Karlsruhe Institute of Technology

Methods and Tools for Semi-automatic Ontology Engineering

Johanna Völker
Postdoctoral Researcher
Institute AIFB, University of Karlsruhe

Forschungszentrum Karlsruhe
in der Helmholtz-Gemeinschaft

Universität Karlsruhe (TH)
Forschungsuniversität • gegründet 1825

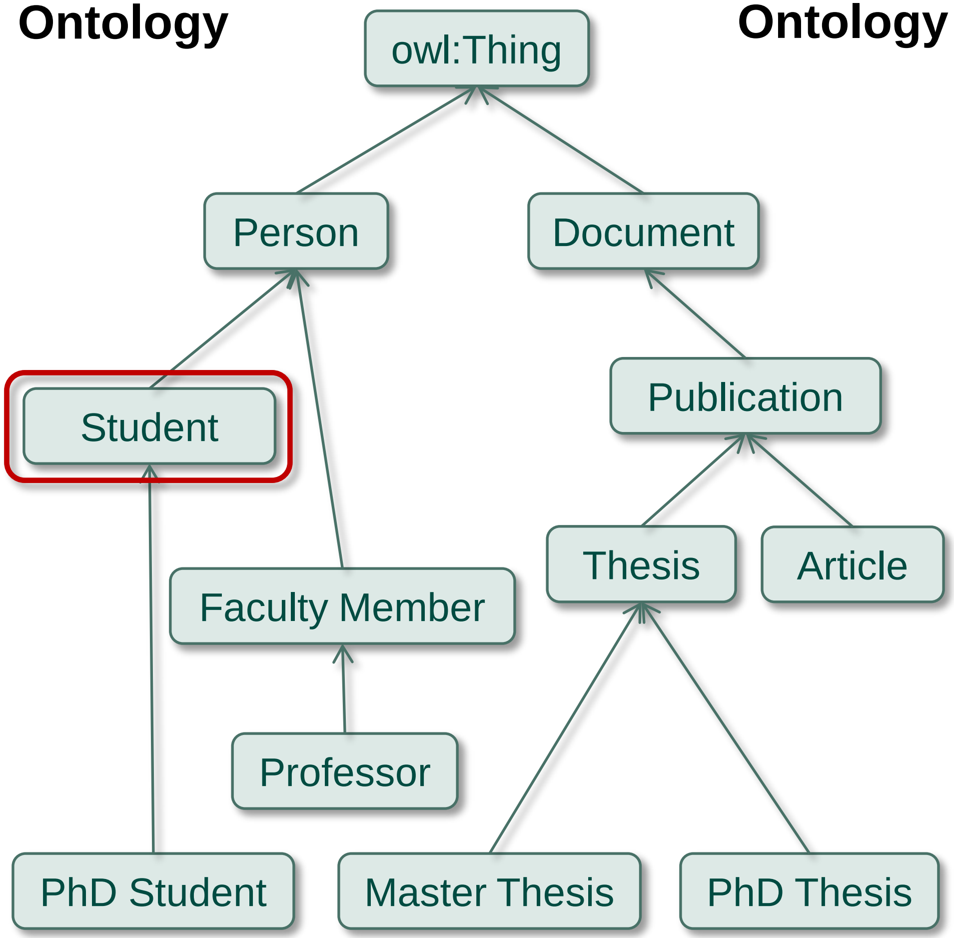
Ontologies

- *Formal specification of a conceptualization**
  - „Formal“: machine-interpretable
  - „Conzeptualization“: abstract model of a domain
- Specification
  - Ontology languages: F-Logic, RDF(S), **OWL** etc.
    - OWL1: W3C standard since 2004
    - Sub-languages: OWL Lite, **OWL DL**, OWL Full
  - Syntactic notations and visualizations
    - Graphs or sets of axioms

* Th. Gruber. *Toward Principles for the Design of Ontologies Used for Knowledge Sharing.* International Journal on Human-Computer Studies. Vol. 43, Issues 5-6, pp. 907-928, November 1995

Ontologies as Formal Specifications

SWRC Ontology



TAMBIS Ontology

Genomic-dna
$\equiv \text{Macromolecular-compound}$
$\sqcap\ \exists\text{polymer-of.Deoxy-nucleotide}$
$\sqcap\ \forall\text{polymer-of.Deoxy-nucleotide}$
$\sqcap\ \exists\text{strandedness.Double-stranded}$
$\sqcap\ {=}1\text{part-of}.\top$
$\sqcap\ \forall\text{part-of.(Nuclear-chromosome}$
$\sqcup\ \text{Mitochondrial-chromosome}$
$\sqcup\ \text{Chloroplast-chromosome)}$

$\top \sqsubseteq \forall\text{strandedness}^{-1}.\text{Nucleic-acid}$

$\text{Lysis} \equiv \text{Reaction} \sqcap \exists\text{lysis-of.Covalent-bond}$

