SHACL: A Description Logic in Disguise

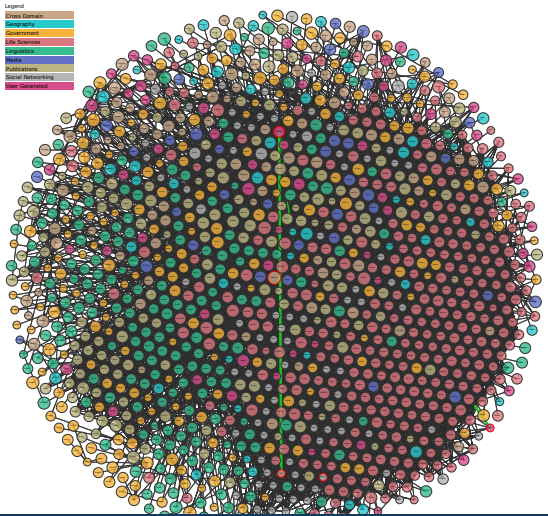
BNAIC 2021

Bart Bogaerts, *Maxime Jakubowski* & Jan Van den Bussche

Vrije Universiteit Brussel & Universiteit Hasselt

November 2021

Semantic Web & Linked Data



- Enormous datasets
- Data quality: constraints
- Shapes Constraint Language (SHACL)

OWL

- Description Logic $\mathcal{SROIQ}$
- Modeling knowledge
- Deductive reasoning
- Example:
  "Every book has a title"

```
Ontology(:BookOntology
  SubClassOf(
    v:Book
    ObjectMinCardinality( 1
      v:title
      owl:Thing)))
```

OWL

- Description Logic $\mathcal{SROIQ}$
- Modeling knowledge
- Deductive reasoning
- Example:
  "Every book has a title"

```
Ontology(:BookOntology
  SubClassOf(
    v:Book
    ObjectMinCardinality( 1
      v:title
      owl:Thing)))
```

SHACL

- [Corman 2018]
- Modeling constraints
- Validation
- Example:
  "Every book has a title"

```
:BookShape
  a sh:PropertyShape;
  sh:path v:title;
  sh:minCount 1.

:BookShape sh:targetClass v:Book.
```

The Wedge

1. Both SHACL and OWL are used for *modeling tasks*
2. OWL has its *logical foundations* in Description Logic
3. SHACL has its *logical foundations* in ??
4. The languages are very similar at their core

> *"OWL was inspired by and designed to exploit 20+ years of research in Description Logics (DL). [...] there is little connection between this research and the practical data modeling needs of the common real world software systems."*

The Wedge

1. Both SHACL and OWL are used for *modeling tasks*
2. OWL has its *logical foundations* in Description Logic
3. SHACL has its *logical foundations* in **Description Logic as well!**
4. The languages are very similar at their core

> *"OWL was inspired by and designed to exploit 20+ years of research in Description Logics (DL). [...] there is little connection between this research and the practical data modeling needs of the common real world software systems."*

SHACL

Shape expressions

$$E ::= p \mid p^- \mid E \cup E \mid E/E \mid E^* \mid E?$$
$$\phi ::= \top \mid s \mid \{c\} \mid \phi \wedge \phi \mid \phi \vee \phi \mid \neg\phi \mid \forall E.\phi \mid \geq_n E.\phi \mid eq(p, E) \mid disj(p, E) \mid closed(Q)$$

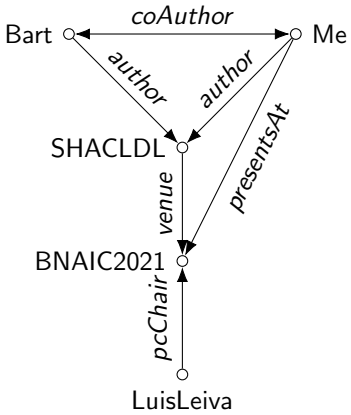
Schema

- Shape definitions: $s \leftarrow \phi$
- Target inclusions: $\phi \subseteq s$

Validation

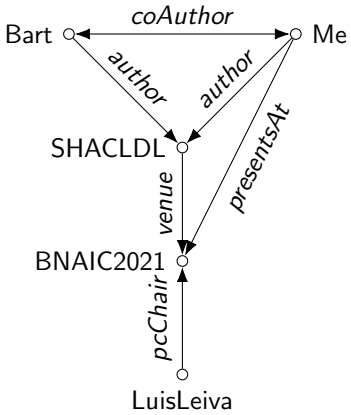
- Given an *RDF graph* G and a shape schema $\mathcal{S}$
- Does G conform to $\mathcal{S}$?

