Exchanging OWL 2 QL Knowledge Bases

Marcelo Arenas
PUC Chile &
Univ. of Oxford, U.K.
marenas@ing.puc.cl

Elena Botoeva
Free U. of Bolzano
Italy
botoeva@inf.unibz.it

Diego Calvanese
Free U. of Bolzano, Italy &
TU Vienna, Austria
calvanese@inf.unibz.it

Vladislav Ryzhikov
Free U. of Bolzano
Italy
ryzhikov@inf.unibz.it

Abstract

Knowledge base exchange is an important problem in the area of data exchange and knowledge representation, where one is interested in exchanging information between a source and a target knowledge base connected through a mapping. In this paper, we study this fundamental problem for knowledge bases and mappings expressed in OWL 2 QL, the profile of OWL 2 based on the description logic *DL-Lite*$_\mathcal{R}$. More specifically, we consider the problem of computing universal solutions, identified as one of the most desirable translations to be materialized, and the problem of computing UCQ-representations, which optimally capture in a target TBox the information that can be extracted from a source TBox and a mapping by means of unions of conjunctive queries. For the former we provide a novel automata-theoretic technique, and complexity results that range from NP to EXPTIME, while for the latter we show NLOGSPACE-completeness.

1 Introduction

Complex forms of information, maintained in different formats and organized according to different structures, often need to be shared between agents. In recent years, both in the data management and in the knowledge representation communities, several settings have been investigated that address this problem from various perspectives: in *information integration*, uniform access is provided to a collection of data sources by means of an ontology (or global schema) to which the sources are mapped [Lenzerini, 2002]; in *peer-to-peer systems*, a set of peers declaratively linked to each other collectively provide access to the information assets they maintain [Kementsietsidis *et al.*, 2003; Adjiman *et al.*, 2006; Fuxman *et al.*, 2006]; in *ontology matching*, the aim is to understand and derive the correspondences between elements in two ontologies [Euzenat and Shvaiko, 2007; Shvaiko and Euzenat, 2013]; finally, in *data exchange*, the information stored according to a source schema needs to be restructured and translated so as to conform to a target schema [Fagin *et al.*, 2005; Barceló, 2009].

The work we present in this paper is inspired by the latter setting, investigated in databases. We study it, however, under the assumption of incomplete information typical of knowledge representation [Arenas *et al.*, 2011]. Specifically, we investigate the problem of *knowledge base exchange*, where a source knowledge base (KB) is connected to a target KB by means of a declarative mapping specification, and the aim is to exchange knowledge from the source to the target by exploiting the mapping. We rely on a framework for KB exchange based on lightweight Description Logics (DLs) of the *DL-Lite* family [Calvanese *et al.*, 2007], recently proposed in [Arenas *et al.*, 2012a; Arenas *et al.*, 2012b]: both source and target are KBs constituted by a DL TBox, representing implicit information, and an ABox, representing explicit information, and mappings are sets of DL concept and role inclusions. Note that in data and knowledge base exchange, differently from ontology matching, mappings are first-class citizens. In fact, it has been recognized that building schema mappings is an important and complex activity, which requires the designer to have a thorough understanding of the source and how the information therein should be related to the target. Thus, several techniques and tools have been developed to support mapping design, e.g., exploiting lexical information [Fagin *et al.*, 2009]. Here, similar to data exchange, we assume that for building mappings the target signature is given, but no further axioms constraining the target knowledge are available. In fact, such axioms are derived from the source KB and the mapping.

We consider two key problems: *(i)* computing *universal solutions*, which have been identified as one of the most desirable translations to be materialized; *(ii)* UCQ-*representability* of a source TBox by means of a target TBox that captures at best the intensional information that can be extracted from the source according to a mapping using union of conjunctive queries. Determining UCQ-representability is a crucial task, since it allows one to use the obtained target TBox to infer new knowledge in the target, thus reducing the amount of extensional information to be transferred from the source. Moreover, it has been noticed that in many data exchange applications users only extract information from the translated data by using specific queries (usually conjunctive queries), so query-based notions of translation specifically tailored to store enough information to answer such queries have been widely studied in the data exchange area [Madhavan and Halevy, 2003; Fagin *et al.*, 2008; Arenas *et al.*, 2009; Fagin and Kolaitis, 2012; Pichler *et al.*, 2013]. For these two problems, we investigate both the task of checking *membership*, where a candidate universal solution (resp., UCQ-representation) is given and one needs to check its correctness, and *non-emptiness*, where the aim is to determine the existence of a universal solution (resp., UCQ-representation).

We significantly extend previous results in several directions. First of all, we establish results for OWL 2 QL [Motik *et al.*, 2012], one of the profiles of the standard Web Ontology Language OWL 2 [Bao *et al.*, 2012], which is based on the DL *DL-Lite*$_\mathcal{R}$. To do so, we have to overcome the difficulty of dealing with null values in the ABox, since these become necessary in the target to represent universal solutions. Also, for the first time, we address disjointness assertions in the TBox, a construct that is part of OWL 2 QL. The main contribution of our work is then a detailed analysis of the computational complexity of both membership and non-emptiness for universal solutions and UCQ-representability. For the non-emptiness problem of universal solutions, previous known results covered only the simple case of *DL-Lite*$_{RDFS}$, the RDFS fragment of OWL 2 QL, in which no new facts can be inferred, and universal solutions always exist and can be computed in polynomial time via a chase procedure (see [Calvanese *et al.*, 2007]). We show that in our case, instead, the problem is PSPACE-hard, hence significantly more complex, and provide an EXPTIME upper bound based on a novel approach exploiting two-way alternating automata. We provide also NP upper bounds for the simpler case of ABoxes without null values, and for the case of the membership problem. As for UCQ-representability, we adopt the notion of UCQ-*representability* introduced in [Arenas *et al.*, 2012a; Arenas *et al.*, 2012b] and extend it to take into account disjointness of OWL 2 QL. For that case we show NLOGSPACE-completeness of both non-emptiness and membership, improving on the previously known PTIME upper bounds.

The paper is organized as follows. In Section 2, we give preliminary notions on DLs and queries. In Section 3, we define our framework of KB exchange and discuss the problem of computing solutions. In Section 4, we overview our contributions, and then we provide our results on computing universal solutions in Section 5, and on UCQ-representability in Section 6. Finally, in Section 7, we draw some conclusions and outline issues for future work.

