Efficient Instance Retrieval over Semi-Expressive Ontologies

TUHH Hamburg University of Technology

Dissertation Presentation

STS Software Technology Systems Institute for Software Systems

Sebastian Wandelt

Hamburg, 6th of October 2011

Chairman: Professor Volker Turau (Hamburg University of Technology)

Reviewer: Professor Ralf Möller (Hamburg University of Technology)

Reviewer: Professor Ian Horrocks (University of Oxford)

Reviewer: Professor Norbert Ritter (University of Hamburg)

Overview

- Motivation
- Research Objectives and Methodology
- Main Contributions and Evaluation
- Discussion and Future Work

Motivation and Background

- Semantic Web
- Ontologies / Description logics
- Reasoning is hard (expressivity vs. scalability)
- Thesis:

  "Instance retrieval for the description logic SHI"

Related Work

- Less expressive description logics
  - DL-Lite [ACKZ09]
- Sound only
  - Triple Stores [AG10], Approximations [RPZ10]
- Sound and complete
  - SHER [DFK09]
  - GCQs/CQs [HM08], [SBPKK07]
  - Rewriting [HMS07, HKRT08], Hypertableau [MSH09]
  - Instance Store [SHT05]
- Neither sound, nor complete
  - Approximations [TGH10]

Research Objectives

- Release the main-memory dependency from DL reasoning systems
- Focus on
  - Semi-expressive DLs (SHI), no datatypes
  - Large ABox, mid-size TBox / RBox
  - Atomic instance retrieval queries
- Prepare for ontology changes

Scientific Methodology

- Practical work
- Well-documented implementation
- Proofs
- Runs on off-the-shelf laptop
  - Intel C3 2.4 GHz, 4 GB RAM, 500GB, Windows 7, Java 6
- Evaluation with benchmark ontology

