

# 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 is represented in RDF
  - BibTeX data
  - FOAF data
  - blog citations
- built around Sesame repository

## Pimp your homepage

The image displays two overlapping browser windows from Mozilla Firefox. The left window, titled 'openacademia:acdd data - Mozilla Firefox', shows a registration page titled 'Step 2'. It contains the following text: 'Please provide some information about yourself (printed in bold are required). Note: If you already have a FOAF profile and know can skip this step. The only thing you need to do to your profile pointing to your BibTeX file.' Below this is a form with fields for 'First Name', 'Last Name', 'E-mail', 'The address of your homepage', 'A photo of yourself', and 'URL of your BibTeX file'. There are 'Submit' and 'Reset' buttons. The right window, titled 'openacademia:html generator - Mozilla Firefox', shows a page titled 'openacademia::html generator'. It contains the text: 'This page allows you to generate a magic code that adds your publication list dynamically to your homepage. You can choose yourself 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. 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.' Below this is a form with a 'BibTeX file URL' field. At the bottom, it says 'Please specify your search criteria. Note: the search is case insensitive.'

# Publication list for homepage

Homepage Michel Klein - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

http://www.cs.vu.nl/~mcklein/

sesame@kocahost sesame@prauw0000 Peter's pubs Welcome to Flink Research vu:openacademia survey@local beta::openacademia survey@prauw

- A browsable FOAF-web of our AI department.
- The production of my solar panels.

**publications**

Selected publications:

- Michel C.A. Klein and Zharko Aleksovski. **Using Lexical and Logical Methods for the Alignment of Medical Terminologies**. In: *Proceedings of the 10th Conference on Artificial Intelligence in Medicine, AIME 05*, Aberdeen, UK, July 25-27, 2005, Summary of AIME 05 publication. [LINK] [BibTeX]
- Michel C.A. Klein. **XML, RDF, and Relatives (short tutorial)**. *IEEE Intelligent Systems, special issue on "Semantic Web Technology" 16 (2)*, page 26-28, March/April, 2001 [LINK] [BibTeX]
- Heiner Stuckenschmidt and Michel C.A. Klein. **Structure-Based Partitioning of Large Concept Hierarchies**. In: *3rd International Semantic Web Conference (ISWC2004)*, Hiroshima, Japan, November, 7 - 11, 2004. See <http://wwwserver.cs.vu.nl/partitioning/> [LINK] [BibTeX]
- Natalya F. Noy and Michel C.A. Klein. **Ontology Evolution: Not the Same as Schema Evolution**. *Knowledge and Information Systems 6 (4)*, page 428-440, July, 2004 [LINK] [BibTeX]
- Zharko Aleksovski and Michel C.A. Klein. **Ontology Mapping using Background Knowledge**. In: *Proceedings of Third International Conference on Knowledge Capture, K-CAP 05*, Banff, Canada, October 2-5, 2005 [LINK] [BibTeX]

See the publication page for the complete list of papers.

The list above is dynamically generated from distributed RDF data, i.e. the BibTeX files of the members of our department (e.g. mine), the foaf-files (e.g. this one) and an RDF file with the specification of a selection. See the BuRST toolsuite for a more extensive explanation.

The slides of some presentations I gave can be found at the presentation page (outdated; please send me an email if you are interested in a specific presentation).

**education**

Courses:

- Knowledge-based Systems**: all information can be found at the BlackBoard site.
- Web-based Knowledge Representation**: information also via its BlackBoard site.

© 2005 Michel Klein

vu: Universiteit amsterdam

Find: jefhn Find Next Find Previous Highlight all Match case Reached end of page, continued from top

Done

# Query interface

openacademia:find - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

http://vu.openacademia.org/

Author mika Title Year 2005 Max all choose source

Type All Group Event Commented Submit

Results Full tagcloud> topics> co-authors> timelines> clustermaps> relations> Add to Google MY Yahoo! RSS 5 results in 0.656s

- Peter Mika, Michel C.A. Klein, and Radu Serban. **Semantics-based Publication Management using RSS and FOAF**. In: *Proceedings of the Semantic Desktop Workshop at the ISWC, Stefan Decker, Jack Park, Dennis Quan and Leo Sauermann (eds.)*, CEUR Workshop Proceedings vol. 175, Galway, Ireland, November, 2005 [BibTeX]  
[http://CEUR-WS.org/Vol-175/6\\_mika\\_svwu\\_final.pdf](http://CEUR-WS.org/Vol-175/6_mika_svwu_final.pdf)

Abstract: Listing references to scientific publications on personal or group homepages is a common practice. Doing this in a consistent and structured manner either requires a lot of discipline or a centralized database. Scientific publication, however, is a distributed activity by nature. We present a completely distributed and RDF-based implementation for disseminating references to scientific publications. Our application only uses existing information sources and allows for different output formats, e.g. HTML, RSS and RDF.

