Linking and Building Ontologies of Linked Data

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

- Vast collection of interlinked information
- Different sources with different schemas
• Interlinked instances in the various domains
• Equivalent instances linked with `owl:sameAs`
Interlinked Instances

<table>
<thead>
<tr>
<th>Source 1</th>
<th>Source 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Schema Level</strong></td>
<td><strong>Schema Level</strong></td>
</tr>
<tr>
<td>PopulatedPlace</td>
<td>City</td>
</tr>
<tr>
<td><strong>Instance Level</strong></td>
<td><strong>Instance Level</strong></td>
</tr>
<tr>
<td>Los Angeles</td>
<td>City of Los Angeles</td>
</tr>
</tbody>
</table>

owl:sameAs
Disjoint Schemas

Source 1

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

<table>
<thead>
<tr>
<th>Instance Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Los Angeles</td>
</tr>
</tbody>
</table>

Source 2

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

<table>
<thead>
<tr>
<th>Instance Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>City of Los Angeles</td>
</tr>
</tbody>
</table>

owl:sameAs

NO LINKS!!
Objective 1: Find Schema Alignments

<table>
<thead>
<tr>
<th>Source 1</th>
<th>Source 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Schema Level</strong></td>
<td></td>
</tr>
<tr>
<td>PopulatedPlace</td>
<td>=</td>
</tr>
<tr>
<td>City</td>
<td></td>
</tr>
<tr>
<td><strong>Instance Level</strong></td>
<td></td>
</tr>
<tr>
<td>Los Angeles</td>
<td>owl:sameAs</td>
</tr>
<tr>
<td>City of Los Angeles</td>
<td></td>
</tr>
</tbody>
</table>
Ontologies of Linked Data

- Ontologies can be highly specialized
  - e.g. DBpedia has classes for *Educational Institutions*, *Bridges*, *Airports*, etc.

- But some can be rudimentary
  - e.g. in Geonames all instances only belong to a single class – ‘Feature’
  - Derived from RDBMS schemas from which Linked Data was generated
• There might not exist exact equivalences between classes in two sources
• Only subset relations possible
• A specialized class can be created by restricting the value of one or more properties

• The following Venn diagram explains a restriction class in Geonames with a restriction on the value of the featureCode property as ‘S.SCH’

Set of all instances in Original Class - \( \text{rdf:type}=\text{Feature} \)

Set of all instances in Restricted Class - \( \text{rdf:type}=\text{Feature} \& \text{featureCode}=\text{S.SCH} \)
Objective 2: Find Alignments Between Restriction Classes

• Find and model specialized descriptions of classes

<table>
<thead>
<tr>
<th>Geonames</th>
<th>DBpedia</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schema Level</td>
<td></td>
</tr>
<tr>
<td>$rdf$type=Feature &amp; featureCode=S.SCH</td>
<td>$rdf$type=Educational Institution</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Instance Level</td>
<td></td>
</tr>
<tr>
<td>University of Southern California</td>
<td>University of Southern California</td>
</tr>
</tbody>
</table>

owl:sameAs
Domains

- Geospatial
  - Dbpedia
  - LinkedGeoData
  - Geonames
- Zoology
  - Geospecies
  - Dbpedia
- Genetics (Bio2RDF)
  - GeneID
  - MGI
Approach

- Aligning Restriction Classes
Approach

- Aligning Restriction Classes

- Find relation between the two restriction classes
  - Equivalent
  - Subset
Extensional Approach to Ontology Alignment

<table>
<thead>
<tr>
<th>Set Representation</th>
<th>Relation</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="disjoint.png" alt="Image" /></td>
<td>Disjoint</td>
</tr>
<tr>
<td><img src="subset_r1.png" alt="Image" /></td>
<td>$r_1 \subseteq r_2$</td>
</tr>
<tr>
<td><img src="subset_r2.png" alt="Image" /></td>
<td>$r_2 \subseteq r_1$</td>
</tr>
<tr>
<td><img src="equal.png" alt="Image" /></td>
<td>$r_1 = r_2$</td>
</tr>
<tr>
<td><img src="not_enough_support.png" alt="Image" /></td>
<td>Not enough support</td>
</tr>
</tbody>
</table>