The proofs are available in an extended technical report accessible at `http://arxiv.org/abs/1304.5810`.

2 Preliminaries

The DLs of the *DL-Lite* family [Calvanese *et al.*, 2007] of light-weight DLs are characterized by the fact that standard reasoning can be done in polynomial time. We adapt here *DL-Lite*$_\mathcal{R}$, the DL underlying OWL 2 QL, and present now its syntax and semantics. Let N_C, N_R, N_a, N_ℓ be pairwise disjoint sets of *concept names*, *role names*, *constants*, and *labeled nulls*, respectively. Assume in the following that $A \in N_C$ and $P \in N_R$; in *DL-Lite*$_\mathcal{R}$, B and C are used to denote basic and arbitrary (or complex) concepts, respectively, and R and Q are used to denote basic and arbitrary (or complex) roles, respectively, defined as follows:

$$\begin{array}{llll} R ::= P \mid P^- & & B ::= A \mid \exists R \\ Q ::= R \mid \neg R & & C ::= B \mid \neg B \end{array}$$

From now on, for a basic role R, we use R^- to denote P^- when $R = P$, and P when $R = P^-$.

A TBox is a finite set of *concept inclusions* $B \sqsubseteq C$ and *role inclusions* $R \sqsubseteq Q$. We call an inclusion of the form $B_1 \sqsubseteq \neg B_2$ or $R_1 \sqsubseteq \neg R_2$ a *disjointness assertion.* An ABox is a finite set of *membership assertions* $B(a)$, $R(a, b)$, where $a, b \in N_a$. In this paper, we also consider extended ABoxes, which are obtained by allowing labeled nulls in membership assertions. Formally, an *extended ABox* is a finite set of membership assertions $B(u)$ and $R(u, v)$, where $u, v \in (N_a \cup N_\ell)$. Moreover, a(n *extended*) *KB* $\mathcal{K}$ is a pair $\langle \mathcal{T}, \mathcal{A} \rangle$, where $\mathcal{T}$ is a TBox and $\mathcal{A}$ is an (extended) ABox.

A *signature* Σ is a finite set of concept and role names. A KB $\mathcal{K}$ is said to be *defined over* (or simply, *over*) Σ if all the concept and role names occurring in $\mathcal{K}$ belong to Σ (and likewise for TBoxes, ABoxes, concept inclusions, role inclusions and membership assertions). Moreover, an *interpretation* $\mathcal{I}$ *of* Σ is a pair $\langle \Delta^\mathcal{I}, \cdot^\mathcal{I} \rangle$, where $\Delta^\mathcal{I}$ is a non-empty domain and $\cdot^\mathcal{I}$ is an interpretation function such that: (1) $A^\mathcal{I} \subseteq \Delta^\mathcal{I}$, for every concept name $A \in \Sigma$; (2) $P^\mathcal{I} \subseteq \Delta^\mathcal{I} \times \Delta^\mathcal{I}$, for every role name $P \in \Sigma$; and (3) $a^\mathcal{I} \in \Delta^\mathcal{I}$, for every constant $a \in N_a$. Function $\cdot^\mathcal{I}$ is extended to also interpret concept and role constructs:

$$\begin{array}{l} (\exists R)^\mathcal{I} = \{x \in \Delta^\mathcal{I} \mid \exists y \in \Delta^\mathcal{I} \text{ such that } (x, y) \in R^\mathcal{I}\}; \\ (P^-)^\mathcal{I} = \{(y, x) \in \Delta^\mathcal{I} \times \Delta^\mathcal{I} \mid (x, y) \in P^\mathcal{I}\}; \\ (\neg B)^\mathcal{I} = \Delta^\mathcal{I} \setminus B^\mathcal{I}; \qquad (\neg R)^\mathcal{I} = (\Delta^\mathcal{I} \times \Delta^\mathcal{I}) \setminus R^\mathcal{I}. \end{array}$$

Note that, consistently with the semantics of OWL 2 QL, we do *not* make the unique name assumption (UNA), i.e., we allow distinct constants $a, b \in N_a$ to be interpreted as the same object, i.e., $a^\mathcal{I} = b^\mathcal{I}$. Note also that labeled nulls are *not* interpreted by $\mathcal{I}$.

Let $\mathcal{I} = \langle \Delta^\mathcal{I}, \cdot^\mathcal{I} \rangle$ be an interpretation over a signature Σ. Then $\mathcal{I}$ is said to satisfy a concept inclusion $B \sqsubseteq C$ over Σ, denoted by $\mathcal{I} \models B \sqsubseteq C$, if $B^\mathcal{I} \subseteq C^\mathcal{I}$; $\mathcal{I}$ is said to satisfy a role inclusion $R \sqsubseteq Q$ over Σ, denoted by $\mathcal{I} \models R \sqsubseteq Q$, if $R^\mathcal{I} \subseteq Q^\mathcal{I}$; and $\mathcal{I}$ is said to satisfy a TBox $\mathcal{T}$ over Σ, denoted by $\mathcal{I} \models \mathcal{T}$, if $\mathcal{I} \models \alpha$ for every $\alpha \in \mathcal{T}$. Moreover, satisfaction of membership assertions over Σ is defined as follows. A *substitution* over $\mathcal{I}$ is a function $h : (N_a \cup N_\ell) \to \Delta^\mathcal{I}$ such that $h(a) = a^\mathcal{I}$ for every $a \in N_a$. Then $\mathcal{I}$ is said to satisfy an (extended) ABox $\mathcal{A}$, denoted by $\mathcal{I} \models \mathcal{A}$, if there exists a substitution h over $\mathcal{I}$ such that:

– for every $B(u) \in \mathcal{A}$, it holds that $h(u) \in B^\mathcal{I}$; and

– for every $R(u, v) \in \mathcal{A}$, it holds that $(h(u), h(v)) \in R^\mathcal{I}$.