$\text{Protein} \sqcap \text{Nucleic-acid} \sqsubseteq \bot$

Holoprotein
$\equiv \text{Macromolecular-compound}$
$\sqcap\ \exists\text{polymer-of.Amino-acid}$
$\sqcap\ \forall\text{polymer-of.Amino-acid}$
$\sqcap\ \geq 1\ \text{has-bound}.\top$
$\sqcap\ \exists\text{has-bound.Prosthetic-group}$

Motivation

- **„Lightweight"** ontologies contain underspecified class descriptions, important classes or properties are missing
  - Logical derivation of **class memberships** and **query answering** are hindered by lack of **expressivity**
- Example: SWRC (*Semantic Web for Research Communities*)
  - Persons, publications, projects etc.
  - Query: „*Is Rudi Studer a **PhD**?*"
  - Class memberships are often determined by…
    - Explicit or implicit disjointness of classes
    - Relations and attributes (z.B. „*Rudi Studer is the **author of a PhD thesis**. He **gives lectures** on semantic technologies.*")
  - **Problem:** Identification and formalization of missing knowledge

Agenda

- Ontologies
- **Ontology Learning**
- Semi-automatic Engineering of Expressive Ontologies
  - Experiments
- Conclusion

Ontology Learning

- *"Ontology learning is a subtask of* ***information extraction****. The goal of ontology learning is to (semi-)automatically extract* ***relevant concepts*** *and* ***relations*** *from a given corpus or other kinds of data sets to form an Ontology."**
- *"Ontology Learning is a mechanism for* ***semi-automatically*** *supporting the ontology engineer in engineering ontologies."***
- *"Ontology Learning aims at the integration of a multitude of disciplines in order to facilitate the construction of ontologies, in particular* ***ontology engineering*** *and* ***machine learning****."****

* Wikipedia 2008/12/15: *http://en.wikipedia.org/wiki/Ontology_learning*

** A. D. Mädche. *Ontology Learning for the Semantic Web*. Dissertation. Universität Karlsruhe, 2001

*** A. D. Mädche, S. Staab. *Ontology Learning*. Handbook of Ontologies in Information Systems, 2004

Tools and Frameworks

Karlsruhe Institute of Technology

Lexical ontology learning: informal or semi-formal data (e.g. texts)

Framework	Institution	Reference
ASIUM	INRIA, Jouy-en-Josas	Faure and Nedellec 1999
TextToOnto	AIFB, University of Karlsruhe	Mädche and Volz 2001
HASTI	Amir Kabir University, Teheran	Shamsfard, Barforoush 2004
OntoLT	DFKI, Saarbrücken	Buitelaar et al. 2004
DOODLE	Shizuoka University	Morita et al. 2004
Text2Onto	AIFB, University of Karlsruhe	Cimiano and Völker 2005
OntoLearn	University of Rome	Velardi et al. 2005
OLE	Brno University of Technology	Novacek and Smrz 2005
OntoGen	Institute Jozef Stefan, Ljubljana	Fortuna et al., 2007
GALeOn	Technical University of Madrid	Manzano-Macho et al. 2008
DINO	DERI, Galway	Novacek et al. 2008
OntoLancs	Lancester University	Gacitua et al. 2008

Tools and Frameworks

Logical ontology learning: formal data (e.g. ontologies)

Framework	Institution	Reference
YINGYANG	University of Bari	Iannone 2006
DL Learner	University of Leipzig	Lehmann 2006
RELExO	AIFB, University of Karlsruhe	Völker and Rudolph 2008
RoLExO	AIFB, University of Karlsruhe	Völker and Rudolph 2008
OntoComp	University of Dresden	Sertkaya 2008

Hybrid implementations

Framework	Institution	Reference
LeDA	AIFB, University of Karlsruhe	Völker et al. 2007
SOFIE	MPI, Saarbrücken	Suchanek et al. 2009
...	...	...

Agenda

- Ontologies
- Ontology Learning
- **Semi-automatic Engineering of Expressive Ontologies**
  - **Experiments**
    - **Pattern-based Refactoring**
    - **Automatic Enrichment**
    - **Logical Refinement**
- Conclusion

Semi-automatic Ontology Engineering
Towards Learning (More) Expressive Ontologies

- **Experiment 1:** Pattern-based refactoring of lightweight ontologies
  - **Methods:** Lexico-syntactic patterns, WordNet
  - **Tools:** Text2Onto, OntoCase (Blomqvist 2007)

lexical

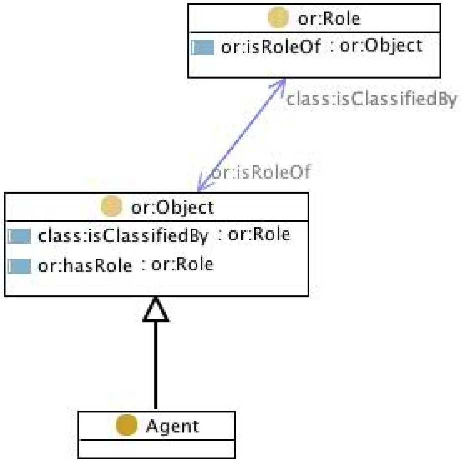
- **Experiment 2:** Automatic enrichment with disjointness axioms
  - **Methods:** Machine learning, lexical and logical features
  - **Tools:** LeDA

