Description Logic Knowledge Base Exchange

Elena Botoeva
supervised by Diego Calvanese

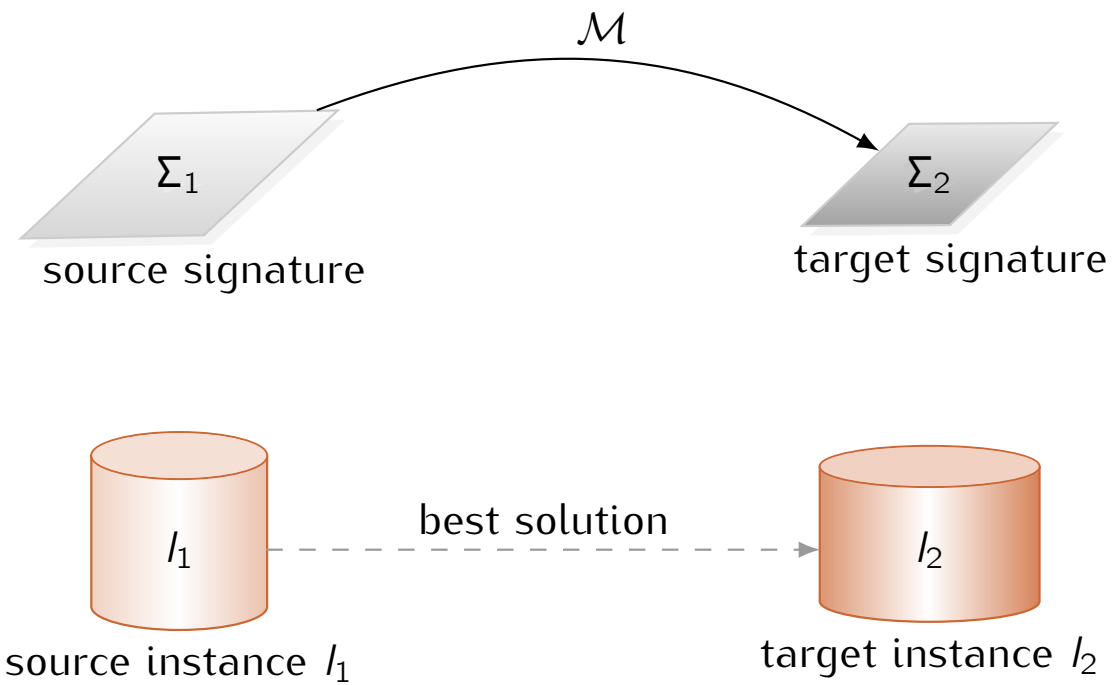
KRDB Research Centre, Free University of Bozen-Bolzano, Italy
lastname@inf.unibz.it

Reasoning Web and Rules Doctoral Consortium
10 September 2012, Vienna

Data Exchange

Problem

given a mapping $\mathcal{M}$ and a source instance I_1,
compute a target instance I_2 that is a *solution* for I_1 under $\mathcal{M}$.



Data Exchange Example

$\mathcal{M}$: $\forall a, t.(\quad AuthorOf(a, t) \quad \rightarrow \quad \exists g.BookInfo(t, a, g) \quad)$

I_1 :

AuthorOf	
nabokov	lolita
tolkien	lotr

Data Exchange Example

$\mathcal{M}$: $\forall a, t.($ $AuthorOf(a, t)$ $\rightarrow$ $\exists g.BookInfo(t, a, g)$ $)$

I_1 :

AuthorOf	
nabokov	lolita
tolkien	lotr

I_2 :