=>Reproducibility and repeatability

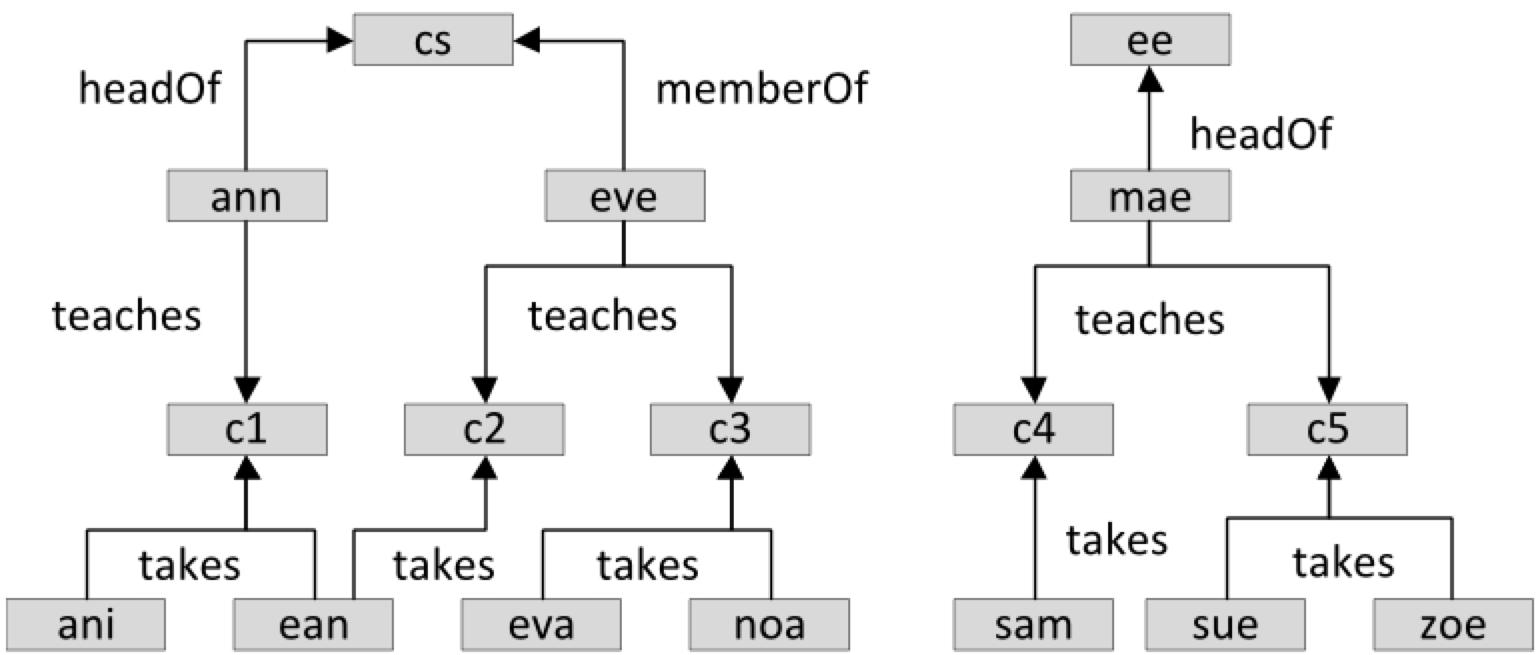
In the following …

- Instance checking
- Instance retrieval
- Ontology changes

Instance Checking

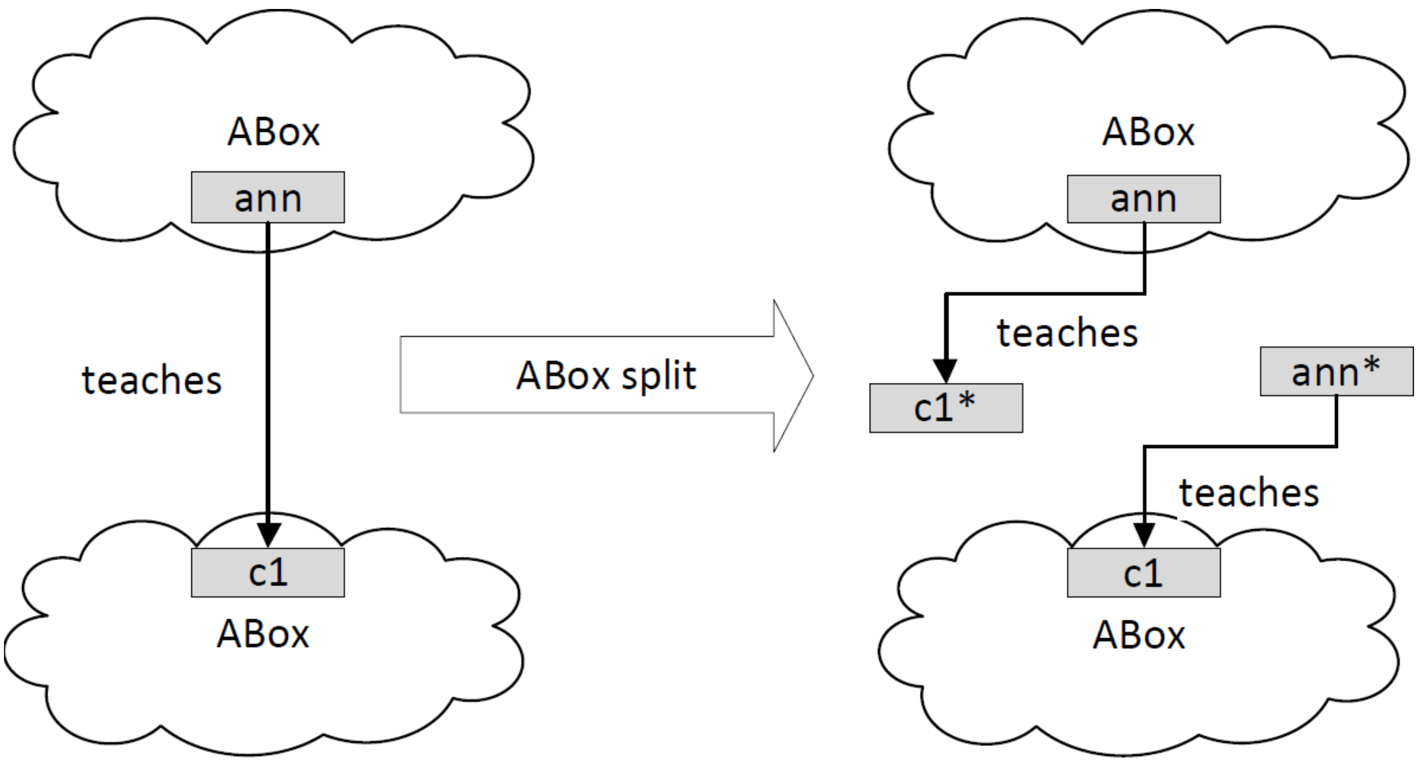
$$
\begin{aligned}
\mathcal{T} = \{&Chair \equiv \exists headOf.Department, Student \equiv \exists takes.Course,\\
&UndergraduateCourse \sqsubseteq Course, Course \sqcap Chair \sqsubseteq \bot, \top \sqsubseteq \forall takes.Course,\\
&\top \sqsubseteq \forall teaches.Course, \exists memberOf.\top \sqsubseteq Professor\}\\
\mathcal{R} = \{&headOf \sqsubseteq memberOf, teaches \equiv isTaughtBy^{-}, Trans(suborgOf)\}\\
\mathcal{A} = \{&Department(cs), Department(ee),\\
&Professor(ann), Professor(eve), Professor(mae),\\
&UndergraduateCourse(c1), UndergraduateCourse(c4),\\
&UndergraduateCourse(c5),\\
&GraduateCourse(c2), GraduateCourse(c3), Student(ani), Student(ean),\\
&Student(eva), Student(noa), Student(sam), Student(sue), Student(zoe),\\
&headOf(ann, cs), memberOf(eve, cs), headOf(mae, ee),\\
&teaches(ann, c1), teaches(eve, c2), teaches(eve, c3),\\
&teaches(mae, c4), teaches(mae, c5),\\
&takes(ani, c1), takes(ean, c1), takes(ean, c2), takes(eva, c3),\\
&takes(noa, c3), takes(sam, c4), takes(sue, c5), takes(zoe, c5)\}
\end{aligned}
$$