hybrid

- **Experiment 3:** Interactive refinement by relational exploration
  - **Methods:** Formal Concept Analysis, OWL reasoning
  - **Tools:** RELExO, RoLExO

logical

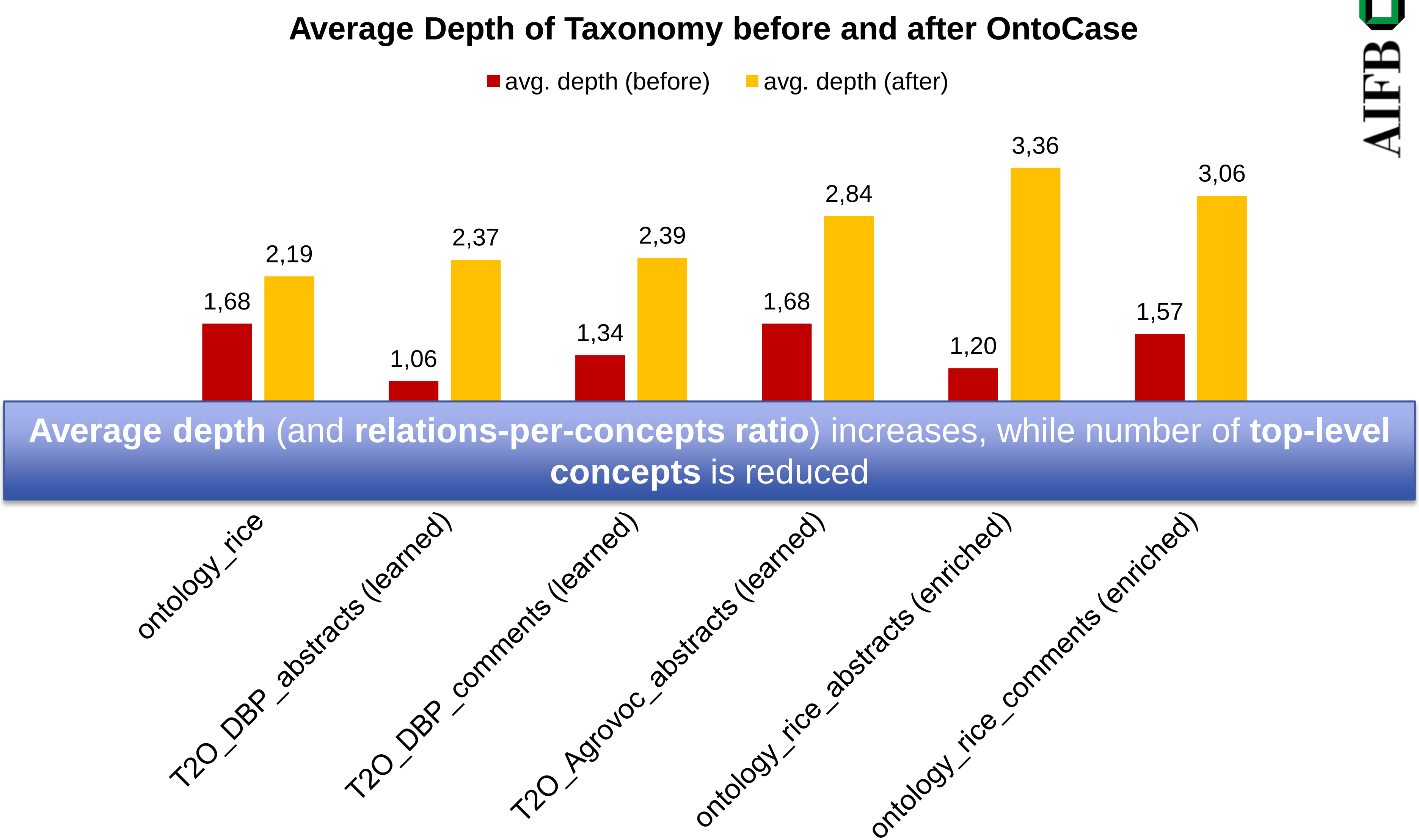
Experiment 1: Refactoring



- **Text2Onto** (Cimiano and Völker 2005)
  - Framework for ontology learning from text
  - Lightweight, semi-formal ontologies, lexical semantics
- **OntoCase** (Blomqvist 2007)
  - Automatic matching of ontology engineering patterns
- **http://ontologydesignpatterns.org**
  - Templates created from best practices in ontology engineering
  - Links between domain and top-level ontologies such as DOLCE
- Which **synergies** can arise from a combination of **ontology learning** and pattern-based **ontology engineering**?*
  - **Hypothesis 1:** Patterns can help to improve the structure and correctness of learned ontologies, even if matched automatically.
  - **Hypothesis 2:** Ontology learning can facilitate the pattern matching process.

