Lab
Constraints and Metrics

Nicoleta Preda
nicoleta@preda.fr

January 31, 2025

1 SHACL Validation of an RDF Graph

SHACL Validation Processor We will use the open-access processor available at the following address: https://shacl.org/playground/.

New Dataset Consider the graph given in the file `rdf-graph.ttl` as an example. Propose SHACL shapes (constraints) to express the following constraints. Verify the validity of the solutions by extending the graph. In the report, it is important to include these extensions in addition to the SHACL constraints.

(Exercise 1.1) Define a constraint indicating that each student must have an email and a Social Security Number (SSN).

(Exercise 1.2) Ensure that the SSN is a string and that the email is a string.

(Exercise 1.3) Add a constraint to enforce that each student must have **exactly** one SSN.

(Exercise 1.4) Add a constraint to specify that a student may have at most one email.

(Exercise 1.5) Specify a constraint to enforce that the email must follow a valid format (e.g., `name@example.com`).

(Exercise 1.6) Add a constraint ensuring that the SSN is unique—two students cannot have the same SSN.

(Exercise 1.7) Add a constraint to enforce that a student must have a supervisor and be affiliated with a `schema:ResearchOrganization`.

(Exercise 1.8) Ensure that a student's supervisor is a person affiliated with a research unit `schema:ResearchOrganizat`

(Exercise 1.9) Define a constraint to enforce that an organization must have a name (`rdfs:label`).

(Exercise 1.10) Ensure that an organization has only one denomination per language.

(Exercise 1.11) All PhD advisors (see the predicate `ex:phdAdviser`) must hold a doctorate.

(Exercise 1.12) All supervisors of a PhD student must be affiliated with the same university as the student.

(Exercise 1.13) At least one of the supervisors of a PhD student must be affiliated with the same university as the student.

(Exercise 1.14) PhD students enrolled in a German university cannot have more than one supervisor.

More Abstract Questions:

(Exercise 1.15) Which of these constraints cannot be expressed in SHACL?

(Exercise 1.16) Propose a test to verify whether the processor supports type inference (see the lecture section *How to specify target nodes?*).

2 Evaluation of Rule Quality Metrics

Objective Formulate and implement in SPARQL calculations to assess the quality of automatic inference rules.

Exercise 3.1 Define a metric to identify functional predicates or paths in an RDF graph, corresponding intuitively to functional dependencies in a relational database.

Exercise 3.2 Translate the defined formula into a SPARQL query to compute this metric on an RDF graph.

Testing Platforms The platforms **Wikidata** and **YAGO** provide SPARQL endpoints accessible online:

- Wikidata SPARQL Endpoint
- YAGO SPARQL Endpoint

To explore the available predicates on Wikidata, consult the complete list of properties: Wikidata Property List.