

Abstraction Refinement for Ontology Materialization

Birte Glimm¹ Yevgeny Kazakov¹ Thorsten Liebig²
Trung-Kien Tran¹ Vincent Vialard²

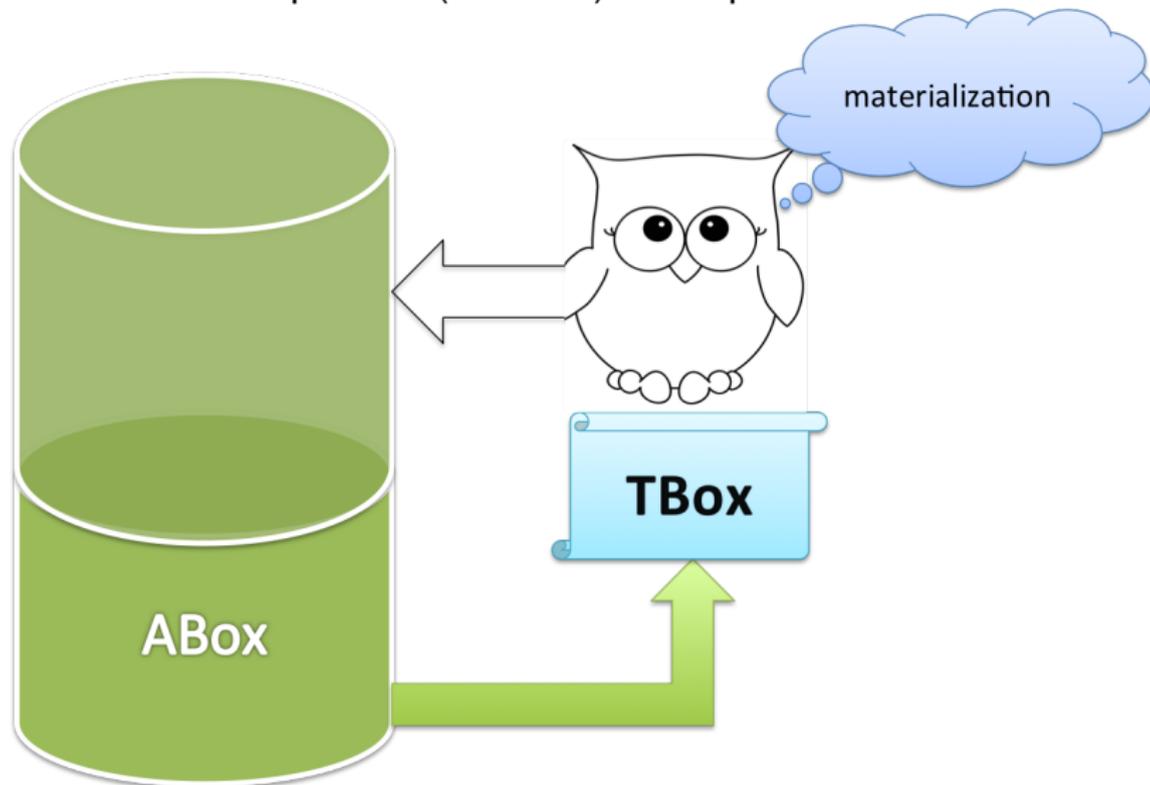
¹University of Ulm, Germany,

²derivo GmbH, Germany

October 2014

Motivation

Compute all (entailed) concept assertions



Motivation

Many reasoning steps are repeated



Motivation

Many reasoning steps are repeated



\exists likes.Fish \sqsubseteq FishLover

Motivation

Many reasoning steps are repeated



Representative cat



\exists likes.Fish \sqsubseteq FishLover

Motivation

Many reasoning steps are repeated



Representative cat



$\exists \textit{likes.Fish} \sqsubseteq \textit{FishLover}$

- Partition individuals into equivalence classes
- Compute concept assertions for just on representative

Simple Merging Does Not Work

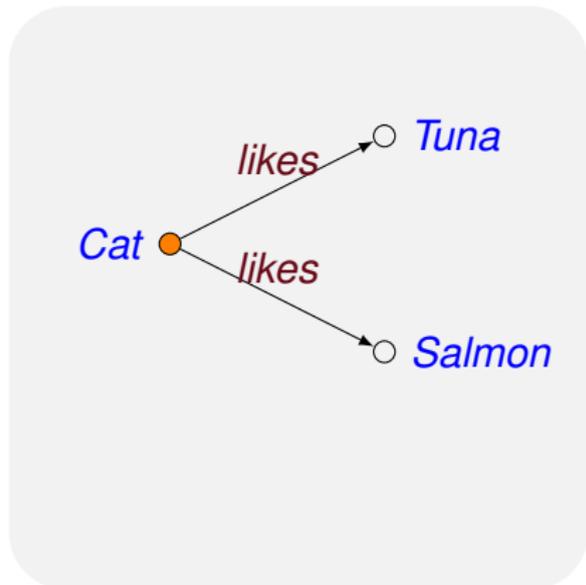
Cat ● *likes* → ○ *Tuna*

Cat ● *likes* → ○ *Salmon*

Simple Merging Does Not Work

Cat ● *likes* → ○ Tuna

Cat ● *likes* → ○ Salmon

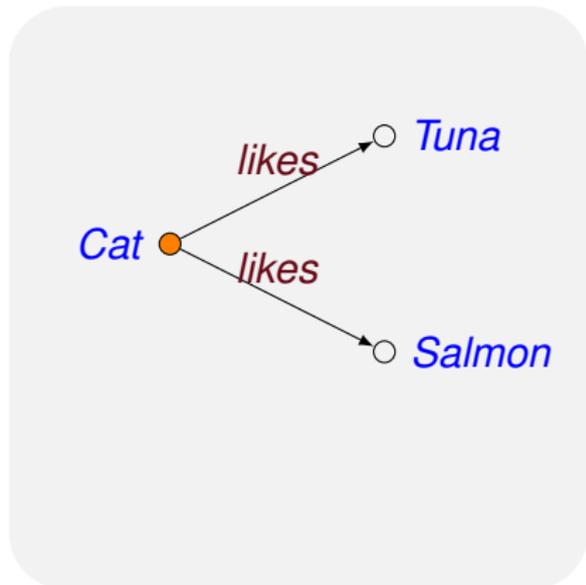


Simple Merging Does Not Work

Cat ● *likes* → ○ *Tuna*

Cat ● *likes* → ○ *Salmon*

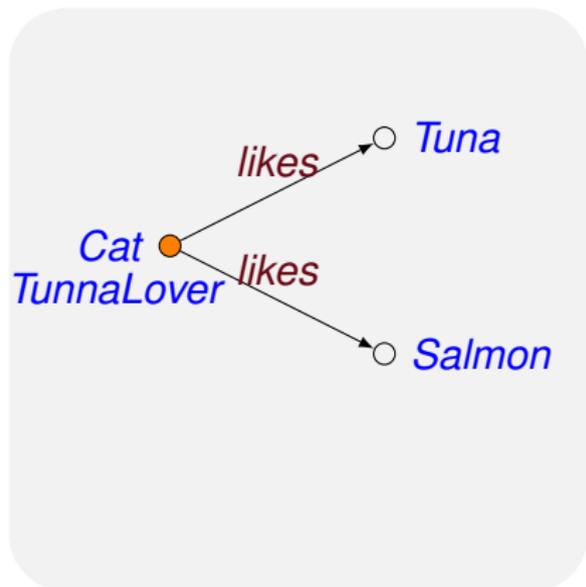
$\exists \textit{likes}. \textit{Tuna} \sqsubseteq \textit{TunaLover}$



Simple Merging Does Not Work



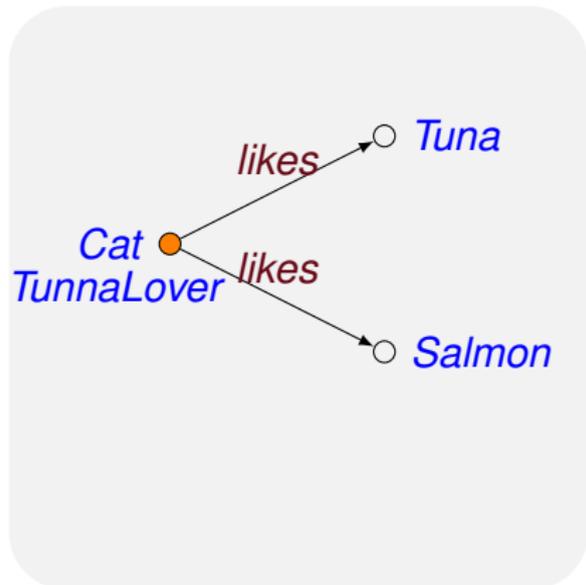
$\exists \textit{likes}. \textit{Tuna} \sqsubseteq \textit{TunaLover}$



Simple Merging Does Not Work

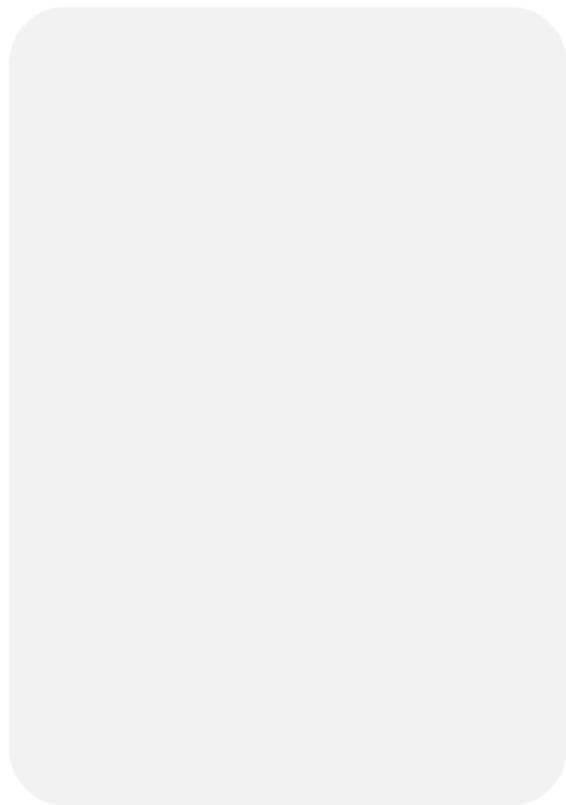
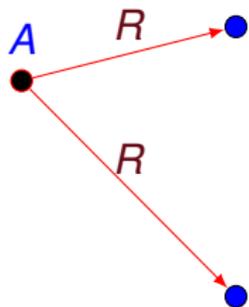


