Rough Description Logics for Modeling Uncertainty in Instance Unification

Stefan Schlobach
Vrije Universiteit Amsterdam

joint work with Peter Mika and Michel Klein

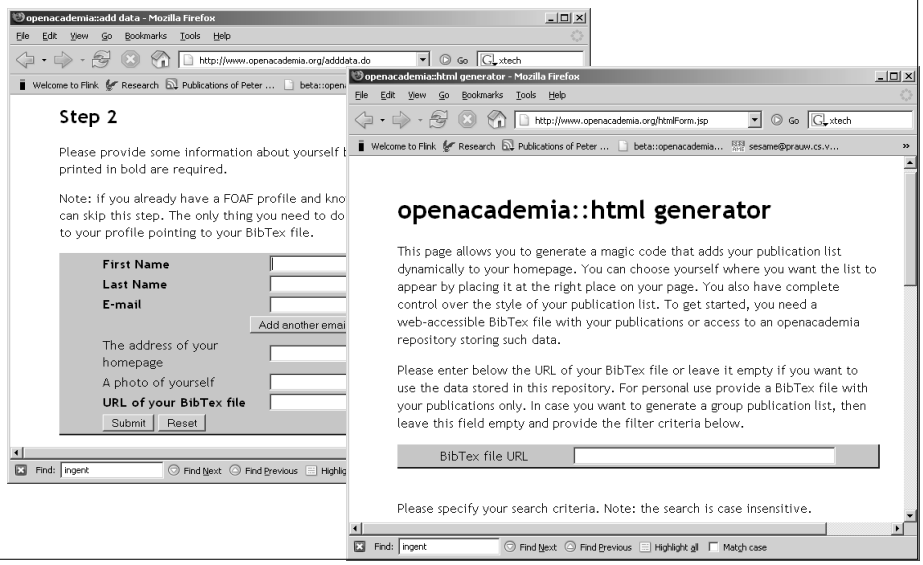
Contents

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

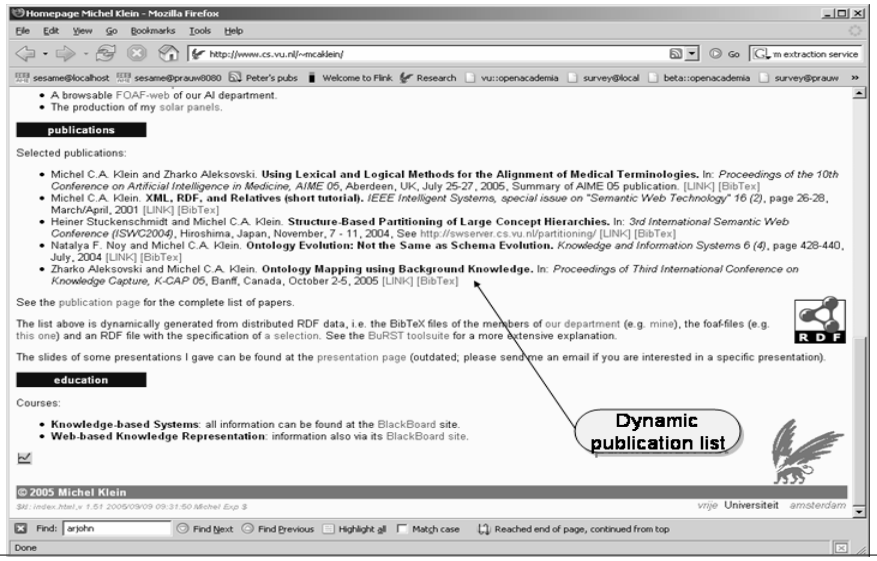
openacademia

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

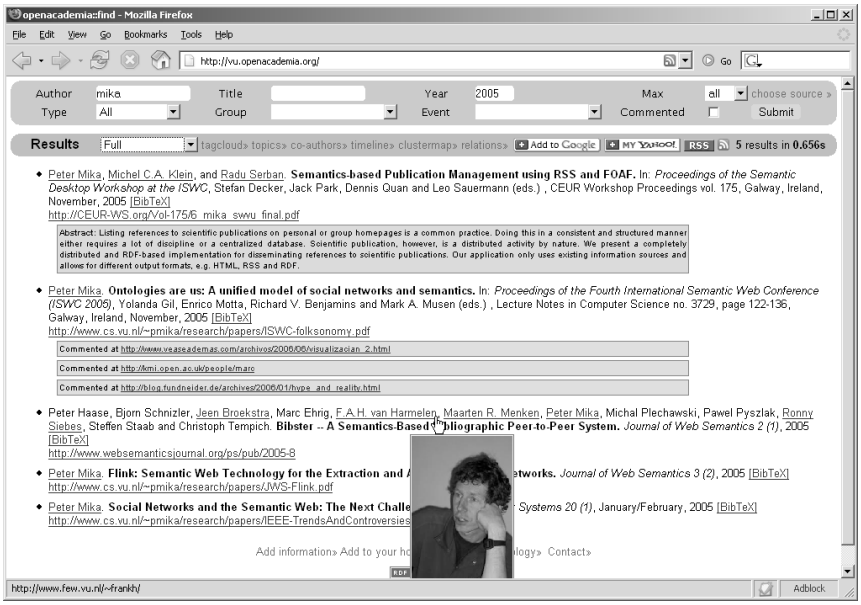
Pimp your homepage



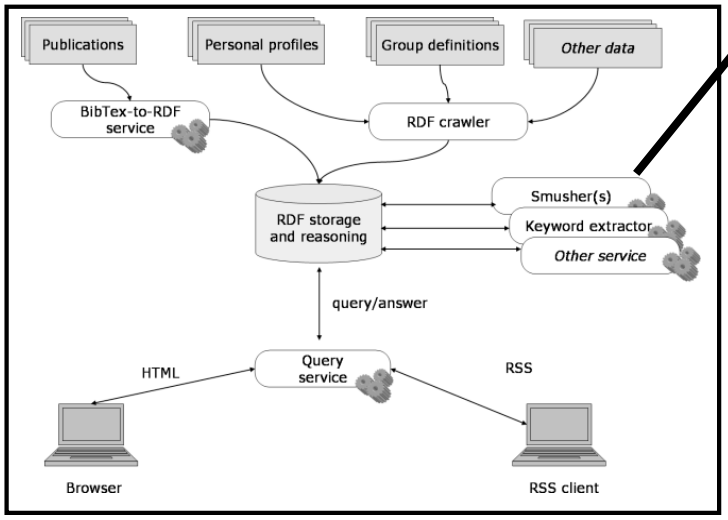
Publication list for homepage



Query interface



Architecture



Introduces Vagueness

Requires Instance unification

Contents

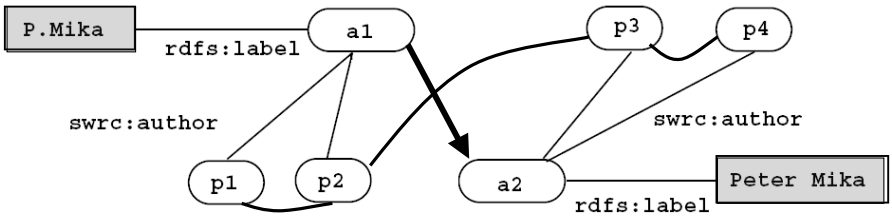
- openacademia
  - functions, architecture
- Instance unification: what, why and how
- Rough Set Theory and RoughDL
- Querying with approximations in openacademia
- Conclusions

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 **`owl:sameAs`** statements
  - drawbacks:
    - often an overcommitment
      - there only is some evidence for equivalence
    - possibly logical inconsistencies
    - source of statements is lost (context)