ABox Modularization



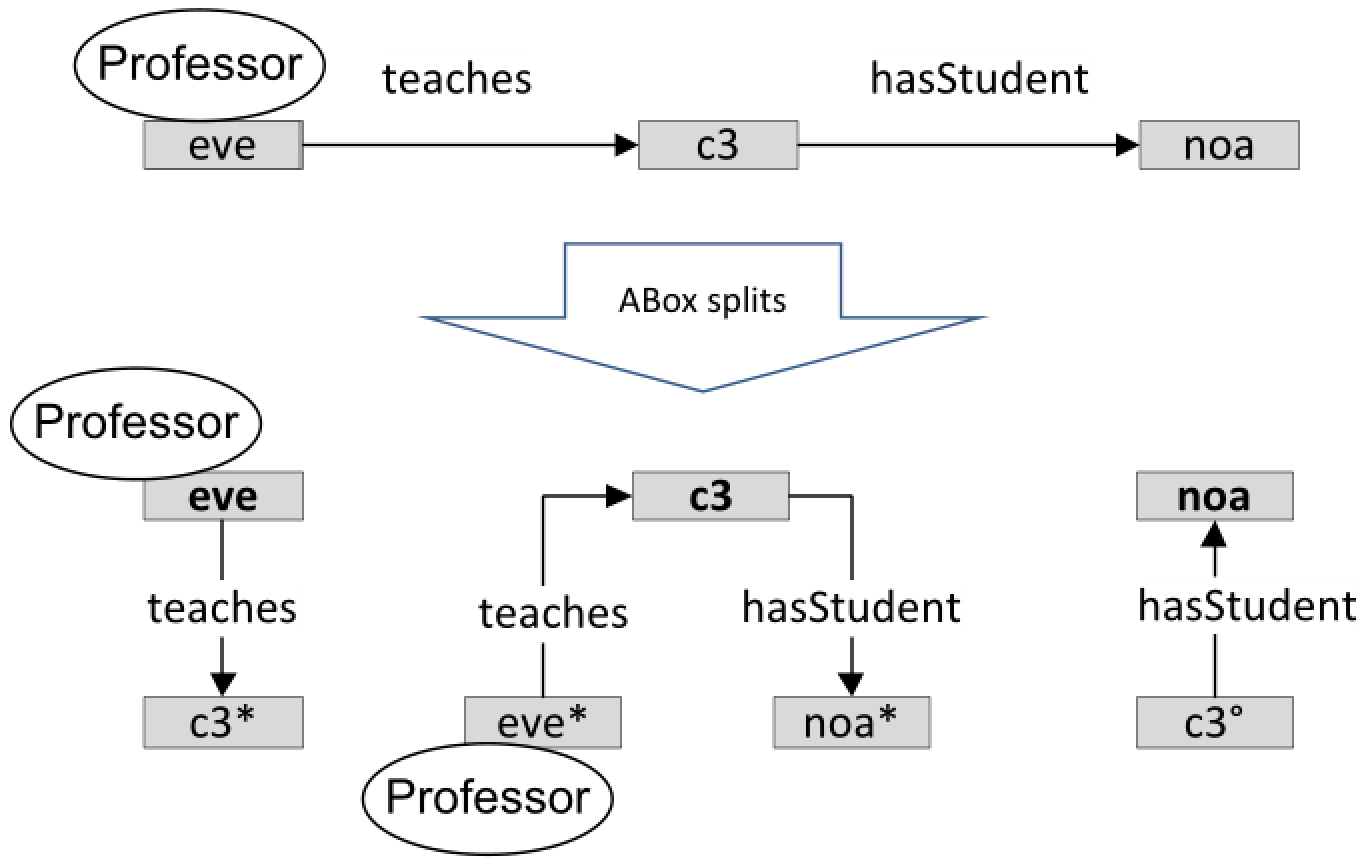
ABox Split

- Break up a role assertion:



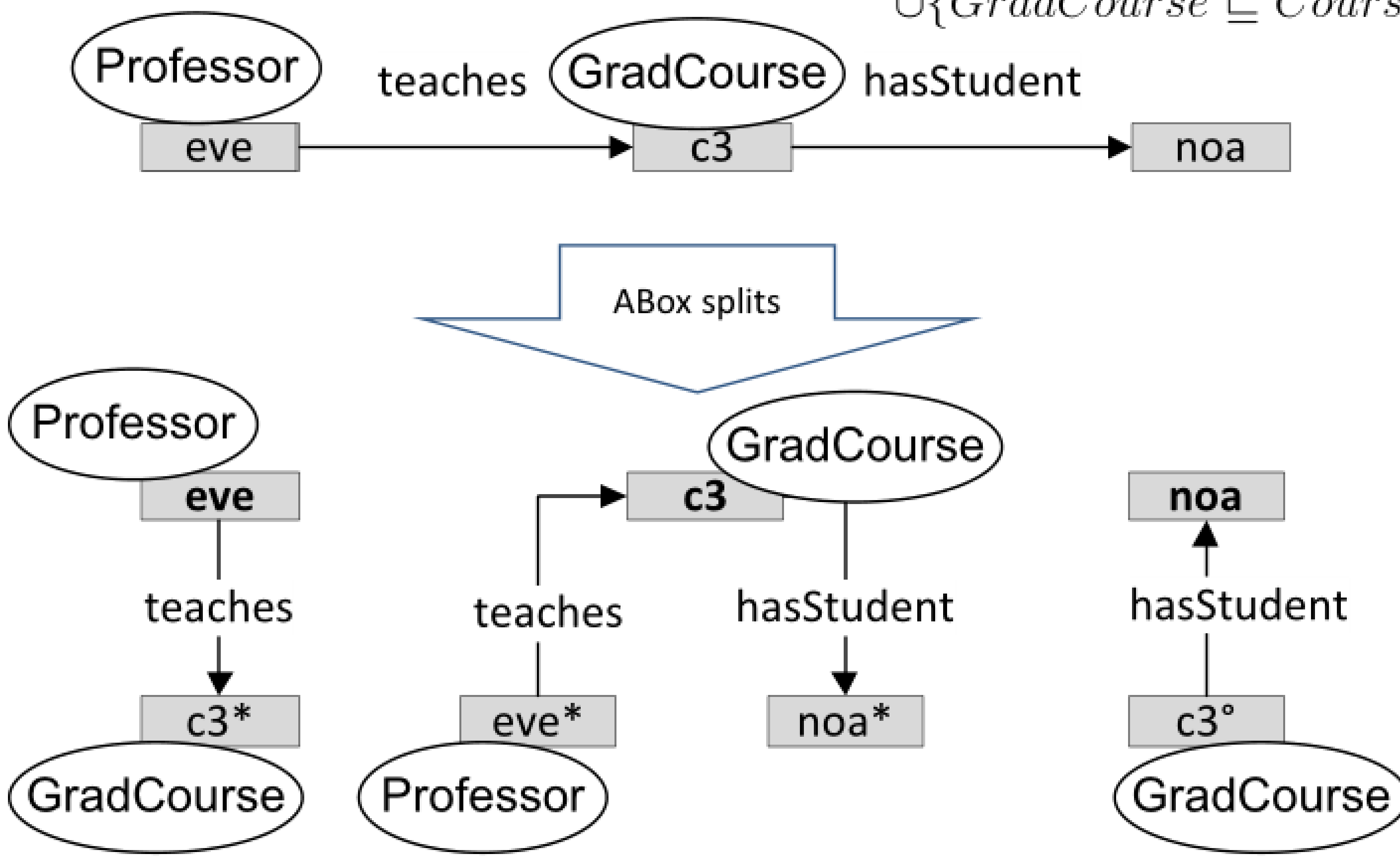
ABox Split – Active Students?