Finally, $\mathcal{I}$ is said to *satisfy* a(n extended) KB $\mathcal{K} = \langle \mathcal{T}, \mathcal{A} \rangle$, denoted by $\mathcal{I} \models \mathcal{K}$, if $\mathcal{I} \models \mathcal{T}$ and $\mathcal{I} \models \mathcal{A}$. Such $\mathcal{I}$ is called a *model* of $\mathcal{K}$, and we use $\textsc{Mod}(\mathcal{K})$ to denote the set of all models of $\mathcal{K}$. We say that $\mathcal{K}$ is *consistent* if $\textsc{Mod}(\mathcal{K}) \neq \emptyset$.

As is customary, given an (extended) KB $\mathcal{K}$ over a signature Σ and a membership assertion or an inclusion α over Σ, we use notation $\mathcal{K} \models \alpha$ to indicate that for every interpretation $\mathcal{I}$ of Σ, if $\mathcal{I} \models \mathcal{K}$, then $\mathcal{I} \models \alpha$.

2.1 Queries and certain answers

A k-ary query q over a signature Σ, with $k \geq 0$, is a function that maps every interpretation $\langle \Delta^\mathcal{I}, \cdot^\mathcal{I} \rangle$ of Σ into a k-ary relation $q^{\mathcal{I}} \subseteq (\Delta^{\mathcal{I}})^k$. In particular, if $k = 0$, then q is said to be a Boolean query, and $q^{\mathcal{I}}$ is either a relation containing the empty tuple $()$ (representing the value true) or the empty relation (representing the value false). Given a KB $\mathcal{K}$ over Σ, the set of *certain answers* to q over $\mathcal{K}$, denoted by $cert(q, \mathcal{K})$, is defined as:

$$\bigcap_{\mathcal{I}\in\textsc{Mod}(\mathcal{K})}\{(a_1,\ldots,a_k) \mid \{a_1,\ldots,a_k\}\subseteq N_a \text{ and } (a_1^{\mathcal{I}},\ldots,a_k^{\mathcal{I}})\in q^{\mathcal{I}}\},$$

Notice that the certain answer to a query does *not* contain labeled nulls. Besides, notice that if q is a Boolean query, then $cert(q, \mathcal{K})$ evaluates to true if $q^{\mathcal{I}}$ evaluates to true for every $\mathcal{I} \in \textsc{Mod}(\mathcal{K})$, and it evaluates to false otherwise.

A *conjunctive query* (CQ) *over a signature* Σ is a formula of the form $q(\vec{x}) = \exists\vec{y}.\, \varphi(\vec{x}, \vec{y})$, where $\vec{x}, \vec{y}$ are tuples of variables and $\varphi(\vec{x}, \vec{y})$ is a conjunction of atoms of the form $A(t)$, with A a concept name in Σ, and $P(t, t')$, with P a role name in Σ, where each of t, t' is either a constant from N_a or a variable from $\vec{x}$ or $\vec{y}$. Given an interpretation $\mathcal{I} = \langle\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}}\rangle$ of Σ, the answer of q over $\mathcal{I}$, denoted by $q^{\mathcal{I}}$, is the set of tuples $\vec{a}$ of elements from $\Delta^{\mathcal{I}}$ for which there exist a tuple $\vec{b}$ of elements from $\Delta^{\mathcal{I}}$ such that $\mathcal{I}$ satisfies every conjunct in $\varphi(\vec{a}, \vec{b})$. A union of conjunctive queries (UCQ) over a signature Σ is a formula of the form $q(\vec{x}) = \bigvee_{i=1}^{n} q_i(\vec{x})$, where each q_i $(1 \leq i \leq n)$ is a CQ over Σ, whose semantics is defined as $q^{\mathcal{I}} = \bigcup_{i=1}^{n} q_i^{\mathcal{I}}$.

3 Exchanging OWL 2 QL Knowledge Bases

We generalize now, in Section 3.1, the setting proposed in [Arenas *et al.*, 2011] to OWL 2 QL, and we formalize in Section 3.2 the main problems studied in the rest of the paper.

3.1 A knowledge base exchange framework for OWL 2 QL

Assume that Σ_1, Σ_2 are signatures with no concepts or roles in common. An inclusion $E_1 \sqsubseteq E_2$ is said to be *from* Σ_1 *to* Σ_2, if E_1 is a concept or a role over Σ_1 and E_2 is a concept or a role over Σ_2. A mapping is a tuple $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$, where $\mathcal{T}_{12}$ is a TBox consisting of inclusions from Σ_1 to Σ_2 [Arenas *et al.*, 2012a]. Recall that in this paper, we deal with *DL-Lite*$_{\mathcal{R}}$ TBoxes only, so $\mathcal{T}_{12}$ is assumed to be a set of *DL-Lite*$_{\mathcal{R}}$ concept and role inclusions. The semantics of such a mapping is defined in [Arenas *et al.*, 2012a] in terms of a notion of satisfaction for interpretations, which has to be extended in our case to deal with interpretations not satisfying the UNA (and, more generally, the standard name assumption). More specifically, given interpretations $\mathcal{I}$, $\mathcal{J}$ of Σ_1 and Σ_2, respectively, pair $(\mathcal{I}, \mathcal{J})$ *satisfies* TBox $\mathcal{T}_{12}$, denoted by $(\mathcal{I}, \mathcal{J}) \models \mathcal{T}_{12}$, if (*i*) for every $a \in N_a$, it holds that $a^{\mathcal{I}} = a^{\mathcal{J}}$, (*ii*) for every concept inclusion $B \sqsubseteq C \in \mathcal{T}_{12}$, it holds that $B^{\mathcal{I}} \subseteq C^{\mathcal{J}}$, and (*iii*) for every role inclusion $R \sqsubseteq Q \in \mathcal{T}_{12}$, it holds that $R^{\mathcal{I}} \subseteq Q^{\mathcal{J}}$. Notice that the connection between the information in $\mathcal{I}$ and $\mathcal{J}$ is established through the constants that move from source to target according to the mapping. For this reason, we require constants to be interpreted in the same way in $\mathcal{I}$ and $\mathcal{J}$, i.e., they preserve their meaning when they are transferred. Besides, notice that this is the only restriction imposed on the domains of $\mathcal{I}$ and $\mathcal{J}$ (in particular, we require neither that $\Delta^{\mathcal{I}} = \Delta^{\mathcal{J}}$ nor that $\Delta^{\mathcal{I}} \subseteq \Delta^{\mathcal{J}}$). Finally, $\textsc{Sat}_{\mathcal{M}}(\mathcal{I})$ is defined as the set of interpretations $\mathcal{J}$ of Σ_2 such that $(\mathcal{I}, \mathcal{J}) \models \mathcal{T}_{12}$, and given a set $\mathcal{X}$ of interpretations of Σ_1, $\textsc{Sat}_{\mathcal{M}}(\mathcal{X})$ is defined as $\bigcup_{\mathcal{I}\in\mathcal{X}} \textsc{Sat}_{\mathcal{M}}(\mathcal{I})$.