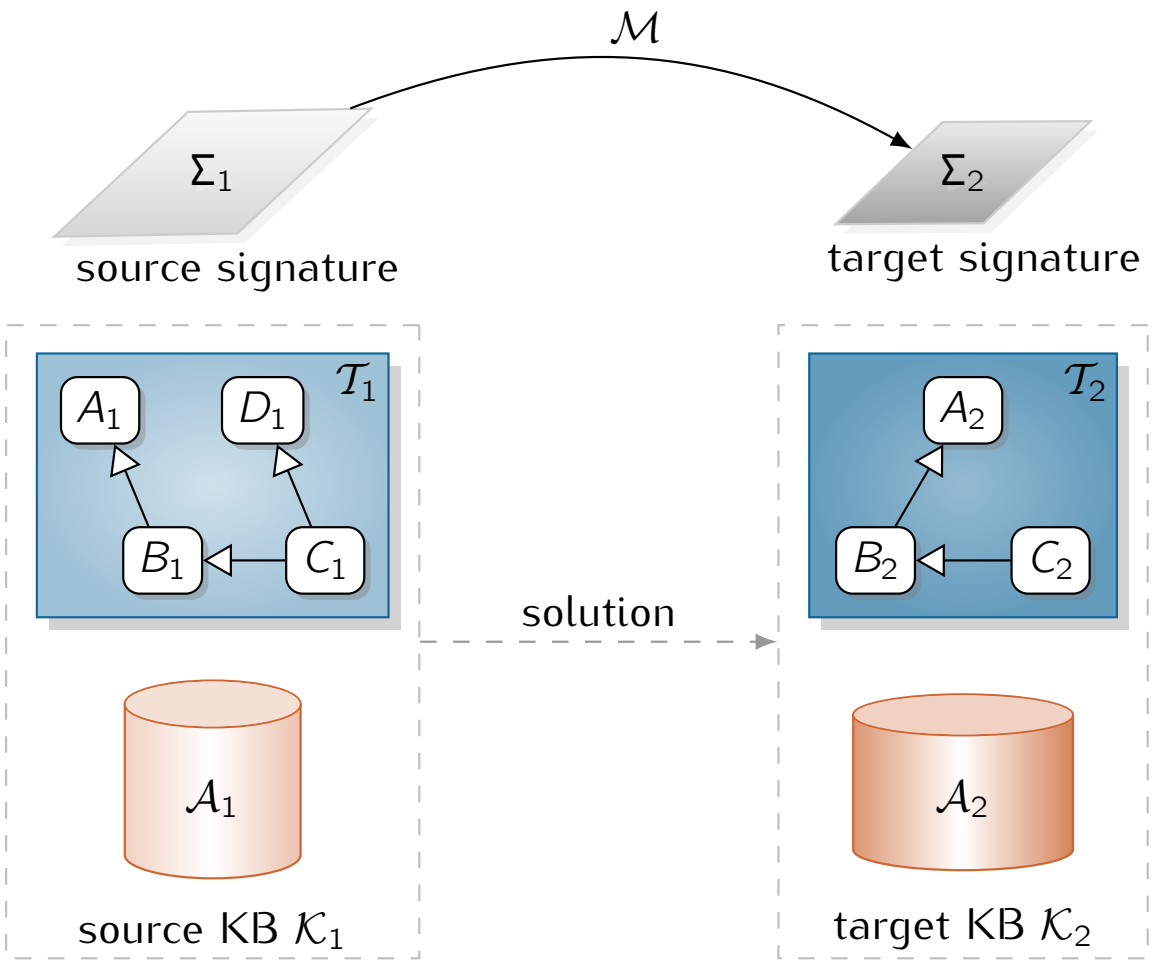
BookInfo		
lolita	nabokov	n_1
lotr	tolkien	n_2

I_2 is a universal solution for I_1 under $\mathcal{M}$.

Knowledge Base Exchange

Problem

given a mapping $\mathcal{M}$ and a source knowledge base (KB) $\mathcal{K}_1$, compute a target KB $\mathcal{K}_2$ that is a *solution* for $\mathcal{K}_1$ under $\mathcal{M}$.



Knowledge Base Exchange: Example

$\mathcal{M}$:

$$\exists \mathit{AuthorOf}^- \sqsubseteq \exists \mathit{BookGenre}$$
$$\mathit{AuthorOf}^- \sqsubseteq \mathit{WrittenBy}$$
$$\mathit{TaxNumber} \sqsubseteq \mathit{SSN}$$

$\mathcal{T}_1$:

$$\exists \mathit{AuthorOf} \sqsubseteq \mathit{Author}$$
$$\mathit{Author} \sqsubseteq \exists \mathit{TaxNumber}$$

$\mathcal{A}_1$:

AuthorOf	
nabokov	lolita
tolkien	lotr

Knowledge Base Exchange: Example

$\mathcal{M}$:

$$\exists \mathit{AuthorOf}^- \sqsubseteq \exists \mathit{BookGenre}$$
$$\mathit{AuthorOf}^- \sqsubseteq \mathit{WrittenBy}$$
$$\mathit{TaxNumber} \sqsubseteq \mathit{SSN}$$

