RELATIONAL DATABASE SCHEMA TO ONTOLOGY MAPPING APPROACHES

by
Yassaman Zand Moghaddam and Joe D. Horton
November 10th 2010
Outline

- A brief description of mapping and correspondence
- Correspondence Patterns
- Database to ontology mapping process
  - A naïve solution
- Relational database to ontology mapping approaches
- Ontology matching
- Matching evaluation parameters
- The idea
- Evaluation method
- An overview of the MapOnto tool
What do “Mapping” and “Correspondence” mean?

- **Mapping** is the process of finding relationships or correspondences between entities.
- **Correspondence** is the relation holding or supposed to hold according to a particular mapping algorithm between entities.
- Each correspondence set a bridge between entities.
Correspondence patterns (Scharffe & Fensel, 2008) are like templates to model ontology alignments (The Alignment is the output of ontology mapping). They capture regularities recurring when aligning ontologies.

In many applications and domains, specific correspondence patterns exist.

✓ Relational database to ontology mapping is one of those applications.
Correspondence patterns identified in relational database to ontology mapping application:

- Direct Mapping
- Join/Union Mapping
- Projection
- Selection
- Value transformation
- Combinations
At a conceptual level a database and an ontology are semantically related and correspondences are established between the database components and the ontology components.
Database to Ontology Mapping Approaches

Figure 1. Classification of database to ontology mapping approaches (Ghawi & Cullot, 2007)
Discovering mappings between a relational database schema and an ontology usually has two phases:

- Searching for simple mappings between entities in the relational database schema and the ontology.
- Constructing complex compositions based on simple mappings.
A naïve solution

- Find the “shortest” connection between concepts.
- It involves:
  - Finding the minimum spanning tree(s) connecting the corresponded concepts
  - Encoding the tree(s) into formulas
An example of the naïve solution
(An, Borgida, & Mylopoulos, 2005)
An example of the naïve solution

**Simple correspondences:**
- $T: \text{Employee.ssn} \rightarrow O: \text{Employee.hasSsn}$
- $T: \text{Employee.name} \rightarrow O: \text{Employee.hasName}$
- $T: \text{Employee.dept} \rightarrow O: \text{Department.hasDeptNumber}$
- $T: \text{Employee.proj} \rightarrow O: \text{Worksite.hasNumber}$

**Expected formula:**
- $T: \text{Employee(ssn,name,dept,proj) :-}$
- $O: \text{Employee(x1), O:hasSsn(x1,ssn), O:hasName(x1,name),}$
- $O: \text{Department(x2), O:works_for(x1,x2), O:hasDeptNumber(x2,dept), O:Worksite(x3),}$
- $O: \text{works_on(x1,x3), O:hasNumber(x3,proj).}$
In some cases, because of the known relational schema design rules, the spanning tree may not provide the desired semantics for a table.

- **Project (name, manager)** -> “manager” is functionally dependent on “name”

  \[ T : \text{Project.name} \rightarrow O : \text{Worksite.hasName} \]

  \[ T : \text{Project.manager} \rightarrow O : \text{Employee.hasSsn} \]
We need to find a functional connection from “Worksite” to “Employee”
Mapping approaches

Some of the approaches that address the problem of mapping databases to ontologies are:

- RDB2Onto (Laclov’k, 2007)
- DB2OWL (Cullot, Ghawi, & Ytongnon, 2007)
- Relational OWL (Prez & Conrad, 2005)
- VisAVis (Konstantinou, Spanos, Chalas, Solidakis, & Mitrou, 2006)
- Discovering Simple Mappings Between Relational Database Schema and Ontologies (Hu & Qu, 2007)
Mapping approaches

- Another Literature review on the topic “Database to Ontology Mapping” in (Sataya S. et al., 2009), classifies the reviewed approaches based on:

  - Mapping Creation
  - Mapping Representation and Accessibility
  - Mapping Implementation
  - Query Implementation
  - Application Domain
  - Data Integration
Ontology matching

Figure 2. Classification of ontology matching techniques (Euzenat & Shvaiko, 2007)

Relational Database Schema to Ontology Mapping Approaches
Ontology matching

Figure 2. Classification of ontology matching techniques (Euzenat & Shvaiko, 2007)

Relational Database Schema to Ontology Mapping Approaches
Ontology matching

Figure 3. Extracted from Classification of ontology matching techniques (Euzenat & Shvaiko, 2007)

Matching Techniques

Element-level
- Syntactic
  - String-based
    - name similarity, description similarity, global namespace
  - Constraint-based
    - type similarity, key properties
- Structure-level
  - Syntactic