The main problem studied in the knowledge exchange area is the problem of translating a KB according to a mapping, which is formalized through several different notions of translation (for a thorough comparison of different notions of solutions see [Arenas *et al.*, 2012a]). The first such notion is the concept of solution, which is formalized as follows. Given a mapping $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$ and KBs $\mathcal{K}_1$, $\mathcal{K}_2$ over Σ_1 and Σ_2, respectively, $\mathcal{K}_2$ is a *solution* for $\mathcal{K}_1$ under $\mathcal{M}$ if $\textsc{Mod}(\mathcal{K}_2) \subseteq \textsc{Sat}_{\mathcal{M}}(\textsc{Mod}(\mathcal{K}_1))$. Thus, $\mathcal{K}_2$ is a solution for $\mathcal{K}_1$ under $\mathcal{M}$ if every interpretation of $\mathcal{K}_2$ is a valid translation of an interpretation of $\mathcal{K}_1$ according to $\mathcal{M}$. Although natural, this is a mild restriction, which gives rise to the stronger notion of universal solution. Given $\mathcal{M}$, $\mathcal{K}_1$ and $\mathcal{K}_2$ as before, $\mathcal{K}_2$ is a *universal solution* for $\mathcal{K}_1$ under $\mathcal{M}$ if $\textsc{Mod}(\mathcal{K}_2) = \textsc{Sat}_{\mathcal{M}}(\textsc{Mod}(\mathcal{K}_1))$. Thus, $\mathcal{K}_2$ is designed to exactly represent the space of interpretations obtained by translating the interpretations of $\mathcal{K}_1$ under $\mathcal{M}$ [Arenas *et al.*, 2012a]. Below is a simple example demonstrating the notion of universal solutions. This example also illustrates some issues regarding the absence of the UNA, which has to be given up to comply with the OWL 2 QL standard, and regarding the use of disjointness assertions.

Example 3.1 *Assume* $\mathcal{M} = (\{F(\cdot), G(\cdot)\}, \{F'(\cdot), G'(\cdot)\}, \mathcal{T}_{12})$*, where* $\mathcal{T}_{12} = \{F \sqsubseteq F', G \sqsubseteq G'\}$*, and let* $\mathcal{K}_1 = \langle\mathcal{T}_1, \mathcal{A}_1\rangle$*, where* $\mathcal{T}_1 = \{\}$ *and* $\mathcal{A}_1 = \{F(a), G(b)\}$*. Then the ABox* $\mathcal{A}_2 = \{F'(a), G'(b)\}$ *is a universal solution for* $\mathcal{K}_1$ *under* $\mathcal{M}$*.*

Now, if we add a seemingly harmless disjointness assertion $\{F \sqsubseteq \neg G\}$ *to* $\mathcal{T}_1$*, we obtain that* $\mathcal{A}_2$ *is no longer a universal solution (not even a solution) for* $\mathcal{K}_1$ *under* $\mathcal{M}$*. The reason for that is the lack of the UNA on the one hand, and the presence of the disjointness assertion in* $\mathcal{T}_1$ *on the other hand. In fact, the latter forces* a *and* b *to be interpreted differently in the source. Thus, for a model* $\mathcal{J}$ *of* $\mathcal{A}_2$ *such that* $a^{\mathcal{J}} = b^{\mathcal{J}}$ *and* $F'^{\mathcal{J}} = G'^{\mathcal{J}} = \{a^{\mathcal{J}}\}$*, there exists no model* $\mathcal{I}$ *of* $\mathcal{K}_1$ *such that* $(\mathcal{I}, \mathcal{J}) \models \mathcal{T}_{12}$ *(which would require* $a^{\mathcal{I}} = a^{\mathcal{J}}$ *and* $b^{\mathcal{I}} = b^{\mathcal{J}}$*). In general, there exists no universal solution for* $\mathcal{K}_1$ *under* $\mathcal{M}$*, even though* $\mathcal{K}_1$ *and* $\mathcal{T}_{12}$ *are consistent with each other.*

A second class of translations is obtained in [Arenas *et al.*, 2012a] by observing that solutions and universal solutions are too restrictive for some applications, in particular when one only needs a translation storing enough information to properly answer some queries. For the particular case of UCQ, this gives rise to the notions of UCQ-solution and universal UCQ-solution. Given a mapping $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$, a KB $\mathcal{K}_1 = \langle\mathcal{T}_1, \mathcal{A}_1\rangle$ over Σ_1 and a KB $\mathcal{K}_2$ over Σ_2, $\mathcal{K}_2$ is a UCQ-*solution* for $\mathcal{K}_1$ under $\mathcal{M}$ if for every query $q \in$ UCQ over Σ_2: $cert(q, \langle\mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1\rangle) \subseteq cert(q, \mathcal{K}_2)$, while $\mathcal{K}_2$ is a *universal* UCQ-*solution* for $\mathcal{K}_1$ under $\mathcal{M}$ if for every query $q \in$ UCQ over Σ_2: $cert(q, \langle\mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1\rangle) = cert(q, \mathcal{K}_2)$.

