Rough Description Logics for Modeling Uncertainty in Instance Unification

Stefan Schlobach
Vrije Universiteit Amsterdam

joint work with Peter Mika and Michel Klein

Contents

• openacademia
  – functions, architecture
• Instance unification: what, why and how
• Rough Description Logics
• Querying with approximations in openacademia
• Conclusions
openacademia

• distributed citation management
• completely standards-based
• all data in represented in RDF
  – BibTeX data
  – FOAF data
  – blog citations
• built around Sesame repository

Pimp your homepage

This page allows you to generate a magic code that adds your publication list dynamically to your homepage. You can choose where you want the list to appear by placing it at the right place on your page. You also have complete control over the style of your publication list. To get started, you need a web-accessible BibTeX file with your publications or access to an openacademia repository storing such data.

openacademia: html generator

Please enter below the URL of your BibTeX file or leave it empty if you want to use the data stored in this repository. For personal use provide a BibTeX file with your publications only. In case you want to generate a group publication list, then leave this field empty and provide the filter criteria below.

Please specify your search criteria. Note: the search is case insensitive.
Publication list for homepage

Query interface
Architecture

Contents

• openacademia
  – functions, architecture
• Instance unification: what, why and how
• Rough Set Theory and RoughDL
• Querying with approximations in openacademia
• Conclusions
Instance Unification: What?

*Combining different resources that represent the same object*

*also: object co-identification, coreference resolution*

Example: resource a1 and a2

Instance Unification: Why?

- duplicates negatively influence precision and ranking
  - Peter Mika and P.Mika each have 2 publications instead of 4
- information might be missed
  - queries for publications of “Peter” only retrieves half of the results
Instance Unification: How?

- Mainly NLP techniques to discover duplicates
- Usual implementation in SW context:
  - adding $\text{owl:sameAs}$ statements
  - drawbacks:
    - often an overcommitment
      - there only is some evidence for equivalence
    - possibly logical inconsistencies
    - source of statements is lost (context)

Contents

- openacademia
  - functions, architecture
- Instance unification: what, why and how
- Rough Set Theory and RoughDL
- Querying with approximations in openacademia
- Conclusions
Vagueness in Classical DL

• How can vagueness be represented in classical DL?
  \textbf{Definitely} \_Author \lor \textbf{Author} \lor \textbf{Possibly} \_Author

• Let us look at a more complex case

\[
\begin{align*}
\text{Possibly} \_\text{SW} \_\text{Author} & \lor \text{Possibly} \_\text{Author} \\
\lor & \\
\text{SW} \_\text{Author} & \lor \text{Author} \\
\lor & \\
\text{Definitively} \_\text{SW} \_\text{Author} & \lor \text{Definitively} \_\text{Author}
\end{align*}
\]

Rough Description Logics: intuition

\textbf{Definitive Authors}: all similar persons are Authors

\textbf{Possible Authors}: there is a similar persons which is an Author
Rough Description Logics: formal issues

- Syntax
  - For every concept $C_2$ RDL also the upper approximation $\overline{C}_2$ RDL and the lower approximation $\underline{C}_2$ RDL.

- Semantics
  - $\overline{C} = \{i \cup j | j \cup (i,j) \cup R^* \cap j \cup C\}$
  - $\underline{C} = \{i \cup j | j \cup (i,j) \cup R^* \cup j \cup C\}$
  - where $R^*$ is a unique and fixed equivalence (similarity) relation

- Reasoning
  - In the presence of reflexivity, symmetry and transitivity + 9 and 8 in DL reasoning in RDL reduces to DL.
  - Translate: $(\overline{C})^* = 9 \circ \cdot C$; $(\underline{C})^* = 8 \circ \cdot C$ and reason classically

Some Rough DL consequences

- $T = \{\ldots, \text{Septic v Inf, Septic v OF, Septic v MOF}\}$
Some Rough DL consequences

- \( T = \{\ldots, \text{Septic} \lor \text{Inf}, \text{Septic} \lor \text{OF}, \text{Septic} \lor \text{MOF}\} \)

- Possibly septic pats are definitely infected:
  \( T \models \text{Septic} \lor \text{Inf} \).
- Possibly septic pats have possible organ failure:
  \( T \models \text{Septic} \lor \text{OF} \).

More Rough DL consequences

- *There are no definitely non-typical sepsis patients.*
  - \( T \models \{\text{NTS} \lor \text{Septic} \land \text{Sepsis}\} \rightarrow \text{NTS} = ? \)
- *Approximations of approximations are equivalent to the approximations themselves*
  - \( \text{Septic} \equiv \text{Septic} \)

- **Comment:** these are simple S5 consequences!
RoughOWL

• Two new operators:
  – upperApproximation (sim-relation Description)
  – lowerApproximation (sim-relation Description)

*E.g.: A is an upper approximation of B when using property X as similarity relation*

• Conservative extension of OWL
  – standard reasoning services are sufficient
    • Racer, FaCT++, Pellet

• Up to now, we only use RoughOWL for querying
  – No representations & deduction in RoughOWL
  – Because of our inference engine (Sesame)

Contents

• openacademia
  – functions, architecture
• Instance unification: what, why and how
• Rough Set Theory and RoughDL
• Querying with approximations in openacademia
• Conclusions
Querying in openacedemia

- Four types of similarity relations between foaf:Person instances are added, i.e. when two instances have:
  - same value for an inverse functional property
  - exactly equal full name
  - exactly equal last name and same initial
  - similar fullname

RoughOWL → SeRQL

- RoughOWL concepts are translated into SeRQL
  - querying can be used for instance checking because RDF is simple enough
  - lower approximation requires CWA

```
SELECT distinct Pub
FROM {Pub} swrc:author {Person},
     {Person} oa:nameSimilarTo3
{<http://www.cs.vu.nl/~pmika/pubs.bib#marta_sabou>}
```
Some Results

- Querying for publications of http://www.cs.vu.nl/~pmika/pubs.bib#marta_sabou
- Direct:
  - 3 results
- Inverse functional
  - 3 results including a link to homepage
- Exact match on fullname
  - 7 results
- Exact match on lastname and initial
  - 8 results
- Fuzzy match on fullname
  - 9 results (including “Martha Sabou” in a BibTeX file from CWI)
Conclusions

• openacademia shows usability of SW and Web technology
  – many others did before 😊
• RoughOWL allows for instance unification using gradually weakening notions of similarity
  – conceptually nice
  – gives user control over the extent of unification
  – Potential to be used in modeling
• Using OWL semantics in RDF repositories requires careful consideration