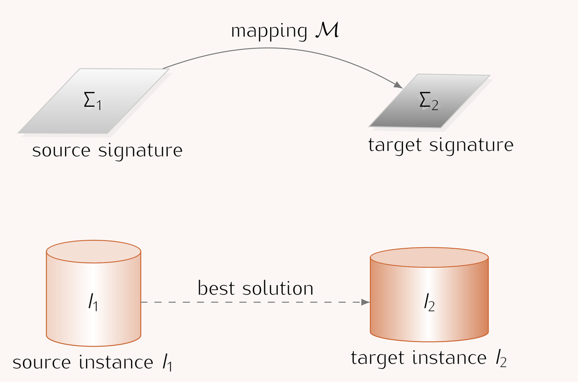
# Description Logic Knowledge Base Exchange

Elena Botoeva supervised by Diego Calvanese
KRDB Research Centre, Free University of Bozen-Bolzano, Italy
botoeva@inf.unibz.it

## Data Exchange vs. Knowledge Base Exchange

Knowledge Base Exchange is a special case of Data Exchange with incomplete information.

### Data Exchange

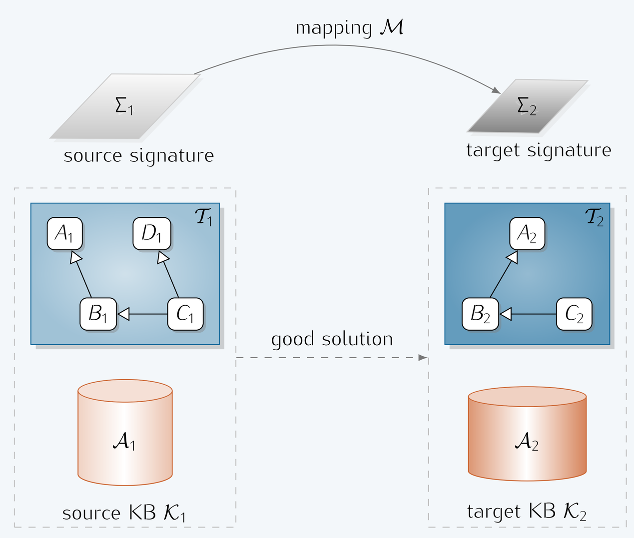



Each database instance is *complete*, i.e., every fact is either true or false, therefore, represents *exactly one* possible instance (itself).

#### Data Exchange Example

Let $\mathcal{M} = \{\ \forall a, t.(\text{AuthorOf}(a, t) \rightarrow \exists g.\text{BookInfo}(t, a, g))\ \}$,
$I_1 = \{\ \text{AuthorOf}(\text{nabokov}, \text{lolita}), \text{AuthorOf}(\text{tolkien}, \text{lotr})\ \}$.
Then $I_2 = \{\ \text{BookInfo}(\text{lolita}, \text{nabokov}, n_1), \text{BookInfo}(\text{lotr}, \text{tolkien}, n_2)\ \}$
is a universal solution.

### Knowledge Base Exchange




A Description Logic (DL) Knowledge Base consists of

- *incomplete data* given as an ABox and
- *structural rules* for completing data given as a TBox,

therefore, is a compact representation of (possibly infinitely) *many* actual instances.

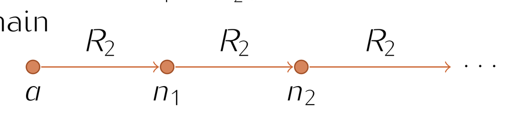
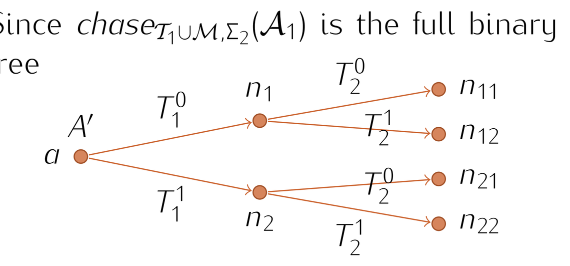
#### Knowledge Base Exchange Example

