

## Distributed Enterprise Systems (CO3409) Lab 19: RDF Datasets and SPARQL

### 19.1 LODCat 2022 Survey

The [RDF dataset survey](#) launched recently at the University of Paderborn gives you an opportunity to examine representative datasets and at the same time support research and development on the annotation and documentation of data.

Go through the [survey](#). At each step, you will be given a dataset to download. The files are in *N-triples format*, which is compliant with TTL format, but restricted to just showing a sequence of triples, *e.g.*, as follows:

```
<http://data.semanticweb.org/organization/german-research-center-for-artificial-intelligence-dfki> <http://www.w3.org/1999/02/22-
<http://data.semanticweb.org/organization/german-research-center-for-artificial-intelligence-dfki> <http://xmlns.com/foaf/0.1/nar
```

In full turtle notation, having introduced the appropriate prefixes, the above might instead have been rendered as:

```
orgdata:german-research-center-for-artificial-intelligence-dfki a foaf:Organization;
  foaf:name "DFKI"^^xs:string.
```

Look at each RDF dataset; you can use a text editor for this, or if you have a working installation of [Protégé](#), that tool might be helpful as well. Observe: What individuals, concepts, and relations do you find in each case? On this basis, which of the topic annotations suggested as alternatives within the survey is the least suitable?

Given the dataset `Oca`, which of the following 4 topics does **not** fit to the dataset?

- College basketball**  
state, basketball, conference, university, ncaa, kentucky, carolina, tournament, man, college
- Philosophy**  
language, study, social, theory, university, work, culture, history, press, philosophy
- Information**  
use, system, software, user, datum, computer, Include, information, support, service
- University**  
university, college, research, science, institute, professor, award, work, study, society

Choose one of the following answers

Topic 1

Topic 2

Topic 3

Topic 4

No answer

Please enter your comment here:

The screenshot shows the Protégé ontology editor. The main window displays the class hierarchy on the left, with 'Proceedings-Paper-Reference' selected. The right pane shows the annotations for a specific instance, including 'has-author', 'has-date', 'has-page-numbers', and 'has-title'. The title is 'The chronology of the Viking Age in the Isle of Man'.

### 19.2 SPARQL querying Wikidata

[https://home.bawue.de/~horsch/teaching/co3409/tutorial/Lab\\_19\\_RDF\\_SPARQL.html](https://home.bawue.de/~horsch/teaching/co3409/tutorial/Lab_19_RDF_SPARQL.html)

The [Wikidata SPARQL end point](#) is a good device for training yourself in the practical use of SPARQL. The documentation contains a long [list of query examples](#). The IRIs used by Wikidata are resolvable, employing the following prefixes:

```
@prefix wd: <https://wikidata.org/wiki/>
@prefix wdt: <https://wikidata.org/wiki/Property:>
```

