Diego Calvanese Wolfgang Fischl Reinhard Pichler Emanuel Sallinger Mantas Šimkus
Free University of Bozen-Bolzano Vienna University of Technology

Capturing Functional Dependencies and Relational Schemas in RDFS

Motivation
- Enormous growth in Semantic Web data
- Most Semantic Web data originates from relational databases
- Normal forms for relational databases are important
  **Goal:** Is there a Normal Form for Semantic Web data?

Semantic Web
- Data is stored in RDF graphs extended with RDFS
- Focus on DL-LiteRDFS with disjointness

Relational Databases
- Data is stored in n-ary relations
- Functional dependencies (FDs) ensure consistency
- Normal forms avoid redundancy

Direct Mapping of Relational Data to RDFS Graph Data
- Each database instance corresponds to a model of the ontology and vice versa.

Example

<table>
<thead>
<tr>
<th>Table: loc</th>
<th>room</th>
<th>house</th>
<th>addr</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entrance</td>
<td>White House</td>
<td>1600 PA Av</td>
<td></td>
</tr>
<tr>
<td>Oval Office</td>
<td>White House</td>
<td>1600 PA Av</td>
<td></td>
</tr>
</tbody>
</table>

1. Translate schema information
   - Concept disjointness:
     functionality (funct has room) (funct has house)
   - Role typing:
     has room has house has addr
   - Translation does not preserve satisfaction of dependencies.

2. Translate data
   - Avoided by storing addr directly to the house.

Example

Consider the FD \( \{ A, B \} \rightarrow C \) and
its translation to a pid
\( C'(has_{C}^{-} \cdot has_{A}, has_{C}^{-} \cdot has_{B}). \)
Translation does not preserve satisfaction of dependencies.

Definition
An RDF graph is in RDF Normal Form (RNF) if and only if all tids can be substituted by tids with depth 1.

Theorem
A relational database is in BCNF if the RDF graph with tids (translated with the direct mapping) is in RNF.

Theorem
Deciding whether an RDF graph is in RNF is feasible in polynomial time.

Tree-based identification constraints (tids)

Definition
A tid is an expression constructed using the following grammar:
\[ \tau ::= \epsilon | \sigma \cdot \tau | (\tau, \tau) \]
where \( \sigma \) is an ordinary role or a test role of the form \( C'\), and \( \tau \) is a concept.

Example
The tid \( C'(has_{C}^{-} \cdot (has_{A}, has_{B})) \) captures the FD \( \{ A, B \} \rightarrow C \).

Theorem
Tree-based identification constraints are able to express FDs over RDF graphs.