Let $\mathcal{M} = \{\ \exists\text{AuthorOf}^- \sqsubseteq \exists\text{BookGenre}, \text{AuthorOf}^- \sqsubseteq \text{WrittenBy}, \text{TaxNumber} \sqsubseteq \text{SSN}\ \}$,
$\mathcal{T}_1 = \{\ \exists\text{AuthorOf} \sqsubseteq \text{Author}, \text{Author} \sqsubseteq \exists\text{TaxNumber}\ \}$,
$\mathcal{A}_1 = \{\ \text{AuthorOf}(\text{nabokov}, \text{lolita}), \text{AuthorOf}(\text{tolkien}, \text{lotr})\ \}$.
Then $\mathcal{A}_2 = \{\ \text{WrittenBy}(\text{lolita}, \text{nabokov}), \text{WrittenBy}(\text{lotr}, \text{tolkien}), \text{SSN}(\text{nabokov}, m_1), \text{SSN}(\text{tolkien}, m_2), \text{BookGenre}(\text{lolita}, m_3), \text{BookGenre}(\text{lotr}, m_4)\ \}$
is a universal solution.

We consider exchange of Description Logic KBs: each KB is constituted by a TBox and an ABox, and mapping is a set of DL inclusions from the source signature to the target signature. We start with lightweight DLs *DL-Lite*$_{\mathcal{R}}$ and *DL-Lite*$_{RDFS}$.

## Universal Solutions vs. Universal UCQ-Solutions

Universal UCQ-Solutions are the *preferred* solutions in our setting.

| Example | Universal Solutions | Universal UCQ-Solutions |
|---|---|---|
| • simplest example<br>$\mathcal{M} = \{\ A_1 \sqsubseteq A_2, B_1 \sqsubseteq B_2\ \}$,<br>$\mathcal{T}_1 = \{\ B_1 \sqsubseteq A_1\ \}$,<br>$\mathcal{A}_1 = \{\ B_1(a)\ \}$. | • universal solutions cannot have a non-empty TBox<br>Since axiom $B_2 \sqsubseteq A_2$ is not satisfied by $\mathcal{I}_1 = \{B_2(a), B_2(b), A_2(a)\}$, a $\Sigma_2$-model of $\langle \mathcal{T}_1 \cup \mathcal{M}, \mathcal{A}_1 \rangle$. | • universal UCQ-solutions can have a non-empty TBox<br>$\mathcal{K}_2 = \langle \mathcal{T}_2, \mathcal{A}_2 \rangle$ with $\mathcal{T}_2 = \{\ B_2 \sqsubseteq A_2\ \}$, $\mathcal{A}_2 = \{\ B_2(a)\ \}$, is a univ. UCQ-solution for $\mathcal{K}_1$ under $\mathcal{M}$. |
| • infinite $R_2$-chain<br>$\mathcal{M} = \{\ R_1 \sqsubseteq R_2\ \}$,<br>$\mathcal{T}_1 = \{\ A_1 \sqsubseteq \exists R_1, \exists R_1^- \sqsubseteq \exists R_1\ \}$,<br>$\mathcal{A}_1 = \{\ A_1\ (a)\ \}$. | • a universal solution does not exist<br>Since $chase_{\mathcal{T}_1 \cup \mathcal{M}, \Sigma_2}(\mathcal{A}_1)$ is the following chain<br><br>and there exists no ABox homomorphically equivalent to this chain. | • a universal UCQ-solution exists<br>$\mathcal{K}_2 = \langle \mathcal{T}_2, \mathcal{A}_2 \rangle$ with $\mathcal{T}_2 = \{\exists R_2^- \sqsubseteq \exists R_2\}$, $\mathcal{A}_2 = \{R_2(a, n_1)\}$. is a univ. UCQ-solution for $\mathcal{K}_1$ under $\mathcal{M}$. |
| • full binary tree (exponential)<br>$\mathcal{M} = \{\ S_i^k \sqsubseteq T_i^k, A \sqsubseteq A'\ \}$,<br>$\mathcal{T}_1 = \{\ A \sqsubseteq \exists S_0^k, \exists {S_{i-1}^k}^- \sqsubseteq \exists S_i^k\ \}$,<br>$\mathcal{A}_1 = \{\ A_1\ (a)\ \}$, | • universal solutions are exponential in the size of $\mathcal{K}_1$ and $\mathcal{M}$<br>Since $chase_{\mathcal{T}_1 \cup \mathcal{M}, \Sigma_2}(\mathcal{A}_1)$ is the full binary tree<br><br>and the smallest universal solution. | • universal UCQ-solutions are of polynomial size<br>$\mathcal{K}_2 = \langle \mathcal{T}_2, \mathcal{A}_2 \rangle$ with $\mathcal{T}_2 = \{\ A' \sqsubseteq \exists T_0^k, \exists {T_{i-1}^k}^- \sqsubseteq \exists T_i^k\ \}$, $\mathcal{A}_2 = \{A'(a)\}$. is a univ. UCQ-solution for $\mathcal{K}_1$ under $\mathcal{M}$. |