$\exists \textit{likes}. \textit{Tuna} \sqsubseteq \textit{TunnaLover}$



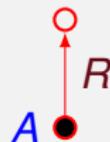
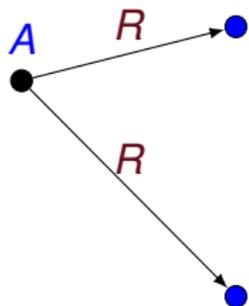
Abstraction

Partition individuals by:
concepts, pre/successor roles



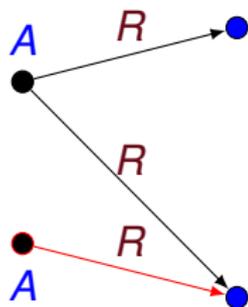
Abstraction

Partition individuals by:
concepts, pre/successor roles



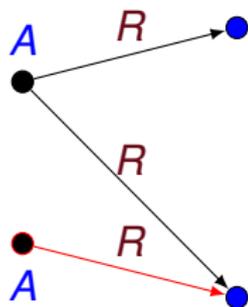
Abstraction

Partition individuals by:
concepts, pre/successor roles



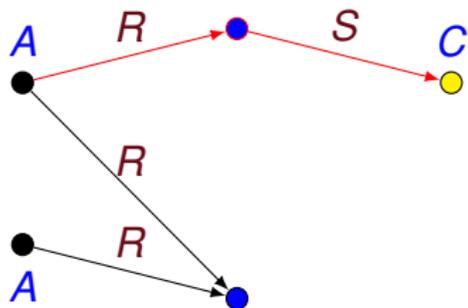
Abstraction

Partition individuals by:
concepts, pre/successor roles



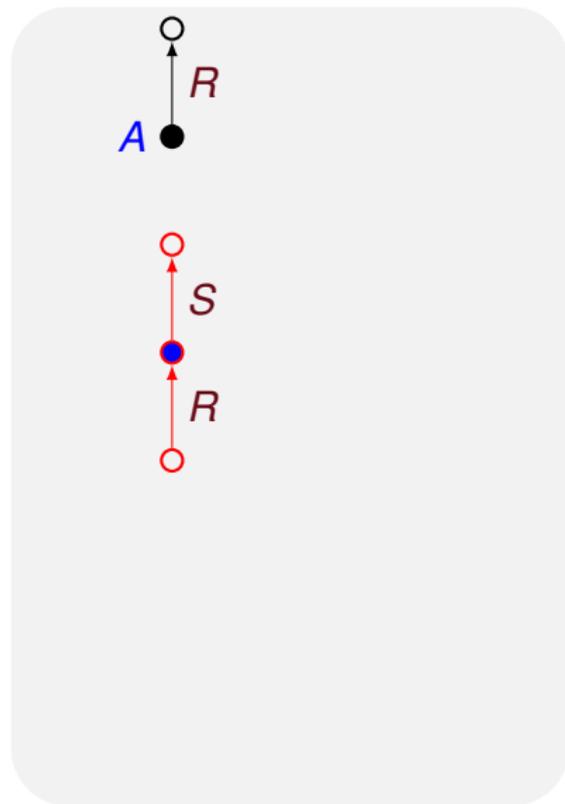
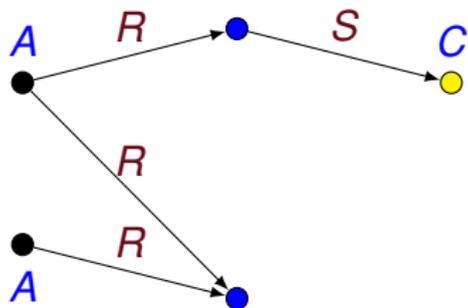
Abstraction

Partition individuals by:
concepts, pre/successor roles



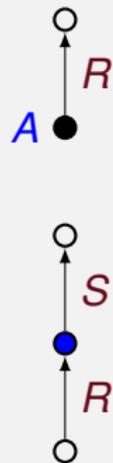
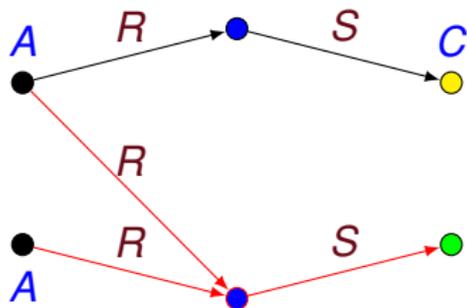
Abstraction

Partition individuals by:
concepts, pre/successor roles



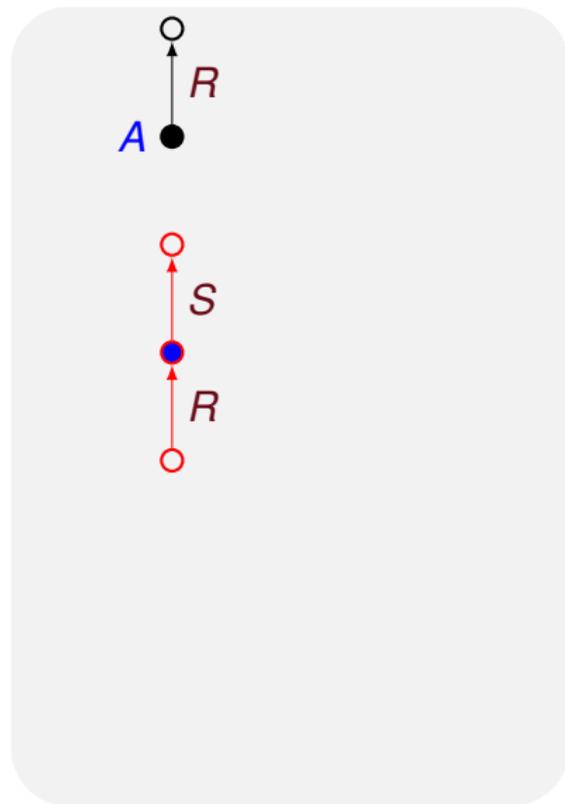
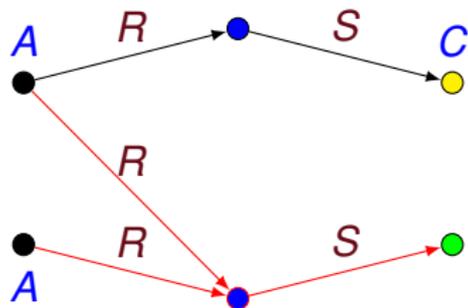
Abstraction

Partition individuals by:
concepts, pre/successor roles



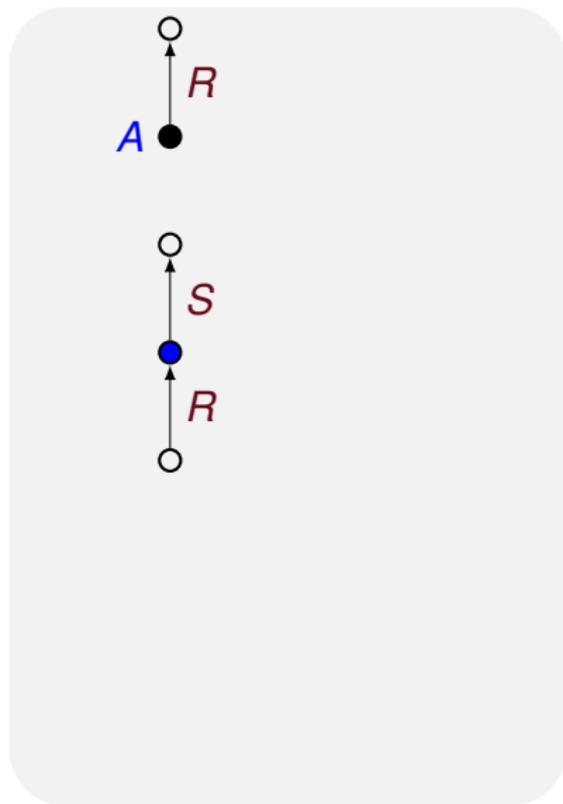
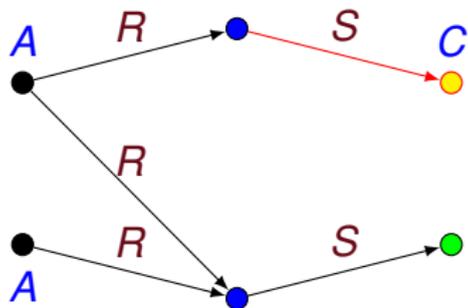
Abstraction

Partition individuals by:
concepts, pre/successor roles



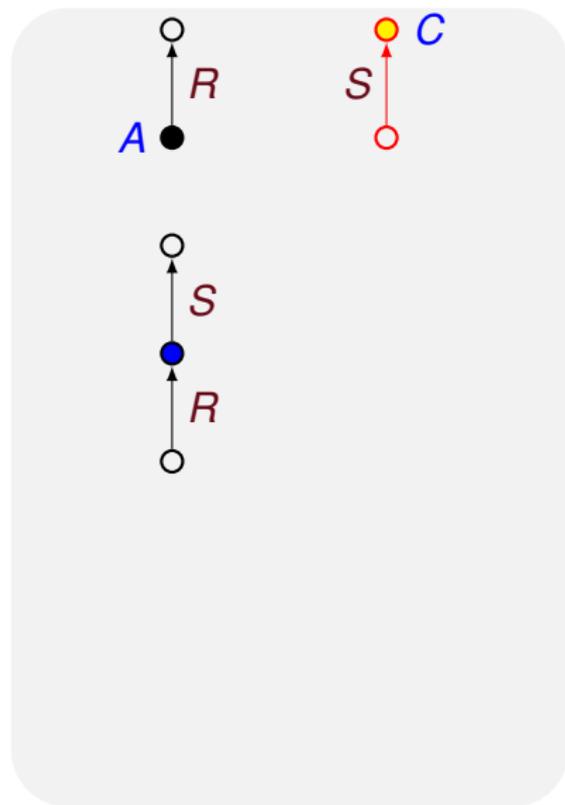
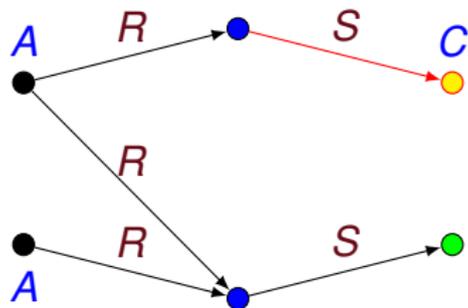
Abstraction