$\mathcal{T}_1$:

$$\exists \mathit{AuthorOf} \sqsubseteq \mathit{Author}$$
$$\mathit{Author} \sqsubseteq \exists \mathit{TaxNumber}$$

$\mathcal{A}_1$:

AuthorOf	
nabokov	lolita
tolkien	lotr

$\mathcal{A}_2$:

WrittenBy	
lolita	nabokov
lotr	tolkien

SSN	
nabokov	m_1
tolkien	m_2

BookGenre	
lolita	m_3
lotr	m_4

$\mathcal{A}_2$ is a universal solution for $\langle \mathcal{T}_1, \mathcal{A}_1 \rangle$ under $\mathcal{M}$.

Knowledge Base Exchange: Example

$\mathcal{M}$:

$$\exists \mathit{AuthorOf}^- \sqsubseteq \exists \mathit{BookGenre}$$
$$\mathit{AuthorOf}^- \sqsubseteq \mathit{WrittenBy}$$
$$\mathit{TaxNumber} \sqsubseteq \mathit{SSN}$$

$\mathcal{T}_1$:

$$\exists \mathit{AuthorOf} \sqsubseteq \mathit{Author}$$
$$\mathit{Author} \sqsubseteq \exists \mathit{TaxNumber}$$

$\mathcal{A}_1$:

AuthorOf	
nabokov	lolita
tolkien	lotr

Knowledge Base Exchange: Example

$\mathcal{M}$:

$$\exists \mathit{AuthorOf}^- \sqsubseteq \exists \mathit{BookGenre}$$
$$\mathit{AuthorOf}^- \sqsubseteq \mathit{WrittenBy}$$
$$\mathit{TaxNumber} \sqsubseteq \mathit{SSN}$$

$\mathcal{T}_1$:

$$\exists \mathit{AuthorOf} \sqsubseteq \mathit{Author}$$
$$\mathit{Author} \sqsubseteq \exists \mathit{TaxNumber}$$

$\mathcal{A}_1$:

AuthorOf	
nabokov	lolita
tolkien	lotr

$\mathcal{T}_2$:

$$\exists \mathit{WrittenBy}^- \sqsubseteq \exists \mathit{SSN}$$
$$\exists \mathit{WrittenBy} \sqsubseteq \exists \mathit{BookGenre}$$

$\mathcal{A}_2$:

WrittenBy	
lolita	nabokov
lotr	tolkien

$\langle \mathcal{T}_2, \mathcal{A}_2 \rangle$ is a universal-UCQ solution for $\langle \mathcal{T}_1, \mathcal{A}_1 \rangle$ under $\mathcal{M}$.

Reasoning Problems

We are interested in

- computing universal solutions

  Does the core exist?
  (Is there a finite subset of $chase_{\mathcal{T}_1 \cup \mathcal{M}, \Sigma_2}(\mathcal{A}_1)$ homomorphically equivalent to $chase_{\mathcal{T}_1 \cup \mathcal{M}, \Sigma_2}(\mathcal{A}_1)$)

  - PSpace-hard
  - in ExpTime

- computing universal-UCQ solutions

  Is $chase_{\mathcal{T}_2}(\mathcal{A}_2)$ homomorphically equivalent to $chase_{\mathcal{T}_1 \cup \mathcal{M}, \Sigma_2}(\mathcal{A}_1)$?