$$\{Professor \sqsubseteq \forall teaches.Course, Course \sqsubseteq \forall hasStudent.ActiveStudent\}$$



ABox Split – Active Students?

$$\{Professor \sqsubseteq \forall teaches.Course, Course \sqsubseteq \forall hasStudent.ActiveStudent\}$$
$$\cup \{GradCourse \sqsubseteq Course\}$$

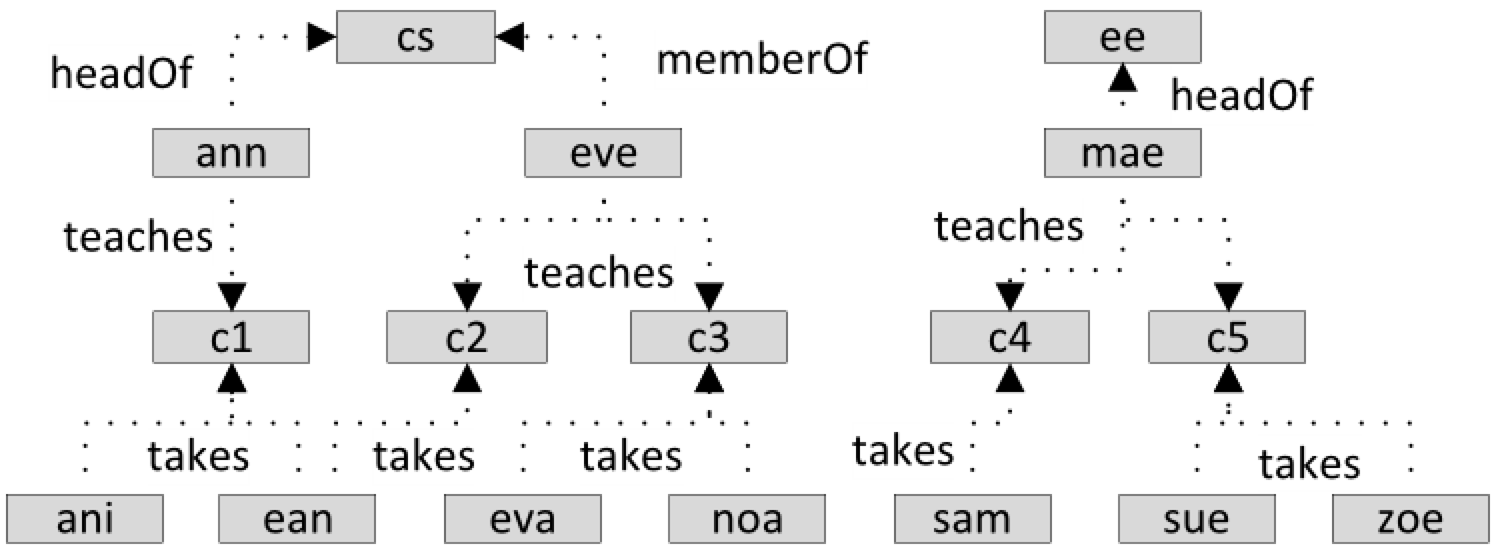


ABox Split - Criterion

Definition ($\mathcal{SHI}$-splittability of Role Assertions)**:**
Given a $\mathcal{SHI}$-ontology $\mathcal{O} = \langle \mathcal{T}, \mathcal{R}, \mathcal{A} \rangle$ and a role assertion $R(a, b)$, we say that $R(a, b)$ is *$\mathcal{SHI}$-splittable with respect to $\mathcal{O}$* if

