Finding Commonalities in Linked Open Data

Simona Colucci[1], Silvia Giannini[2], Francesco M. Donini[1]

[1] – DISUCOM
Università della Tuscia
Viterbo, Italy

[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:

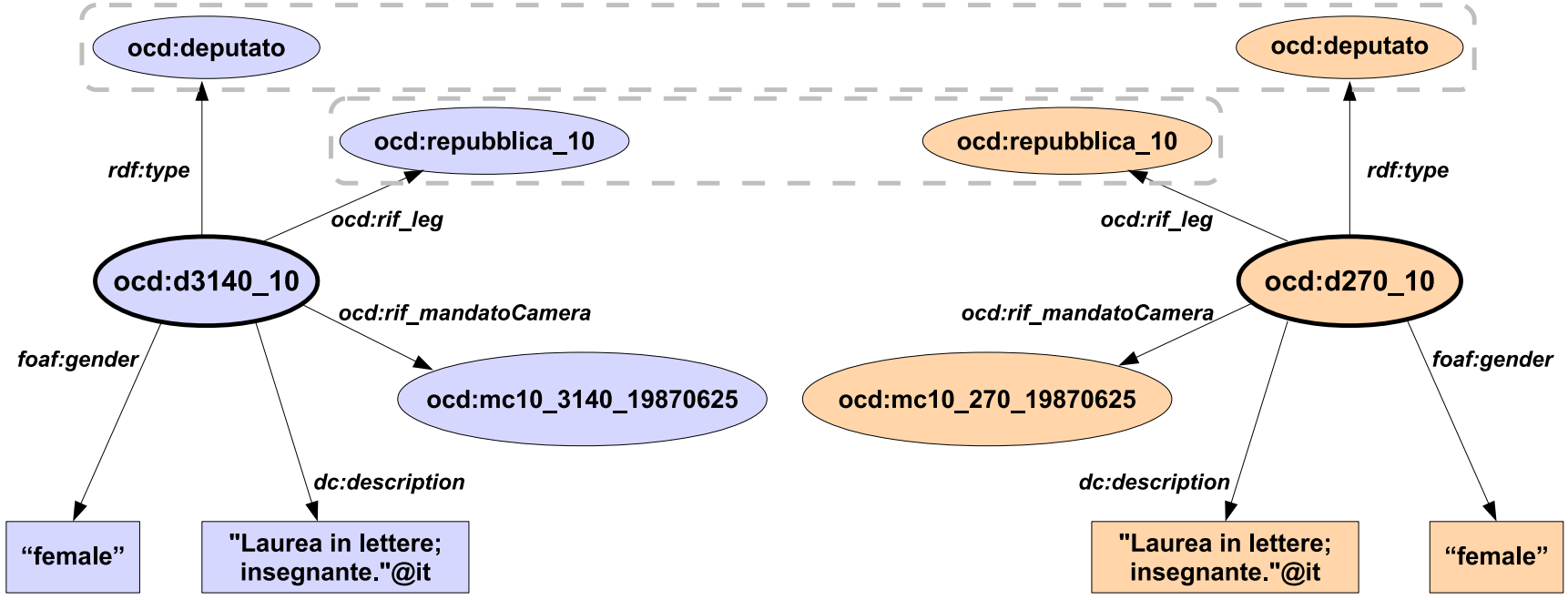
$$T_a \models T_{cs}[cs \mapsto a] \quad \text{and} \quad T_b \models T_{cs}[cs \mapsto b]$$