Finally, a last class of solutions is obtained in [Arenas *et al.*, 2012a] by considering that users want to translate as much of the knowledge in a TBox as possible, as a lot of effort is put in practice when constructing a TBox. This observation gives rise to the notion of UCQ-representation [Arenas *et al.*, 2012a], which formalizes the idea of translating a source TBox according to a mapping. Next, we present an alternative formalization of this notion, which is appropriate for our setting where disjointness assertions are considered.[1] Assume that $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$ and $\mathcal{T}_1$, $\mathcal{T}_2$ are TBoxes over Σ_1 and Σ_2, respectively. Then $\mathcal{T}_2$ is a UCQ-*representation* of $\mathcal{T}_1$ under $\mathcal{M}$ if for every query $q \in$ UCQ over Σ_2 and every ABox $\mathcal{A}_1$ over Σ_1 that is consistent with $\mathcal{T}_1$:

$$cert(q, \langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle) = \bigcap_{\substack{\mathcal{A}_2 \,:\, \mathcal{A}_2 \text{ is an ABox over } \Sigma_2 \text{ that} \\ \text{is a UCQ-solution for } \mathcal{A}_1 \text{ under } \mathcal{M}}} cert(q, \langle \mathcal{T}_2, \mathcal{A}_2 \rangle). \qquad (\dagger)$$

Notice that in the previous definition, $\mathcal{A}_2$ is said to be a UCQ-solution for $\mathcal{A}_1$ under $\mathcal{M}$ if the KB $\langle \emptyset, \mathcal{A}_2 \rangle$ is a UCQ-solution for the KB $\langle \emptyset, \mathcal{A}_1 \rangle$ under $\mathcal{M}$. Let us explain the intuition behind the definition of the notion of UCQ-representation. Assume that $\mathcal{T}_1$, $\mathcal{T}_2$, $\mathcal{M}$ satisfy (†). First, $\mathcal{T}_2$ captures the information in $\mathcal{T}_1$ that is translated by $\mathcal{M}$ and that can be extracted by using a UCQ, as for every ABox $\mathcal{A}_1$ over Σ_1 that is consistent with $\mathcal{T}_1$ and every UCQ q over Σ_2, if we choose an arbitrary UCQ-solution $\mathcal{A}_2$ for $\mathcal{A}_1$ under $\mathcal{M}$, then it holds that $cert(q, \langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle) \subseteq cert(q, \langle \mathcal{T}_2, \mathcal{A}_2 \rangle)$. Notice that $\mathcal{A}_1$ is required to be consistent with $\mathcal{T}_1$ in the previous condition, as we are interested in translating data that make sense according to $\mathcal{T}_1$. Second, $\mathcal{T}_2$ does not include any piece of information that can be extracted by using a UCQ and it is not the result of translating the information in $\mathcal{T}_1$ according to $\mathcal{M}$. In fact, if $\mathcal{A}_1$ is an ABox over Σ_1 that is consistent with $\mathcal{T}_1$ and q is a UCQ over Σ_2, then it could be the case that $cert(q, \langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle) \subsetneq cert(q, \langle \mathcal{T}_2, \mathcal{A}_2^\star \rangle)$ for some UCQ-solution $\mathcal{A}_2^\star$ for $\mathcal{A}_1$ under $\mathcal{M}$. However, the extra tuples extracted by query q are obtained from the extra information in $\mathcal{A}_2^\star$, as if we consider a tuple $\vec{a}$ that belong to $cert(q, \langle \mathcal{T}_2, \mathcal{A}_2 \rangle)$ for every UCQ-solution $\mathcal{A}_2$ for $\mathcal{A}_1$ under $\mathcal{M}$, then it holds that $\vec{a} \in cert(q, \langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle)$.

Example 3.2 *Assume that* $\mathcal{M} = (\{F(\cdot), G(\cdot), H(\cdot), D(\cdot)\}, \{F'(\cdot), G'(\cdot), H'(\cdot)\}, \mathcal{T}_{12})$*, where* $\mathcal{T}_{12} = \{F \sqsubseteq F', G \sqsubseteq G', H \sqsubseteq H'\}$*, and let* $\mathcal{T}_1 = \{F \sqsubseteq G\}$*. As expected, TBox* $\mathcal{T}_2 = \{F' \sqsubseteq G'\}$ *is a* UCQ*-representation of* $\mathcal{T}_1$ *under* $\mathcal{M}$*. Moreover, we can add the inclusion* $D \sqsubseteq \neg H'$ *to* $\mathcal{T}_{12}$*, and* $\mathcal{T}_2$ *will still remain a* UCQ*-representation of* $\mathcal{T}_1$ *under* $\mathcal{M}$*. Notice that in this latter setting, our definition has to deal with some ABoxes* $\mathcal{A}_1$ *that are consistent with* $\mathcal{T}_1$ *but not with* $\mathcal{T}_1 \cup \mathcal{T}_{12}$*, for instance* $\mathcal{A}_1 = \{H(a), D(a)\}$ *for some constant* a*. In those cases, Equation* (†) *is trivially satisfied, since* $\textsc{Mod}(\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle) = \emptyset$ *and the set of* UCQ*-solutions for* $\mathcal{A}_1$ *under* $\mathcal{M}$ *is empty.*

3.2 On the problem of computing solutions

Arguably, the most important problem in knowledge exchange [Arenas *et al.*, 2011; Arenas *et al.*, 2012a], as well as in data exchange [Fagin *et al.*, 2005; Kolaitis, 2005], is the task of computing a translation of a KB according to a mapping. To study the computational complexity of this task for the different notions of solutions presented in the previous section, we introduce the following decision problems. The *membership* problem for universal solutions (resp. universal UCQ-solutions) has as input a mapping $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$ and KBs $\mathcal{K}_1$, $\mathcal{K}_2$ over Σ_1 and Σ_2, respectively. Then the question to answer is whether $\mathcal{K}_2$ is a universal solution (resp. universal UCQ-solution) for $\mathcal{K}_1$ under $\mathcal{M}$. Moreover, the membership problem for UCQ-representations has as input a mapping $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$ and TBoxes $\mathcal{T}_1$, $\mathcal{T}_2$ over Σ_1 and Σ_2, respectively, and the question to answer is whether $\mathcal{T}_2$ is a UCQ-representation of $\mathcal{T}_1$ under $\mathcal{M}$.

In our study, we cannot leave aside the existential versions of the previous problems, which are directly related with the problem of computing translations of a KB according to a mapping. Formally, the *non-emptiness* problem for universal solutions (resp. universal UCQ-solutions) has as input a mapping $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$ and a KB $\mathcal{K}_1$ over Σ_1. Then the question to answer is whether there exists a universal solution (resp. universal UCQ-solution) for $\mathcal{K}_1$ under $\mathcal{M}$. Moreover, the non-emptiness problem for UCQ-representations has as input a mapping $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$ and a TBox $\mathcal{T}_1$ over Σ_1, and the question to answer is whether there exists a UCQ-representation of $\mathcal{T}_1$ under $\mathcal{M}$.