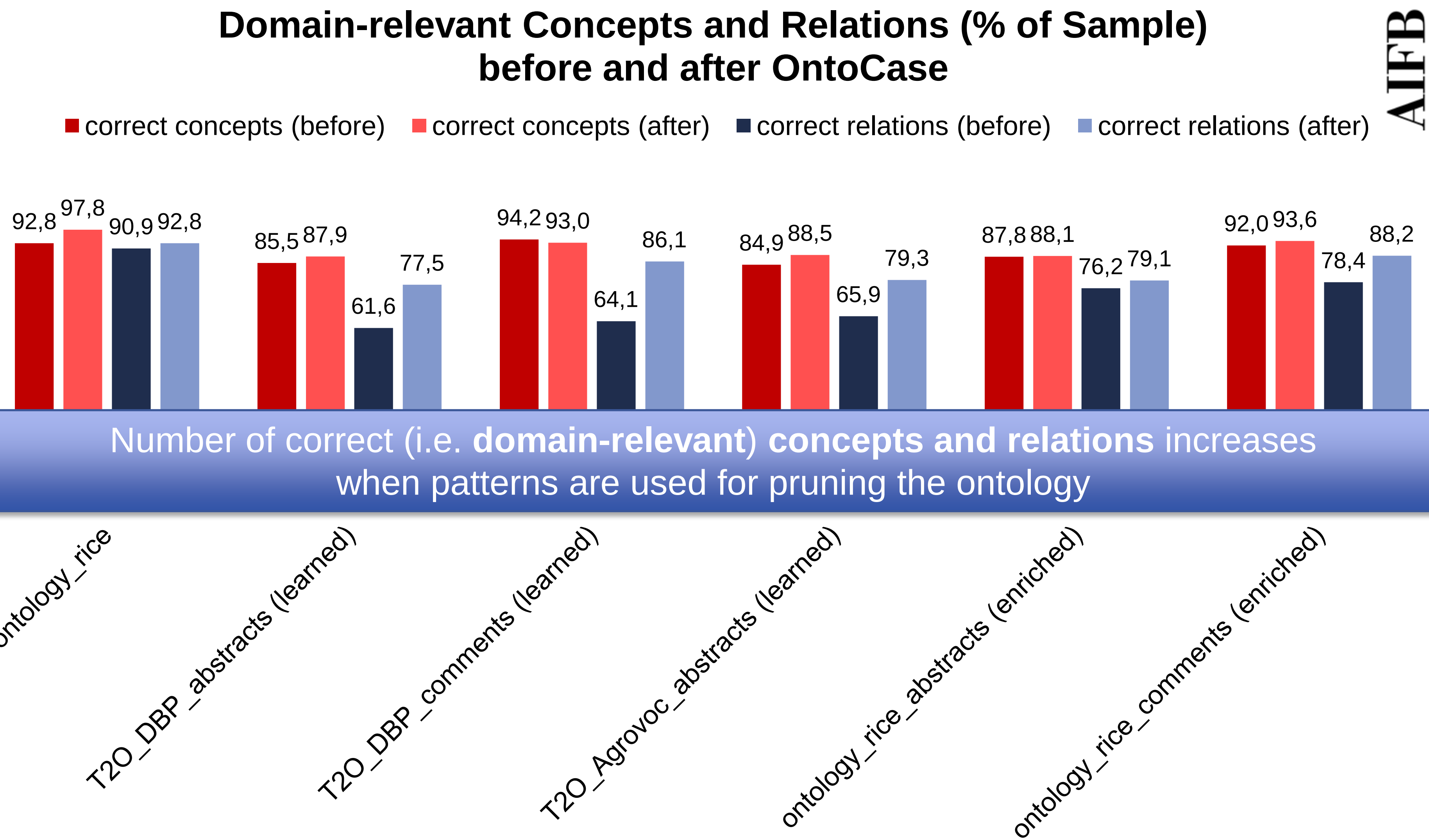
* Johanna Völker and Eva Blomqvist. *Evaluation of Methods for Contextualized Learning of Networked Ontologies*, NeOn Deliverable 3.8.2, February 2009

