in Wikidata, similar to the [citation needed] in Wikipedia. And just like the cited references in each scholarly article itself. The citation network is general seen as an essential part of (doing) science, even without citation intention annotation. Nowadays, citations are mostly open, but this took very serious lobbying by the Initiative for Open Citations and not every publisher reacted immediately. But now that they are open, projects like OpenCitations are making this citation network FAIR.

Therefore, when an article is cited as reference in Wikidata, I think that the articles (and other research output) cited in that article is part of the reference. After all, it is really hard to understand any article without the details in the cited articles. So, getting these citations between article into Wikidata deepens the knowledge captured by Wikidata. Of course, Wikidata is also one of the few places where we can capture the citation intentions at all.

Adding these citations manually is cumbersome but sometimes needed as these citations are not open or not FAIR yet. Fortunately, in many cases we can automate the process, for which I wrote a Bacting-cased script. Until recently, the script takes as input a single DOI or a list of DOIs as input, and for each DOI looks up in OpenCitations if it cites other article DOIs and is cited by other DOIs. For the cited and citing DOIs it checks if those are in Wikidata and (only) if they are in Wikidata, then it create QuickStatements. The result can look like this:

The script also needs a OpenCitation token, which you can get here. This is how I run the code from the command line (with the token in the **TOKEN** environment variable), for a single DOI:

```
groovy quickstatements.groovy -t ${TOKEN} -d 10.1002/JLAC.18721620110 | tee
output.qs
```

A list of DOIs is provided as a text file, with one DOI on one line. I then use the `-l` parameter (oh, here DOIs of works by Shyamala Gopalan, mother of Kamala Harris):

```
groovy quickstatements.groovy -t ${TOKEN} -l harris_dois.txt | tee output.qs
```

chem-bla-ics

But last weekend I created a new feature. To enrich the profiles of authors, for example Nobel Prize winners, mothers of, or famous chemists, previously I would create a list of DOIs, now I have the script do that:

So, today I could add the citation network for any arbitraty author, e.g. Carolyn Bertozzi, I just pass the Wikidata QID:

```
groovy quickstatements.groovy -t ${TOKEN} -a Q7442 | tee output.qs
```

I can imagine that in the future the script will have more such options, to do the same for many authors at some affiliation, or all DOIs for a certain journal.