- Peter Mika. **Ontologies are us: A unified model of social networks and semantics**. In: *Proceedings of the Fourth International Semantic Web Conference (ISWC 2005)*, Yolanda Gil, Enrico Motta, Richard V. Benjamins and Mark A. Musen (eds.), Lecture Notes in Computer Science no. 3729, page 122-136, Galway, Ireland, November, 2005 [BibTeX]  
<http://www.cs.vu.nl/~pmika/research/papers/ISWC-folksonomy.pdf>

Commented at [http://www.vseasidemas.com/archives/2006/01/visualizacija\\_n\\_2.html](http://www.vseasidemas.com/archives/2006/01/visualizacija_n_2.html)

Commented at <http://hmi.open.ac.uk/people/marek>

Commented at [http://blog.fundneider.de/archives/2006/01/hvps\\_and\\_reality.htm](http://blog.fundneider.de/archives/2006/01/hvps_and_reality.htm)

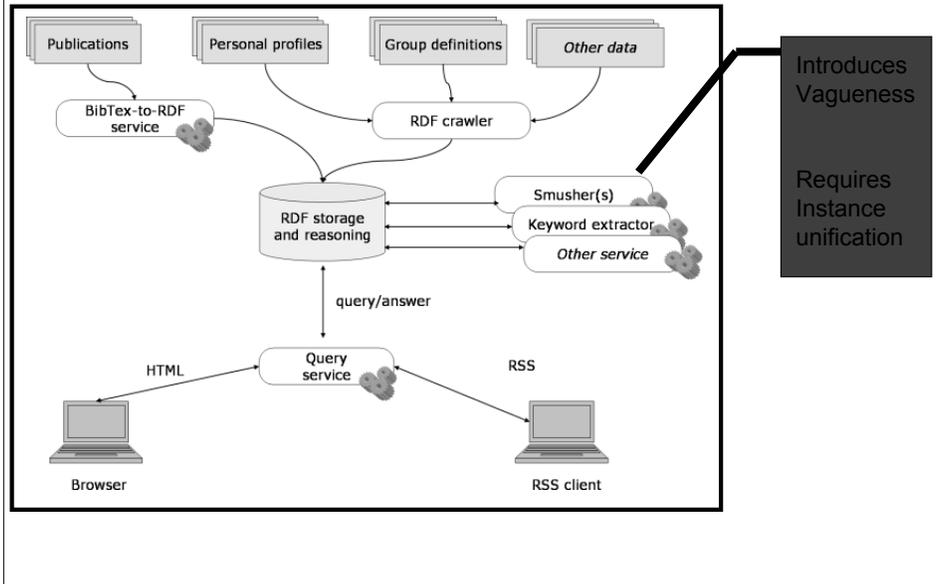
- Peter Haase, Bjorn Schinzier, Jeen Broekstra, Marc Ehrig, F.A.H. van Harmelen, Maarten R. Menken, Peter Mika, Michal Plechawski, Pawel Pyszlak, Romy Siebes, Steffen Staab and Christoph Tempich. **Bibster - A Semantics-Based Bibliographic Peer-to-Peer System**. *Journal of Web Semantics 2 (1)*, 2005 [BibTeX]  
<http://www.websemanticsjournal.org/ps/pub/2005-8>
- Peter Mika. **Flink: Semantic Web Technology for the Extraction and Analysis of Social Networks**. *Journal of Web Semantics 3 (2)*, 2005 [BibTeX]  
<http://www.cs.vu.nl/~pmika/research/papers/IWS-Flink.pdf>
- Peter Mika. **Social Networks and the Semantic Web: The Next Challenge**. *Knowledge and Information Systems 20 (1)*, January/February, 2005 [BibTeX]  
<http://www.cs.vu.nl/~pmika/research/papers/IEEE-TrendsAndControversies>