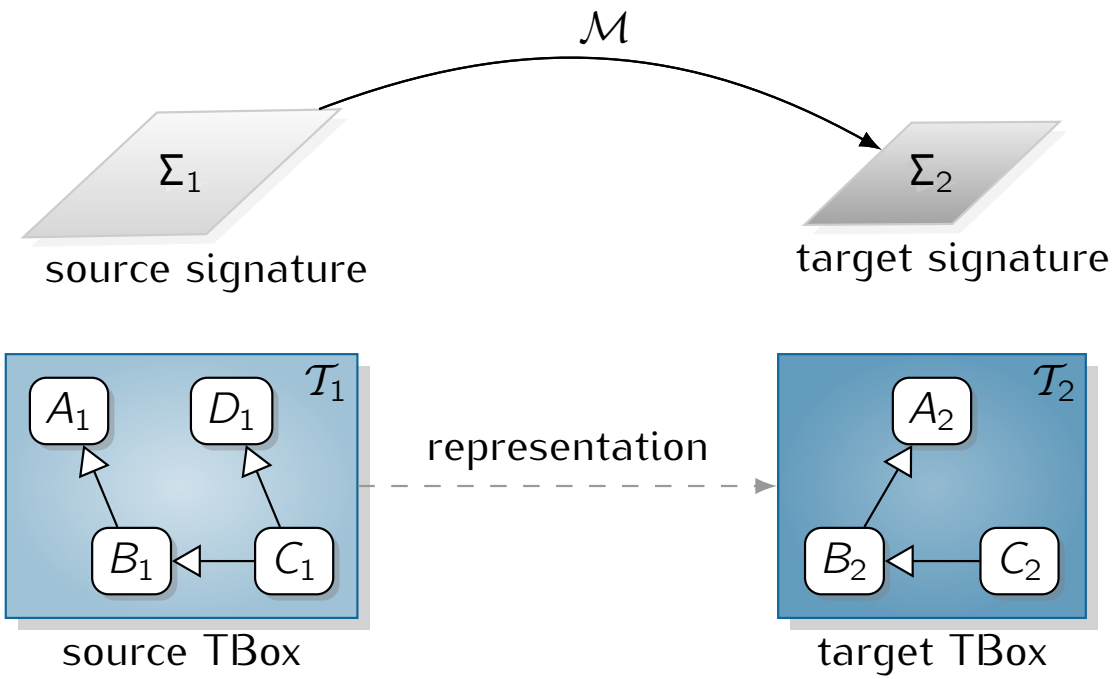
- maximizing the implicit knowledge in the target

  Universal solutions vs. Universal-UCQ solutions

Representability Problem

Problem

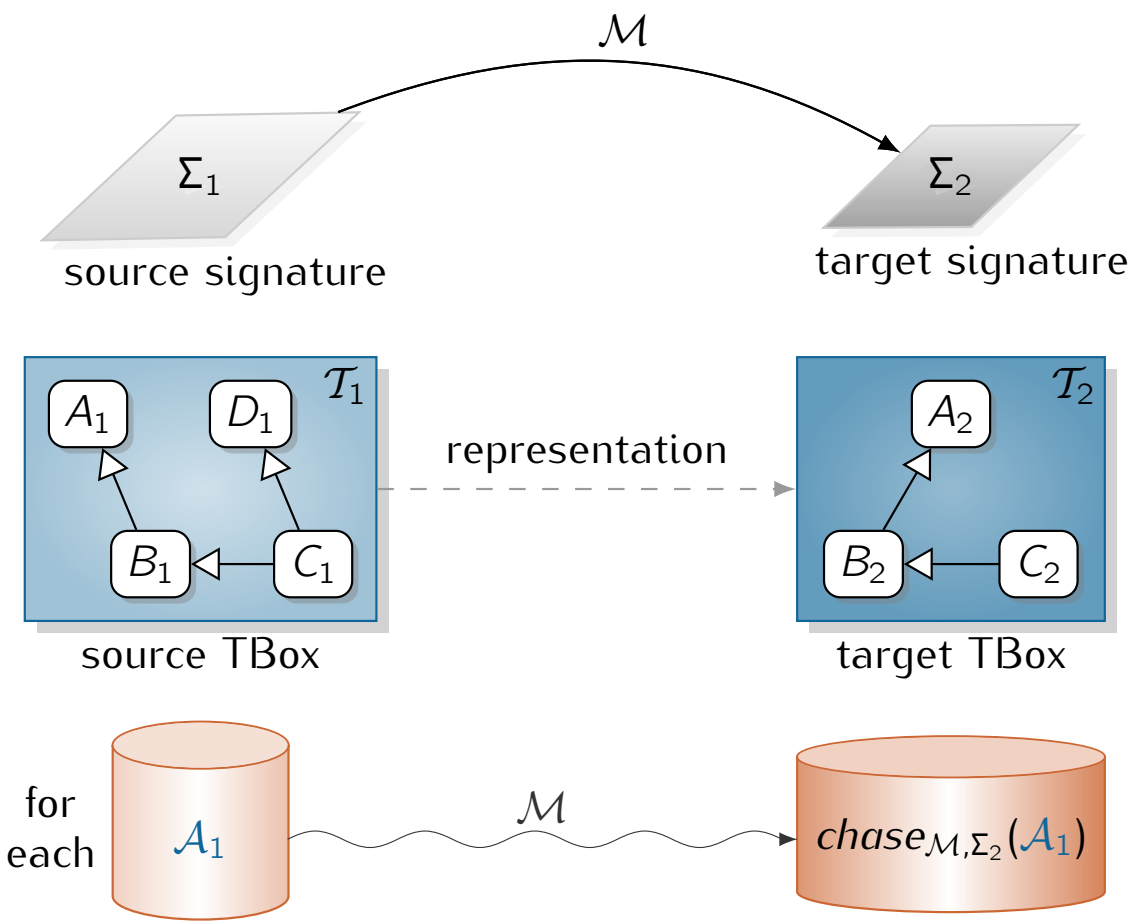
given a mapping $\mathcal{M}$ and a source TBox $\mathcal{T}_1$,
compute (if exists) a target TBox $\mathcal{T}_2$ such that for each source ABox $\mathcal{A}_1$,
$\langle \mathcal{T}_2, chase_{\mathcal{M},\Sigma_2}(\mathcal{A}_1) \rangle$ is a universal UCQ-solution for $\langle \mathcal{T}_1, \mathcal{A}_1 \rangle$ under $\mathcal{M}$.



