

Lab session: Computer science conferences in Wikidata

The goal of this lab session is to explore Wikidata in the topic of computer science conferences and discover how to contribute to it. First, you should get acquainted with possible properties for a computer science conference by checking out some entities, e.g. ICDT (Q25106323) or IJCAI (Q863401) or PODS (Q7661890).

Note that, in other areas, you would use the Showcase items¹, but Wikidata currently doesn't have one for scientific conferences.

Adding topics (and finding entities)

Scientific conference usually have a topic, which should be expressed with property P921, but it is often not indicated.

It is a useful property. For instance, here is how one can use it:

- Write a SPARQL query to find conferences (instances of subclasses of Q47258130).
- Write a SPARQL query to find conferences about theoretical computer science (Q2878974).
- Write a SPARQL query to find conferences about theoretical computer science or a subfield of it.

Now, let us see how to find cases where it is missing:

- Write a SPARQL query to find conferences without a known topic (P921).
- Pick one conference (or several) and find a suitable topic for them by searching in Wikidata for a suitable entity.
- Add it to the conference by creating a fact (“add statement”).

You can check that your changes are reflected when re-evaluating your SPARQL queries (but this may not be immediate, because of caching, and because of replication lag).

Website (and adding a literal value with a qualifier)

Scientific conferences often have a website, but it is sometimes missing from Wikidata. Let us see how to add it, and discover the notion of qualifiers.

- Write a SPARQL query to find conferences without a website (P856).
- Pick one conference (or several) which is missing a website, and add it.
- Don't forget to fill in the qualifier (P407); you will get a warning prompting you to do it once you have saved.

Inception (and adding references)

We will now add information (which may be harder to find) about the year at which a conference started to exist (P571).

- Write a SPARQL query to find conferences without this property (you may want to restrict to conferences in computer science or a subarea, and/or conferences with a known website).
- Play Sherlock Holmes and find the information about when the conference was first held, either on the Web or on Wikipedia. If you can't find it, try a different conference.
- Add the information to the entity, and add a reference for the claim that you created:
 - If you found it on a Wikipedia, add P143 with value Q328 (or similarly if imported from Wikipedia in a different language), and P4656 with a link to the page where you found it
 - If you found it on the Web, add P854 with the URL where you found it.

¹https://www.wikidata.org/wiki/Wikidata:Showcase_items

Open access (and creating entities)

- Write a SPARQL query on query.wikidata.org to find out the possible values for the property P6954, both in general, and in the case of academic conference series.
- Pick a conference with a known open access status².
- Find the conference in Wikidata if it exists, otherwise create it: <https://www.wikidata.org/wiki/Special:NewItem>
- Complete the value for P6954.

Bulk editing with OpenRefine (advanced)

Instead of manually copying datasets in Wikidata, you can import information automatically using tools such as OpenRefine³. Here is one possible idea to follow to automatically import inception dates from DBLP for Wikidata conferences already linked to DBLP.

- Write a SPARQL query to retrieve all DBLP venue IDs known in Wikidata.
- Download the DBLP dataset⁴ (including the `dblp.xml` file!).
- Parse this file to extract, for every DBLP venue ID, the first year for which an article was published for it (which we assume is the inception date). Be sure to parse the XML file using a SAX parser (it is too large to fit in memory), and to parse it using the entities in `dblp.xml`.
- Produce a CSV file containing, for every DBLP venue ID, the identifier, the conference name, and the inception year.
- Download and run OpenRefine, and import the file that you created.
- Prepare a bulk edit of Wikidata with the statements⁵ that you have prepared, sourced to the corresponding DBLP page.
- Do not run the bulk edit unless you are reasonably sure that it will not clash with existing data in Wikidata (added, e.g., by another student who was faster than you ;)).

Other ideas for editing tasks

Here are other ideas of sources from which to import information about scientific conferences (and journals) into Wikidata:

- The page <https://conferences-computer.science/>
- The list https://en.wikipedia.org/wiki/List_of_computer_science_journals
- The top Google Scholar venues: https://scholar.google.com/citations?view_op=top_venues
- WikiCFP: <http://www.wikicfp.com/cfp/>

²You can use, e.g., the list by Pierre Senellart: <https://pierre.senellart.com/other/open-access/>

³<https://openrefine.org/>

⁴<https://dblp.org/faq/1474679.html>

⁵You should follow this tutorial: www.wikidata.org/wiki/Wikidata:Tools/OpenRefine/Editing/Tutorials/Basic_editing