Partition individuals by:
concepts, pre/successor roles



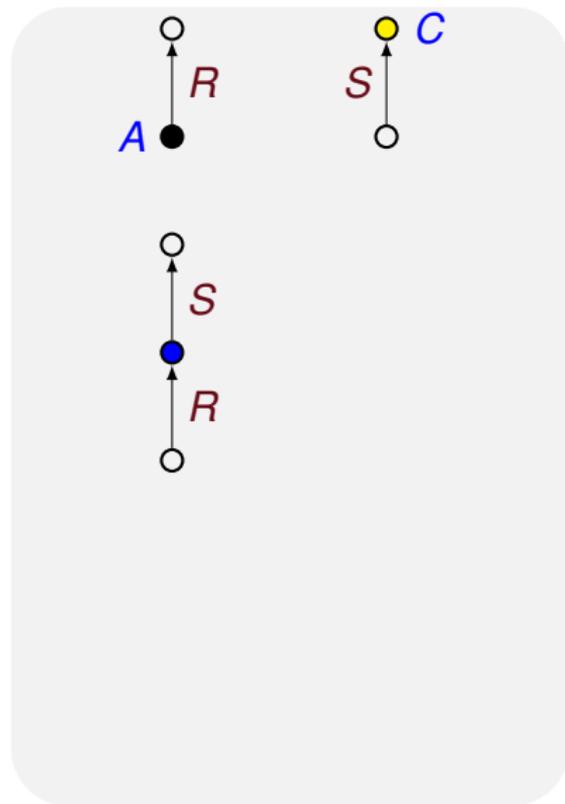
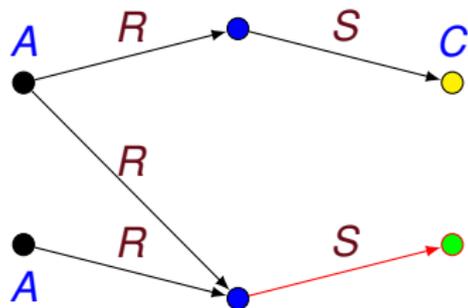
Abstraction

Partition individuals by:
concepts, pre/successor roles



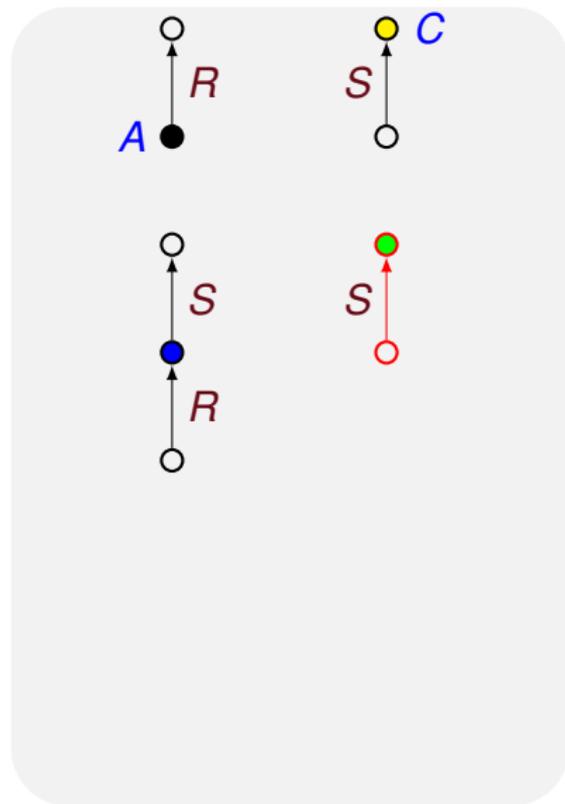
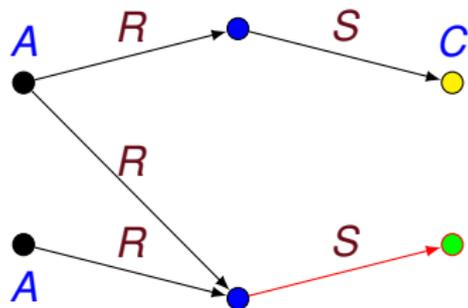
Abstraction

Partition individuals by:
concepts, pre/successor roles



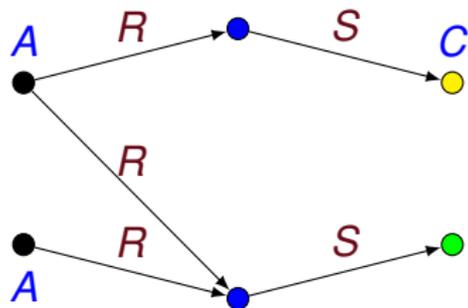
Abstraction

Partition individuals by:
concepts, pre/successor roles

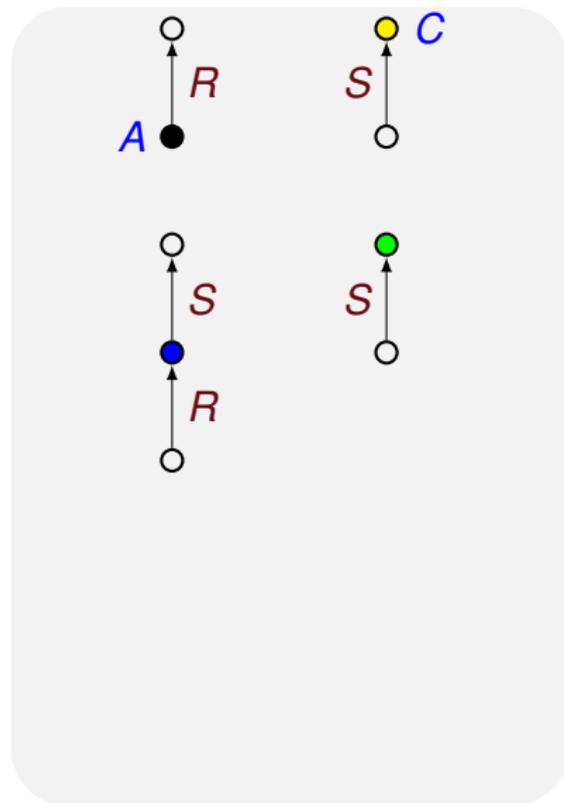


Abstraction

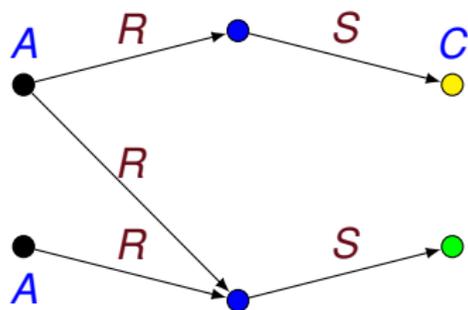
Partition individuals by:
concepts, pre/successor roles