Relational Database Schema to Ontology Mapping Approaches

Granularity/Input Interpretation
- Basic techniques
  - Graph-based
    - graph homomorphism, path, children, leaves
- Kind of input
  - Terminological
  - Internal
  - Structural
Matching evaluation parameters

- Precision = \( \frac{|B|}{|B| + |C|} \)

- Recall = \( \frac{|B|}{|A| + |B|} \)

- F-Measure = \( \frac{|B|}{(1 - \alpha) \times |A| + |B| + \alpha \times |C|} \) \( 0 \leq \alpha \leq 1 \)

- Overall = \( 1 - \frac{|A| + |C|}{|A| + |B|} \)

1: True Correspondences
2: Derived Correspondences
A: False Negative   C: False Positive
B: True Positive    D: True Negative

Relational Database Schema to Ontology Mapping Approaches
The idea

**Inputs:**
- RDB Schema
- Ontology

**Intermediate Graph**

**Matching (Structure-based)**

**Outputs:**
- A set of correspondences
- A refined set of correspondences

Some string-based checks

Relational Database Schema to Ontology Mapping Approaches
An isomorphism of graphs $G$ and $H$ is a bijection between the vertex sets of $G$ and $H$ such that any two vertices $u$ and $v$ of $G$ are adjacent in $G$ if and only if $f(u)$ and $f(v)$ are adjacent in $H$.

$G$: $\begin{array}{c}
 a & g \\
 b & h \\
 c & i \\
 d & j \\
\end{array}$

$H$: $\begin{array}{c}
 1 & 2 \\
 5 & 6 \\
 8 & 7 \\
 4 & 3 \\
\end{array}$

$f: V(G) \rightarrow V(H)$

- $f(a) = 1$
- $f(b) = 6$
- $f(c) = 8$
- $f(d) = 3$
- $f(g) = 5$
- $f(h) = 2$
- $f(i) = 4$
- $f(j) = 7$

An isomorphism between $G$ and $H$:
String-based techniques are often used in order to match names and name descriptions of entities. These techniques consider strings as sequences of letters in an alphabet. In general, they are based on the intuition: the more similar the strings, the more likely they are to represent the same concepts. Usually, distance functions map a pair of strings to a real number, where a smaller value of the real number indicates a greater similarity between the strings. Examples: prefix, suffix, edit, and n-gram distances.
Evaluation Method

One method can be the repetition of the process in reverse which means:

✓ construction of the relational database schema and the ontology from intermediate graphs
✓ comparison of results with original ones (inputs).
✓ If they are equivalent, there is a high probability that the algorithm can work correctly on any inputs.
MapOnto Tool (An, Borgida, & Mylopoulos, 2006)
A system for constructing complex mappings between ontologies and relational or XML database schemas.

- The inputs of the tool are:
  - An ontology specified in an ontology representation language like OWL
  - A relational or XML database schema
  - A set of simple correspondences

- The output of the system is:
  - An ordered list of complex mapping formulas expressed in Horn clauses.

- The issue is:
  - Simple mappings between the relational database schema and the ontology are specified manually.
Datasets that were used in evaluation of the MapOnto tool are in different real world domains and were implemented independently.

<table>
<thead>
<tr>
<th>Database Schema</th>
<th>#Tables</th>
<th>#Columns</th>
<th>Ontology</th>
<th>#Nodes</th>
<th>#Links</th>
</tr>
</thead>
<tbody>
<tr>
<td>UTCS Department</td>
<td>8</td>
<td>32</td>
<td>Academic Department</td>
<td>62</td>
<td>1913</td>
</tr>
<tr>
<td>VLDB Conference</td>
<td>9</td>
<td>38</td>
<td>Academic Conference</td>
<td>27</td>
<td>143</td>
</tr>
<tr>
<td>DBLP Bibliography</td>
<td>5</td>
<td>27</td>
<td>Bibliographic Data</td>
<td>75</td>
<td>1178</td>
</tr>
<tr>
<td>OBSERVER Project</td>
<td>8</td>
<td>115</td>
<td>Bibliographic Data</td>
<td>75</td>
<td>1178</td>
</tr>
<tr>
<td>Country</td>
<td>6</td>
<td>18</td>
<td>CIA factbook</td>
<td>52</td>
<td>125</td>
</tr>
</tbody>
</table>

Table1. Characteristics of Schemas and ontologies for the experiments (An, Borgida, & Mylopoulos, 2006).
Thank You
References


References


References


References