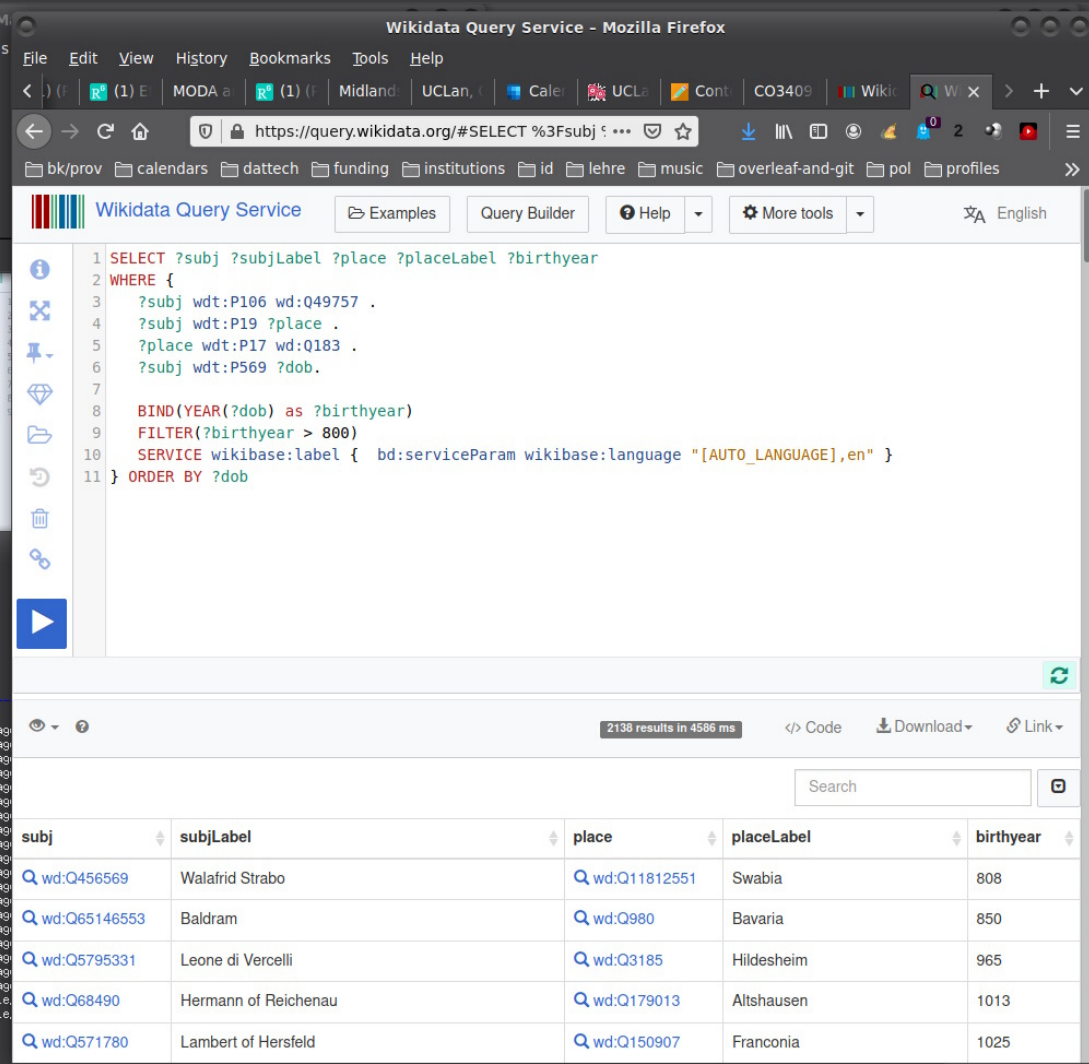
Accordingly, consider the following query from the [list of examples](#):

```
# Birth places of German poets
#
SELECT ?subj ?subjLabel ?place ?placeLabel ?birthyear
WHERE {
  ?subj wdt:P106 wd:Q49757 .
  ?subj wdt:P19 ?place .
  ?place wdt:P17 wd:Q183 .
  ?subj wdt:P569 ?dob.

  BIND(YEAR(?dob) as ?birthyear)
  FILTER(?birthyear > 800)
  SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGUAGE],en" }
} ORDER BY ?dob
```

For example, in the first triple, [wdt:P106](#) expands to <https://wikidata.org/wiki/Property:P106>, which is an object property labelled "occupation." The whole triple has the meaning "?subj [has the occupation poet](#)."

Practice formulating your own queries; for example, try asking for a table of Nobel laureates who are/were affiliated with the [University of Manchester](#).



The screenshot shows the Wikidata Query Service interface. The query editor contains the following SPARQL query:

```
1 SELECT ?subj ?subjLabel ?place ?placeLabel ?birthyear
2 WHERE {
3   ?subj wdt:P106 wd:Q49757 .
4   ?subj wdt:P19 ?place .
5   ?place wdt:P17 wd:Q183 .
6   ?subj wdt:P569 ?dob.
7
8   BIND(YEAR(?dob) as ?birthyear)
9   FILTER(?birthyear > 800)
10  SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGUAGE],en" }
11 } ORDER BY ?dob
```

The results table shows 5 rows of data:

subj	subjLabel	place	placeLabel	birthyear
<a href="#">Q456569</a>	Walafrid Strabo	<a href="#">Q11812551</a>	Swabia	808
<a href="#">Q65146553</a>	Baldram	<a href="#">Q980</a>	Bavaria	850
<a href="#">Q5795331</a>	Leone di Vercelli	<a href="#">Q3185</a>	Hildesheim	965
<a href="#">Q68490</a>	Hermann von Reichenau	<a href="#">Q179013</a>	Altshausen	1013
<a href="#">Q571780</a>	Lambert of Hersfeld	<a href="#">Q150907</a>	Franconia	1025

If you are interested in feedback on your work, send an email to [Aaron Bryant and Martin Horsch](#).