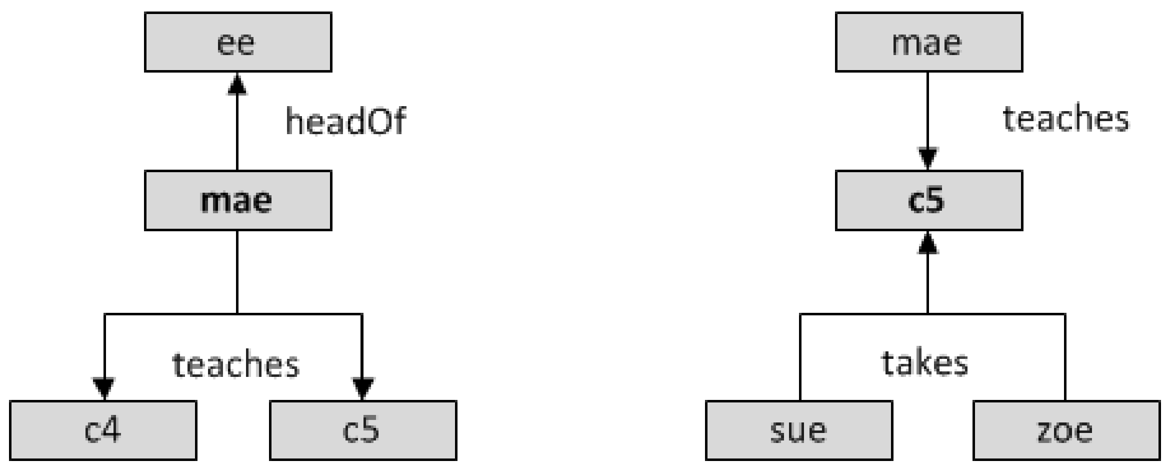
1. there exists no transitive role R_2 with respect to $\mathcal{O}$, such that $\mathcal{O} \models R \sqsubseteq R_2$,
2. for each $C \in \mathit{extinfo}^{\forall}_{\mathcal{T},\mathcal{R}}(R)$
   - $C = \bot$ or
   - there exists a concept description C_2, such that $C_2(b) \in \mathcal{A}$ and $\mathcal{T} \models C_2 \sqsubseteq C$ or
   - there exists a concept description C_2, such that $C_2(b) \in \mathcal{A}$ and $\mathcal{T} \models C \sqcap C_2 \sqsubseteq \bot$
3. for each $C \in \mathit{extinfo}^{\forall}_{\mathcal{T},\mathcal{R}}(R^{-})$
   - $C = \bot$ or
   - there exists a concept description C_2, such that $C_2(a) \in \mathcal{A}$ and $\mathcal{T} \models C_2 \sqsubseteq C$ or
   - there exists a concept description C_2, such that $C_2(a) \in \mathcal{A}$ and $\mathcal{T} \models C \sqcap C_2 \sqsubseteq \bot$.