Representability Problem

Problem

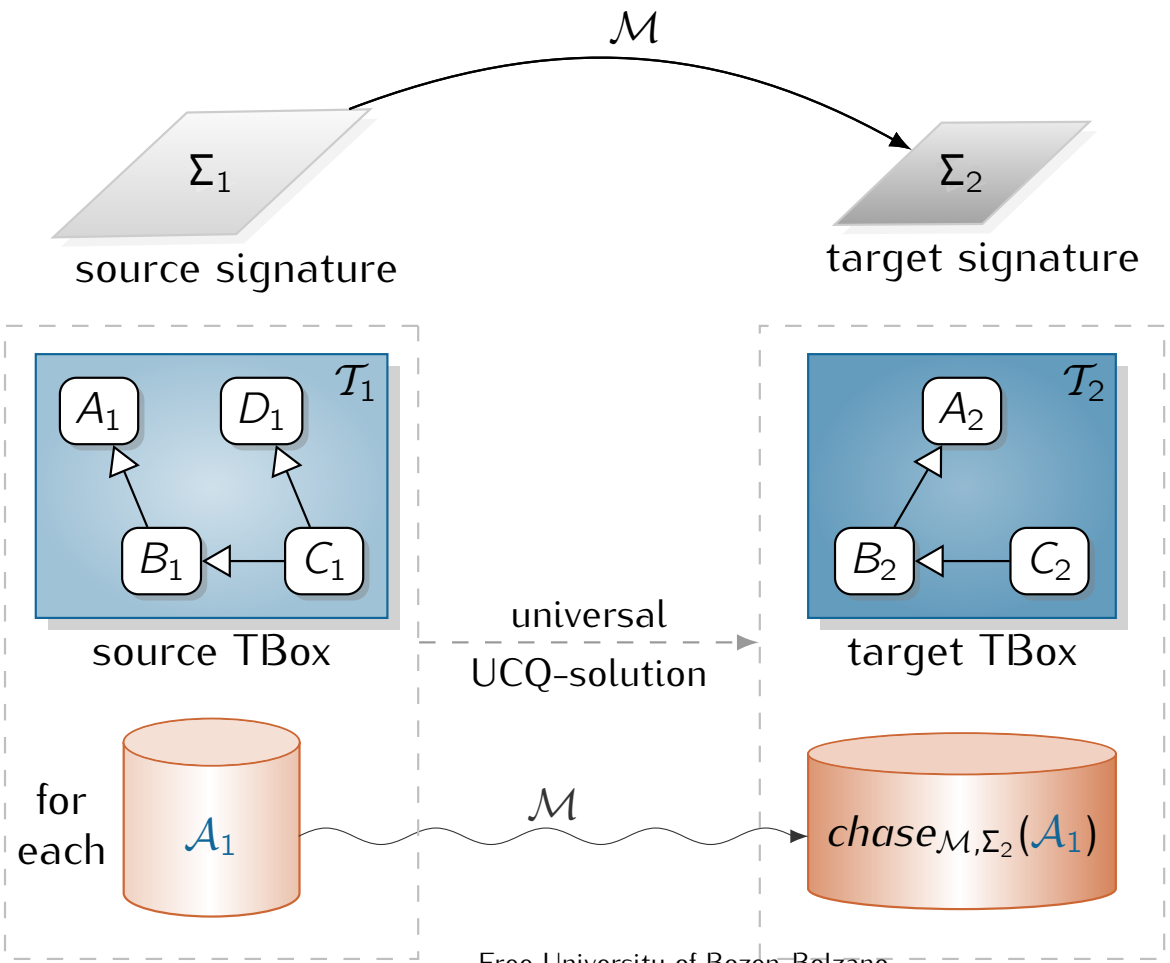
given a mapping $\mathcal{M}$ and a source TBox $\mathcal{T}_1$,
compute (if exists) a target TBox $\mathcal{T}_2$ such that for each source ABox $\mathcal{A}_1$,
$\langle \mathcal{T}_2, \mathit{chase}_{\mathcal{M},\Sigma_2}(\mathcal{A}_1) \rangle$ is a universal UCQ-solution for $\langle \mathcal{T}_1, \mathcal{A}_1 \rangle$ under $\mathcal{M}$.



