A Logic-based approach to Named-Entity Disambiguation in the Web of Data

Silvia Giannini, Simona Colucci,
Francesco M. Donini, Eugenio Di Sciascio

Politecnico di Bari
Università della Tuscia - Viterbo

RDF

- RDF is the language for Semantic Web
- It comes with...
  - A grammar (with three syntaxes)
    - <subject> <predicate> <object>
  - Semantics (three: Simple, RDF, RDFS)
    - Extensional (set-based)
  - Entailment Relation & Deductive Rules
    - (three, depending on the semantics)

Is entailment the only KR service needed for RDF data?

- Seems not:
- Several processing proposed on RDF data::
  - Similarity
  - Clustering
  - Resource comparison
  - ….

Another KR service for RDF

Pair resources with their relevant triples
<resource, triples>

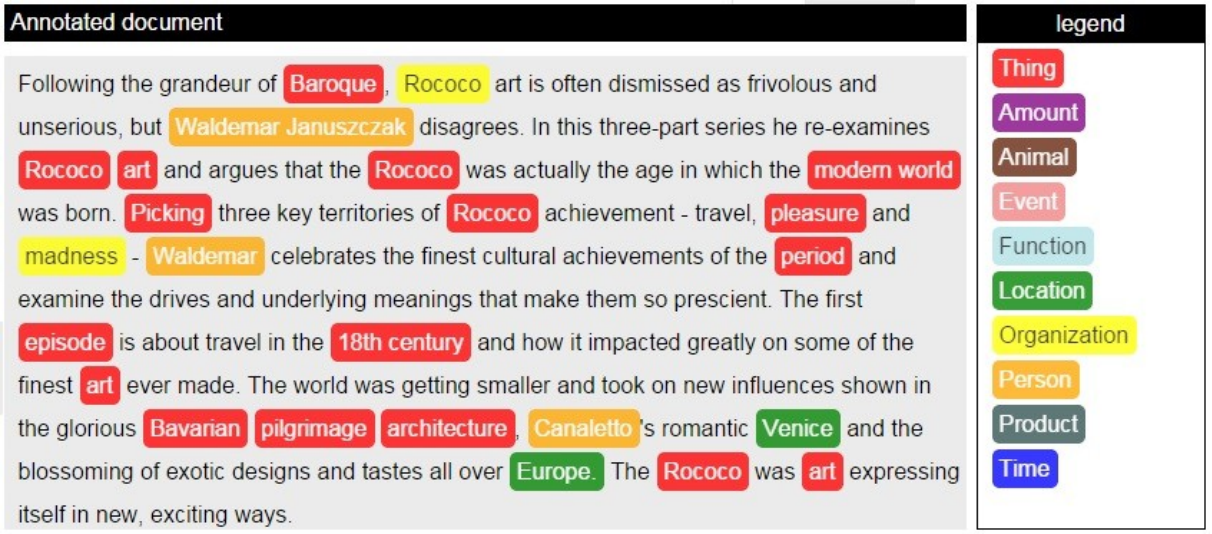
The ***Least Common Subsumer***

of two RDF resources <a,Ta>, <b,Tb> is:

- A resource <r,Tr> such that
  - Both <a,Ta>, <b,Tb> entail <r,Tr>
  - <r,Tr> is the most specific set with this property

Application of LCS (this paper)

- Named Entity Linking:
- Link parts of a text to RDF resources



- When several resources can be linked to the same text, how to disambiguate?

LCS for disambiguation?

- Resources in the same phrase share a common context, (partly) evidenced by LCS
- The more specific their LCS, the more plausible is their choice
- Try with *"I prefer an Orange to a Mandarin"*
  - *Orange*: fruit --- or --- telephone company?
  - *Mandarin*: Chinese language --- or --- fruit?

An experiment with Babelfy

Three miles below St. Petersburg, at a point where the Mississippi River was a trifle over a mile wide,... (M. Twain, “Tom Sawyer”)