Key:
- ![Image](disjoint.png) Set of instance pairs where both $r_1$ and $r_2$ holds
- ![Image](subset_r1.png) Set of instances from $O_1$ where $r_1$ holds
- ![Image](subset_r2.png) Set of instances from $O_2$ where $r_2$ holds
- ![Image](equal.png) Set of instances from $O_2$ paired to instances from $O_1$
- ![Image](not_enough_support.png) Instance pairs where both $r_1$ and $r_2$ holds
- ![Image](not_enough_support.png) Instance pairs where $r_1$ holds
Instances belonging to a restriction class also belong to parent restriction class
  
  e.g. restrictions from Geonames below

This also results in a hierarchy in the alignments, which our algorithm exploits
Exploration of Hypotheses Search Space (LinkedGeoData with DBpedia)

Seed hypotheses generation

Seed hypothesis pruning (owl:Thing covers all instances)

- (rdf:type=lgd:country)
  (rdf:type=owl:Thing)

- (lgd:gnis%3AST_alpha=NJ)
  (dbpedia:Place#type=http://dbpedia.org/resource/City_(New_Jersey))

Prune as no change in the extension set

Pruning on empty set $r_2=\emptyset$

- (rdf:type=lgd:node)
  (rdf:type=dbpedia:BodyOfWater)

- (rdf:type=lgd:node)
  (rdf:type=dbpedia:PopulatedPlace)

- (rdf:type=lgd:node)
  (dbpedia:Place#type=dbpedia:City)

- (rdf:type=lgd:node)
  (dbpedia:Place#type=dbpedia:City & rdf:type=owl:Thing)

- (rdf:type=lgd:node)
  (dbpedia:Place#type=dbpedia:City & rdf:type=dbpedia:City)
1. Prune seed hypothesis if either restriction covers all instances in that source

---

Seed hypothesis pruning (owl:Thing covers all instances)

(rdf:type=lgd:country)
(rdf:type=owl:Thing)

(rdf:type=lgd:node)
(rdf:type=dbpedia:BodyOfWater)

(rdf:type=lgd:node)
(rdf:type=dbpedia:PopulatedPlace)

(rdf:type=lgd:node)
(rdf:type=dbpedia:BodyOfWater & dbpedia:Place#type=dbpedia:City)

(lgd:gnis%3AST_alpha=NJ)
(dbpedia:Place#type=http://dbpedia.org/resource/City_(New_Jersey))

(rdf:type=lgd:node)
(dbpedia:Place#type=dbpedia:City)

(rdf:type=lgd:node)
(dbpedia:Place#type=dbpedia:City & rdf:type=owl:Thing)

Prune as no change in the extension set

Pruning on empty set
$r_2=\emptyset$
2. Number of instance pairs supporting hypothesis must be above a threshold.

Seed hypotheses generation

- (rdf:type=lgd:country)
  (rdf:type=owl:Thing)

Seed hypothesis pruning (owl:Thing covers all instances)

- (rdf:type=lgd:node)
  (rdf:type=dbpedia:BodyOfWater)

- (rdf:type=lgd:node)
  (rdf:type=dbpedia:PopulatedPlace)

- (rdf:type=lgd:node)
  (dbpedia:Place#type=dbpedia:City)

- (rdf:type=lgd:node)
  (dbpedia:Place#type=dbpedia:City & rdf:type=owl:Thing)

Pruning on empty set

\[ r_2 = \emptyset \]

- (rdf:type=lgd:node)
  (rdf:type=dbpedia:BodyOfWater & dbpedia:Place#type=dbpedia:City)

- (rdf:type=lgd:node)
  (rdf:type=dbpedia:PopulatedPlace & dbpedia:Place#type=dbpedia:City)

- (rdf:type=lgd:node)
  (dbpedia:Place#type=dbpedia:City)

Prune as no change in the extension set

- (lgd:gnis%3AST_alpha=NJ)
  (dbpedia:Place#type=http://dbpedia.org/resource/City_(New_Jersey))