Abstract ABox

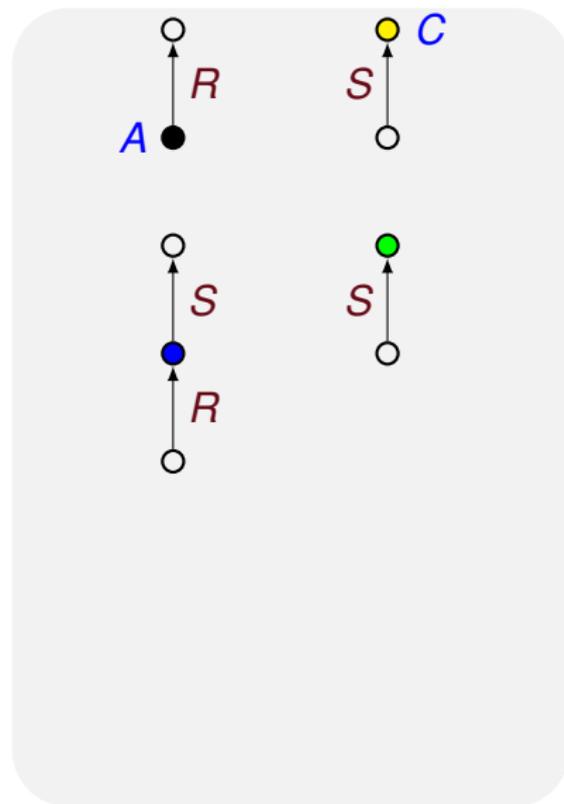


Materializing the Abstract ABox

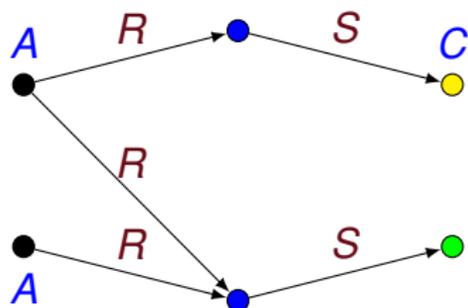


$$A \sqsubseteq \forall R.B$$

Abstract ABox

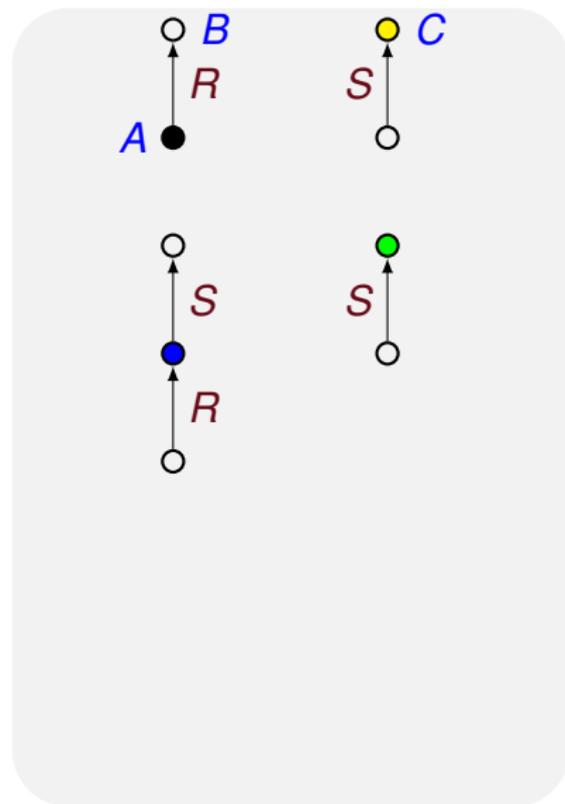


Materializing the Abstract ABox

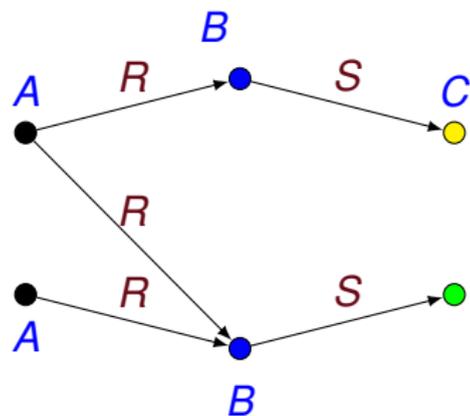


$$A \sqsubseteq \forall R.B$$

Abstract ABox

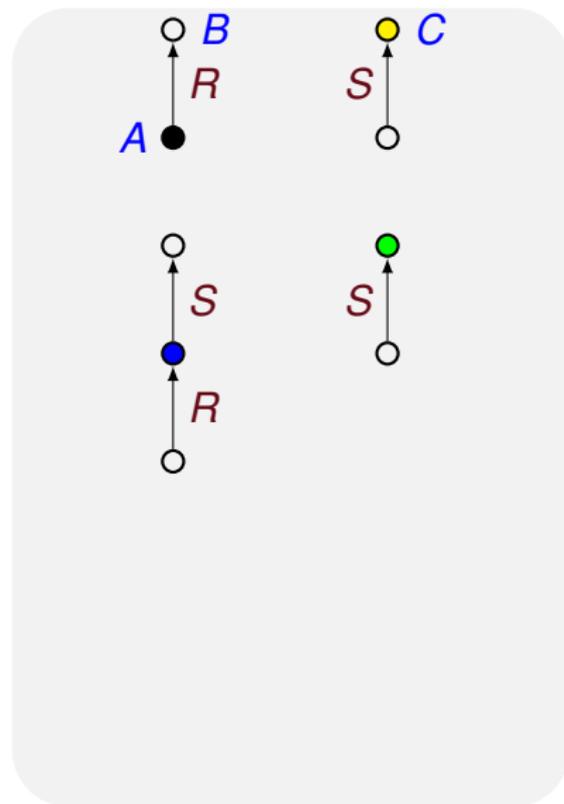


Materializing the Abstract ABox

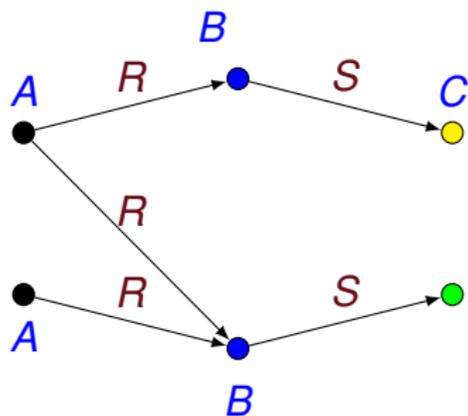


$$A \sqsubseteq \forall R.B$$

Abstract ABox

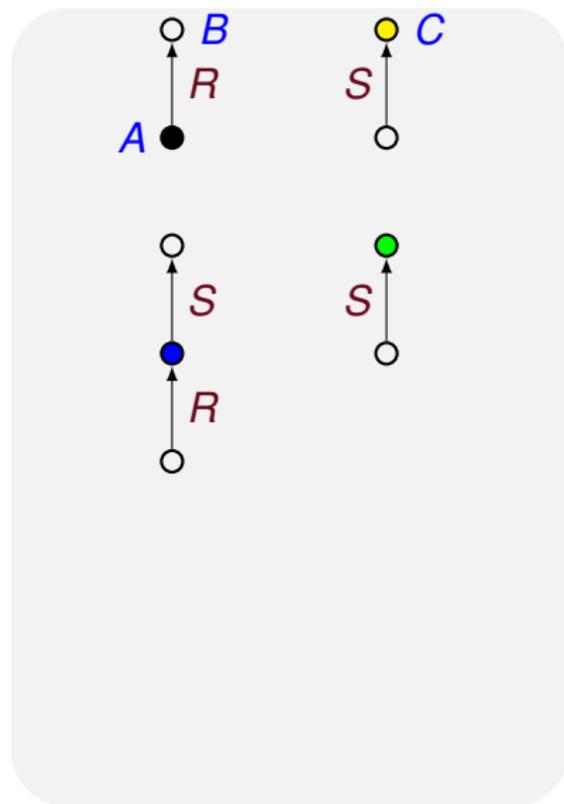


Materializing the Abstract ABox

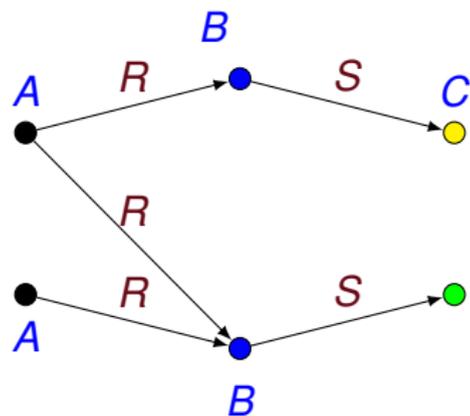


$$A \sqsubseteq \forall R.B$$
$$\exists S.C \sqsubseteq C$$

Abstract ABox

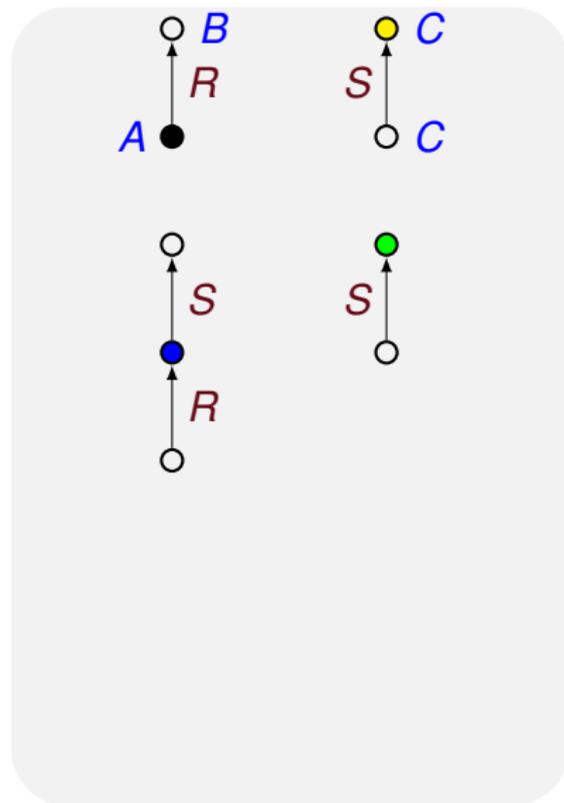


Materializing the Abstract ABox

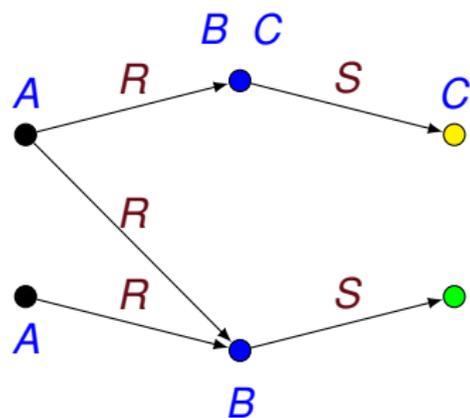


$$A \sqsubseteq \forall R.B$$
$$\exists S.C \sqsubseteq C$$

Abstract ABox

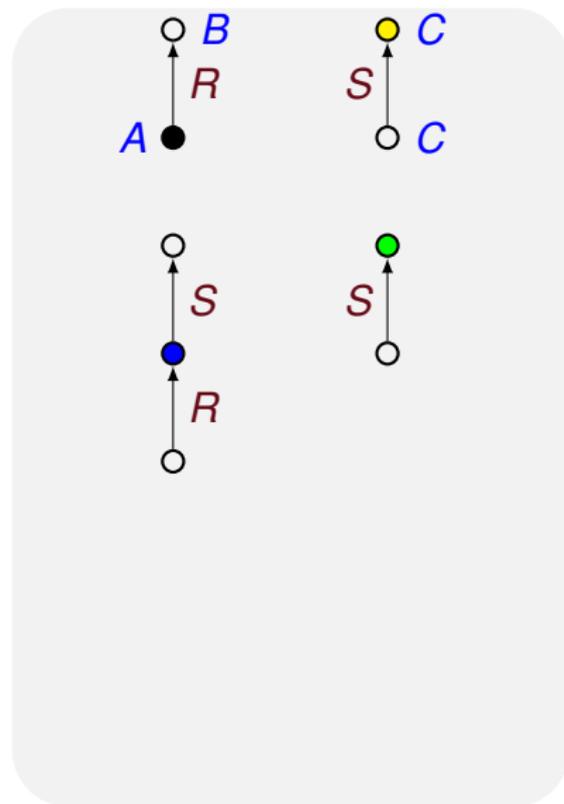


