Introduction to SHACL

Jan Van den Bussche

Data Science Institute, Hasselt University

Shapes Constraint Language (SHACL)

- A language for expressing integrity constraints on RDF graphs
- W3C standard

1. RDF graph
2. Shape
3. Constraint

RDF graph

- directed graph with labels on edges
- edge labels are called “properties”
- edge $x \rightarrow y$ with label p:
  - x is called the “subject”
  - y is called the “object”
  - p is called the “property”

- In RDF:
  - properties can also be nodes
    - SHACL, however, is oblivious to this
  - nodes can be of different kinds (IRI, blank, literal)
    - SHACL provides tests for this

Shape

- a unary query over RDF graphs
- a property of nodes of RDF graphs
  - node under consideration is called “focus node”

- Examples: let x denote the focus node
  - “x has a phone property, but no email”
  - “x has at least five managed-by edges”
  - “x has a path of friend-edges to the CEO of Apple”
  - “x has no other properties than name, address, and birthdate”

Constraints

- We are slightly generalizing SHACL here
- Constraints are expressed as **inclusions** between shapes:

$$\varphi_1 \subseteq \varphi_2$$

where φ_1 and φ_2 are shapes

  - referred here to as left-hand shape (**lhs**) and right-hand shape (**rhs**)

- "Every node satisfying φ_1 also satisfies φ_2"

Examples of inclusions

- “Every node of type Person has a phone or email property”
  - lhs: nodes with an edge labeled type to node Person
  - rhs: nodes with a phone or email property
- “Different nodes never have the same email”
  - lhs: nodes with an incoming email edge
  - rhs: nodes that do not have two or more incoming email edges
- “Every Mathematician has a finite Erdös number”
  - lhs: nodes of type Mathematician
  - rhs: nodes that can reach node Erdös by a property path $(\text{author}^{-}/\text{author})^{*}$

Targets

- In reality, SHACL does not have these arbitrary left-hand shapes
- Instead, right-hand shapes are associated to “targets”
- Targets are simple shapes of four kinds:
  - **node**: a constant node
  - **class** : nodes with a path $(\text{type}/\text{subclass}^*)$ to some constant
  - **subjects-of**: nodes with an outgoing edge of some label
  - **objects-of**: nodes with an incoming edge of some label

SHACL

- SHACL, a language for expressing shapes
  - and associating them with targets to form inclusion constraints
- The syntax is in RDF!
- “Shapes graph”
- Description-logic syntax was introduced
  - Corman et al. ISWC 2018; Andresel et al. WWW 2020
- Extended by Jakubowski (Delva et al. EDBT 2023)
  - equivalent to full SHACL
  - SLS parser https://github.com/MaximeJakubowski/sls_project

SHACL Logical Syntax

$$
\begin{aligned}
E &::= p \mid E^{-} \mid E/E \mid E \cup E \mid E^{*} \mid E? \\
F &::= E \mid \mathrm{id} \\
\phi &::= \top \mid \bot \mid hasShape(s) \mid test(t) \mid hasValue(c) \\
&\quad \mid eq(F, p) \mid disj(F, p) \mid closed(P) \\
&\quad \mid lessThan(E, p) \mid lessThanEq(E, p) \mid uniqueLang(E) \\
&\quad \mid \neg\phi \mid \phi \wedge \phi \mid \phi \vee \phi \\
&\quad \mid \geq_n E.\phi \mid \leq_n E.\phi \mid \forall E.\phi
\end{aligned}
$$

Some research on the theory of SHACL

- Expressiveness (Bogaerts et al. ICDT 2023)
- Satisfiability, containment
  - Pareti et al.; Leinberger et al.; both ISWC 2020
- Recursion
  - Corman et al., Andresel et al.
  - Bogaerts & Jakubowski ICLP 2021
  - Chmurovic et al. Datalog 2022
- Logical entailment from ontologies (Ahmetaj et al. ECAI 2023)

SHACL engines

- pySHACL
- Apache Jena
- TopBraid
- some more engines, e.g.,
  - **shaclex** (Labra)
  - Cem Okulmus

Some SHACL systems research

- SHACL2SPARQL (Corman et al. ISWC 2019)
- TravSHACL (Figuera et al. WWW 2021)
- Magic Sets optimization (Ahmetaj et al. VLDB 2022)

Some research on new applications of SHACL

- mining SHACL shapes
  - Rabbani et al. VLDB 2023
- semantic SPARQL optimization (Rabbani et al. EDBT 2021)
- inferring shapes (Dimou)
- explaining & repairing SHACL constraint violations
  - Ahmetaj et al., KR 2021, ISWC 2022
- provenance, shape fragments (Delva et al. EDBT 2023)
- access control (?)

Other approaches

- Many of the mentioned research topics have also been pursued for other approaches
  - ShEx
  - property graphs
  - and more
- see this seminar!

Upload your slides!

- Homepage of this Dagstuhl seminar
  - google [Dagstuhl shapes theory implementation]
- Follow link to “Materials” site
  - only accessible for participants
- We’ll also have to make a report