- so far, we consider only *simple entailment*

Example: LOD Chamber of Deputies

10th Legislature: Find commonalities between deputies

Nilde Iotti and Tina Anselmi

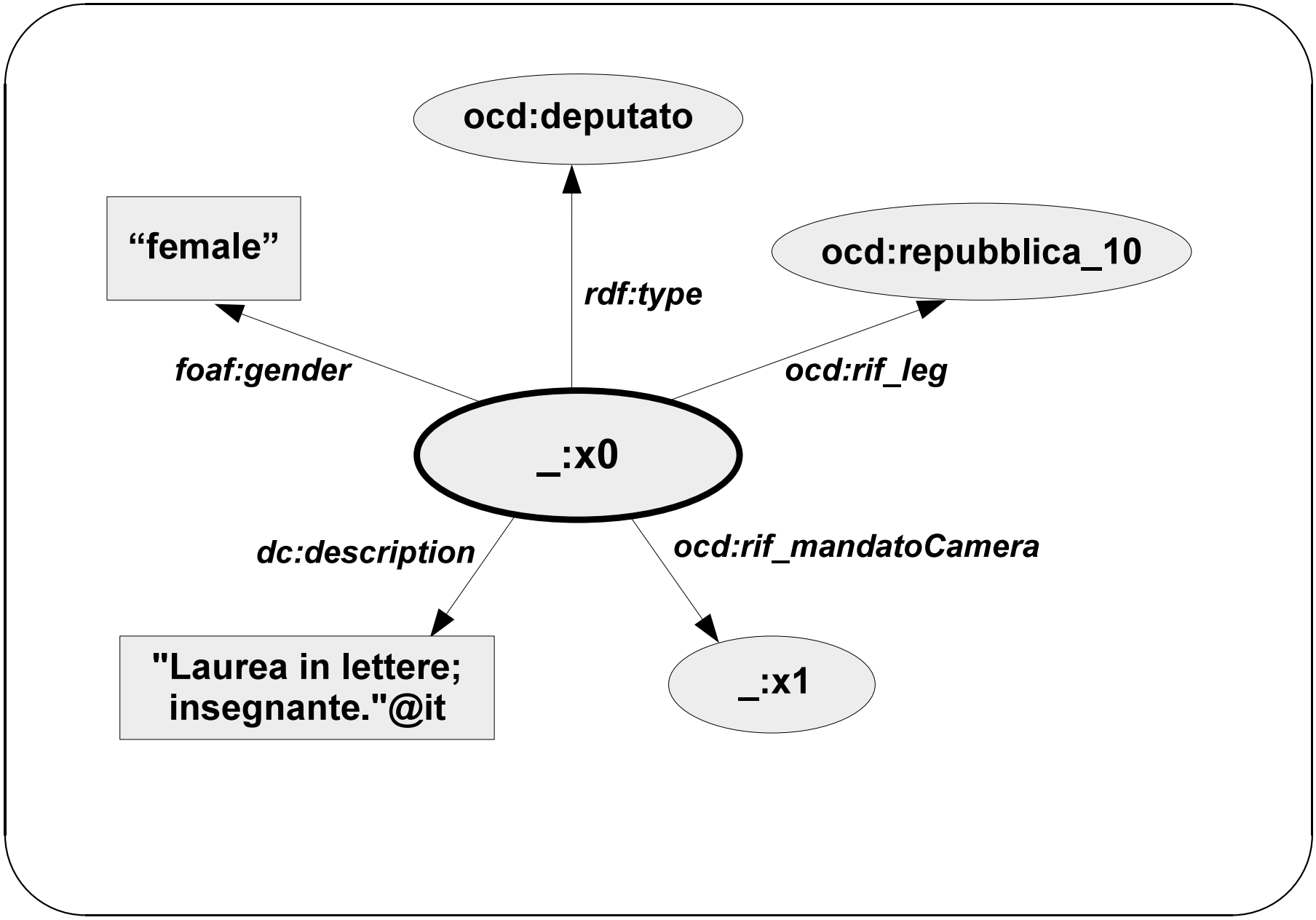


Computing a CS of two resources

- 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

    $\longrightarrow$ same resource in t
  - if different resources $\longrightarrow$ blank node in t

Example (ctd.): computed CS



Filtering triples

- Not all triples are relevant
- filter by a *characteristic function* σ
- σ based on:
  - *dataset*
  - *distance* from the resource
  - *predicate* in the triple
  - other criteria (it depends on the application)