4 Our contributions

In Section 3.2, we have introduced the problems that are studied in this paper. It is important to notice that these problems are defined by considering only KBs (as opposed to extended KBs), as they are the formal counterpart of OWL 2 QL. Nevertheless, as shown in Section 5, there are natural examples of OWL 2 QL specifications and mappings where null values are needed when constructing solutions. Thus, we also study the problems defined in Section 3.2 in the case where translations can be extended KBs. It should be noticed that the notions of solution, universal solution, UCQ-solution, universal UCQ-solution, and UCQ-representation have to be enlarged to consider extended KBs, which is straightforward to do. In particular, given a mapping $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$ and TBoxes $\mathcal{T}_1$, $\mathcal{T}_2$ over Σ_1 and Σ_2, respectively, $\mathcal{T}_2$ is said to be a UCQ-representation of $\mathcal{T}_1$ under $\mathcal{M}$ in this extended setting if in Equation (†), $\mathcal{A}_2$ is an extended ABox over Σ_2 that is a UCQ-solution for $\mathcal{A}_1$ under $\mathcal{M}$.

The main contribution of this paper is to provide a detailed analysis of the complexity of the membership and non-emptiness problems for the notions of universal solution and UCQ-representation. In Figure 1, we provide a summary of the main results in the paper, which are explained in more detail in Sections 5 and 6. It is important to notice that these results considerably extend the previous known results about these problems [Arenas *et al.*, 2012a; Arenas *et al.*, 2012b]. In the first place, the problem of computing universal solutions was studied in [Arenas *et al.*, 2012a] for the case of *DL-Lite*$_{RDFS}$, a fragment of *DL-Lite*$_{\mathcal{R}}$ that allows neither for inclusions of the form $B \sqsubseteq \exists R$ nor for disjointness assertions. In that case, it is straightfor-

[1] If disjointness assertions are not allowed, then this new notion can be shown to be equivalent to the original formalization of UCQ-representation proposed in [Arenas *et al.*, 2012a].

Membership	ABoxes	extended ABoxes
Universal solutions	in NP	NP-complete
UCQ-representations	NLogSpace-complete	

Non-emptiness	ABoxes	extended ABoxes
Universal solutions	in NP	PSpace-hard, in ExpTime
UCQ-representations	NLogSpace-complete	

Figure 1: Complexity results obtained in the paper about the membership and non-emptiness problems.

ward to show that every source KB has a universal solution that can be computed by using the chase procedure [Calvanese *et al.*, 2007]. Unfortunately, this result does not provide any information about how to solve the much larger case considered in this paper, where, in particular, the non-emptiness problem is not trivial. In fact, for the case of the notion of universal solution, all the lower and upper bounds provided in Figure 1 are new results, which are not consequences of the results obtained in [Arenas *et al.*, 2012a]. In the second place, a notion of UCQ-representation that is appropriate for the fragment of *DL-Lite*$_\mathcal{R}$ not including disjointness assertions was studied in [Arenas *et al.*, 2012a; Arenas *et al.*, 2012b]. In particular, it was shown that the membership and non-emptiness problems for this notion are solvable in polynomial time. In this paper, we considerably strengthen these results: *(i)* by generalizing the definition of the notion of UCQ-representation to be able to deal with OWL 2 QL, that is, with the entire language *DL-Lite*$_\mathcal{R}$ (which includes disjointness assertions); and *(ii)* by showing that the membership and non-emptiness problems are both NLogSpace-complete in this larger scenario.

It turns out that reasoning about universal UCQ-solutions is much more intricate. In fact, as a second contribution of our paper, we provide a PSpace lower bound for the complexity of the membership problem for the notion of universal UCQ-solution, which is in sharp contrast with the NP and NLogSpace upper bounds for this problem for the case of universal solutions and UCQ-representations, respectively (see Figure 1). Although many questions about universal UCQ-solutions remain open, we think that this is an interesting first result, as universal UCQ-solutions have only been investigated before for the very restricted fragment *DL-Lite*$_{RDFS}$ of *DL-Lite*$_\mathcal{R}$ [Arenas *et al.*, 2012a], which is described in the previous paragraph.

5 Computing universal solutions

In this section, we study the membership and non-emptiness problems for universal solutions, in the cases where nulls are not allowed (Section 5.1) and are allowed (Section 5.2) in such solutions. But before going into this, we give an example that shows the shape of universal solutions in *DL-Lite*$_\mathcal{R}$.

Example 5.1 *Assume that* $\mathcal{M} = (\{F(\cdot), S(\cdot,\cdot)\}, \{G'(\cdot)\}, \{\exists S^- \sqsubseteq G'\})$*, and let* $\mathcal{K}_1 = \langle \mathcal{T}_1, \mathcal{A}_1 \rangle$*, where* $\mathcal{T}_1 = \{F \sqsubseteq \exists S\}$ *and* $\mathcal{A}_1 = \{F(a)\}$*. Then a natural way to construct a universal solution for* $\mathcal{K}_1$ *under* $\mathcal{M}$ *is to 'populate' the target with all implied facts (as it is usually done in data exchange). Thus, the ABox* $\mathcal{A}_2 = \{G'(n)\}$*, where* n *is a labeled null, is a universal solution for* $\mathcal{K}_1$ *under* $\mathcal{M}$ *if nulls are allowed. Notice that here, a universal solution with non-extended ABoxes does not exist: substituting* n *by any constant is too restrictive, ruining universality.*

Example 5.2 *Now, assume* $\mathcal{M} = (\{F(\cdot), S(\cdot,\cdot), T(\cdot,\cdot)\}, \{S'(\cdot,\cdot)\}, \{S \sqsubseteq S', T \sqsubseteq S'\})$*, and* $\mathcal{K}_1 = \langle \mathcal{T}_1, \mathcal{A}_1 \rangle$*, where* $\mathcal{T}_1 = \{F \sqsubseteq \exists S, \exists S^- \sqsubseteq \exists S\}$ *and* $\mathcal{A}_1 = \{F(a), T(a,a)\}$*. In this case, we cannot use the same approach as in Example 5.1 to construct a universal solution, as now we would need of an infinite number of labeled nulls to construct such a solution. However, as* S *and* T *are transferred to the same role* S'*, it is possible to use constant* a *to represent all implied facts. In particular, in this case* $\mathcal{A}_2 = \{S'(a,a)\}$ *is a universal solution for* $\mathcal{K}_1$ *under* $\mathcal{M}$*.*

5.1 Universal solutions without null values

We explain here how the NP upper bound for the non-emptiness problem for universal solutions is obtained, when ABoxes are not allowed to contain null values.