- (lgd:country)
  (dbpedia:Place#type=dbpedia:Country)

- (lgd:node)
  (dbpedia:Place#type=dbpedia:City)

- (lgd:node)
  (dbpedia:Place#type=dbpedia:City & rdf:type=owl:Thing)
3. Prune if the added constraint does not change the extension

Seed hypotheses generation

- (rdf:type=lgd:country)
  (rdf:type=owl:Thing)

Seed hypothesis pruning (owl:Thing covers all instances)

- (rdf:type=lgd:node)
  (rdf:type=dbpedia:BodyOfWater)

- (rdf:type=lgd:node)
  (rdf:type=dbpedia:PopulatedPlace)

- (rdf:type=lgd:node)
  (rdf:type=dbpedia:BodyOfWater & dbpedia:Place#type=dbpedia:City)

- (lgd:gnis%3AST_alpha=NJ)
  (dbpedia:Place#type=http://dbpedia.org/resource/City_(New_Jersey))

- (lgd:country)
  (owl:Thing)
  (lgd:node)
  (dbpedia:PopulatedPlace)
  (lgd:node)
  (dbpedia:BodyOfWater)
  (dbpedia:Place#type=dbpedia:City)

Prune as no change in the extension set

- (rdf:type=lgd:node)
  (dbpedia:Place#type=dbpedia:City & rdf:type=owl:Thing)

Pruning on empty set

- $r_2 = \emptyset$

- (rdf:type=lgd:node)
  (rdf:type=dbpedia:PopulatedPlace & dbpedia:Place#type=dbpedia:City)
Lexicographic ordering provides a systematic search by pruning hypotheses with reverse order.
Relaxed Scoring

- Compensates for missing, inconsistent in the data

<table>
<thead>
<tr>
<th>Set Representation</th>
<th>Relation</th>
<th>P =</th>
<th>R =</th>
<th>P'</th>
<th>R'</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Disjoint</td>
<td>= 0</td>
<td>= 0</td>
<td>≤ 0.01</td>
<td>≤ 0.01</td>
</tr>
<tr>
<td></td>
<td>r₁ ∩ r₂</td>
<td>&lt; 1</td>
<td>= 1</td>
<td>&gt; 0.01</td>
<td>≥ 0.90</td>
</tr>
<tr>
<td></td>
<td>r₂ ∩ r₁</td>
<td>= 1</td>
<td>&lt; 1</td>
<td>≥ 0.90</td>
<td>&gt; 0.01</td>
</tr>
<tr>
<td></td>
<td>r₁ = r₂</td>
<td>= 1</td>
<td>= 1</td>
<td>≥ 0.90</td>
<td>≥ 0.90</td>
</tr>
<tr>
<td></td>
<td>Not enough support</td>
<td>0 &lt; P &lt; 1</td>
<td>0 &lt; R &lt; 1</td>
<td>0.01 &lt; P' &lt; 0.90</td>
<td>0.01 &lt; R' &lt; 0.90</td>
</tr>
</tbody>
</table>
Post-processing: Removing Implied Alignments

**GEONAMES restriction**
- geonames:featureCode=geonames:S,SCH
- geonames:featureCode=geonames:S,SCH &
  geonames:inCountry=geonames:US

**DBPEDIA restriction**
- rdf:type=dbpedia: EducationalInstitution

---

**Key:**
- \( r_i \supseteq r_j \): Subset relations \( (r_i \subset r_j) \) found by the algorithm.
- \( r_i \rightarrow r_j \): Implied subset relations.
- \( r'_i \rightarrow r_j \): Subset relation by construction.
- T: Transitivity in subset relations.
  One relation can be eliminated.
- C: Cycle in subset relations. Hence, all classes are equivalent.
- ×: Relation eliminated by the rule.
- ✓: Relation retained by the rule.