Contents

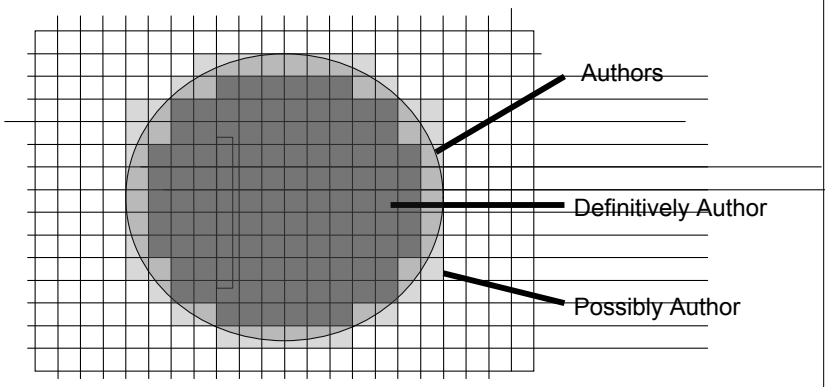
- openacademia
  - functions, architecture
- Instance unification: what, why and how
- Rough Set Theory and RoughDL
- Querying with approximations in openacademia
- Conclusions

Vagueness in Classical DL

- How can vagueness be represented in classical DL?
  Definitely_Author **v** Author **v Possibly**_Author
- Let us look at a more complex case

Possibly_SW_Author	???	**Possibly**_Author
v		v
SW_Author	v	Author
v		v
Definitively_SW_Author	???	**Definitively**_Author

Rough Description Logics: intuition



Definitive Authors: all similar persons are Authors

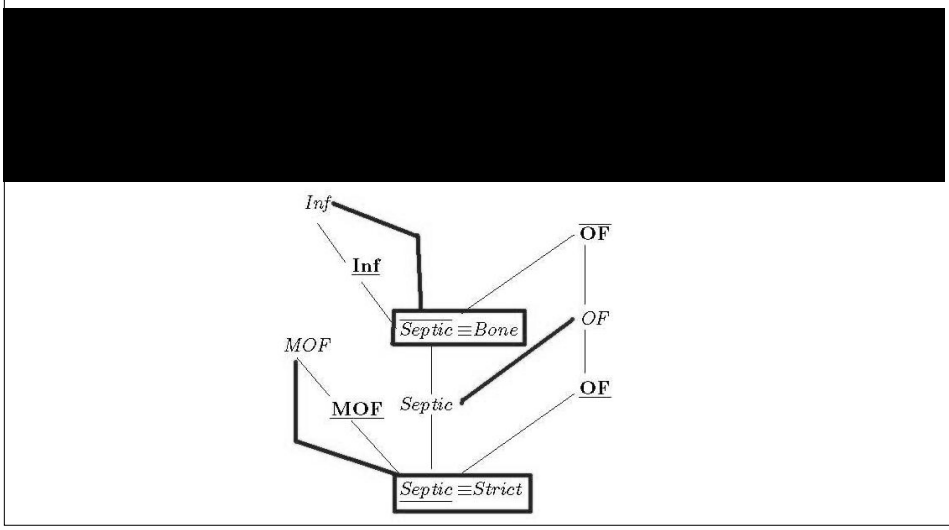
Possible Authors: there is a similar persons which is an Author

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
  - $\overline{C}^I$ = {i **2** U **j 9** j**2** U:(i,j) **2** $R^{»}$ & j**2** C^I}
  - $\underline{C}^I$ = {i **2** U **j 8** j**2** U:(i,j) **2** $R^{»}$ **!** j**2** C^I}
  - 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: $(\overline{C})^t$=**9** $r^{»}$.C; $(\underline{C})^t$=**8** $r^{»}$.C and reason classically

