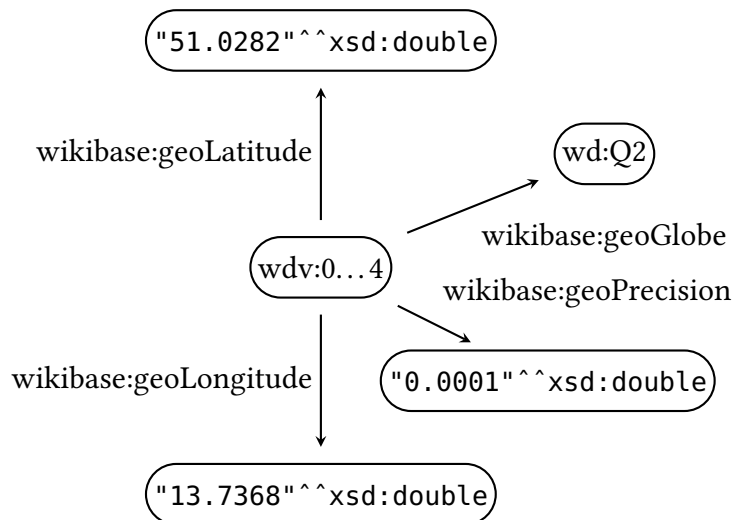# Exercise Sheet 12: Knowledge Graph Quality and Validation
Maximilian Marx, Markus Krötzsch
Knowledge Graphs, 2020-01-21, Winter Term 2019/2020

Exercise 12.1. The following extract from Wikidata shows how geographic coordinates are encoded. Develop a SHACL schema that validates statement values for geographic coordinates in Wikidata.

Hint: Refer to the RDF Dump Format description[1] for details on the encoding.

Exercise 12.2. Show that deciding whether a given RDF graph is valid with respect to some fixed ShEx schema is NP-hard by reducing from 3-colourability.

Hint: You can use the RDF Shape playground[2] to test ShEx validation.

Exercise 12.3. Wikidata Property Constraints[3] are a mechanism to specify how properties should be used on Wikidata. As an example, an Inverse Constraint[4] specifies that every statement for a given property must have a matching statement in the reverse direction using some other property (e.g., every "mother" statement must have a matching "child" statement).

Use the VLog4j client[1] and the Wikidata Query Service[5] to find statements violating an Inverse Constraint:

- write a SPARQL query to find all Inverse Constraints and the related properties,

- write a SPARQL query that finds violating statements for a given pair of forward and inverse properties,

- write a rules program that combines these two SPARQL data sources to obtain all statements violating Inverse Constraints.

---

[1] https://www.mediawiki.org/wiki/Wikibase/Indexing/RDF_Dump_Format#Globe_coordinate

[2] http://rdfshape.weso.es/validate

[3] https://www.wikidata.org/wiki/Help:Property_constraints_portal

[4] https://www.wikidata.org/wiki/Help:Property_constraints_portal/Inverse

[1] https://github.com/knowsys/vlog4j/wiki/Standalone-client

[5] https://query.wikidata.org

Hint: Finding all violations for all inverse constraints might take a long time. For testing, limit your queries to, e.g., 10 pairs of properties. To achieve that for VLog4j data sources, note that you can nest a subquery inside a graph pattern.

Exercise 12.4. Participants and winners of sports tournaments are modelled in Wikidata using properties P1334 ("participant of") and P2522 ("victory").

Write a program that, using the Wikidata Query Service,[5] extracts the subgraph of Wikidata where there is an edge from vertex w to vertex v if v is a participant of some tournament with winner w, and produces as output two files containing

- the graph in METIS graph format (cf. Exercise sheet 1), and

- and a dictionary mapping every vertex ID to the English label of the corresponding Wikidata item (with each line being of the form n,"l", where n is the vertex ID and l is the item label), respectively.