sheXer

Automatic Extraction of Shape Expressions In Wikidata

Contact: Daniel Fernández-Álvarez
fernandezalvdaniel@uniovi.es
WESO Research Group

Wikidata ❤ ShEx

... and they are such a wonderful couple! 😍

What is ShEx?

The **Shape Expressions (ShEx)** language describes sturctures called shapes in RDF graphs.

As you may know, ShEx has been recently adopted by Wikidata as a tool for describing groups of entities of different domain models...

Toy example of a shape human (Q5) with some comments:

```
<human> {
  wdt:P31 [wd:Q5]; # humans are instance of (P31) the entity human (Q5)
  wdt:P21 [wd:Q6581097 wd:Q6581072 wd:Q1097630 wd:Q1052281 wd:Q2449503 wd:Q48270];
  # their gender (P21) is one of male (Q6581097), female (Q6581072), intersex (Q1097630), non-binary (Q48270)...
  wdt:P19 IRI; # They have one place of birth (P19)
  wdt:P106 IRI * ; They can have several occupations (P106) }
```

You can check **Wikiproject ShEx** using the following QR:

TIP: Instead of handcrafting each shape from scratch, you can use automatic processes to detect the most common features of a set of entities. For such a goal, we provide the free tool **sheXer**.

sheXer workflow

1. **Select the graph** to be considered: a local RDF file, some subgraph collected using an endpoint...
2. **Choose the shapes** you want to extract and the entities related to them. Typically, each set is composed by instances of a class, but custom node agrupattions can be selected as well.
3. **Run and get** the shapes!

How to use sheXer

You can try an easy-to-use **online demo** deployed at shexer.weso.es or the **library 'shexer' for Python**.

Online demo | Library at GitHub

Main current features

- Shapes compilant with the defined standards
- Scores of trustworthiness for each constraint
- Detection of swapped constraints
- Publicly available library and source code (Python) at GitHub
- Algorithm compatible with any RDF graph, not just Wikidata
- Recognition of IRIs and different kinds of literals
- Value sets in triples with instance of (P31)
- Many confifuration options for adapting the process to your needs
- Shapes interlinkage

Universidad de Oviedo