Some Rough DL consequences

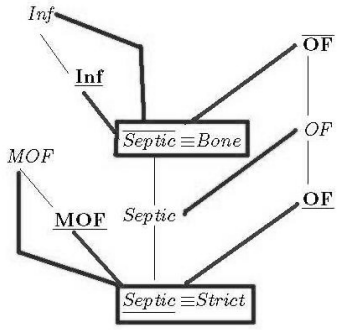
- T = {…, $\overline{\text{Septic}}$ **v** Inf, Septic **v** OF, $\underline{\text{Septic}}$ **v** MOF}



Some Rough DL consequences

- T = {…, $\overline{\text{Septic}}$ **v** Inf, Septic **v** OF, $\underline{\text{Septic}}$ **v** MOF}

- Possibly septic pats are definitely infected:
  T ² Septic **v** $\underline{\text{Inf}}$.
- Possibly septic pats have possible organ failure:
  T ² Septic **v** OF.



More Rough DL consequences

- *There are no definitely non-typical sepsis patients*.
  - T **[** {NTS **v** Septic **u :** Sepsis} ² $\underline{\text{NTS}}$ = **?**
- *Approximations of approximations are equivalent to the approximations themselves*
  - $\underline{\text{Septic}}$ ≡ $\underline{\text{Septic}}$

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

RoughOWL

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

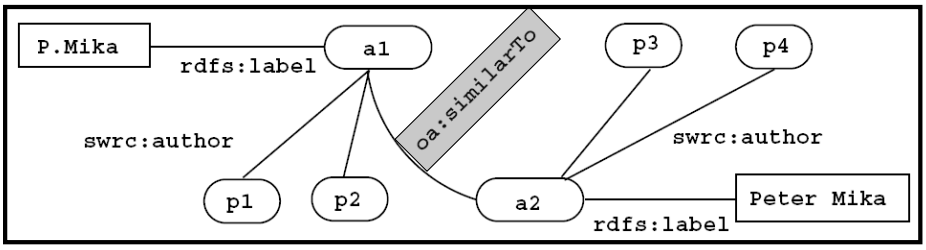
  E.g.: A is an upper approximation of B when using property X as similarity relation
- Conservative extension of OWL
  - standard reasoning services are sufficient
    - Racer, FaCT++, Pellet
- Up to now, we only use RoughOWL for querying
  - No representations & deduction in RoughOWL
  - Because of our 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



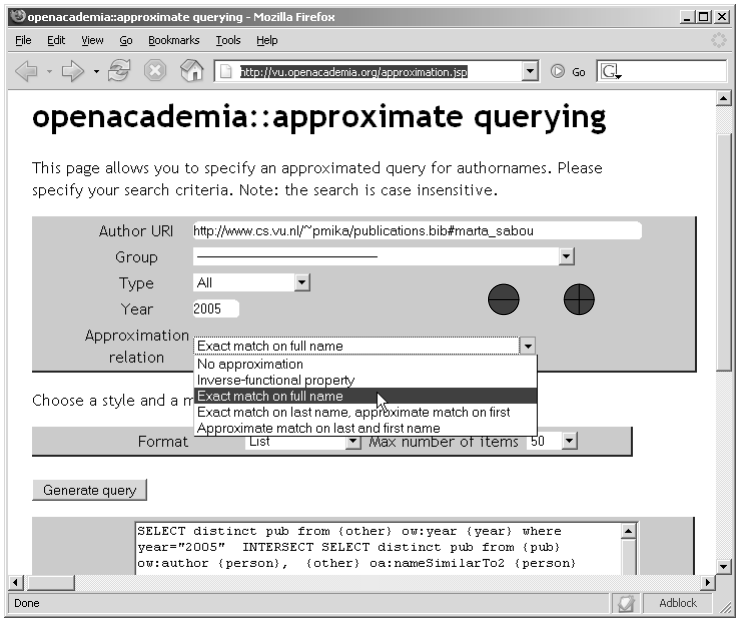
RoughOWL → SeRQL

- 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