Materializing the Abstract ABox

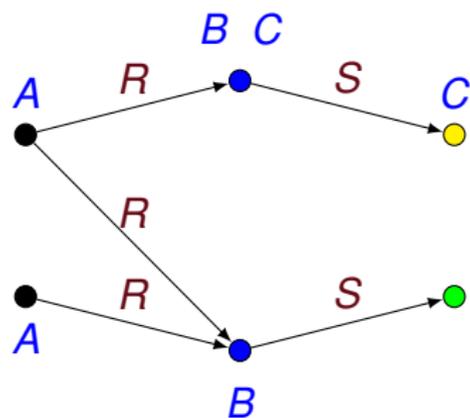


$$A \sqsubseteq \forall R.B$$
$$\exists S.C \sqsubseteq C$$

Abstract ABox

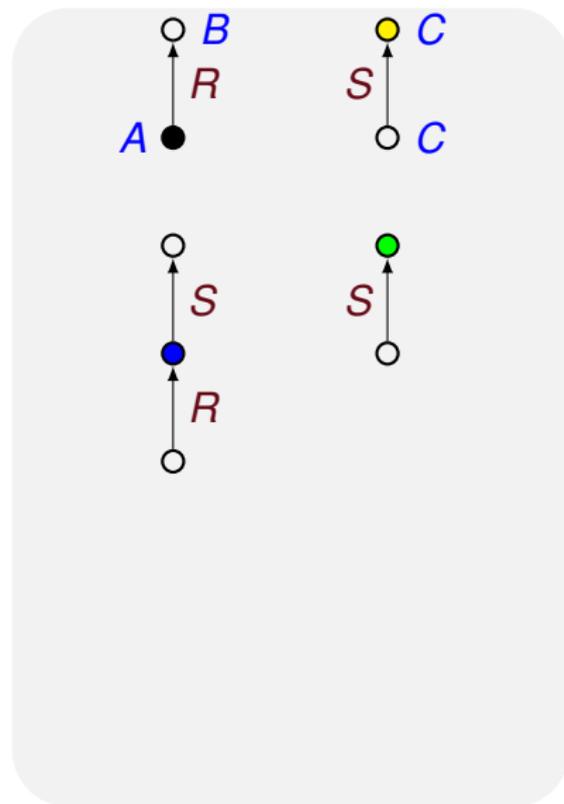


Materializing the Abstract ABox

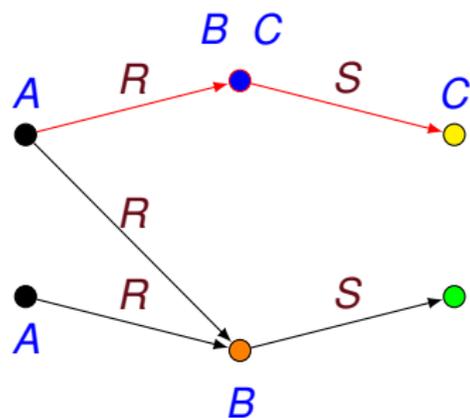


$$A \sqsubseteq \forall R.B$$
$$\exists S.C \sqsubseteq C$$
$$B \sqcap C \sqsubseteq D$$

Abstract ABox

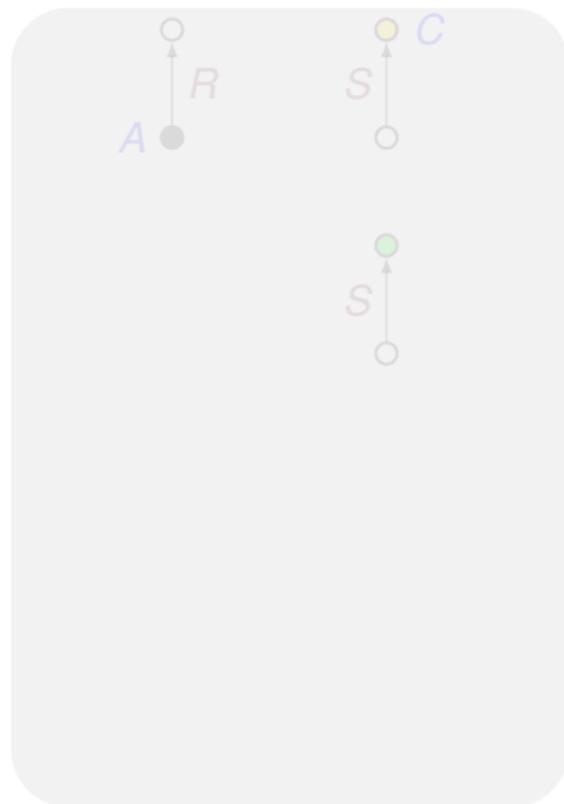


Refinement

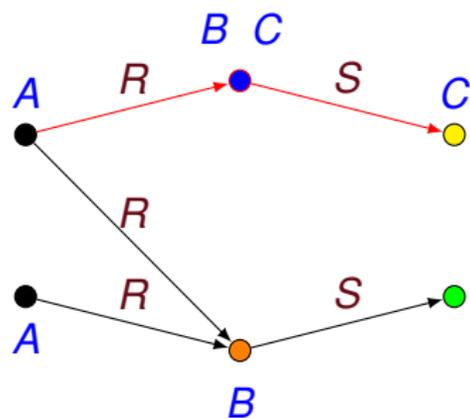


$$A \sqsubseteq \forall R.B$$
$$\exists S.C \sqsubseteq C$$
$$B \sqcap C \sqsubseteq D$$

Abstract ABox

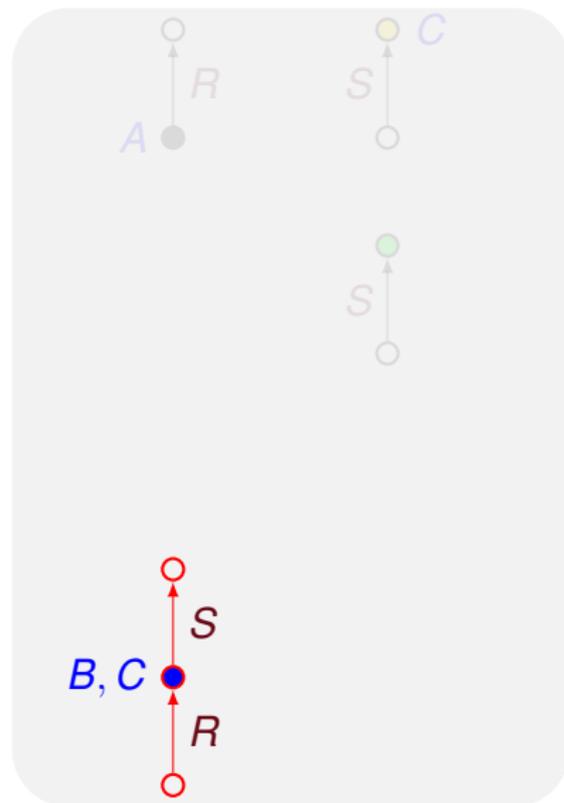


Refinement

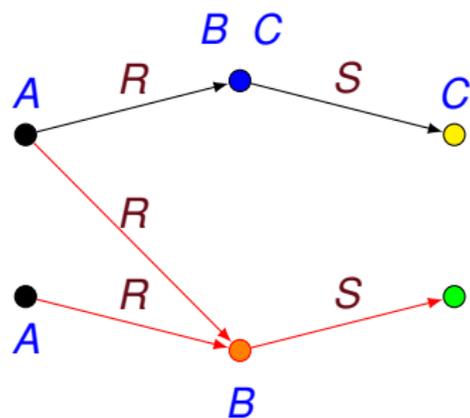


$$A \sqsubseteq \forall R.B$$
$$\exists S.C \sqsubseteq C$$
$$B \sqcap C \sqsubseteq D$$

Abstract ABox

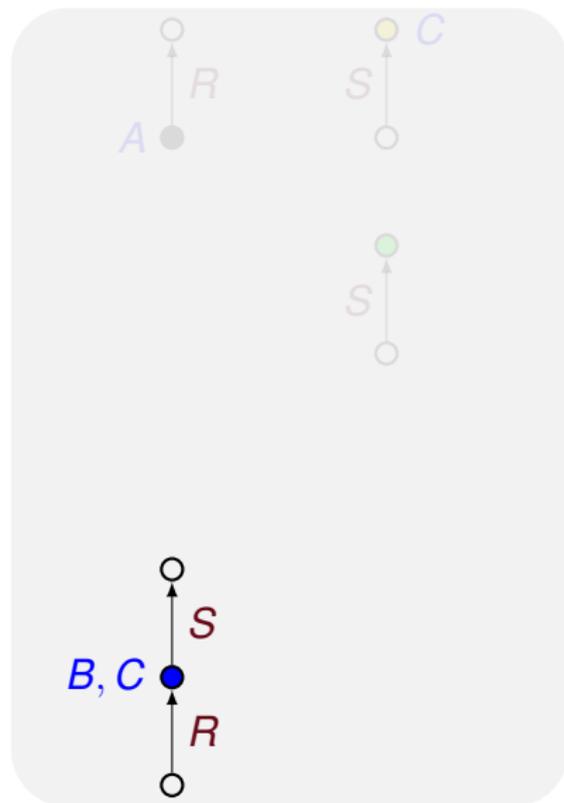


Refinement

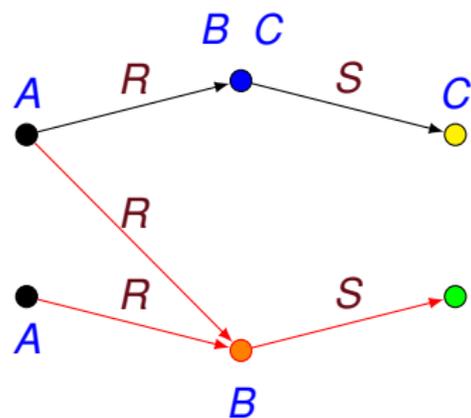


$$A \sqsubseteq \forall R.B$$
$$\exists S.C \sqsubseteq C$$
$$B \sqcap C \sqsubseteq D$$