## Open Problems and Future Work

Problems that remain open

- 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,
- computing the minimal mapping $\mathcal{M}^*$ for a weakly representable $\mathcal{T}_1$, such that $\mathcal{T}_1$ is representable under $\mathcal{M}^*$.

Plans for future work:

- implement the representability algorithm,
- implement a simple prototype for KB exchange,
- 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.

## (Computing) KB Solutions

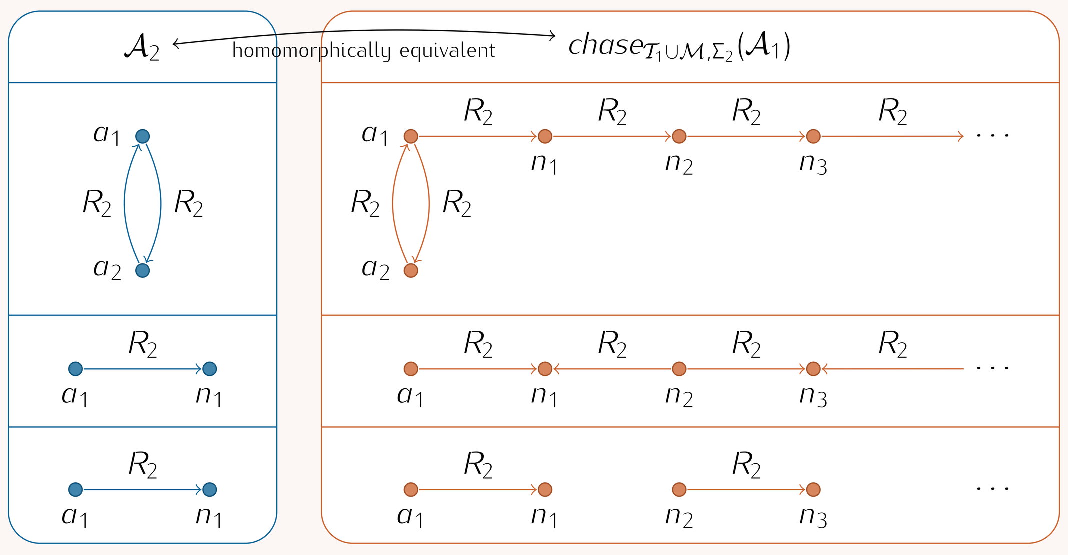
The basic reasoning problem is to compute a universal solution.

- $\mathcal{K}_2$ is said to be a *universal solution* for $\mathcal{K}_1$ under a mapping $\mathcal{M}$ if

$$\text{Mod}(\mathcal{K}_2) = \text{Sat}_{\mathcal{M}}(\text{Mod}(\mathcal{K}_1))$$

- $\mathcal{K}_2 = \langle \emptyset, \mathcal{A}_2 \rangle$ is a universal solution for $\mathcal{K}_1 = \langle \mathcal{T}_1, \mathcal{A}_1 \rangle$ under $\mathcal{M}$ iff

$$\mathcal{A}_2 \leftrightarrow chase_{\mathcal{T}_1 \cup \mathcal{M}, \Sigma_2}(\mathcal{A}_1)$$




- Deciding existence of a universal solution in the case of *DL-Lite*$_{\mathcal{R}}$ is
  → PSpace-hard (reduction from satisfiability of QBF) and
  → in ExpTime (using two-way alternating automatas).

We are also interested in query-solutions for a class of queries $\mathcal{Q}$.