Add information> Add to your h...> Blog> Contact>

http://www.few.vu.nl/~frankh/

AdBlock

# Architecture



# Contents

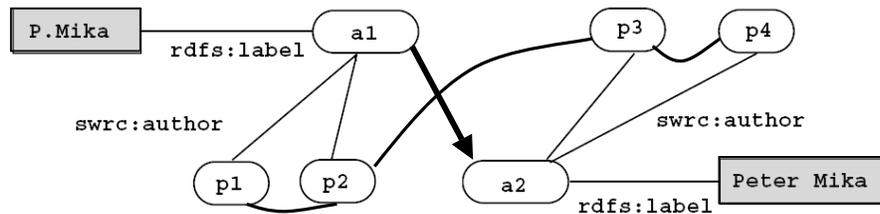
- openacademia
  - functions, architecture
- Instance unification: what, why and how
- Rough Set Theory and RoughDL
- Querying with approximations in openacademia
- Conclusions

## 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?

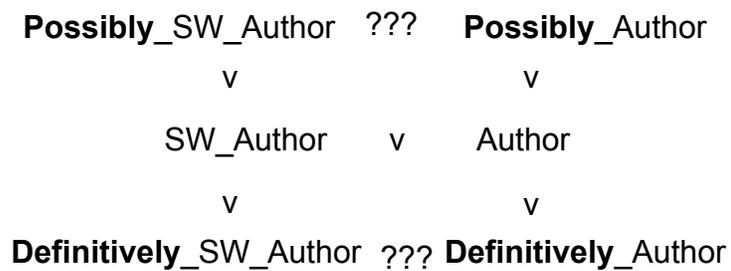
- Mainly NLP techniques to discover duplicates
- Usual implementation in SW context:
  - adding owl : sameAs statements
  - drawbacks:
    - often an overcommitment
      - there only is some evidence for equivalence
    - possibly logical inconsistencies
    - source of statements is lost (context)

## Contents

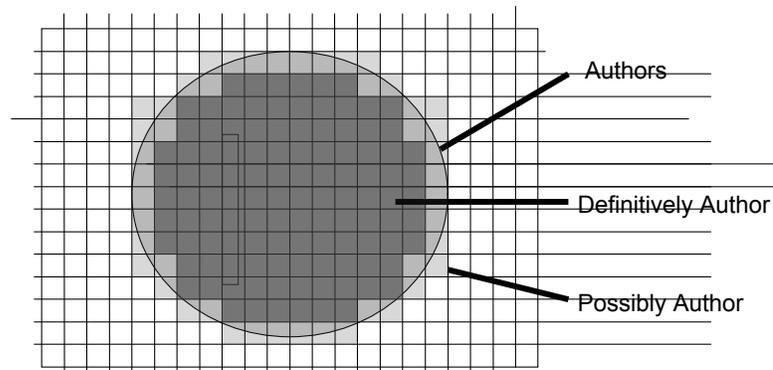
- openacademia
  - functions, architecture
- Instance unification: what, why and how
- Rough Set Theory and RoughDL
- Querying with approximations in openacademia
- Conclusions

# Vagueness in Classical DL

- How can vagueness be represented in classical DL?  
**Definitely**\_Author  $\sqsubset$  Author  $\sqsubset$  **Possibly**\_Author
- Let us look at a more complex case



## Rough Description Logics: intuition



**Definitive Authors:** all similar persons are Authors

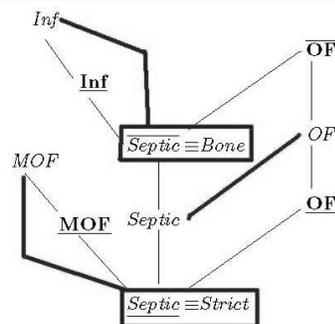
**Possible Authors:** there is a similar persons which is an Author

## Rough Description Logics: formal issues

- Syntax
  - For every concept  $C$  in RDL also the *upper approximation*  $\overline{C}$  in RDL and the *lower approximation*  $\underline{C}$  in RDL.
- Semantics
  - $\overline{C} = \{i \in U \mid \exists j \in U: (i,j) \in R^* \ \& \ j \in C\}$
  - $\underline{C} = \{i \in U \mid \exists j \in U: (i,j) \in R^* \ \& \ j \in C\}$
  - 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:  $(\overline{C})^{\mathbf{9}} \substack{r^* \\ \supset} C$ ;  $(\underline{C})^{\mathbf{8}} \substack{r^* \\ \supset} C$  and reason classically

