

PRACTICAL ASPECTS OF QUERY REWRITING FOR OWL 2

Héctor Pérez-Urbina, Ian Horrocks, and Boris Motik

Oxford University
Computing Laboratory

OWL: Experiences and Directions Workshop
October 2009





DATA ACCESS WITH OWL 2 QL



DATA ACCESS WITH OWL 2 QL

$Q(x) \leftarrow \text{teaches}(x, y)$

Teacher	\sqsubseteq	$\exists \text{teaches}$
Professor	\sqsubseteq	Teacher
$\exists \text{hasTutor}^-$	\sqsubseteq	Professor

PROFESSOR(*name, office, phone*)

STUDENT(*name, major, tutor*)



DATA ACCESS WITH OWL 2 QL

$Q(x) \leftarrow \text{teaches}(x, y)$

Teacher $\sqsubseteq \exists \text{teaches}$
Professor $\sqsubseteq \text{Teacher}$
 $\exists \text{hasTutor}^- \sqsubseteq \text{Professor}$

PROFESSOR(*name, office, phone*)

STUDENT(*name, major, tutor*)

Professor \mapsto SELECT name
FROM PROFESSOR
hasTutor \mapsto SELECT name, tutor
FROM STUDENT



DATA ACCESS WITH OWL 2 QL

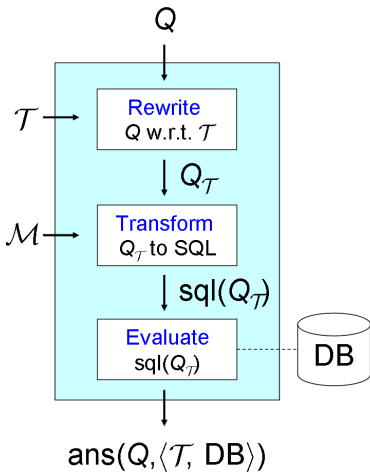
$Q(x) \leftarrow \text{teaches}(x, y)$

Teacher	\sqsubseteq	$\exists \text{teaches}$
Professor	\sqsubseteq	Teacher
$\exists \text{hasTutor}^-$	\sqsubseteq	Professor

PROFESSOR(*name, office, phone*)

STUDENT(*name, major, tutor*)

Professor	\mapsto	SELECT name FROM PROFESSOR
hasTutor	\mapsto	SELECT name, tutor FROM STUDENT



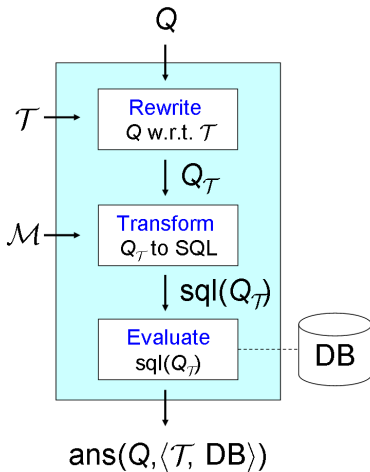


DATA ACCESS WITH OWL 2 QL

$Q(x) \leftarrow \text{teaches}(x, y)$

Teacher	\sqsubseteq	$\exists \text{teaches}$
Professor	\sqsubseteq	Teacher
$\exists \text{hasTutor}^-$	\sqsubseteq	Professor

CGLLR rewriting algorithm by
Calvanese et al.





DATA ACCESS WITH OWL 2 QL

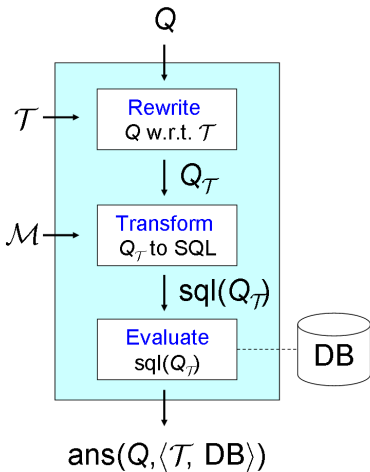
$Q(x) \leftarrow \text{teaches}(x, y)$

Teacher	\sqsubseteq	$\exists \text{teaches}$
Professor	\sqsubseteq	Teacher
$\exists \text{hasTutor}^-$	\sqsubseteq	Professor