Instance Checking: Individual Islands

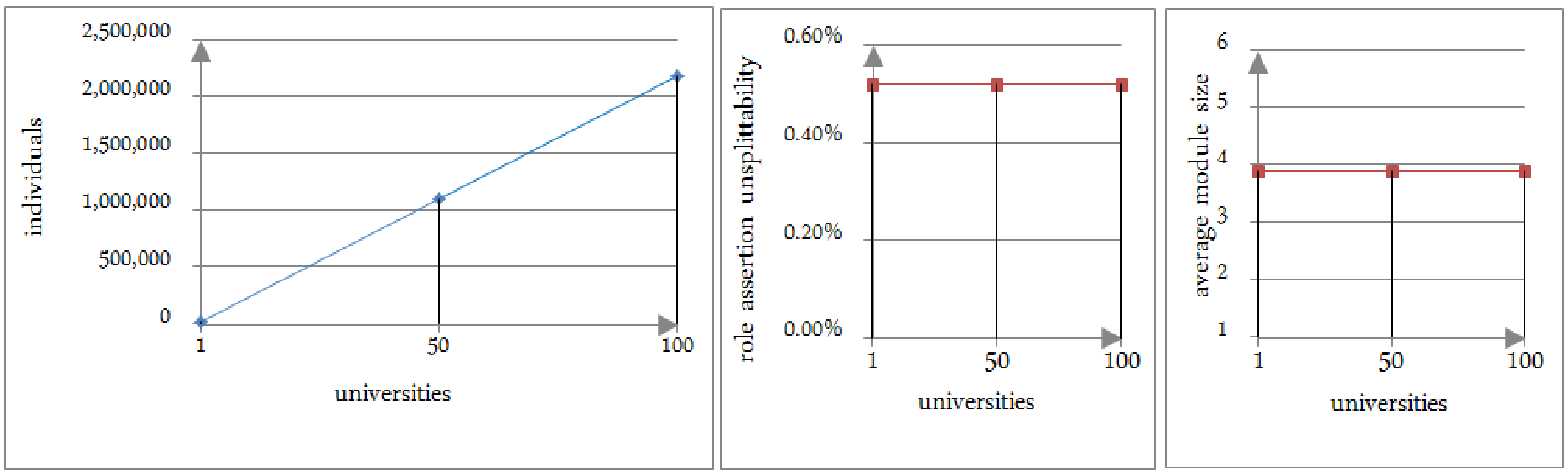


Instance Checking: Individual Islands

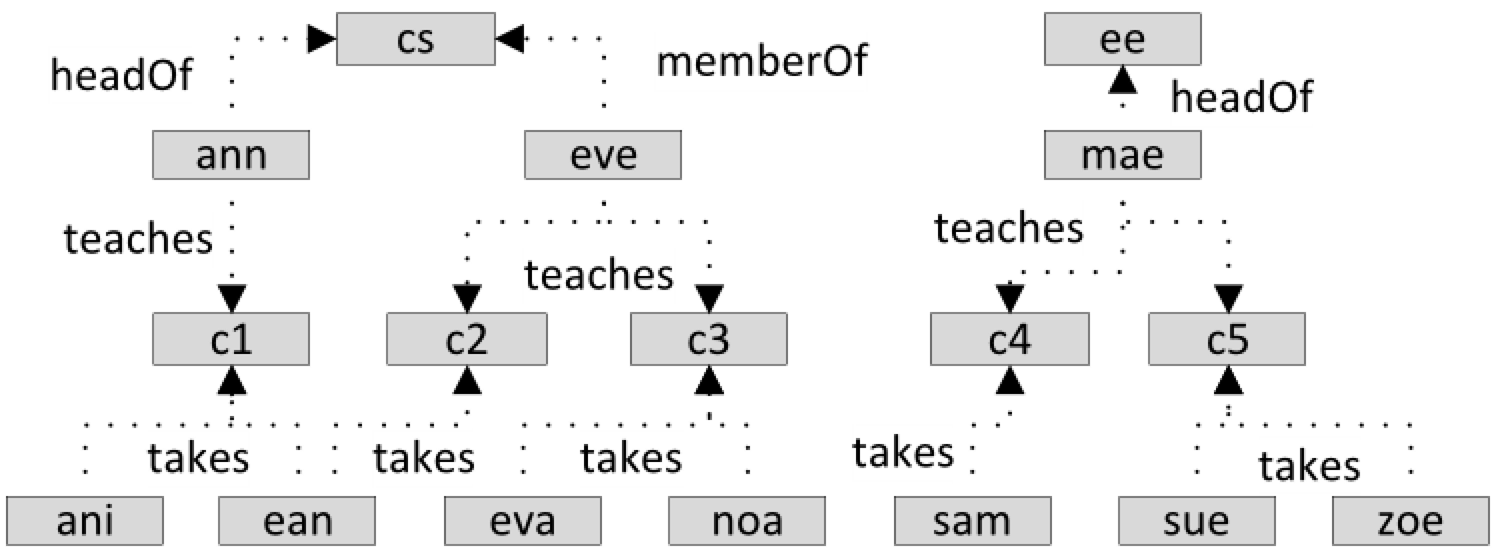
- Small modules fitting into main memory
- Note: we do not have to perform ABox splits in practice!