Evaluation: Experiment 1



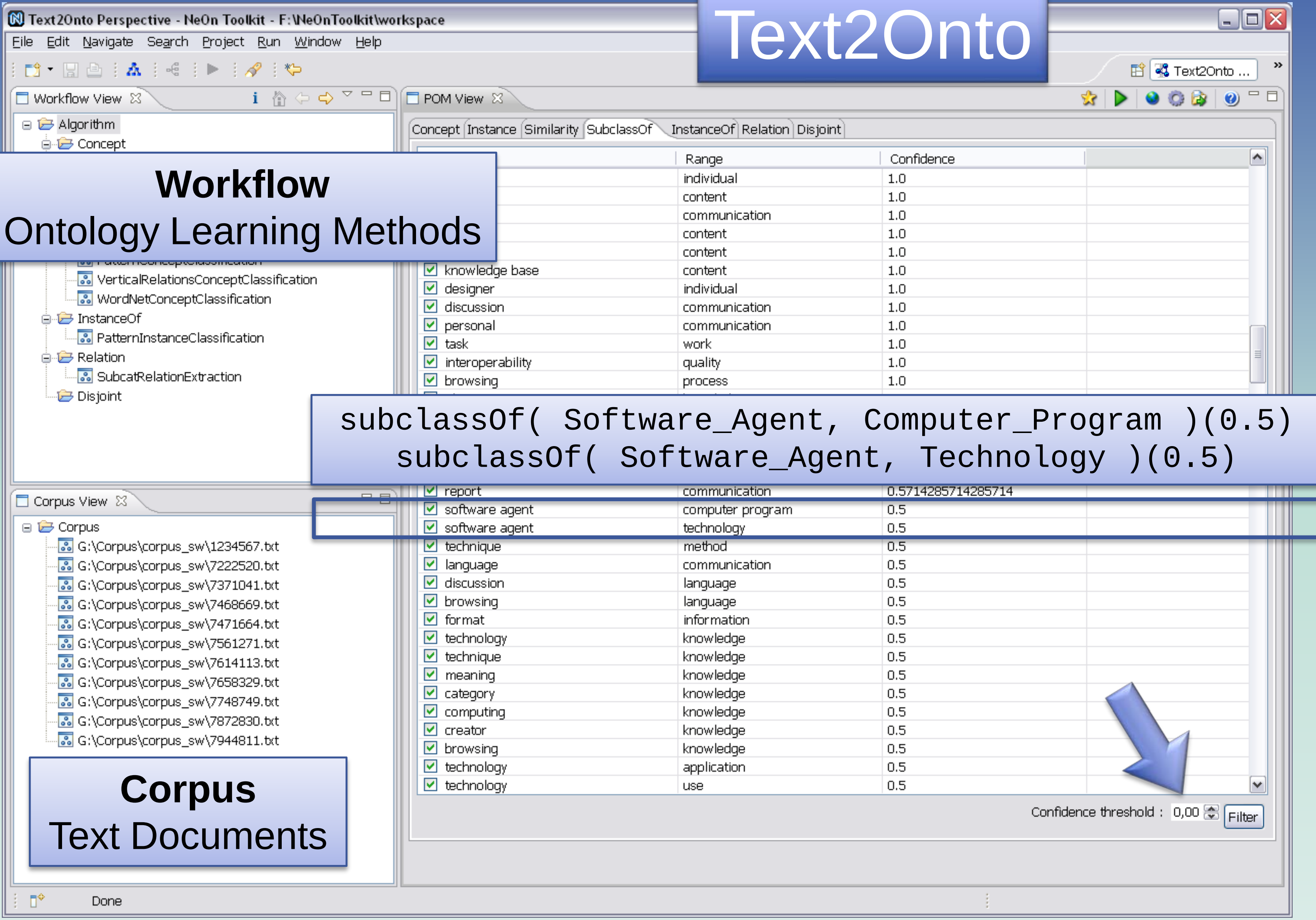
Average depth (and **relations-per-concepts ratio**) increases, while number of **top-level concepts** is reduced

Evaluation: Experiment 1



Number of correct (i.e. **domain-relevant**) **concepts and relations** increases when patterns are used for pruning the ontology

