

***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

Annotated document	legend
<p>Following the grandeur of Baroque, Rococo art is often dismissed as frivolous and unserious, but Waldemar Januszczak disagrees. In this three-part series he re-examines Rococo art and argues that the Rococo was actually the age in which the modern world was born. Picking three key territories of Rococo achievement - travel, pleasure and madness - Waldemar celebrates the finest cultural achievements of the period and examine the drives and underlying meanings that make them so prescient. The first episode is about travel in the 18th century and how it impacted greatly on some of the finest art ever made. The world was getting smaller and took on new influences shown in the glorious Bavarian pilgrimage architecture, Canaletto's romantic Venice and the blossoming of exotic designs and tastes all over Europe. The Rococo was art expressing itself in new, exciting ways.</p>	<p>Thing</p> <p>Amount</p> <p>Animal</p> <p>Event</p> <p>Function</p> <p>Location</p> <p>Organization</p> <p>Person</p> <p>Product</p> <p>Time</p>

- 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")

The screenshot shows a web browser window with the Babelfy™ interface. The search bar contains the text "tom sawyer text". Below the search bar, the original text is displayed with highlighted words: "miles", "below", "St. Petersburg", "point", "where the", and "Mississippi River". Below each highlighted word is a tooltip providing a definition or image:

- miles**: Il miglio è un'unità di misura di lunghezza, il cui valore, variabile da 1 a 2 km circa, 
- St. Petersburg**: San Pietroburgo con circa 5 milioni di abitanti, è la seconda città della Russia per 
- point**: A geometric element that has position but no extension
- Mississippi River**: Il Mississippi è un fiume dell'America settentrionale che attraversa la parte 