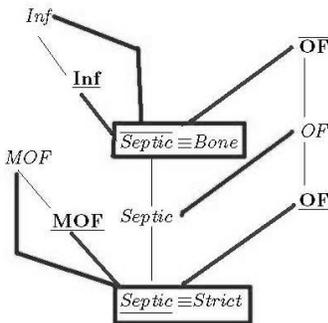
## Some Rough DL consequences

- $T = \{\dots, \overline{\text{Septic}} \vee \text{Inf}, \text{Septic} \vee \text{OF}, \underline{\text{Septic}} \vee \text{MOF}\}$



## Some Rough DL consequences

- $T = \{\dots, \overline{\text{Septic}} \vee \text{Inf}, \text{Septic} \vee \text{OF}, \underline{\text{Septic}} \vee \text{MOF}\}$
- Possibly septic pats are definitely infected:  
 $T \models \text{Septic} \vee \underline{\text{Inf}}$ .
- Possibly septic pats have possible organ failure:  
 $T \models \text{Septic} \vee \text{OF}$ .



## More Rough DL consequences

- *There are no definitely non-typical sepsis patients.*
  - $T \models \{\text{NTS} \vee \text{Septic} \vee \text{Sepsis}\} \models \underline{\text{NTS}} = ?$
- *Approximations of approximations are equivalent to the approximations themselves*
  - $\underline{\text{Septic}} \equiv \underline{\text{Septic}}$
- **Comment:** these are simple S5 consequences!

# RoughOWL

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

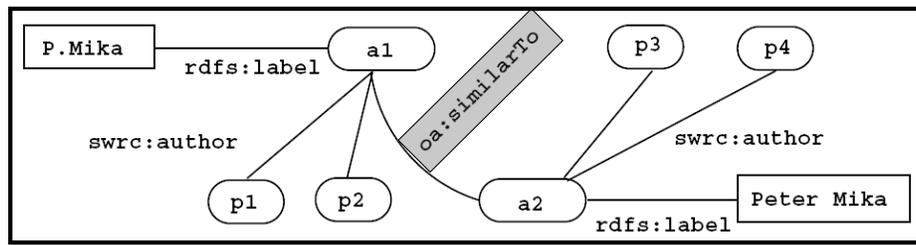
*E.g.: A is an upper approximation of B when using property X as similarity relation*
- Conservative extension of OWL
  - standard reasoning services are sufficient
    - Racer, FaCT++, Pellet
- Up to now, we only use RoughOWL for querying
  - No representations & deduction in RoughOWL
  - Because of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



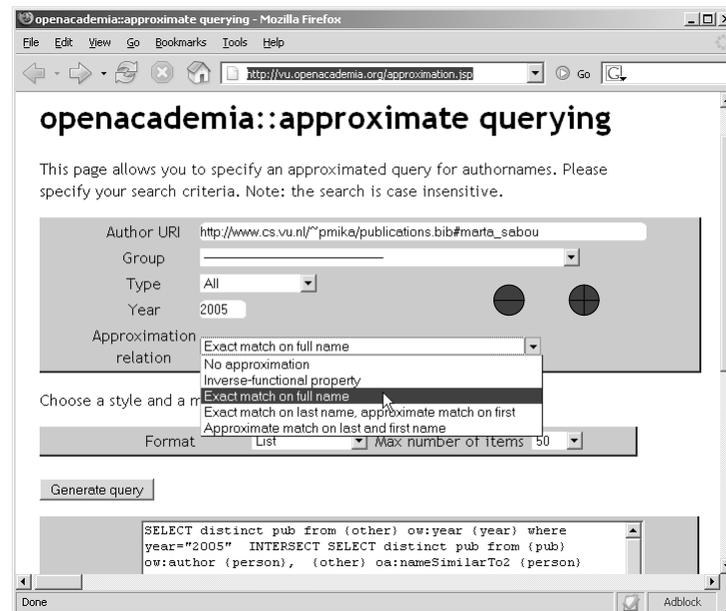
## RoughOWL → SeRQL

- 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/publications.bib#marta\\_sabou](http://www.cs.vu.nl/~pmika/publications.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