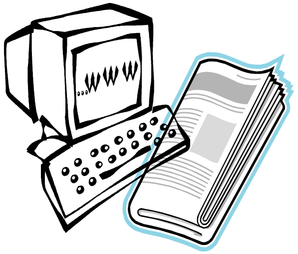
Text2Onto



http://www.neon-toolkit.org

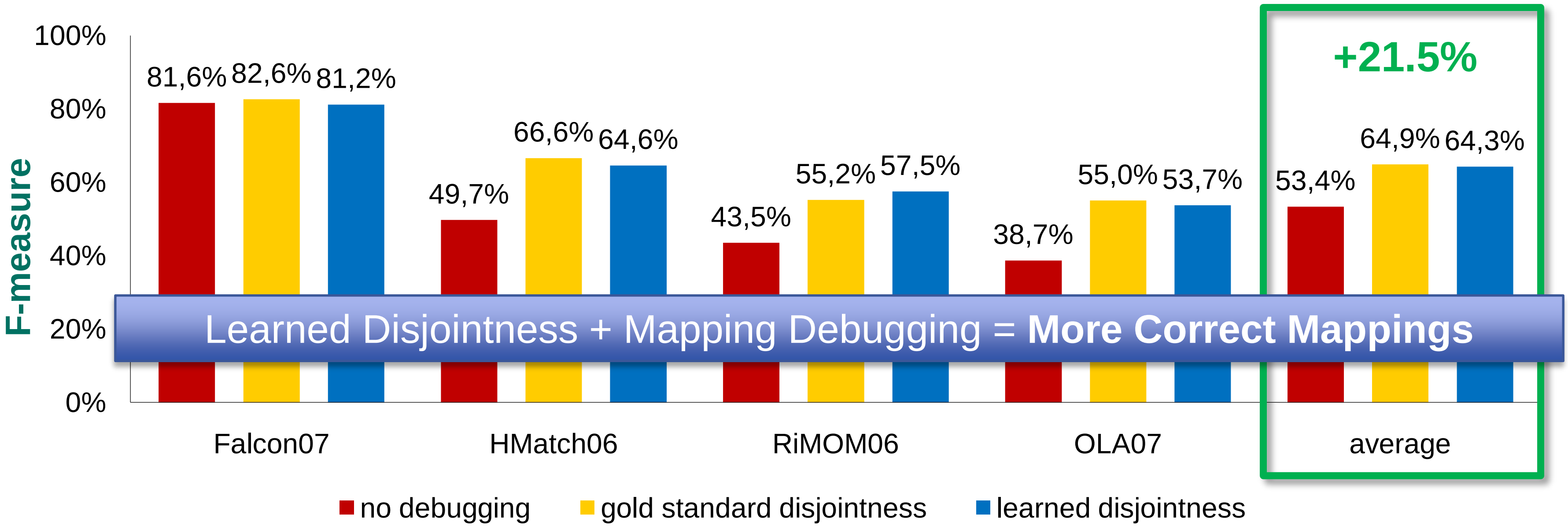
Experiment 2: Enrichment

- **LeDA** (Völker et al. 2007)
  - Classification-based approach with lexical and logical features
  - e.g. $Person \sqsubseteq \neg University$
- **Mapping debugging** (Meilicke and Stuckenschmidt 2007)
  - Automatically detect and remove incorrect mappings
- **Hypothesis**
  - **Ontology enrichment** facilitates **mapping debugging**, thus helps to improve the quality of automatically generated mappings
- **Experiment***
  - **Data set:** Conference ontologies (OAEI)
    - No disjointness axioms
    - Manually added disjointness axioms (**gold standard**)
    - Automatically generated disjointness axioms



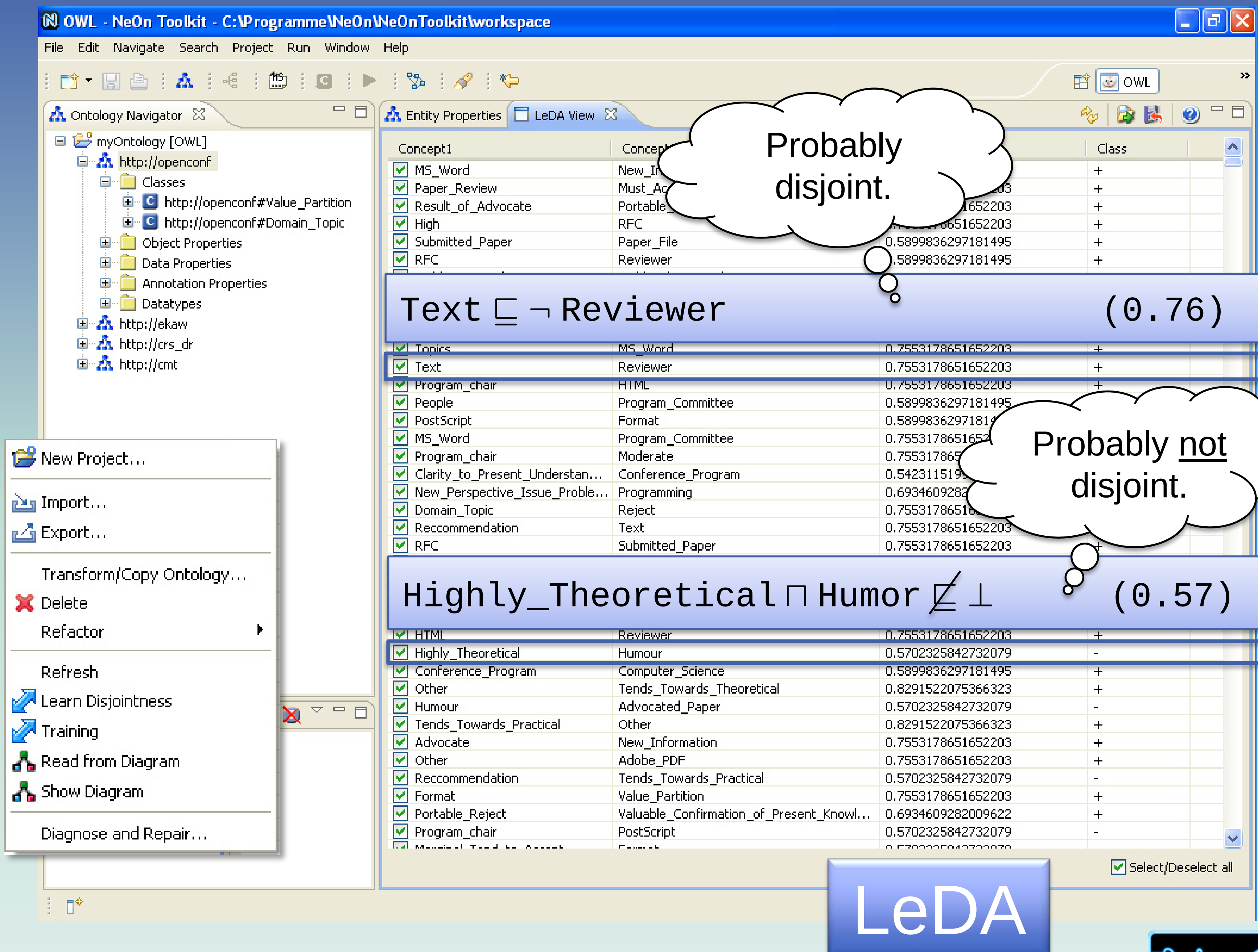
* Chr. Meilicke, J. Völker, H. Stuckenschmidt. *Debugging Mappings between Lightweight Ontologies.* International Conf. on Knowledge Engineering and Knowledge Management (EKAW), 2008 (*best paper award*)