Abstract ABox

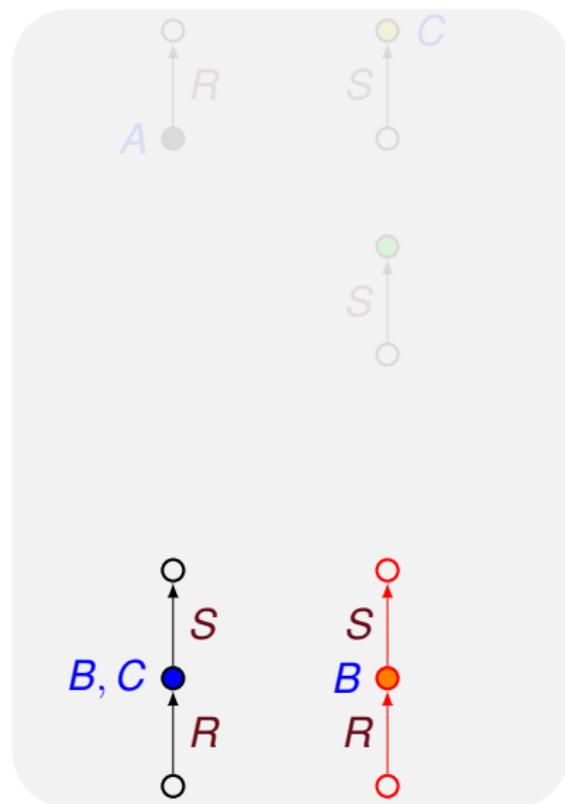


Refinement

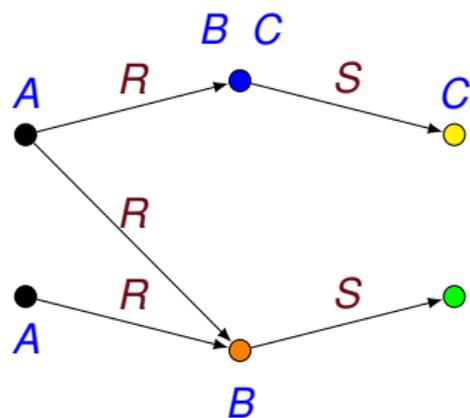


$$\begin{aligned} A &\sqsubseteq \forall R.B \\ \exists S.C &\sqsubseteq C \\ B \sqcap C &\sqsubseteq D \end{aligned}$$

Abstract ABox



Materializing the Abstract ABox

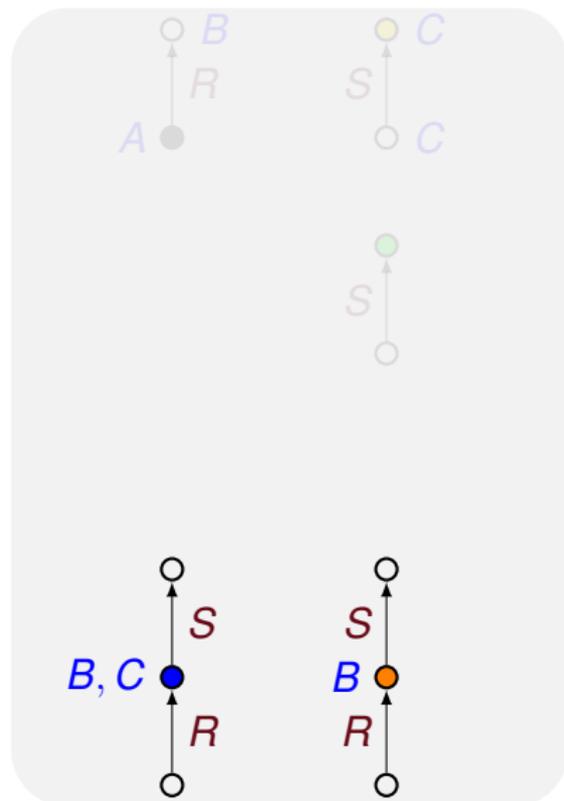


$$A \sqsubseteq \forall R.B$$

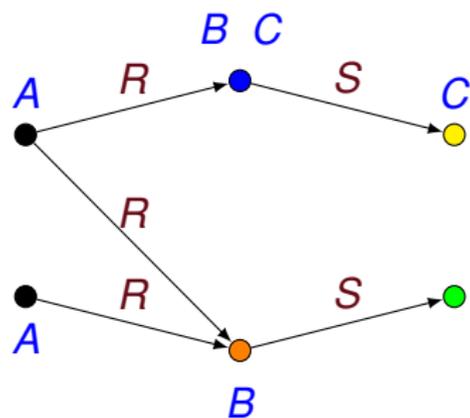
$$\exists S.C \sqsubseteq C$$

$$B \sqcap C \sqsubseteq D$$

Abstract ABox



Materializing the Abstract ABox

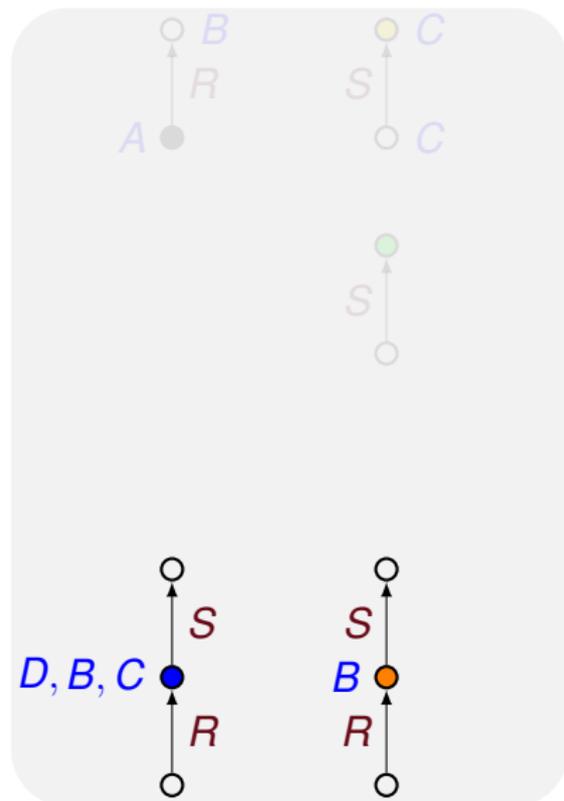


$$A \sqsubseteq \forall R.B$$

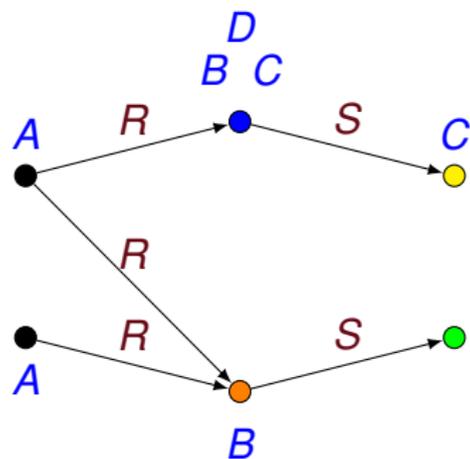
$$\exists S.C \sqsubseteq C$$

$$B \sqcap C \sqsubseteq D$$

Abstract ABox

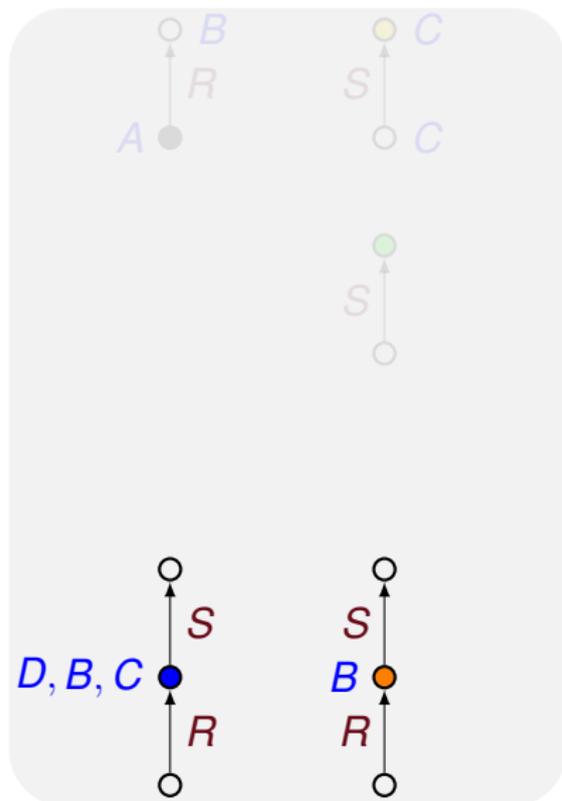


Materializing the Abstract ABox

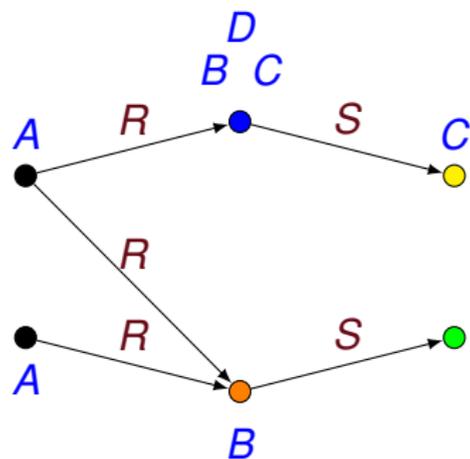


$$\begin{aligned} A &\sqsubseteq \forall R.B \\ \exists S.C &\sqsubseteq C \\ B \sqcap C &\sqsubseteq D \end{aligned}$$