Instance Checking: Evaluation

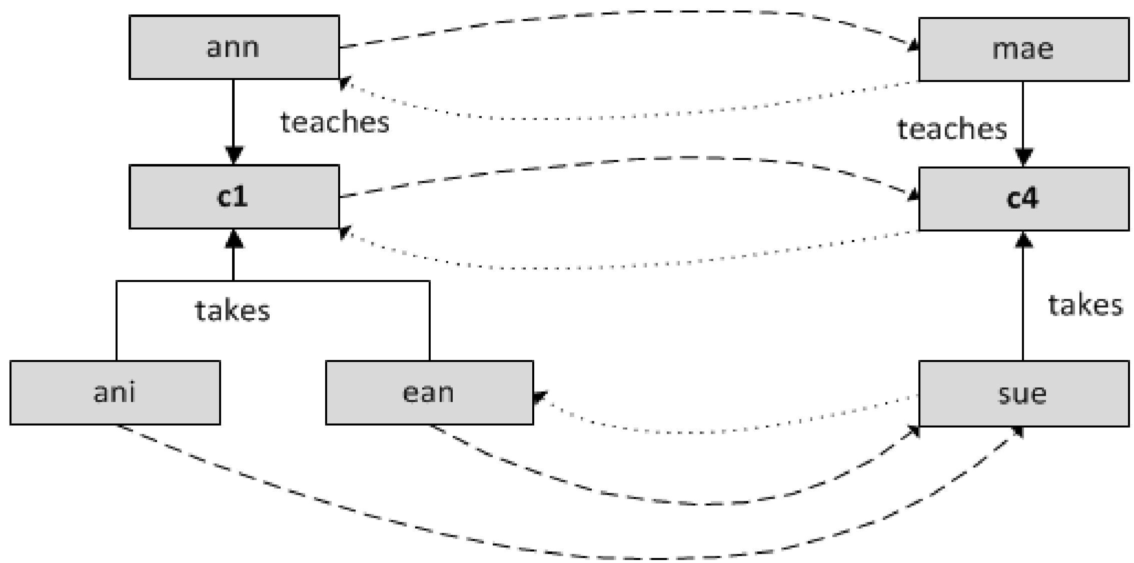


Instance Retrieval



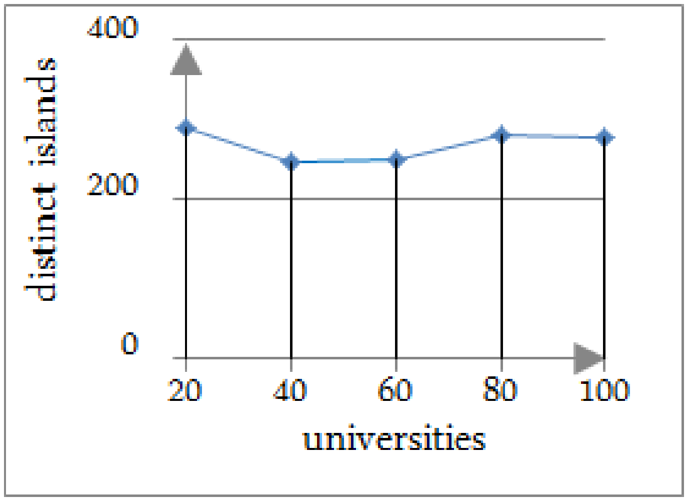
Instance Retrieval: Similarity

- Many (small) islands are similar to each other
- => use of homomorphisms

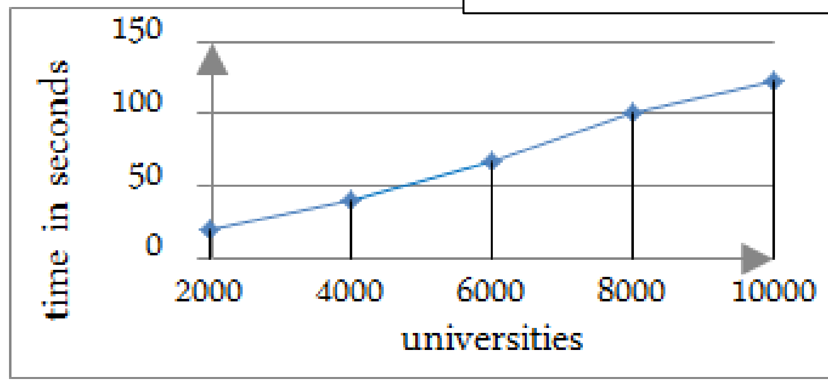


- Example: 9 instead of 17 instance checks

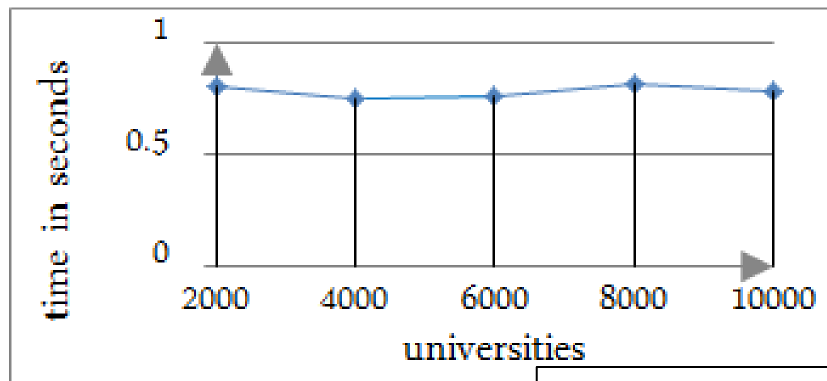
Instance Retrieval: Evaluation



IR: Chair? [SGH10]



LUBM(10000)=
1.4 billion ABox assertions

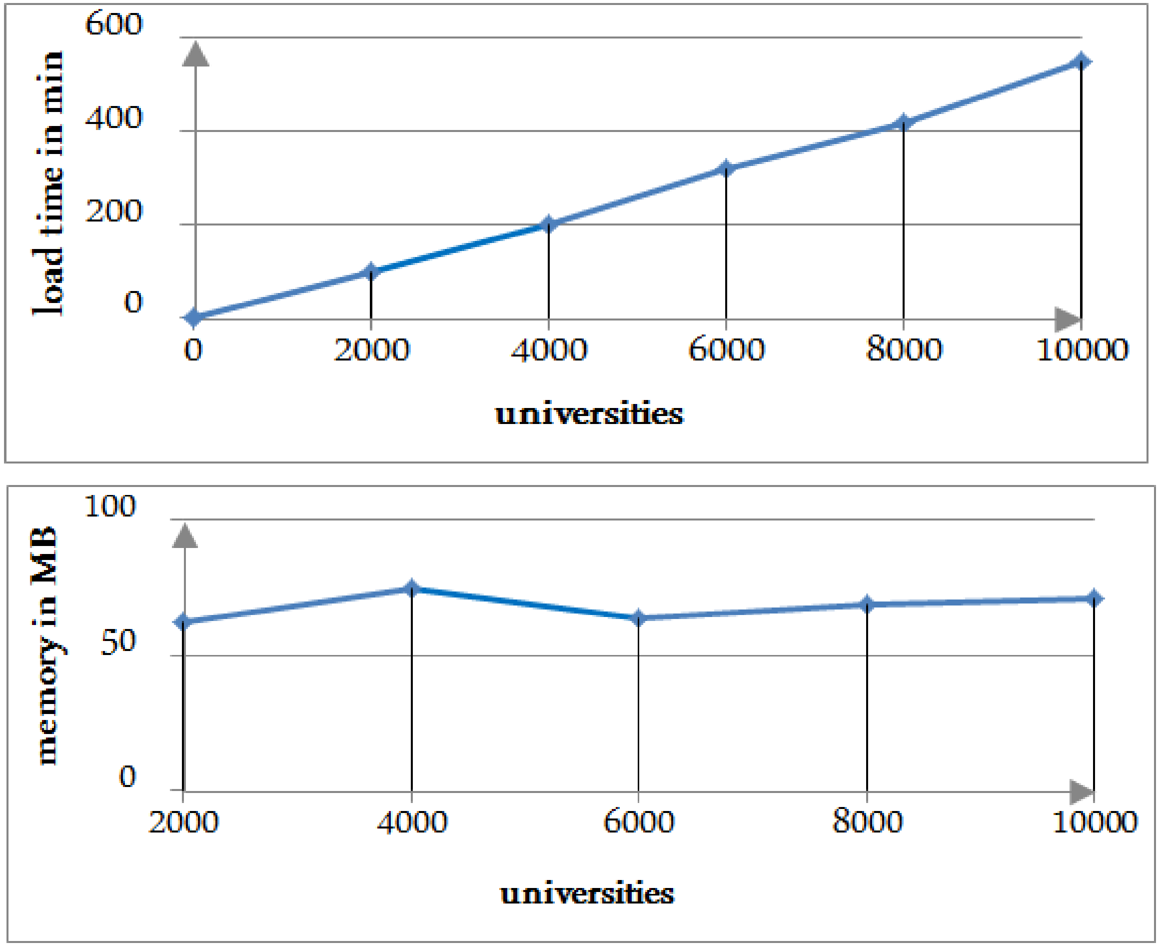


IR: University?