---

Keep the simpler definition &
Remove the implied definition
Removing Implied Alignments

Cascading

Key:
- \( r_i \rightarrow r_j \): Subset relations \((r_i \subseteq r_j)\) found by the algorithm.
- \( r_i \rightarrow r_j \): Implied subset relations.
- \( r'_i \rightarrow r_j \): Subset relation by construction.
- T: Transitivity in subset relations. One relation can be eliminated.
- C: Cycle in subset relations. Hence, all classes are equivalent.
- \( \times \): Relation eliminated by the rule.
- \( \checkmark \): Relation retained by the rule.
<table>
<thead>
<tr>
<th>#</th>
<th>LINKEDGEODATA restriction</th>
<th>DBPEDIA restriction</th>
<th>Relation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>rdf:type=lgd:node</td>
<td>rdf:type/owl:Thing</td>
<td>$r_1 = r_2$</td>
</tr>
<tr>
<td>2</td>
<td>rdf:type=lgd:aerodrome</td>
<td>rdf:type=dbpedia:Airport</td>
<td>$r_1 = r_2$</td>
</tr>
<tr>
<td>3</td>
<td>rdf:type=lgd:island</td>
<td>rdf:type=dbpedia:Island</td>
<td>$r_1 = r_2$</td>
</tr>
<tr>
<td>4</td>
<td>lgd:gnis_3AST_alpha=NJ</td>
<td>dbpedia:Place#type=<a href="http://dbpedia.org/resource/City_(New_Jersey)">http://dbpedia.org/resource/City_(New_Jersey)</a></td>
<td>$r_1 = r_2$</td>
</tr>
<tr>
<td>5</td>
<td>rdf:type=lgd:village</td>
<td>rdf:type=dbpedia:PopulatedPlace</td>
<td>$r_1 \subseteq r_2$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>#</th>
<th>GEONAMES restriction</th>
<th>DBPEDIA restriction</th>
<th>Relation</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>geonames:featureClass=geonames:P</td>
<td>rdf:type=dbpedia:PopulatedPlace</td>
<td>$r_1 = r_2$</td>
</tr>
<tr>
<td>7</td>
<td>geonames:featureClass=geonames:H</td>
<td>rdf:type=dbpedia:BodyOfWater</td>
<td>$r_1 = r_2$</td>
</tr>
<tr>
<td>8</td>
<td>geonames:parentFeature=<a href="http://sws.geonames.org/3174618/">http://sws.geonames.org/3174618/</a></td>
<td>dbpedia:City_region=<a href="http://dbpedia.org/resource/Lombardy">http://dbpedia.org/resource/Lombardy</a></td>
<td>$r_1 = r_2$</td>
</tr>
<tr>
<td>9</td>
<td>geonames:featureCode=geonames:S.SCH</td>
<td>rdf:type=dbpedia:EducationalInstitution</td>
<td>$r_1 = r_2$</td>
</tr>
<tr>
<td>10</td>
<td>geonames:featureCode=geonames:S.SCH &amp;</td>
<td></td>
<td>$r_1 = r_2$</td>
</tr>
<tr>
<td></td>
<td>geonames:inCountry=geonames:US</td>
<td></td>
<td>$r_1 \subseteq r_2$</td>
</tr>
<tr>
<td>11</td>
<td>geonames:featureCode=geonames:T.MT</td>
<td>rdf:type=dbpedia:Mountain</td>
<td>$r_1 \subseteq r_2$</td>
</tr>
<tr>
<td>#</td>
<td>GEOSPECIES restriction</td>
<td>DBPEDIA restriction</td>
<td>Relation</td>
</tr>
<tr>
<td>----</td>
<td>----------------------------------------------------</td>
<td>---------------------------------------------------</td>
<td>----------</td>
</tr>
<tr>
<td>12</td>
<td>geospecies:inKingdom=<a href="http://lod.geospecies.org/kingdoms/Aa">http://lod.geospecies.org/kingdoms/Aa</a></td>
<td>rdf:type=dbpedia:Animal</td>
<td>r₁ = r₂</td>
</tr>
<tr>
<td>13</td>
<td>geospecies:hasOrderName=Lepidoptera</td>
<td>dbpedia:order=<a href="http://dbpedia.org/resource/Lepidoptera">http://dbpedia.org/resource/Lepidoptera</a></td>
<td>r₁ = r₂</td>
</tr>
<tr>
<td>14</td>
<td>geospecies:hasOrderName=Lepidoptera</td>
<td>dbpedia:kingdom=<a href="http://dbpedia.org/resource/Animal">http://dbpedia.org/resource/Animal</a> &amp; dbpedia:order=<a href="http://dbpedia.org/resource/Lepidoptera">http://dbpedia.org/resource/Lepidoptera</a></td>
<td>r₁ = r₂</td>
</tr>
<tr>
<td>15</td>
<td>geospecies:hasGenusName=Falco</td>
<td>dbpedia:genus=<a href="http://dbpedia.org/resource/Falcon">http://dbpedia.org/resource/Falcon</a></td>
<td>r₁ = r₂</td>
</tr>
<tr>
<td>16</td>
<td>geospecies:hasOrderName=Primates</td>
<td>dbpedia:order=<a href="http://dbpedia.org/resource/Primates">http://dbpedia.org/resource/Primates</a></td>
<td>r₂ ⊆ r₁</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>#</th>
<th>GEOSPECIES restriction</th>
<th>GEOSPECIES restriction</th>
<th>Relation</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>geospecies:hasKingdomName=Animalia</td>
<td>geospecies:inKingdom=<a href="http://lod.geospecies.org/kingdoms/Aa">http://lod.geospecies.org/kingdoms/Aa</a></td>
<td>r₁ = r₂</td>
</tr>
<tr>
<td>21</td>
<td>geospecies:hasClassName=Insecta</td>
<td>geospecies:inClass=<a href="http://lod.geospecies.org/bioclasses/aQado">http://lod.geospecies.org/bioclasses/aQado</a></td>
<td>r₂ ⊆ r₁</td>
</tr>
<tr>
<td>22</td>
<td>geospecies:inFamily=<a href="http://lod.geospecies.org/families/amTJ9">http://lod.geospecies.org/families/amTJ9</a></td>
<td>geospecies:hasSubfamilyName=Sigmodontinae</td>
<td>r₂ ⊆ r₁</td>
</tr>
<tr>
<td>#</td>
<td>MGI restriction</td>
<td>GENEID restriction</td>
<td>Relation</td>
</tr>
<tr>
<td>----</td>
<td>--------------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------</td>
<td>----------</td>
</tr>
<tr>
<td>17</td>
<td>bio2rdf:subType=Pseudogene</td>
<td>bio2rdf:subType=pseudo</td>
<td>$r_1 = r_2$</td>
</tr>
<tr>
<td>18</td>
<td>bio2rdf:subType=Pseudogene &amp; mgi:genomeStart=17</td>
<td>geneid:chromosome=17 &amp; bio2rdf:subType=pseudo</td>
<td>$r_1 = r_2$</td>
</tr>
<tr>
<td>19</td>
<td>bio2rdf:chromosomePosition=1.00 &amp; mgi:genomeStart=4</td>
<td>geneid:chromosome=4 &amp; bio2rdf:subType=pseudo</td>
<td>$r_2 \subset r_1$</td>
</tr>
</tbody>
</table>
Results: Alignments Found