- $\mathcal{K}_2$ is said to be a *universal $\mathcal{Q}$-solution* for $\mathcal{K}_1$ under $\mathcal{M}$, if

$$\forall q \in \mathcal{Q} \text{ over } \Sigma_2,\ cert(q, \langle \mathcal{T}_1 \cup \mathcal{M}, \mathcal{A}_1 \rangle) = cert(q, \mathcal{K}_2)$$

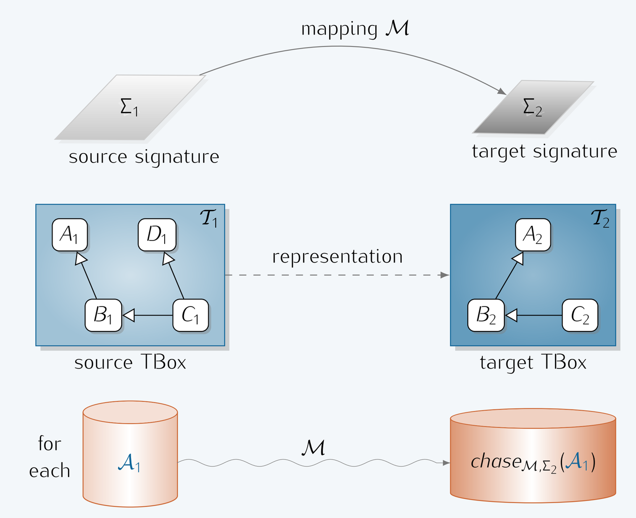
- $\mathcal{K}_2 = \langle \mathcal{T}_2, \mathcal{A}_2 \rangle$ is a univ. UCQ-solution for $\mathcal{K}_1 = \langle \mathcal{T}_1, \mathcal{A}_1 \rangle$ under $\mathcal{M}$ iff

$$chase_{\mathcal{T}_2}(\mathcal{A}_2) \leftrightarrow chase_{\mathcal{T}_1 \cup \mathcal{M}, \Sigma_2}(\mathcal{A}_1)$$

- The same complexity results hold for universal UCQ-solutions.

## Representability Problem

We want to maximize implicit knowledge in the target.




*$\mathcal{T}_1$ is representable under $\mathcal{M}$ if there exists $\mathcal{T}_2$ s.t. for each 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 $\mathcal{M}$, and $\mathcal{T}_2$ is called representation of $\mathcal{T}_1$ under $\mathcal{M}$.*

Let $\mathcal{M} = \{\ A_1 \sqsubseteq A_2, B_1 \sqsubseteq B_2\ \}$ and $\mathcal{T}_1 = \{\ B_1 \sqsubseteq A_1\ \}$. Then $\mathcal{T}_1$ is representable under $\mathcal{M}$ and $\mathcal{T}_2 = \{\ B_2 \sqsubseteq A_2\ \}$ is a representation of $\mathcal{T}_1$ under $\mathcal{M}$.

#### The representability problem for *DL-Lite*$_{\mathcal{R}}$ TBoxes

is decidable in PTime.

If $\mathcal{T}_1$ is representable under $\mathcal{M}$, then we have an algorithm to construct *universal UCQ-solutions of polynomial size*. Otherwise, a more relaxed notion of representability can be used.

*$\mathcal{T}_1$ is weakly representable under $\mathcal{M}$ if there exists $\mathcal{M}^*$, such that $\mathcal{M} \subseteq \mathcal{M}^*$, $\mathcal{M} \cup \mathcal{T}_1 \models \mathcal{M}^*$ and $\mathcal{T}_1$ is representable under $\mathcal{M}^*$.*

Let $\mathcal{M} = \{\ A_1 \sqsubseteq A_2\ \}$ and $\mathcal{T}_1 = \{\ B_1 \sqsubseteq A_1\ \}$. Then $\mathcal{T}_1$ is weakly representable under $\mathcal{M}$: consider $\mathcal{M}^* = \{\ A_1 \sqsubseteq A_2, B_1 \sqsubseteq A_2\ \}$ and $\mathcal{T}_2 = \{\}$.

#### The weak representability problem for *DL-Lite*$_{\mathcal{R}}$ TBoxes

is decidable in PTime.

## Publications

[1] M. Arenas, E. Botoeva, and D. Calvanese. Knowledge base exchange. In *Proc. of DL 2011*, volume 745 of *CEUR*, ceur-ws.org, 2011.

[2] M. Arenas, E. Botoeva, D. Calvanese, V. Ryzhikov, and E. Sherkhonov. Exchanging description logic knowledge bases. In *Proc. of KR 2012*, 2012.

[3] M. Arenas, E. Botoeva, D. Calvanese, V. Ryzhikov, and E. Sherkhonov. Representability in dl-liter knowledge base exchange. In *Proc. of the 25th Int. Workshop on Description Logics (DL 2012)*, 2012.