A Complex Alignment Benchmark: GeoLink Dataset

Lu Zhou, Michelle Cheatham, Adila Krisnadhi, Pascal Hitzler
Data Semantics Laboratory
Wright State University
ISWC 2018 Resources Track, Date: 10/10/2018
Ontology alignment (OA) is about mapping one ontology onto another. Think data integration.

Until recently, OA was heavily pre-occupied with just finding ways to map classes onto classes.

[Cheatham & H, ISWC 2013] demonstrated that the state of the art does hardly better than clever string matching.

Now there is growing awareness that we really need to study complex mapping rules.
What “rules” do we need?

Even for one-one class-to-class mappings it’s difficult to come up with a good benchmark, as humans disagree what the correct mapping would be.

This is even more tricky when it’s about complex mapping rules.

As a consequence, we have not yet really understood what “types” of mapping rules are in fact needed.

Creating synthetic benchmarks (to drive OA research on complex alignment) runs the risk of falling far from reality.
GeoLink: Completed NSF EarthCube Building Blocks project concerning the integration of ocean science data repositories.

Integration was done via an overarching ontology (schema), which was made for the purpose, the GeoLink Modular Ontology (GMO) [Krisnadhi+, ISWC 2015].

GMO followed ontology modeling best practices using ontology design patterns. It is a rather complicated ontology.
To make things easier for data providers and application developers, we developed a simplified version of the GMO, called the GeoLink Base Ontology (GBO).

GBO is smaller and straightforward.

Since it was constructed from the GMO, we know exactly how they align.

In fact, by design, the alignment was to be used to populate the GMO through the GBO.
We thus have:

- GMO – complex ontology
- GBO – simplified version
- A complex alignment between them which is by design the reference alignment.

We did not initially plan to propose this as a complex OA benchmark.

However it turns out that this natural alignment contains mappings which look rather odd from the perspective of the OA field’s preconceptions.
Geolink benchmark

The original reference alignment has 111 axioms.

- Some of them are not expressible in OWL.
- Some of them are not expressible in EDOAL (tool-supported OA language).
- Some of them were left out of this year’s OAEI benchmark as they were deemed too much out of scope for the current state of investigations.

Consequently, the OAEI benchmark has only 66 axioms.
• Size of the ontologies

<table>
<thead>
<tr>
<th>Ontology</th>
<th>Classes</th>
<th>Object Properties</th>
<th>Data Properties</th>
</tr>
</thead>
<tbody>
<tr>
<td>GeoLink Base Ontology</td>
<td>40</td>
<td>149</td>
<td>49</td>
</tr>
<tr>
<td>GeoLink Modular Ontology</td>
<td>156</td>
<td>124</td>
<td>46</td>
</tr>
</tbody>
</table>
12 types of correspondences

<table>
<thead>
<tr>
<th>Pattern</th>
<th>Occurrences</th>
<th>Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class Equivalence</td>
<td>10</td>
<td>1:1</td>
</tr>
<tr>
<td>Class Subsumption</td>
<td>2</td>
<td>1:1</td>
</tr>
<tr>
<td>Property Equivalence</td>
<td>7</td>
<td>1:1</td>
</tr>
<tr>
<td>Property Equivalence Inverse</td>
<td>5</td>
<td>1:1</td>
</tr>
<tr>
<td>Class Typecasting Equivalence</td>
<td>4</td>
<td>1:n</td>
</tr>
<tr>
<td>Class Typecasting Subsumption</td>
<td>1</td>
<td>1:n</td>
</tr>
<tr>
<td>Property Typecasting Subsumption</td>
<td>5</td>
<td>1:n</td>
</tr>
<tr>
<td>Property Typecasting Subsumption Inverse</td>
<td>5</td>
<td>1:n</td>
</tr>
<tr>
<td>Typed Property Chain Equivalence</td>
<td>26</td>
<td>m:n</td>
</tr>
<tr>
<td>Typed Property Chain Equivalence Inverse</td>
<td>17</td>
<td>m:n</td>
</tr>
<tr>
<td>Typed Property Chain Subsumption</td>
<td>17</td>
<td>m:n</td>
</tr>
<tr>
<td>Typed Property Chain Subsumption Inverse</td>
<td>12</td>
<td>m:n</td>
</tr>
</tbody>
</table>

Simple

Complex
Dataset

- 1:1 Simple Correspondences
  - Class Equivalence
  - Class Subsumption (next slide)
  - Property Equivalence
  - Property Equivalence Inverse

Formal Pattern: $C_1(x) \iff C_2(x)$
Example: Award($x$) $\iff$ FundingAward($x$)

Formal Pattern: $C_1(x) \rightarrow C_2(x)$
Example: GeoFeature($x$) $\rightarrow$ Place($x$)

Formal Pattern: $p_1(x, y) \iff p_2(x, y)$
Example: hasAward($x, y$) $\iff$ fundedBy($x, y$)

Formal Pattern: $p_1(x, y) \iff p_2(y, x)$
Example: isAwardOf($x, y$) $\iff$ fundedBy($y, x$)
Dataset

- 1:n Complex Correspondences
  - Class Typecasting Equivalence
  - Class Typecasting Subsumption
  - Property Typecasting Subsumption
  - Property Typecasting Subsumption Inverse

Formal Pattern: $C_1(x) \leftrightarrow \text{rdfs:subClassOf}(x, C_2)$
Example: PlaceType($x$) $\leftrightarrow \text{rdfs:subClassOf}(x, \text{Place})$

Formal Pattern: $C_1(x) \rightarrow \text{rdfs:subClassOf}(x, C_2)$
Example: GeoFeatureType($x$) $\rightarrow \text{rdfs:subClassOf}(x, \text{Place})$

Formal Pattern: $p_1(x, y) \rightarrow \text{rdfs:subClassOf}(y, C_2)$
Example: hasPlaceType($x, y$) $\rightarrow \text{rdfs:subClassOf}(y, \text{Place})$


Class Subsumption

Formal Pattern: $p_1(x, y) \rightarrow \text{rdfs:subClassOf}(x, C_2)$
Example: isGeoFeatureTypeOf($x, y$) $\rightarrow \text{rdfs:subClassOf}(x, \text{Place})$

Class Equivalence

Graphical diagram showing relationships and equivalences between classes and types in a dataset.
Dataset

• m:n Complex Correspondences
  • Typed Property Chain Equivalence
  • Typed Property Chain Equivalence Inverse
• m:n Complex Correspondences
  • Typed Property Chain Subsumption
  • Typed Property Chain Subsumption Inverse
• Representation
  • EDOAL
    • More expressive than standard alignment format
  • Rule Syntax
    • Easy for humans to understand

• Data publication
  • Download link (CC-BY): http://doi.org/10.6084/m9.figshare.5907172
  • OAEI complex track website

• HermiT reasoner has been applied to the ontologies independently to check the satisfiability
Conclusion and Future Work

• **Conclusion**
  • Presentation of two ontologies from real-world case to support data representation, sharing, integration, and discovery.
  • Creation of an alignment between these two ontologies includes simple and complex alignment.
  • Publication of the benchmark alignment in both rule syntax and EDOAL format at a persistent URL under a CC-BY license.

• **Future Work**
  • Complex track in OAEI (ongoing)
  • Create an automated complex alignment to detect the complex correspondences and evaluate the performance.
Thank you, Questions?