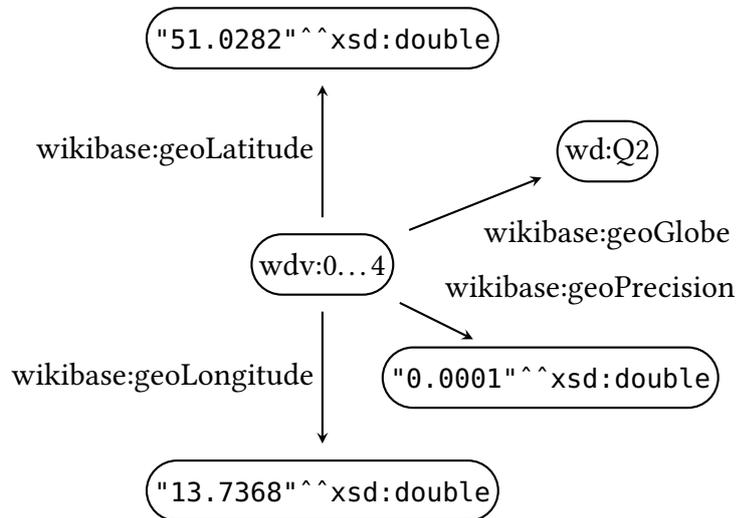


## Exercise Sheet 11: Knowledge Graph Quality and Validation

Maximilian Marx, Markus Krötzsch

Knowledge Graphs, 2019-01-15, Winter Term 2018/2019

**Exercise 11.1.** 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<sup>1</sup> for details on the encoding.

**Exercise 11.2.** 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<sup>2</sup> to test ShEx validation.

**Exercise 11.3.** Wikidata Property Constraints<sup>3</sup> are a mechanism to specify how properties should be used on Wikidata. As an example, an Inverse Constraint<sup>4</sup> 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 Wikidata Query Service<sup>5</sup> to find statements violating an Inverse Constraint:

- write a SPARQL query to find all Inverse Constraints and the related properties, and
- extend this to find statements violating these constraints.

**Hint:** To avoid exceeding the available memory, use LIMIT/OFFSET to restrict the query to 5 constraints at a time. You may also need to disable the query optimiser (using `hint:Query hint:optimizer "None" .`) and manually enforce the join order to prevent this query from timing out.

<sup>1</sup>[https://www.mediawiki.org/wiki/Wikibase/Indexing/RDF\\_Dump\\_Format#Globe\\_coordinate](https://www.mediawiki.org/wiki/Wikibase/Indexing/RDF_Dump_Format#Globe_coordinate)

<sup>2</sup><http://rdfshape.weso.es/validate>

<sup>3</sup>[https://www.wikidata.org/wiki/Help:Property\\_constraints\\_portal](https://www.wikidata.org/wiki/Help:Property_constraints_portal)

<sup>4</sup>[https://www.wikidata.org/wiki/Help:Property\\_constraints\\_portal/Inverse](https://www.wikidata.org/wiki/Help:Property_constraints_portal/Inverse)

<sup>5</sup><https://query.wikidata.org>

**Exercise 11.4.** 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<sup>5</sup> 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.