Ontology Changes

- Syntactic ontology updates
  - Keep complex data structures updated
- Non-atomic queries
  - As long as the query-concept does not contain existential constraints - and does not change the role hierarchy, nothing has to be recomputed (individual islands would only become more small)!
  - In the other case, new role assertions can become unsplittable!

Ontology Changes: Evaluation



Ontology Changes: TBox / RBox

- Hard to find *representative* update
  - From adding: $FullProfessor \sqsubseteq Professor$
  - ... over removing: $Student \equiv \exists takes.Course$
  - ... to (high impact) RBox-updates

Analysis

- Pro:
  - Very good for ontologies with many (mainly) integrity constraints
  - ABox updates are *local* and usually *fast*
- Con:
  - *Computational* ontologies
  - Complex updates of the terminology can be slow

Conclusions / Scientific Contributions

- ABox modularization techniques

  Sebastian Wandelt, Ralf Möller: Island Reasoning for ALCHI Ontologies, FOIS 2008

- Optimized instance retrieval

  Sebastian Wandelt et.al.: Towards Scalable Instance Retrieval over Ontologies, J. of Software and Informatics 2010

- Parallelization of instance retrieval

  Sebastian Wandelt, Ralf Möller: Distributed Island-Based Query Answering for Expressive Ontologies, DL 2010

- Updating data structures under changes to the ontology

  Sebastian Wandelt, Ralf Möller: Updatable Island Reasoning for ALCHI-ontologies, KEOD 2009

- Instance retrieval can be solved for LUBM(10000)

  Sebastian Wandelt, Ralf Möller: Sound and Complete Instance Retrieval for 1 Billion ABox Assertions, SSWS 2011

Future Work

- Optimization of retrieval process from the database
- More expressive query languages
- More expressive ontology languages
- Evaluation on more real world datasets

Not in competition with DL reasoner … results help them to deal with large ontologies more efficiently!

Questions / Discussion

References I

- **[ACKZ09]**: Alessandro Artale, Diego Calvanese, Roman Kontchakov, and Michael Zakharyaschev, The DL-Lite Family and Relations, J. of Artificial Intelligence Research, 2009
- **[AG10]**: AllegroGraph RDFStore Web 3.0's Database , http://www.franz.com/agraph/allegrograph/
- **[DFK09**]: Julian Dolby, Achille Fokoue, Aditya Kalyanpur, Edith Schonberg, Kavitha Srinivas: Scalable highly expressive reasoner (SHER). J. Web Sem. 7(4): 357-361 (2009)
- **[HKRT08]**: Pascal Hitzler, Markus Krötzsch, Sebastian Rudolph, Tuvshintur Tserendorj. Approximate OWL Instance Retreival with Screech. Dagstuhl-Seminar, Logic and Probability for Scene Interpretation, 2008
- **[HMS07]**: Ullrich Hustadt, Boris Motik, and Ulrike Sattler. Reasoning in Description Logics by a Reduction to Disjunctive Datalog. Journal of Automated Reasoning, 39(3):351–384, 2007.
- **[HM08]**: V. Haarslev and R. Möller. On the Scalability of Description Logic Instance Retrieval. Journal of Automated Reasoning, 41(2):99–142, 2008.
- [**MSH09]**: Boris Motik, Rob Shearer, and Ian Horrocks. Hypertableau Reasoning for Description Logics. Journal of Artificial Intelligence Research, 36:165–228, 2009

References II

- **[RPZ10]**: Yuan Ren, Jeff Z. Pan and Yuting Zhao. Towards Soundness Preserving Approximation for ABox Reasoning of OWL2. The International Description Logic Workshop (DL2010). 2010
- **[SBPKK07]**: E. Sirin et.al.. Pellet: A practical OWL-DL reasoner. Web Semantics 2007
- **[SGH10]**: Giorgos Stoilos, Bernardo Cuenca Grau, and Ian Horrocks. How Incomplete is your Semantic Web Reasoner? In Proc. of the 20th Nat. Conf. on Artificial Intelligence (AAAI 10), pages 1431-1436. AAAI Publications, 2010
- **[SHT05]**: Sean Bechhofer, Ian Horrocks, and Daniele Turi. The OWL Instance Store: System Description. In Proc. of the 20th Int. Conf. on Automated Deduction (CADE-20), Lecture Notes in Artificial Intelligence, pages 177-181. Springer, 2005
- **[TGH10]**: Tuvshintur Tserendorj, Stephan Grimm, Pascal Hitzler. Approximate Instance Retrieval on Ontologies. In: P. Garcia Bringas, A. Hameurlain, G. Quirchmayr, Database and Expert Systems Applications, 21st International Conference, DEXA 2010, Bilbao, Spain, August 30 - September 3, 2010, Proceedings, Part I. Springer Lecture Notes in Computer Science Vol. 6261, 2010