SHACL examples



$BNAICAuthor \leftarrow \exists author/venue.\{BNAIC2021\}$

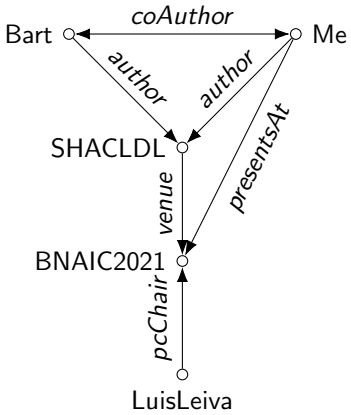
SHACL examples



$BNAICAuthor \leftarrow \exists author/venue.\{BNAIC2021\}$

$\exists presentsAt.\{BNAIC2021\} \subseteq BNAICAuthor$

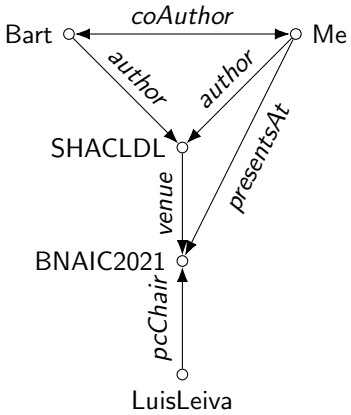
SHACL examples



$BNAICAuthor \leftarrow \exists author/venue.\{BNAIC2021\}$
$NotBNAICAuthor \leftarrow \neg BNAICAuthor$

$\exists presentsAt.\{BNAIC2021\} \subseteq BNAICAuthor$

SHACL examples



$BNAICAuthor \leftarrow \exists author/venue.\{BNAIC2021\}$
$NotBNAICAuthor \leftarrow \neg BNAICAuthor$

$\exists presentsAt.\{BNAIC2021\} \subseteq BNAICAuthor$
$\{LuisLeiva\} \subseteq NotBNAICAuthor$

Description Logics

Ontology / Knowledge Base

- Terminology (TBox): what are the concepts and their relations?
- Assertions (ABox): what is the known information?

Description Logics

Ontology / Knowledge Base

- Terminology (TBox): what are the concepts and their relations?
- Assertions (ABox): what is the known information?
- Example:

TBox:

$$\mathit{Author} \sqsubseteq \mathit{Human} \sqcap \exists \mathit{hasWritten} \,.\, \mathit{Publication}$$

ABox:

$$\mathit{Author} \,:\, \mathit{tolkien}$$

$$\mathit{hasWritten} \,:\, (\mathit{tolkien}, \mathit{fotr})$$

SHACL as a Description Logic

- TBox is a finite set of shape inclusions, given by the shape schema
  - Definitions: :BookShape $\equiv \exists$:title . $\top$
  - Targeting: $\exists$:writtenBy . $\top \sqsubseteq$:BookShape
- There is no ABox

SHACL as a Description Logic

- TBox is a finite set of shape inclusions, given by the shape schema
  - Definitions: :BookShape $\equiv \exists$:title . $\top$
  - Targeting: $\exists$:writtenBy . $\top \sqsubseteq$:BookShape
- There is no ABox

... but what then does the RDF graph *represent*?

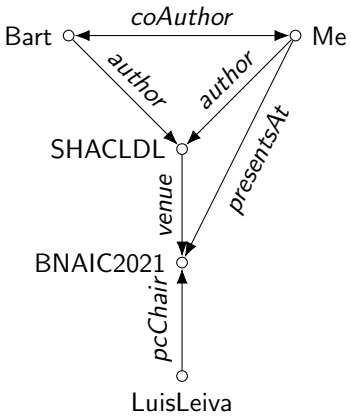
What's in an RDF graph?

- A graph is a finite set of *facts*
- A fact is of the form $p(a, b)$ with p a property name and a, b nodes of G.