- Equivalences, Subset alignments before and after removing implied alignments

<table>
<thead>
<tr>
<th>Source 1 \ Source 2</th>
<th>#(r₁ = r₂) total</th>
<th>#(r₁ = r₂) best matches</th>
<th>#(r₁ ⊆ r₂) before</th>
<th>#(r₁ ⊆ r₂) after</th>
<th>#(r₂ ⊆ r₁) before</th>
<th>#(r₂ ⊆ r₁) after</th>
</tr>
</thead>
<tbody>
<tr>
<td>LinkedGeoData</td>
<td>158</td>
<td>152</td>
<td>2528</td>
<td>1837</td>
<td>1804</td>
<td>1627</td>
</tr>
<tr>
<td>Geonames</td>
<td>31</td>
<td>19</td>
<td>809</td>
<td>400</td>
<td>1384</td>
<td>1247</td>
</tr>
<tr>
<td>Geospecies</td>
<td>509</td>
<td>420</td>
<td>9112</td>
<td>2294</td>
<td>6098</td>
<td>4455</td>
</tr>
<tr>
<td>MGI</td>
<td>10</td>
<td>9</td>
<td>2031</td>
<td>1869</td>
<td>3594</td>
<td>2070</td>
</tr>
<tr>
<td>Geospecies</td>
<td>94</td>
<td>88</td>
<td>1550</td>
<td>1201</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
Datasets: http://www.isi.edu/integration/data/LinkedData

