Foundations of Semantic Web Technologies

Tutorial 7

Dörthe Arndt

WS 2022/23

Exercise 7.1. Wikidata is a dataset proposed to better organise the data of Wikipedia (used for lists, tables, infoboxes, etc.) in a central, language-agnostic, structured repository. Wikidata is used not only by Wikipedia, but also Google[1], Alexa[2], Siri[3], etc. The dataset is published using Semantic Web standards and provides a SPARQL endpoint that we will use for today's lab to try out some SPARQL queries.

Navigate to `http://query.wikidata.org/`. Here you will find the SPARQL interface for Wikidata. Try this query, which returns the Spanish names of Chilean cities:

```
SELECT ?cityLabel
WHERE {
  ?city wdt:P31/wdt:P279* wd:Q515 ;
  wdt:P17 wd:Q298 ;
  rdfs:label ?cityLabel .
  FILTER (lang(?cityLabel)="es")
}
```

Note the following:

- Rather than use identifiers like `ex:Bejing`, which are biased to a particular language , Wikidata uses numeric identifiers like `wd:Q956`; as an additional benefit, when a city changes its name from Peking to Bejing, for example, it is not necessary to change the identifier or maintain multiple identifiers. The downside is that the query is more difficult to read but you can hover over the IRI term in the query interface to get an explanation.
- In order to find identifiers for entities (like `wd:Q515`) you can write `wd:` in the interface, hold Ctrl and press Spacebar; this opens a search box were you can type (e.g.) `chile` and get suggested identifiers. On the other hand, for properties, you can type `wdt:` and then hold Ctrl and press Spacebar.
- Wikidata uses its own properties for some properties already defined by the standards. The property `wdt:P31` denotes INSTANCE-OF, which serves the same role as `rdf:type`. The property `wdt:P279` denotes SUBCLASS-OF which serves the same role as `rdfs:subClassOf`. The expression `wdt:P31/wdt:P279*` is a property path that matches nodes connected to `?city` by one `wdt:P31` edge and zero-or-many `wdt:P279` edges; this pattern matches instances of classes and their transitive sub-classes.
- Wikidata provides human-readable labels for entities in many languages; the filter asks for Spanish labels.

In the following queries you should return names in English where applicable (you'll find more results for English).[4] You should return distinct results in each case. You can assume that each entity has one English label (using `rdfs:label`), but you should not assume that each English label is unique to that entity. Note that if your favourite entity of type x provides empty results, you should select your other favourite entity of type x.

[1] `https://static.googleusercontent.com/media/research.google.com/en//pubs/archive/44818.pdf`
[2] `https://www.wired.com/story/inside-the-alexa-friendly-world-of-wikidata/`
[3] `https://io9.gizmodo.com/siri-erroneously-told-people-stan-lee-was-dead-1827322243`
[4] Use `FILTER(lang(?someLabel)="en")`

(a) Find the names of video games set on Mars.[5]

(b) Find the names of the *other* videogames in the same series as your favourite video game (the results should not include your favourite video game).[6]

(c) Find the name of the highest grossing video game and its revenue.[7]

(d) Find the names of video games directed by women and optionally their direct sequel; return the name of the video game, the director, and the sequel (if any).[8]

(e) Find the names of video game series with at least 25 video game characters; return also the number of characters (descending order) and the name of a random character.[9]

(f) Find the names of exclusive video games for the Sega Dreamcast (only released on that platform) and how many *transitive* sequels they had (returning 0 if they had none[10]).[11]

(g) Find the names of video game characters that appear in more than one video game series and how many different video games they have appeared in.[12]

[5] Not Danté.
[6] Not Dante.
[7] Not Dante.
[8] Not Dante.
[9] Not Dante.
[10] Some sequels might be missing ... you could add them to Wikidata if you wished!
[11] Not Dante.
[12] ...