Assume given a mapping $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$ and a KB $\mathcal{K}_1 = \langle \mathcal{T}_1, \mathcal{A}_1 \rangle$ over Σ_1. To check whether $\mathcal{K}_1$ has a universal solution under $\mathcal{M}$, we use the following non-deterministic polynomial-time algorithm. First, we construct an ABox $\mathcal{A}_2$ over Σ_2 containing every membership assertion α such that $\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle \models \alpha$, where α is of the form either $B(a)$ or $R(a,b)$, and a, b are constants mentioned in $\mathcal{A}_1$. Second, we guess an interpretation $\mathcal{I}$ of Σ_1 such that $\mathcal{I} \models \mathcal{K}_1$ and $(\mathcal{I}, \mathcal{U}_{\mathcal{A}_2}) \models \mathcal{T}_{12}$, where $\mathcal{U}_{\mathcal{A}_2}$ is the interpretation of Σ_2 naturally corresponding[2] to $\mathcal{A}_2$. The correctness of the algorithm is a consequence of the facts that:

a) there exists a universal solution for $\mathcal{A}_1$ under $\mathcal{M}$ if and only if $\mathcal{A}_2$ is a solution for $\mathcal{A}_1$ under $\mathcal{M}$; and

b) $\mathcal{A}_2$ is a solution for $\mathcal{A}_1$ under $\mathcal{M}$ if and only if there exists a model $\mathcal{I}$ of $\mathcal{K}_1$ such that $(\mathcal{I}, \mathcal{U}_{\mathcal{A}_2}) \models \mathcal{T}_{12}$.

Moreover, the algorithm can be implemented in a non-deterministic polynomial-time Turing machine given that: *(i)* $\mathcal{A}_2$ can be constructed in polynomial time; *(ii)* if there exists a model $\mathcal{I}$ of $\mathcal{K}_1$ such that $(\mathcal{I}, \mathcal{U}_{\mathcal{A}_2}) \models \mathcal{T}_{12}$, then there exists a model of $\mathcal{K}_1$ of polynomial-size satisfying this condition; and *(iii)* it can be checked in polynomial time whether $\mathcal{I} \models \mathcal{K}_1$ and $(\mathcal{I}, \mathcal{U}_{\mathcal{A}_2}) \models \mathcal{T}_{12}$.

In addition, in this case, the membership problem can be reduced to the non-emptiness problem, thus, we have that:

Theorem 5.3 *The non-emptiness and membership problems for universal solutions are in* NP.

The exact complexity of these problems remains open. In fact, we conjecture that these problems are in PTime.

We conclude by showing that reasoning about universal UCQ-solutions is harder than reasoning about universal solutions, which can be explained by the fact that TBoxes have

[2] Interpretation $\mathcal{U}_{\mathcal{A}_2}$ can be defined as the Herbrand model of $\mathcal{A}_2$ extended with fresh domain elements to satisfy assertions of the form $\exists R(a)$ in $\mathcal{A}_2$.

bigger impact on the structure of universal UCQ-solutions rather than of universal solutions. In fact, by using a reduction from the validity problem for quantified Boolean formulas, similar to a reduction in [Konev *et al.*, 2011], we are able to prove the following:

Theorem 5.4 *The membership problem for universal* UCQ-*solutions is* PSPACE-*hard.*

5.2 Universal solutions with null values

We start by considering the non-emptiness problem for universal solutions with null values, that is, when extended ABoxes are allowed in universal solutions. As our first result, similar to the reduction above, we show that this problem is PSPACE-hard, and identify the inclusion of inverse roles as one of the main sources of complexity.

To obtain an upper bound for this problem, we use *two-way alternating automata on infinite trees (2ATA)*, which are a generalization of nondeterministic automata on infinite trees [Vardi, 1998] well suited for handling inverse roles in *DL-Lite*$_{\mathcal{R}}$. More precisely, given a KB $\mathcal{K}$, we first show that it is possible to construct the following automata:

- $\mathbb{A}_{\mathcal{K}}^{can}$ is a 2ATA that accepts trees corresponding to the canonical model of $\mathcal{K}$ [3] with nodes arbitrary labeled with a special symbol G;
- $\mathbb{A}_{\mathcal{K}}^{mod}$ is a 2ATA that accepts a tree if its subtree labeled with G corresponds to a tree model $\mathcal{I}$ of $\mathcal{K}$ (that is, a model forming a tree on the labeled nulls); and
- $\mathbb{A}_{\mathit{fin}}$ is a (one-way) non-deterministic automaton that accepts a tree if it has a finite prefix where each node is marked with G, and no other node in the tree is marked with G.

Then to verify whether a KB $\mathcal{K}_1 = \langle \mathcal{T}_1, \mathcal{A}_1 \rangle$ has a universal solution under a mapping $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$, we solve the non-emptiness problem for an automaton $\mathbb{B}$ defined as the product automaton of $\pi_{\Gamma_{\mathcal{K}}}(\mathbb{A}_{\mathcal{K}}^{can})$, $\pi_{\Gamma_{\mathcal{K}}}(\mathbb{A}_{\mathcal{K}}^{mod})$ and $\mathbb{A}_{\mathit{fin}}$, where $\mathcal{K} = \langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle$, $\pi_{\Gamma_{\mathcal{K}}}(\mathbb{A}_{\mathcal{K}}^{can})$ is the projection of $\mathbb{A}_{\mathcal{K}}^{can}$ on a vocabulary $\Gamma_{\mathcal{K}}$ not mentioning symbols from Σ_1, and likewise for $\pi_{\Gamma_{\mathcal{K}}}(\mathbb{A}_{\mathcal{K}}^{mod})$. If the language accepted by $\mathbb{B}$ is empty, then there is no universal solution for $\mathcal{K}_1$ under $\mathcal{M}$, otherwise a universal solution (possibly of exponential size) exists, and we can compute it by extracting the ABox encoded in some tree accepted by $\mathbb{B}$. Summing up, we get:

Theorem 5.5 *If extended ABoxes are allowed in universal solutions, then the non-emptiness problem for universal solutions is* PSPACE-*hard and in* EXPTIME.