Representability Problem

Problem

given a mapping $\mathcal{M}$ and a source TBox $\mathcal{T}_1$,
compute (if exists) a target TBox $\mathcal{T}_2$ such that for each source ABox $\mathcal{A}_1$,
$\langle \mathcal{T}_2, \mathit{chase}_{\mathcal{M},\Sigma_2}(\mathcal{A}_1) \rangle$ is a universal UCQ-solution for $\langle \mathcal{T}_1, \mathcal{A}_1 \rangle$ under $\mathcal{M}$.



Representability: Example

$\mathcal{M}$:

$$\exists \mathit{AuthorOf}^- \sqsubseteq \exists \mathit{BookGenre}$$
$$\mathit{AuthorOf}^- \sqsubseteq \mathit{WrittenBy}$$
$$\mathit{TaxNumber} \sqsubseteq \mathit{SSN}$$

$\mathcal{T}_1$:

$$\exists \mathit{AuthorOf} \sqsubseteq \exists \mathit{TaxNumber}$$

$\mathcal{T}_2$:

$$\exists \mathit{WrittenBy}^- \sqsubseteq \exists \mathit{SSN}$$
$$\exists \mathit{WrittenBy} \sqsubseteq \exists \mathit{BookGenre}$$

$\mathcal{T}_2$ is a representation of $\mathcal{T}_1$ under $\mathcal{M}$.

Representability: Example

$\mathcal{M}$:

$$\exists AuthorOf^- \sqsubseteq \exists BookGenre$$
$$AuthorOf^- \sqsubseteq WrittenBy$$
$$TaxNumber \sqsubseteq SSN$$

$\mathcal{T}_1$:

$$\exists AuthorOf \sqsubseteq \exists TaxNumber$$

$\mathcal{T}_2$:

$$\exists WrittenBy^- \sqsubseteq \exists SSN$$
$$\exists WrittenBy \sqsubseteq \exists BookGenre$$

$\mathcal{A}_1$:

AuthorOf	
nabokov	lolita

TaxNumber	
smith	000

$chase_{\mathcal{M},\Sigma_2}(\mathcal{A}_1)$:

WrittenBy	
lolita	nabokov

SSN	
smith	000

BookGenre	
lolita	n_1

$\mathcal{T}_2$ is a representation of $\mathcal{T}_1$ under $\mathcal{M}$.

Representability Problem

Theorem

Representability problem is decidable in polynomial time.

Note that given $\langle \mathcal{T}_1, \mathcal{A}_1 \rangle$ and $\mathcal{M}$, if $\mathcal{T}_1$ is representable under $\mathcal{M}$, then we can construct a universal UCQ-solution of polynomial size.

Note also that the query inseparability problem for *DL-Lite*$_\mathcal{R}$ TBoxes is PSpace-hard.

Open Problems and Future Work

Open problems

- the exact computational complexity of computing (universal) solutions,
- computing a universal solution in presence of disjointness assertions in the mapping,
- computing a universal UCQ-solution,

Future work:

- implementation of the representability algorithm,
- study KB exchange for more expressive/other languages, such as *DL-Lite*$_{\mathcal{R}}$ with $\exists R.A$, *DL-Lite*$_{horn}$, and $\mathcal{EL}$,
- study composition and inversion of mappings.

Thank you
for your attention!