Discovering Concept Coverings in Aligning Ontologies of Linked Data

Rahul Parundekar, Craig A. Knoblock and Jose-Luis Ambite
{parundek,knoblock}@usc.edu, ambite@isi.edu
University of Southern California
Web of Linked Data

- Different sources with different schemas
- Equivalent instances in the different domains connected with \textit{owl:sameAs}

Example: Geospatial Domain
Interlinked instances with disjoint schemas

Source 1

<table>
<thead>
<tr>
<th>Schema Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Populated Place</td>
</tr>
</tbody>
</table>

| Instance Level |
| Los Angeles |

Source 2

<table>
<thead>
<tr>
<th>Schema Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>City</td>
</tr>
</tbody>
</table>

| Instance Level |
| City of Los Angeles |

owl:sameAs

NO LINKS!!
Problem: Disconnected Ontologies

• Only a small number of Ontologies are linked

• Existing Concepts may not be sufficient for exhaustive set of alignments

• Alignments are necessary for the Interoperability goal of the Semantic Web
Represents set of instances belonging to Class A
Represents set of instances belonging to Class B

Class A is disjoint from Class B
Class A is equivalent to Class B

Class A is subset of Class B
Class B is subset of Class A

Solution: Generate Alignments using Linked Data – Extensional Approach
Aligning Restriction Classes Using Extensional Approach

featureClass = P

\( \text{Img}(r_1) \cap r_2 \sim 1 \)

\( \frac{|\text{Img}(r_1) \cap r_2|}{|\text{Img}(r_1)|} \sim 1 \)

\( \frac{|\text{Img}(r_1) \cap r_2|}{|r_2|} \sim 1 \)

rdf:type = PopulatedPlace
But there is a pattern to be explored in the subset relations

Let’s look at 3 of the subset relations we found…
1) Schools in GeoNames are Educational Institutions in DBpedia

featureCode=S.SCH

class rdf:type=School

School rdf:type=EducationalInstitution
2) Colleges in GeoNames are Educational Institutions in DBpedia

featureCode=S.SCH

featureCode=S.SCHC

rdf:type=EducationalInstitution
3) Universities in GeoNames are Educational Institutions in DBpedia

- `featureCode=S.SCH`
- `featureCode=S.SCHC`
- `featureCode=S.UNIV`
Taken by themselves, the subset relations are not useful.

- featureCode=S.SCH
- rdf:type=EducationalInstitution
- featureCode=S.SCHC
- featureCode=S.UNIV
Using **featureCode** property as a hint, we form a **Union** of concepts

\[
\text{featureCode} = \text{S.SCH} \quad \cup \quad \text{featureCode} = \text{S.SCHC} \quad \cup \quad \text{featureCode} = \text{S.UNIV}
\]
We Can Find Concept Coverings by Extensional Comparison (Contribution 1)

featureCode=S.SCH

featureCode=S.SCHC

featureCode=S.UNIV


rdf:type=EducationalInstitution
Results & Conclusion

We found 7096 Alignments including Outliers for 5 pairs of sources aligned in Geospatial, Zoology & Genetics domains from 77966 subset relations.

Example Alignments from DBpedia alignment with Geonames:

<table>
<thead>
<tr>
<th>Larger Concept</th>
<th>Concepts Covered</th>
<th>Support</th>
<th>Outliers</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>rdf:type = dbpedia:Educational Institution</code></td>
<td><code>geonames:featureCode= {S.SCH, S.SCHC, S.UNIV}</code></td>
<td>396 out of 404 (R'U=0.98)</td>
<td>S.BLDG (3/122), S.EST (1/13), S.LIBR (1/7), S.HSP (1/31), S.MUS (1/43)</td>
</tr>
<tr>
<td><code>dbpedia:country = dbpedia:Spain</code></td>
<td><code>geonames:countryCode = {ES}</code></td>
<td>3917 out of 3918 (R'U=0.99)</td>
<td>IT (1/7635)</td>
</tr>
</tbody>
</table>

**Contribution 2:** Finding Outliers
QUESTIONS? SEND US AN EMAIL

{parundek, knoblock}@usc.edu, ambite@isi.edu