Ontology-Based RDF Integration of Heterogeneous Data

Maxime Buron, François Goasdoué, Ioana Manolescu, Marie-Laure Mugnier
EDBT 2020
Contributions

1. More powerful integration setting:
   - Global-Local-As-View mappings in an OBDA context
   - Queries on the data and the ontology

2. Two novel query answering methods: based on the amount of reasoning performed offline and at query time
Global-Local-As-View Mapping

GLAV mapping

Data graph:
$q_2(x)$ head

Data sources:
$q_1(x)$ body
Global-Local-As-View Mapping Example
### RDFS Ontology

<table>
<thead>
<tr>
<th>data graph</th>
<th>$q_2(x)$</th>
<th>ontology</th>
</tr>
</thead>
<tbody>
<tr>
<td>data sources</td>
<td>$q_1(x)$</td>
<td></td>
</tr>
</tbody>
</table>
RDFS Entailment of Data Triples

saturated graph

$\mathbf{R}_{data}$

data graph

$q_2(x)$

ontology

$q_1(x)$

data sources
RDFS Entailment of Ontological Triples

\[ q_1(x) \xrightarrow{R_{\text{data}}} q_2(x) \xrightarrow{R_{\text{onto}}} \]

saturated graph \quad saturated ontology

data graph \quad ontology

data sources
Query Answering Problem

saturated graph

data graph

data sources

\[ q_1(x) \]

\[ q_2(x) \]

Ontology-Based RDF Integration of Heterogeneous Data, Maxime Buron et al., EDBT 2020
Classical Method: All Reasoning at Query Time

$R_{\text{onto}}$ $R_{\text{data}}$ ontology

query $\rightarrow$ Reformulation $\rightarrow$ answers

$q_1(x)$ $\rightarrow$ Rewriting $\rightarrow$ In mediator evaluation

$q_2(x)$

data graph

data sources
Some reasoning at Query Time Method: Preprocessing
Some reasoning at Query Time Method: Query Time
No Reasoning at Query Time: Pre-Processing

Ontology-Based RDF Integration of Heterogeneous Data, Maxime Buron et al., EDBT 2020
No Reasoning at Query Time: Query Time

- saturated graph
- data graph
- data sources
- saturated ontology
- (x, subclassOf, y)
- query
- answers
- Rewriting
- In mediator evaluation

\[
q_1(x) \rightarrow q_{sat}^{sat}(x) \rightarrow (x, subclassOf, y)
\]
Experiments

- **Software:**
  - *OntoSQL* (reformulation and materialization)
  - *Graal* (rewriting)
  - *Tatooine* (mediation)

- **RDF Integration System:**
  - Extension of BSBM
  - 3863 GLAV mappings
  - RDFS ontology of 2011 triples
  - Induced graph with 108M triples (185M triples when saturated)
  - Two data sources: One *relational* and one *JSON*
Query Answering Times on Heterogeneous Data Sources

- Materialization (MAT) - kind of reference time
- Full reformulation + rewriting (REW-CA)
- Mapping saturation + partial reformulation + rewriting (REW-C)
Conclusion

- Global-Local-As-View mappings in OBDA Context
- Queries on data and ontology
- A new scalable query answering strategy using partial reformulation and saturated mappings

Obi-Wan demo at: http://pages.saclay.inria.fr/maxime.buron/projects/obi-wan/