Abstract ABox



Materialized ABox and Abstract ABox

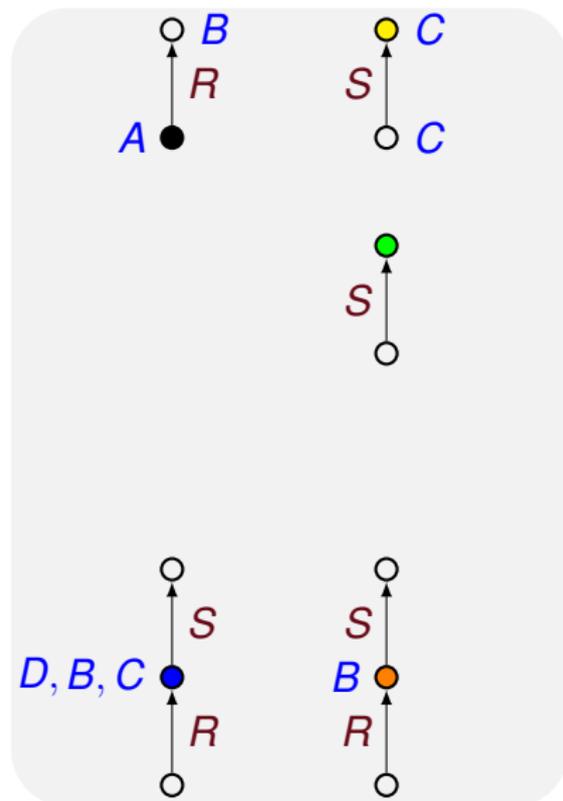


$$A \sqsubseteq \forall R.B$$

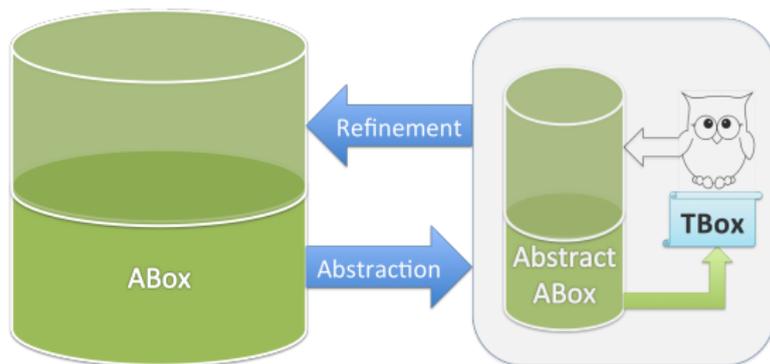
$$\exists S.C \sqsubseteq C$$

$$B \sqcap C \sqsubseteq D$$

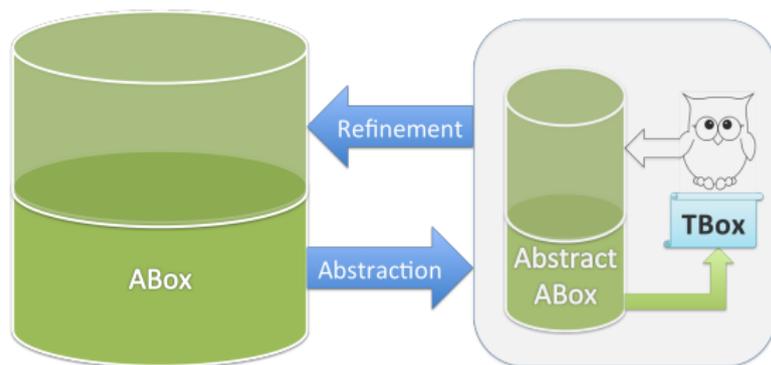
Abstract ABox



Abstraction Refinement

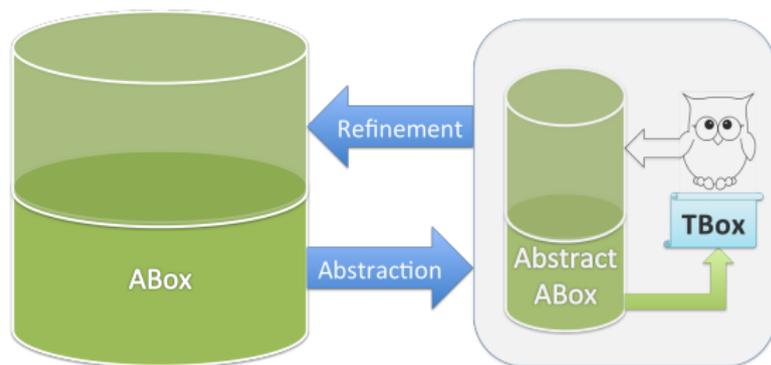


Abstraction Refinement



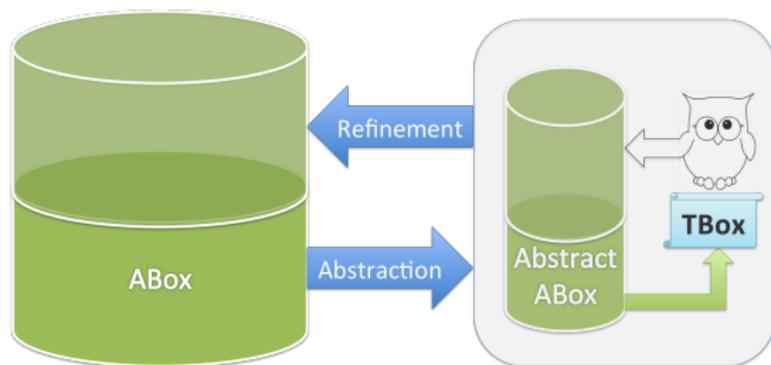
(1) Compute the abstract ABox

Abstraction Refinement



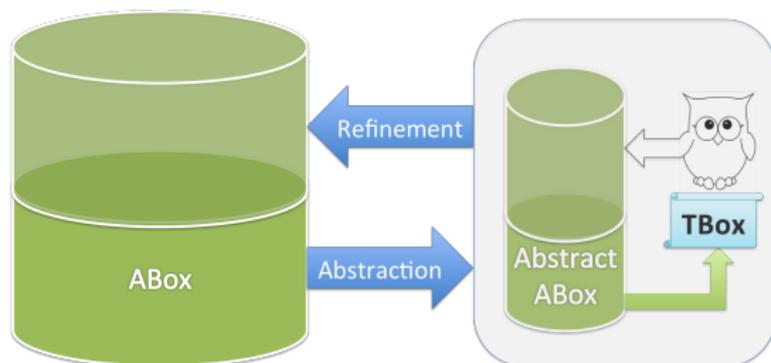
- (1) Compute the abstract ABox
- (2) Materialize the abstract ABox

Abstraction Refinement



- (1) Compute the abstract ABox
- (2) Materialize the abstract ABox
- (3) Add corresponding assertions

Abstraction Refinement

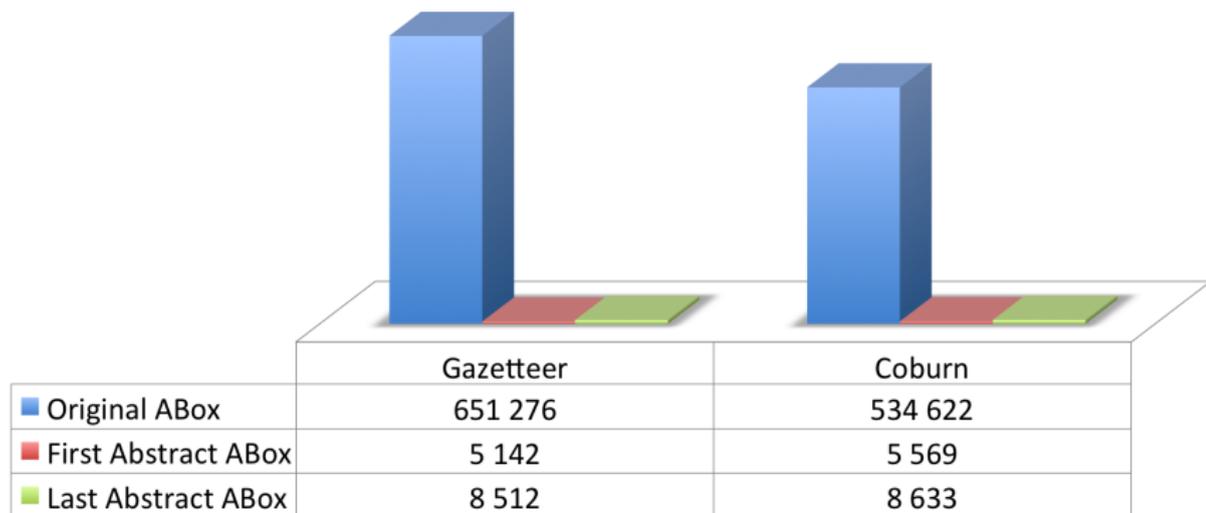


- (1) Compute the abstract ABox
- (2) Materialize the abstract ABox
- (3) Add corresponding assertions

Sound and complete for Horn *ALCHOI*

Evaluation

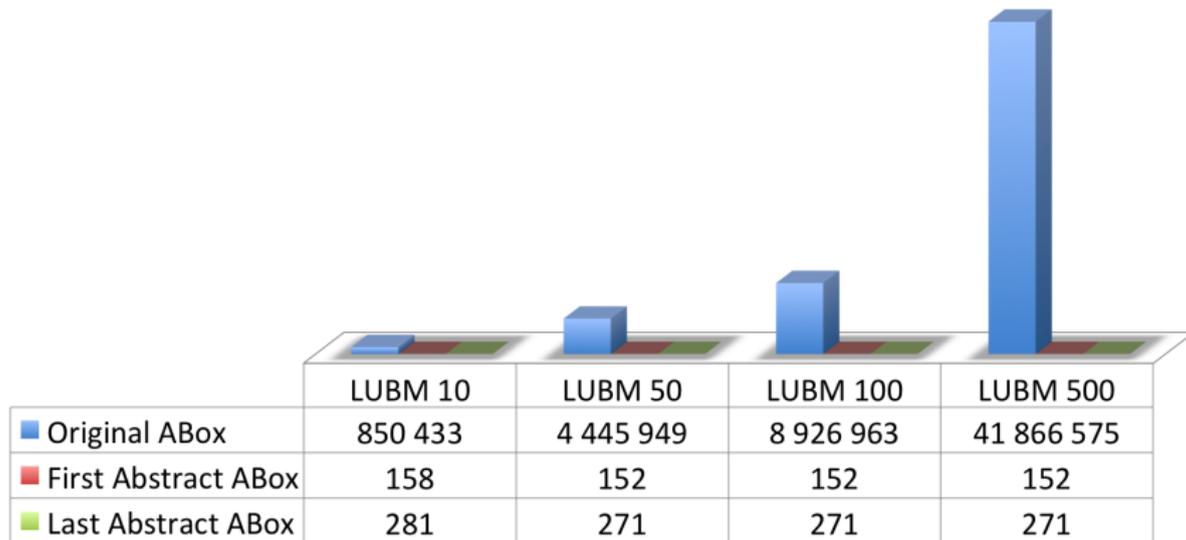
Abstract ABox Size (#Assertions) vs Original ABox Size (#Assertions)