$Q_{\mathcal{T}} = \{$

- $Q(x) \leftarrow \text{teaches}(x, y),$
- $Q(x) \leftarrow \text{Teacher}(x),$
- $Q(x) \leftarrow \text{Professor}(x),$
- $Q(x) \leftarrow \text{hasTutor}(y, x)$

$\}$





DATA ACCESS WITH OWL 2 QL

$Q(x) \leftarrow \text{teaches}(x, y)$

Teacher	\sqsubseteq	$\exists \text{teaches}$
Professor	\sqsubseteq	Teacher
$\exists \text{hasTutor}^-$	\sqsubseteq	Professor

$Q_T = \{$

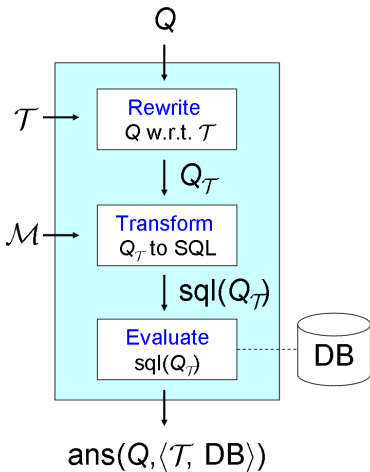
- $Q(x) \leftarrow \text{teaches}(x, y),$
- $Q(x) \leftarrow \text{Teacher}(x),$
- $Q(x) \leftarrow \text{Professor}(x),$
- $Q(x) \leftarrow \text{hasTutor}(y, x)$

$\}$

$\text{sql}(Q_T) =$

```

SELECT name
FROM PROFESSOR
UNION
SELECT Tutor
FROM Student
  
```





SO, ARE WE DONE HERE?



SO, ARE WE DONE HERE?

- $Q_{\mathcal{T}}$ is worst-case **exponential** w.r.t. Q and \mathcal{T}
 - Costly to **compute**
 - Costly (or impossible) to **evaluate**



SO, ARE WE DONE HERE?

- $Q_{\mathcal{T}}$ is worst-case **exponential** w.r.t. Q and \mathcal{T}
 - Costly to **compute**
 - Costly (or impossible) to **evaluate**
- Applications may require constructs that go **beyond** QL
 - Student $\sqcap \exists$ hasSupervisor \sqsubseteq GraduateStudent
 - \exists studies.**Course** \sqsubseteq Student
 - OxfordStudent $\sqsubseteq \exists$ studiesAt.**{OxfordUniversity}**



SO, ARE WE DONE HERE?

- $Q_{\mathcal{T}}$ is worst-case **exponential** w.r.t. Q and \mathcal{T}
 - Costly to **compute**
 - Costly (or impossible) to **evaluate**
- Applications may require constructs that go **beyond** QL
 - Student $\sqcap \exists$ hasSupervisor \sqsubseteq GraduateStudent
 - \exists studies.**Course** \sqsubseteq Student
 - OxfordStudent $\sqsubseteq \exists$ studiesAt.**{OxfordUniversity}**

RQR (RESOLUTION-BASED QUERY REWRITING)

- Handles $\mathcal{ELHI}O^{\neg}$ (most of OWL 2 **EL**)



SO, ARE WE DONE HERE?

- $Q_{\mathcal{T}}$ is worst-case **exponential** w.r.t. Q and \mathcal{T}
 - Costly to **compute**
 - Costly (or impossible) to **evaluate**
- Applications may require constructs that go **beyond** QL
 - Student $\sqcap \exists$ hasSupervisor \sqsubseteq GraduateStudent
 - \exists studies.**Course** \sqsubseteq Student
 - OxfordStudent $\sqsubseteq \exists$ studiesAt.**{OxfordUniversity}**

RQR (RESOLUTION-BASED QUERY REWRITING)

- Handles $\mathcal{ELHI}O^{-}$ (most of OWL 2 **EL**)
- $Q_{\mathcal{T}}$ might be a **datalog** query
- “**Pay-as-you-go**” behavior: extends and generalizes CGLLR