Evaluation: Experiment 2



- Mapping **debugging** based on **gold standard** increases mapping quality considerably (11.5 percentage points = **21.5%**)
  - Significant increase of precision and a small loss of recall
  - But only minor changes for highly precise mappings of top system **Falcon**

- Differences between debugging based on **learned disjointness** and gold standard are only minor (**0.9%** on average)
  - Conclusion: even an **imperfect** set of disjointness axioms can be used to successfully perform mapping debugging

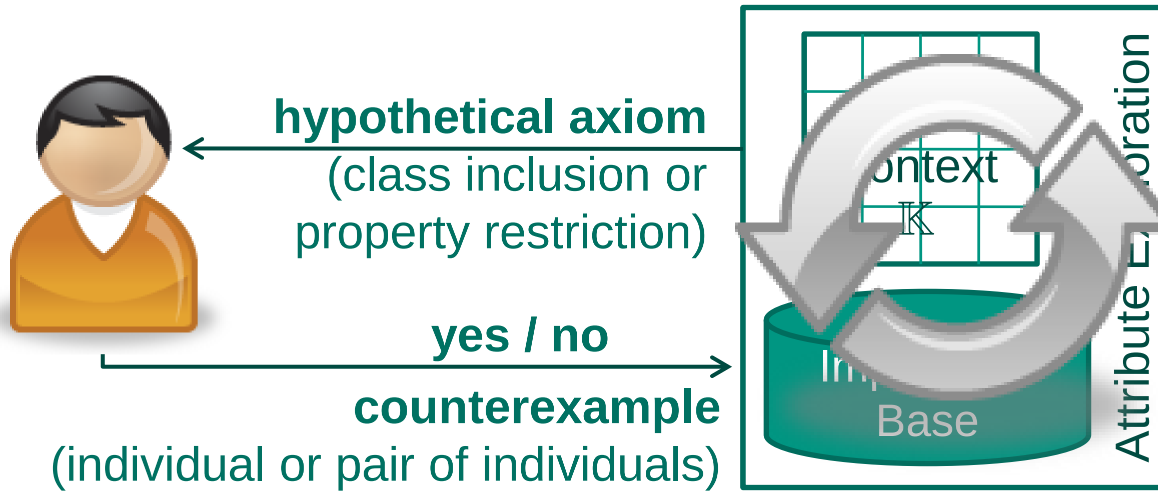


http://www.neon-toolkit.org

Experiment 3: Exploration

- **RELExO and RoLExO** (Völker and Rudolph 2008)
  - Efficient ontology completion by systematic expert interrogation
  - FCA and OWL reasoning to minimize workload for the ontology engineer

Result:
minimal and complete
set of missing axioms
(within a DL fragment)



- **Hypothesis***
  - Relational exploration increases the number of logical conclusions and facilitates query answering

* J. Völker and S. Rudolph. *Fostering Web Intelligence by Semi-automatic OWL Ontology Refinement*
In Proceedings of the 7th International Conference on Web Intelligence (WI). 2008. (*regular paper*)

Evaluation: Experiment 3

- **Acquired axioms**, e.g.
  - *„Every author of a thesis is a graduate."*

    $\text{Thesis} \sqcap \exists \text{has\_author}.\top \sqsubseteq \forall \text{has\_author}.\text{Graduate}$
  - *„Every thesis authored by a PhD student is a master thesis."*

    $\text{Thesis} \sqcap \exists \text{has\_author}.\text{PhDStudent} \sqsubseteq \text{MasterThesis}$