- 1 refinement step for Gazetteer
- 2 refinement steps for Coburn

Evaluation

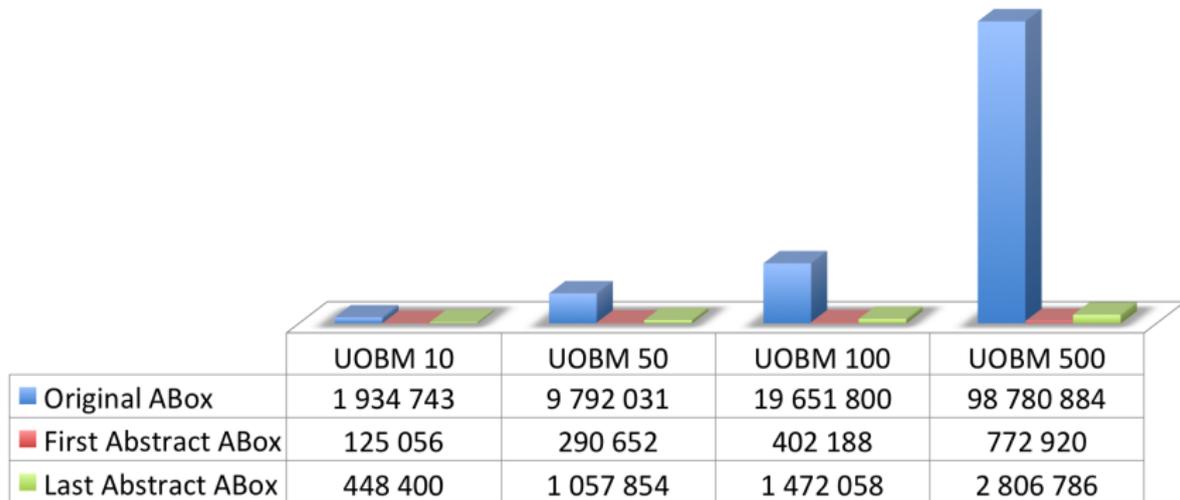
Abstract ABox Size (#Assertions) vs Original ABox Size (#Assertions)



- 1 refinement step for LUBM

Evaluation

Abstract ABox Size (#Assertions) vs Original ABox Size (#Assertions)



- 2 refinement steps for UOBM

▶ Detail

▶ Runtime

Conclusions

- Many (real-life) ontologies with a large ABox
 - ▶ Have a simple TBox
 - ▶ Many individuals are in the same equivalence classes

Conclusions

- Many (real-life) ontologies with a large ABox
 - ▶ Have a simple TBox
 - ▶ Many individuals are in the same equivalence classes
- ⇒ Our approach is particularly efficient

Conclusions

- Many (real-life) ontologies with a large ABox
 - ▶ Have a simple TBox
 - ▶ Many individuals are in the same equivalence classes
- ⇒ Our approach is particularly efficient
- Complete for Horn $ALCHOI$, including RDFS, \mathcal{EL} , \mathcal{QL} , and a large part of OWL 2 RL

Conclusions

- Many (real-life) ontologies with a large ABox
 - ▶ Have a simple TBox
 - ▶ Many individuals are in the same equivalence classes
- ⇒ Our approach is particularly efficient
- Complete for Horn $ALCHOI$, including RDFS, \mathcal{EL} , QL , and a large part of OWL 2 RL
- Future work:
 - ▶ Extend the approach for more expressive languages

Conclusions

- Many (real-life) ontologies with a large ABox
 - ▶ Have a simple TBox
 - ▶ Many individuals are in the same equivalence classes
- ⇒ Our approach is particularly efficient
- Complete for Horn \mathcal{ALCHOI} , including RDFS, \mathcal{EL} , \mathcal{QL} , and a large part of OWL 2 RL
- Future work:
 - ▶ Extend the approach for more expressive languages
 - ▶ Utilize abstract ABoxes for other tasks



Handling large ABoxes: Related Work

- Distributed programming models [Urbani et al., 2012]
- Modularization approach [Wandelt, 2011]
- Summary approach [Dolby et al., 2009]

References

-  Dolby, J., Fokoue, A., Kalyanpur, A., Schonberg, E., and Srinivas, K. (2009).
Scalable highly expressive reasoner (SHER).
J. of Web Semantics, 7(4):357–361.
-  Urbani, J., Kotoulas, S., Maassen, J., van Harmelen, F., and Bal, H. (2012).
WebPIE: A web-scale parallel inference engine using MapReduce.
J. of Web Semantics, 10:59–75.
-  Wandelt, S. (2011).
Efficient instance retrieval over semi-expressive ontologies.
PhD thesis, Hamburg University of Technology.

Evaluation

- Compare the size of the abstract ABoxes w.r.t. the original ABoxes
- Count how many refinement steps
- Tested ontologies
 - ▶ Gazetter, Coburn, LUBM, UOBM
 - ▶ ABoxes with thousands to million assertions

| Ontology | #A | #A _N | #R | #I | #R(a, b) | #A(a) |
|-----------|-----|-----------------|-----|------------|------------|------------|
| Gazetteer | 710 | 711 | 15 | 516 150 | 604 164 | 11 112 |
| Coburn | 719 | 1 161 | 109 | 123 515 | 237 620 | 297 002 |
| LUBM 1 | 43 | 49 | 25 | 17 174 | 49 336 | 18 128 |
| LUBM 10 | 43 | 49 | 25 | 207 426 | 630 753 | 219 680 |
| LUBM 50 | 43 | 49 | 25 | 1 082 818 | 3 298 813 | 1 147 136 |
| LUBM 100 | 43 | 49 | 25 | 2 153 645 | 6 645 928 | 2 281 035 |
| LUBM 500 | 43 | 49 | 25 | 10 189 173 | 31 073 066 | 10 793 509 |
| UOBM 1 | 69 | 90 | 35 | 24 925 | 153 571 | 44 680 |
| UOBM 10 | 69 | 90 | 35 | 242 549 | 1 500 628 | 434 115 |
| UOBM 50 | 69 | 90 | 35 | 1 227 445 | 7 594 996 | 2 197 035 |
| UOBM 100 | 69 | 90 | 35 | 2 462 012 | 15 241 909 | 4 409 891 |
| UOBM 500 | 69 | 90 | 35 | 12 379 113 | 76 612 736 | 22 168 148 |

Evaluation

- The abstract ABoxes are significantly small w.r.t. the original ones
- The ontologies can be materialized with few refinement steps

| Ontology | # of steps | Abstract ABox size | | | | Abstract ABox size (%) | | | |
|-----------|------------|--------------------|-----------|--------------|-----------|------------------------|--------|--------------|-------|
| | | # indiv. | | # assertions | | % indiv. | | % assertions | |
| Gazetteer | 1 | 5 640 | 5 640 | 5 142 | 8 512 | 1.09 | 1.09 | 0.79 | 1.31 |
| Coburn | 2 | 3 992 | 4 059 | 5 569 | 8 633 | 3.23 | 3.29 | 1.04 | 2.16 |
| LUBM 1 | 1 | 139 | 139 | 143 | 254 | 0.81 | 0.81 | 0.21 | 0.38 |
| LUBM 10 | 1 | 154 | 154 | 158 | 281 | 0.07 | 0.07 | 0.02 | 0.03 |
| LUBM 50 | 1 | 148 | 148 | 152 | 271 | 0.01 | 0.01 | 0.003 | 0.006 |
| LUBM 100 | 1 | 148 | 148 | 152 | 271 | 0.007 | 0.007 | 0.002 | 0.003 |
| LUBM 500 | 1 | 148 | 148 | 152 | 271 | 0.001 | 0.001 | 0.001 | 0.001 |
| UOBM 1 | 2 | 25 378 | 39 322 | 31 659 | 101 324 | 101.82 | 157.76 | 15.97 | 51.11 |
| UOBM 10 | 2 | 98 266 | 169 579 | 125 056 | 448 400 | 40.51 | 69.92 | 6.46 | 23.18 |
| UOBM 50 | 2 | 226 176 | 395 956 | 290 652 | 1 057 854 | 18.43 | 32.26 | 2.97 | 10.80 |
| UOBM 100 | 2 | 311 574 | 547 361 | 402 188 | 1 472 058 | 12.66 | 22.23 | 2.05 | 7.49 |
| UOBM 500 | 2 | 596 135 | 1 033 685 | 772 920 | 2 806 786 | 4.82 | 8.35 | 0.78 | 2.84 |

▶ Coburn

▶ LUBM

▶ UOBM

Evaluation: runtime

| | Hermit | | Graphscale | | Graphscale vs Hermit | |
|-----------|--------|------|------------|-----|----------------------|--------|
| | Second | GB | Second | GB | Time | Memory |
| Gazetteer | 421 | 1.5 | 34 | 0.5 | 8% | 35% |
| Coburn | 1583 | 1.0 | 28 | 0.3 | 2% | 25% |
| LUBM 10 | 193 | 1.1 | 26 | 0.4 | 13% | 38% |
| LUBM 50 | 4096 | 5.7 | 98 | 1.2 | 2% | 21% |
| LUBM 100 | 16954 | 11.8 | 174 | 2.6 | 1% | 22% |
| UOBM 10 | 337 | 2.4 | 257 | 1.0 | 76% | 42% |
| UOBM 50 | 8979 | 12.3 | 2324 | 4.3 | 26% | 35% |
| UOBM 100 | 34543 | 23.6 | 4541 | 6.5 | 13% | 27% |

▶ Coburn

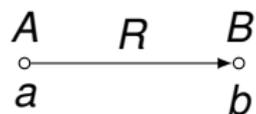
▶ LUBM

▶ UOBM

Disjunctions Make Incompleteness

$$\mathcal{T} = \{B \sqsubseteq C \sqcup D, \exists R.C \sqsubseteq C, A \sqcap C \sqsubseteq \forall R.D\}$$

Original ABox \mathcal{A}



$$\mathcal{A} \cup \mathcal{T} \models D(b)$$

Abstract ABox \mathcal{B}