Clustering with a CS

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

```
SELECT DISTINCT ?x0
        WHERE{
        ?x0 a <http://dati.camera.it/ocd/deputato> .
        ?x0 <http://xmlns.com/foaf/0.1/gender> "female" .
        ?x0 <http://dati.camera.it/ocd/rif_mandatoCamera> ?x1 .
        ...
        }
```

Clustering Deputies—10th Legislature

Seed's URIs	ocd:rif_mandatoCamera	ocd:membro	ocd:aderisce	foaf:gender	dc:description	ocd:rif_ufficioParlamentare	$\|P\|$
$(d3140\_10, d270\_10)$	_:x1	_:x2	_:x3	"female"	"Laurea in lettere; insegnante."@it		2
$(d200023\_10, d22710\_10)$	_:x1	_:x2	_:x3	"female"			81
$(d30010\_10, d17060\_10)$	_:x1	_:x2	_:x3	"male"	"Laurea in giurisprudenza; avvocato"@it		44
$(d20910\_10, d30570\_10)$	_:x1	_:x2	_:x3	"male"		_:x4	148
$(d30140\_10, d60499\_10)$	_:x1	_:x2	_:x3	"male"			398
$(d24780\_10, d31040\_10)$	_:x1		_:x2	"male"			7

Clustering Deputies—1st Legislature

Seed's URIs	ocd:rif_mandatoCamera	ocd:membro	ocd:aderisce	foaf:gender	dc:description	$\|P\|$
$(d19990\_1, d20060\_1)$	_:x1	_:x2	_:x3	"male"	"Laurea in giurisprudenza; avvocato."@it	127
$(d3140\_1, d14290\_1)$	_:x1	_:x2	_:x3	"female"	"Laurea in lettere; insegnante."@it	9
$(d12560\_1, d13120\_1)$	_:x1	_:x2	_:x3	"male"	_:x4	431
$(d26000\_1, d10090\_1)$	_:x1	_:x2	_:x3	"female"	_:x5	35
$(d10800\_1, d25610\_1)$	_:x1	_:x2	_:x3	"male"		9
$(d12140\_1, d8520\_1)$	_:x1		_:x2	_:x3		2

References

In the notes of this slide,
references can be found.

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

References

[Baader and Küsters, 1998] Franz Baader and Ralf Küsters. Computing the least common subsumer and the most specific concept in the presence of cyclic $\mathcal{ALN}$-concept descriptions. In *Proceedings of the Twenty-second German Annual Conference on Artificial Intelligence (KI'98)*, volume 1504 of *Lecture Notes in Computer Science*, pages 129–140. Springer-Verlag, 1998.

[Benatallah *et al.*, 2005] Boualem Benatallah, Mohand S. Hacid, Alain Leger, Christophe Rey, and Farouk Toumani. On automating web services discovery. *Very Large Database Journal*, 14(1):84–96, March 2005.

[Cohen *et al.*, 1992] William W. Cohen, Alex Borgida, and Haym Hirsh. Computing least common subsumers in Description Logics. In William Swartout, editor, *Proceedings of the Tenth National Conference on Artificial Intelligence (AAAI'92)*, pages 754–760. AAAI Press/The MIT Press, 1992.

[Colucci *et al.*, 2008] Simona Colucci, Eugenio Di Sciascio, Francesco M. Donini, and Eufemia Tinelli. Finding informative commonalities in concept collections. In *Proceedings*

of the 17th Conference on Information and Knowledge Management CIKM 2008, pages 807–816. ACM Press, 2008.

[Colucci *et al.*, 2013] Simona Colucci, Francesco M. Donini, and Eugenio Di Sciascio. Common subsumbers in RDF. In Matteo Baldoni, Cristina Baroglio, Guido Boella, and Roberto Micalizio, editors, *AI*IA*, volume 8249 of *Lecture Notes in Computer Science*, pages 348–359. Springer, 2013.