EVALUATION



EVALUATION

NUMBER OF INFERENCES \approx TIME

	REQUIEM (RQR)	C (CGLLR)
Overall	70,846	343,813
Average	2,200	12,066

- REQUIEM: 73% smaller, 0% larger, and 27% equal



EVALUATION

NUMBER OF INFERENCES \approx TIME

	REQUIEM (RQR)	C (CGLLR)
Overall	70,846	343,813
Average	2,200	12,066

- REQUIEM: 73% smaller, 0% larger, and 27% equal

NUMBER OF QUERIES

	REQUIEM (RQR)	C (CGLLR)
Overall	10,931	75,301
Average	289	2,682

- REQUIEM: 83% smaller, 0% larger, and 17% equal



OPTIMIZING THE REWRITINGS

QUERY SUBSUMPTION

$$Q_1 = Q(x) \leftarrow \text{teaches}(x, y)$$

$$Q_2 = Q(x) \leftarrow \text{teaches}(x, y) \wedge \text{Student}(y)$$



OPTIMIZING THE REWRITINGS

QUERY SUBSUMPTION

$Q_1 = Q(x) \leftarrow \text{teaches}(x, y)$

$Q_2 = Q(x) \leftarrow \text{teaches}(x, y) \wedge \text{Student}(y)$

Q_1 **subsumes** Q_2



OPTIMIZING THE REWRITINGS

QUERY SUBSUMPTION

$$Q_1 = Q(x) \leftarrow \text{teaches}(x, y)$$

$$Q_2 = Q(x) \leftarrow \text{teaches}(x, y) \wedge \text{Student}(y)$$

Q_1 **subsumes** Q_2

- Discard **subsumed** queries a posteriori
- **Significant** reduction in the size of the rewritings



OPTIMIZING THE REWRITINGS

QUERY SUBSUMPTION

$$Q_1 = Q(x) \leftarrow \text{teaches}(x, y)$$

$$Q_2 = Q(x) \leftarrow \text{teaches}(x, y) \wedge \text{Student}(y)$$

Q_1 **subsumes** Q_2

- Discard **subsumed** queries a posteriori
- **Significant** reduction in the size of the rewritings
- On the fly: **forward/backward** subsumption
- **Straightforwardly** applicable to RQR



GOING BEYOND QL

CONSIDERATIONS

- Q_T is not guaranteed to be a UCQ: use of **deductive** database systems



GOING BEYOND QL

CONSIDERATIONS

- Q_T is not guaranteed to be a UCQ: use of **deductive** database systems
- Additional **optimizations**
 - Empty EDB predicates pruning
 - Dependency graph pruning



GOING BEYOND QL

CONSIDERATIONS

- Q_T is not guaranteed to be a UCQ: use of **deductive** database systems
- Additional **optimizations**
 - Empty EDB predicates pruning
 - Dependency graph pruning
- **Greedy** unfolding



GOING BEYOND QL

CONSIDERATIONS

- Q_T is not guaranteed to be a UCQ: use of **deductive** database systems
- Additional **optimizations**
 - Empty EDB predicates pruning
 - Dependency graph pruning
- **Greedy** unfolding

EVALUATION

- **Good performance** w.r.t. time and size of the rewritings
- Greedy unfolding produces **UCQs** in many cases



CONCLUSIONS AND FUTURE WORK

CONCLUSIONS

- RQR: **significantly** smaller rewritings in **significantly** fewer steps than existing algorithms
- Amenable to various straightforward **optimizations**
- Use of **databases** for realistic OWL 2 EL ontologies
- **Open source** implementation: REQUIEM¹

¹<http://www.comlab.ox.ac.uk/projects/requiem/>



CONCLUSIONS AND FUTURE WORK

CONCLUSIONS

- RQR: **significantly** smaller rewritings in **significantly** fewer steps than existing algorithms
- Amenable to various straightforward **optimizations**
- Use of **databases** for realistic OWL 2 EL ontologies
- **Open source** implementation: REQUIEM¹

FUTURE WORK

- Evaluate REQUIEM with **databases**

¹<http://www.comlab.ox.ac.uk/projects/requiem/>