We associate to any given graph an interpretation I:

- The domain is the universe of *all nodes*
- Every constant is interpreted as itself
- The interpretation of a property name is fixed by the facts

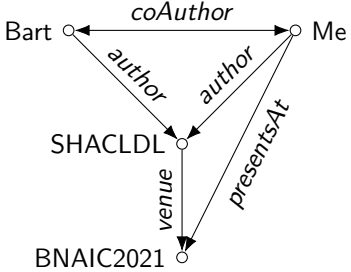
Illustration



$NotBNAICAuthor \leftarrow \neg\exists author/venue.\{BNAIC2021\}$

$\{LuisLeiva\} \subseteq NotBNAICAuthor$

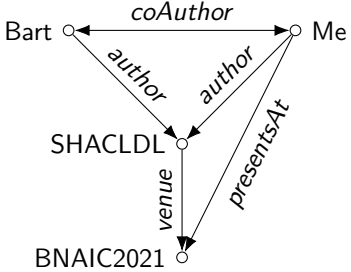
Illustration



$NotBNAICAuthor \leftarrow \neg\exists author/venue.\{BNAIC2021\}$

$\{LuisLeiva\} \subseteq NotBNAICAuthor$

Illustration

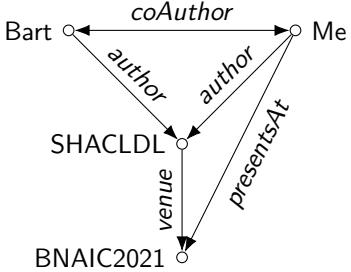


$NotBNAICAuthor \leftarrow \neg\exists author/venue.\{BNAIC2021\}$

$\{LuisLeiva\} \subseteq NotBNAICAuthor$

- $NotBNAICAuthor$ evaluates to $N - \{Bart, Me\}$

Illustration

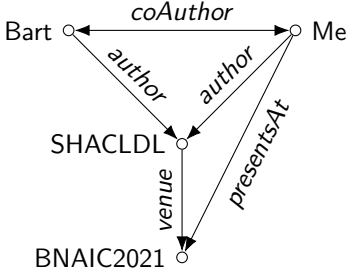


$NotBNAICAuthor \leftarrow \neg\exists author/venue.\{BNAIC2021\}$

$\{LuisLeiva\} \subseteq NotBNAICAuthor$

- $NotBNAICAuthor$ evaluates to $N - \{Bart, Me\}$
  ... because the domain is the universe of all nodes (N)

Illustration

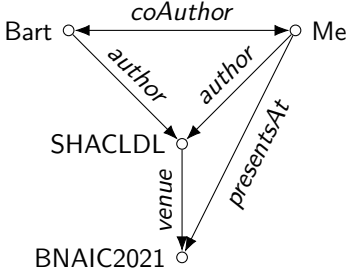


$NotBNAICAuthor \leftarrow \neg\exists author/venue.\{BNAIC2021\}$

$\{LuisLeiva\} \subseteq NotBNAICAuthor$

- $NotBNAICAuthor$ evaluates to $N - \{Bart, Me\}$
  ... because the domain is the universe of all nodes (N)
- $\{LuisLeiva\}$ evaluates to $\{LuisLeiva\}$

Illustration

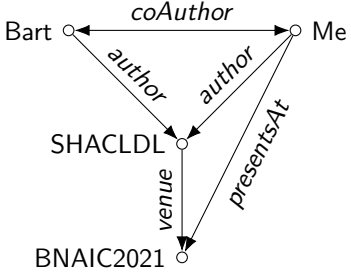


$NotBNAICAuthor \leftarrow \neg\exists author/venue.\{BNAIC2021\}$

$\{LuisLeiva\} \subseteq NotBNAICAuthor$

- $NotBNAICAuthor$ evaluates to $N - \{Bart, Me\}$
  ... because the domain is the universe of all nodes (N)
- $\{LuisLeiva\}$ evaluates to $\{LuisLeiva\}$
  ... because all constants are interpreted as themselves

Illustration



$NotBNAICAuthor \leftarrow \neg\exists author/venue.\{BNAIC2021\}$

$\{LuisLeiva\} \subseteq NotBNAICAuthor$

- $NotBNAICAuthor$ evaluates to $N - \{Bart, Me\}$
  ... because the domain is the universe of all nodes (N)
- $\{LuisLeiva\}$ evaluates to $\{LuisLeiva\}$
  ... because all constants are interpreted as themselves

$\Longrightarrow$ This is also the behavior of **real** SHACL!

SHACL *is* a Description Logic

- It is of value to put emphasis on the formalization of SHACL
  - What does the RDF graph represent?
  - What are the exact semantics of the language?
- The discrepancy between the views on modeling can be summarized as:
  - In OWL, the graph is a *first order-theory* (ABox) and the task is *deduction*
  - In SHACL, the graph is a *first order-interpretation* and task is *model checking*
- We *can* exploit many years of research in Description Logics, e.g.,

Theorem

Consistency of a shape schema is undecidable.