Linking and Building Ontologies of Linked Data

Rahul Pandeydeor, Craig A. Knoblock and Joost-Jans Ambias
University of Southern California,
Information Sciences Institute
4679 Adams Way, Marina del Rey, CA 90292
(pandeydeor,knoblock,ambias@isil.com)

This page provides the dataset used in the paper "Linking and Building Catalogues of Linked Data"

- This dataset is organized as follows:
  - The 5 source pairs, discussed in the paper, each have a compressed file containing the instance pairs used in the algorithm and the alignments generated by the algorithm.
  - Each compressed file contains 3 compressed files ("case files") containing the instance pairs, alignments before and after post-processing. There are two data sources that are being aligned: source1 and source2 (which may be the same) in each of five files.
  - The instance pairs/aligned_pairs file:
    1. Each row in the file represents an instance pair which is a pair of the flattened property-value pairs of the instances from each source (see the paper), where the join is on the property that asserts instance equivalence.
    2. The first line in the case file is the properties in the source it is of the form:
       prop1, prop2, ..., propn, source1, source2, ...
    3. The other lines in the file contain a URI of the instance from the first source, the values of the properties under each of its columns ("? if no value exists) and a similar vector for the URL of the second source.
    4. Post-processing has already been performed on these instances.
  - The alignments/aligned_pairs file:
    1. Each row in the file represents an alignment generated by the algorithm along with the rank that supports the hypothesis.
    2. The first line in the file contains the column headings.
    - The columns in the file are:
      - Restriction class from Ontology 1 (E1): property-value pairs representing restriction class from Source Ontology 1.
      - Restriction class from Ontology 2 (E2): property-value pairs representing restriction class from Source Ontology 2.
      - barnowl: E1 and E2 support score for the alignment from the first source. (See Section paper in Fig. 5 Metrics)
      - "PropRel: E1 and E2 support score for the alignment from the second source. (See PRel paper in Fig. 5 Metrics)
      - RelSource/Emit: E1 subset E2 or E2 subset E1
      - Nid of Intersection
      - Lcss of Intersection
      - One of Intersection
    3. These alignments were produced by the algorithm described in the paper.
    4. Important Note: Alignments still have pending post-processing.
  - The aligned_pairs/aligned_pairs file:
    1. Each row in the file represents an alignment generated by the algorithm after post-processing along with the rank.
    2. The columns in the file are similar to the alignments/aligned_pairs file.
    - These columns are:
      - Restriction class from Ontology 1 (E1): property-value pairs representing restriction class from Source Ontology 1.
      - Restriction class from Ontology 2 (E2): property-value pairs representing restriction class from Source Ontology 2.
      - barnowl: E1 and E2 support score for the alignment from the first source. (See Section paper in Fig. 5 Metrics)
      - "PropRel: E1 and E2 support score for the alignment from the second source. (See PRel paper in Fig. 5 Metrics)
      - RelSource/Emit: E1 subset E2 or E2 subset E1
      - Nid of Intersection
      - Lcss of Intersection
      - One of Intersection
Related Work

• Euzenat et al. – Ontology Matching
  • Terminological
  • Structural
  • Semantic

• FCA-Merge, Duckham et al.
  • Use extensional techniques

• GLUE
  • Uses an extensional technique after performing machine learning operations
Conclusion

• Our algorithm generates alignments, consisting of conjunctions of restriction classes
  • Extensional approach on Linked Data
  • Use of restriction classes
• Alignments based on the actual data
  • We determine the relationships based on the data
  • Schemas of linked sources can be readily modeled and used

• Algorithm also able to
  • Specialize ontologies where original were rudimentary
  • Find complimentary hierarchy across an ontology
Future Work

- How to actually understand these alignments
- Scalability
  - Pre-processing of the sources
  - Faster alignment processing