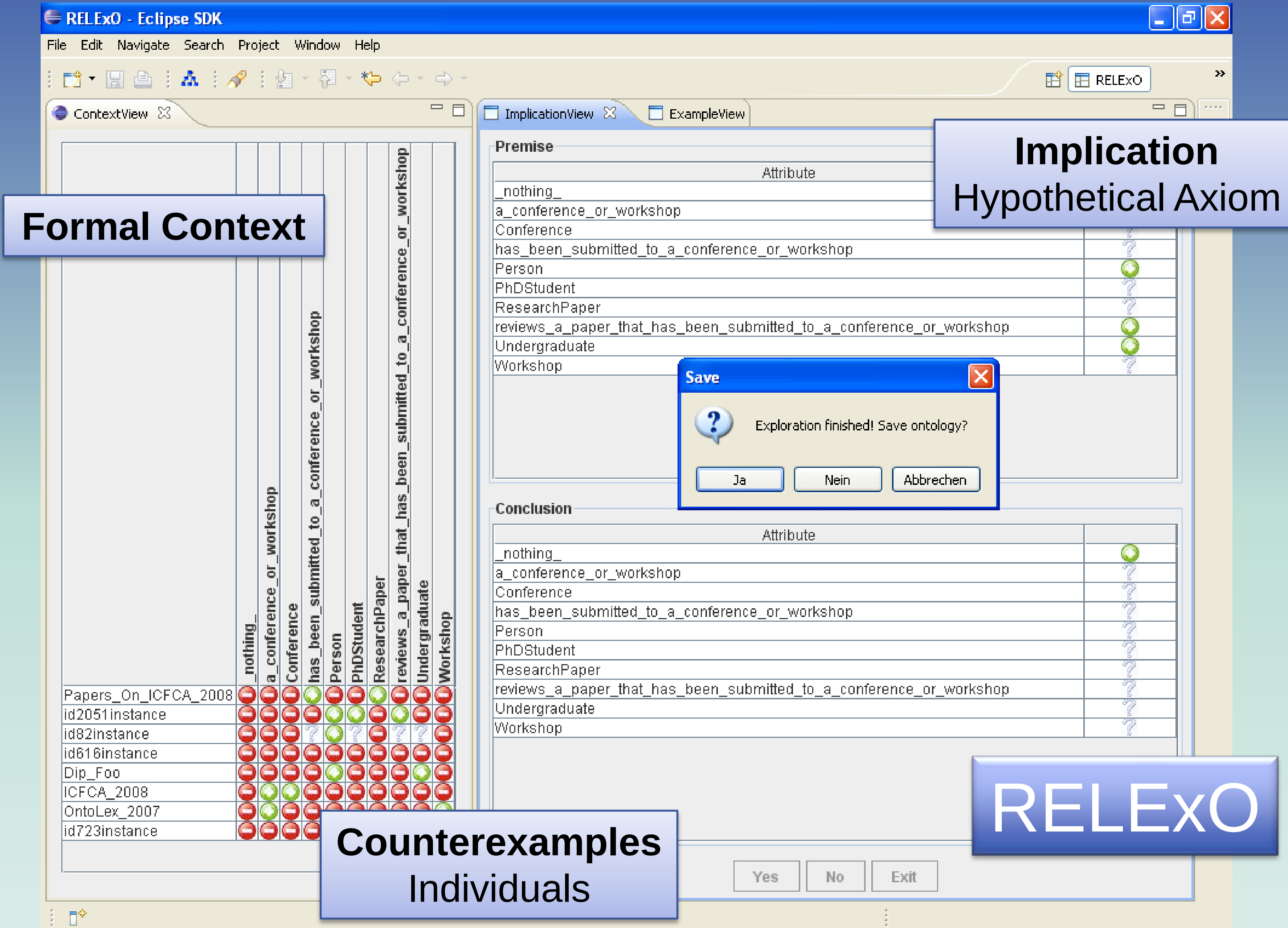
	RELExO (domain)	RELExO (range)	RoLExO	Sum
Answers (Reasoner)	9	8	19	36
Answers (Human)	6	5	13	24
New TBox-Axioms	5	3	4	12
New Individuals	1	2	14	17

Evaluation: Experiment 3

Class	SWRC	RELExO (range)	RELExO (domain)	RoLExO
Article	189	189	189	190
Book	36	36	94	95
MasterThesis	0	0	1	4
PhDThesis	58	58	58	59
Publication	1499	1499	1500	1507
Thesis	58	58	59	63
a_postdoc*	0	63	63	67
FullProfessor	6	6	6	9
Graduate	52	111	111	139
has_written_a_doctoral_thesis	0	63	63	67
Person	1213	1215	1215	1222
PhDStudent**	50	46	46	47
Undergraduate	6	7	7	9
	3167	**3351**	**3412**	**3478**

* New class added during the ontology refinement process. For details see (Völker and Rudolph 2008).

** Automatically retrieved counterexample **PostDoc ⊓ PhDStudent(Peter_Haase)** in response to hypothesis 2 of the exploration. Manual repair of the ontology by generalization of the explicit classification of 3 individuals.



http://www.neon-toolkit.org

Agenda

- Ontologies
- Ontology Learning
- Semi-automatic Engineering of Expressive Ontologies
  - Experiments
- **Conclusion**

Conclusion

- We can support the **construction** of expressive ontologies by ...
  - Ontology engineering **patterns**
  - **Enrichment** of ontologies (e.g. disjointness axioms)
  - Relational **exploration**, i.e. systematic completion
- **Expressivity** helps to ...
  - Add **structure** to flat or sparse ontologies
  - Detect incorrect mappings or **modeling errors**
  - Draw logical **conclusions** (e.g., instance classification)

If we can take advantage of these positive aspects, learning expressive ontologies is not necessarily more difficult!

Johanna Völker. *Learning Expressive Ontologies*. Dissertation, Universität Karlsruhe (TH), Fakultät für Wirtschaftswissenschaften, Dezember 2008

THANK YOU!