Finding Commonalities in Linked Open Data

Simona Colucci\textsuperscript{1}, Silvia Giannini\textsuperscript{2}, Francesco M. Donini\textsuperscript{1}

\textsuperscript{1} – DISUCOM
Università della Tuscia
Viterbo, Italy

\textsuperscript{2} – DEI
Politecnico di Bari
Bari, Italy
Common Subsumers (CS)

—what for?

- learning [Cohen et al., 1992]
- ontology bottom-up construction [Baader and Küsters, 1998]
- web service discovery [Benatallah et al., 2005]
- knowledge management [Colucci et al., 2008]
- now: *clustering* (unsupervised learning) [Colucci et al., 2013]
A definition of CS

- resource $a$, relevant triples $T_a$
- resource $b$, relevant triples $T_b$

A CS of $\langle a, T_a \rangle$ and $\langle b, T_b \rangle$ is a pair $\langle cs, T_{cs} \rangle$ such that:

$$T_a \models T_{cs}[cs \mapsto a] \quad \text{and} \quad T_b \models T_{cs}[cs \mapsto b]$$

- so far, we consider only simple entailment
Example: LOD Chamber of Deputies

10th Legislature: Find commonalities between deputies

Nilde Iotti and Tina Anselmi

Computing a CS of two resources

- joint depth-first exploration of the two RDF-graphs
- for each pair of triples in $T_a \times T_b$, add a triple $t \in T_{cs}$ whose resources are:
  - if resource is the same in $T_a, T_b$ → same resource in $t$
  - if different resources → blank node in $t$
Example (ctd.): computed CS

```
"Laurea in lettere; insegnante."@it
```

```
"female"
```

```
ocd:deputato
```

```
ocd:repubblica_10
```

```
_:x0
```

```
_:x1
```

```
foaf:gender
```

```
rdf:type
```

```
ocd:rif_leg
```

```
dc:description
```

```
ocd:rif_mandatoCamera
```
Filtering triples

- Not all triples are relevant
- filter by a \textit{characteristic function} \( \sigma \)
- \( \sigma \) based on:
  - \textit{dataset}
  - \textit{distance} from the resource
  - \textit{predicate} in the triple
  - other criteria (it depends on the application)
Clustering with a CS

- SPARQL query
- WHERE \{ T_{cs} \text{ [blank nodes } \rightarrow \text{ variables] } \} 
- for the previous example:

```sparql
SELECT DISTINCT ?x0
WHERE{
  ?x0 a <http://dati.camera.it/ocd/deputato> .
  ?x0 <http://xmlns.com/foaf/0.1/gender> "female" .
  ...
}
```
### Clustering Deputies—10th Legislature

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$(d_{3140_10}, d_{270_10})$</td>
<td>_:x1</td>
<td>_:x2</td>
<td>_:x3</td>
<td>&quot;female&quot;</td>
<td>&quot;Laurea in lettere; insegnante.&quot;@it</td>
<td>2</td>
</tr>
<tr>
<td>$(d_{200023_10}, d_{22710_10})$</td>
<td>_:x1</td>
<td>_:x2</td>
<td>_:x3</td>
<td>&quot;female&quot;</td>
<td></td>
<td>81</td>
</tr>
<tr>
<td>$(d_{30010_10}, d_{17060_10})$</td>
<td>_:x1</td>
<td>_:x2</td>
<td>_:x3</td>
<td>&quot;male&quot;</td>
<td>&quot;Laurea in giurisprudenza; avvocato&quot;@it</td>
<td>44</td>
</tr>
<tr>
<td>$(d_{20910_10}, d_{30570_10})$</td>
<td>_:x1</td>
<td>_:x2</td>
<td>_:x3</td>
<td>&quot;male&quot;</td>
<td></td>
<td>148</td>
</tr>
<tr>
<td>$(d_{30140_10}, d_{60499_10})$</td>
<td>_:x1</td>
<td>_:x2</td>
<td>_:x3</td>
<td>&quot;male&quot;</td>
<td></td>
<td>398</td>
</tr>
<tr>
<td>$(d_{24780_10}, d_{31040_10})$</td>
<td>_:x1</td>
<td>_:x2</td>
<td>&quot;male&quot;</td>
<td></td>
<td></td>
<td>7</td>
</tr>
</tbody>
</table>
## Clustering Deputies — 1st Legislature

| Seed’s URIs          | ocd:rif_mandatoCamera | ocd:membro | ocd:aderisce | foaf:gender | dc:description                                                                 | \( |P| \) |
|----------------------|-----------------------|------------|--------------|-------------|--------------------------------------------------------------------------------|------|
| \((d_{19990_1},d_{20060_1})\) | \_:x1 \_:x2 \_:x3 | "male"     |              |             | "Laurea in giurisprudenza; avvocato."@it                                         | 127  |
| \((d_{3140_1},d_{14290_1})\) | \_:x1 \_:x2 \_:x3 | "female"   |              |             | "Laurea in lettere; insegnante."@it                                            | 9    |
| \((d_{12560_1},d_{13120_1})\) | \_:x1 \_:x2 \_:x3 | "male"     |              | \_:x4      |                                                                                 | 431  |
| \((d_{26000_1},d_{10090_1})\) | \_:x1 \_:x2 \_:x3 | "female"   |              | \_:x5      |                                                                                 | 35   |
| \((d_{10800_1},d_{25610_1})\) | \_:x1 \_:x2 \_:x3 | "male"     |              |             |                                                                                 | 9    |
| \((d_{12140_1},d_{8520_1})\) | \_:x1 \_:x2 \_:x3 | "male"     |              |             |                                                                                 | 2    |
In the notes of this slide, references can be found.

Slides are available at
http://sisinflab.poliba.it
References