Interestingly, the membership problem can be solved more efficiently in this scenario, as now the candidate universal solutions are part of the input. In the following theorem, we pinpoint the exact complexity of this problem.

Theorem 5.6 *If extended ABoxes are allowed in universal solutions, then the membership problem for universal solutions is* NP-*complete.*

6 Computing UCQ-representations

In Section 5, we show that the complexity of the membership and non-emptiness problems for universal solutions differ depending on whether ABoxes or extended ABoxes are considered. On the other hand, we show in the following proposition that the use of null values in ABoxes does not make any difference in the case of UCQ-representations. In this proposition, given a mapping $\mathcal{M}$ and TBoxes $\mathcal{T}_1$, $\mathcal{T}_2$, we say that $\mathcal{T}_2$ is a UCQ-representation of $\mathcal{T}_1$ under $\mathcal{M}$ *considering extended ABoxes* if $\mathcal{T}_1$, $\mathcal{T}_2$, $\mathcal{M}$ satisfy Equation (†) in Section 3.1, but assuming that $\mathcal{A}_2$ is an extended ABox over Σ_2 that is a UCQ-solution for $\mathcal{A}_1$ under $\mathcal{M}$.

Proposition 6.1 *A TBox $\mathcal{T}_2$ is a* UCQ-*representation of a TBox $\mathcal{T}_1$ under a mapping $\mathcal{M}$ if and only if $\mathcal{T}_2$ is a* UCQ-*representation of $\mathcal{T}_1$ under $\mathcal{M}$ considering extended ABoxes.*

Thus, from now on we study the membership and non-emptiness problems for UCQ-representations assuming that ABoxes can contain null values.

We start by considering the membership problem for UCQ-representations. In this case, one can immediately notice some similarities between this task and the membership problem for universal UCQ-solutions, which was shown to be PSPACE-hard in Theorem 5.4. However, the universal quantification over ABoxes in the definition of the notion of UCQ-representation makes the latter problem computationally simpler, which is illustrated by the following example.

Example 6.2 *Assume that $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$, where $\Sigma_1 = \{F(\cdot), S_1(\cdot,\cdot), S_2(\cdot,\cdot), T_1(\cdot,\cdot), T_2(\cdot,\cdot)\}$, $\Sigma_2 = \{F'(\cdot), S'(\cdot,\cdot), T'(\cdot,\cdot), G'(\cdot)\}$ and $\mathcal{T}_{12} = \{F \sqsubseteq F', S_1 \sqsubseteq S',\ S_2 \sqsubseteq S', T_1 \sqsubseteq T', T_2 \sqsubseteq T', \exists T_1^- \sqsubseteq G'\}$. Moreover, assume that $\mathcal{T}_1 = \{F \sqsubseteq \exists S_1, F \sqsubseteq \exists S_2, \exists S_1^- \sqsubseteq \exists T_1, \exists S_2^- \sqsubseteq \exists T_2\}$ and $\mathcal{T}_2 = \{F' \sqsubseteq \exists S', \exists S'^- \sqsubseteq \exists T', \exists T'^- \sqsubseteq G'\}$. If we were to verify whether $\langle \mathcal{T}_2, \{F'(a)\} \rangle$ is a universal* UCQ-*solution for $\langle \mathcal{T}_1, \{F(a)\} \rangle$ under $\mathcal{M}$ (which it is in this case), then we would first need to construct the* path *$\pi = \langle F'(a), S'(a,n), T'(n,m), G'(m) \rangle$ formed by the inclusions in $\mathcal{T}_2$, where n, m are fresh null values, and then we would need to explore the translations according to $\mathcal{M}$ of all paths formed by the inclusions in $\mathcal{T}_1$ to find one that matches π.*

On the other hand, to verify whether $\mathcal{T}_2$ is a UCQ-*representation of $\mathcal{T}_1$ under $\mathcal{M}$, one does not need to execute any "backtracking", as it is sufficient to consider independently a polynomial number of pieces $\mathcal{C}$ taken from the paths formed by the inclusions in $\mathcal{T}_1$, each of them of polynomial size, and then checking whether the translation $\mathcal{C}'$ of $\mathcal{C}$ according to $\mathcal{M}$ matches with the paths formed from $\mathcal{C}'$ by the inclusions in $\mathcal{T}_2$. If any of these pieces does not satisfy this condition, then it can be transformed into a witness that Equation (†) is not satisfied, showing that $\mathcal{T}_2$ is not a* UCQ-*representation of $\mathcal{T}_1$ under $\mathcal{M}$ (as we have a universal quantification over the ABoxes over Σ_1 in the definition of* UCQ-*representations). In fact, one of the pieces considered in this case is $\mathcal{C} = \langle T_2(n,m) \rangle$, where n, m are null values, which does not satisfy the previous condition as the translation $\mathcal{C}'$ of $\mathcal{C}$ according to $\mathcal{M}$ is $\langle T'(n,m) \rangle$, and this does not match with the path $\langle T'(n,m), G'(m) \rangle$ formed from $\mathcal{C}'$ by the inclusions in $\mathcal{T}_2$. This particular case is transformed into an ABox*

[3] If $\mathcal{K} = \langle \mathcal{T}, \mathcal{A} \rangle$, then this model essentially corresponds to the chase of $\mathcal{A}$ with $\mathcal{T}$ (see [Konev *et al.*, 2011] for a formal definition).

$\mathcal{A}_1 = \{T_2(b,c)\}$ and a query $q = T'(b,c) \wedge G'(c)$, where b, c are fresh constants, for which we have that Equation (†) is not satisfied.

Notice that disjointness assertions in the mapping may cause $\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle$ to become inconsistent for some source ABoxes $\mathcal{A}_1$ (which will make all possible tuples to be in the answer to every query), therefore additional conditions have to be imposed on $\mathcal{T}_2$. To give more intuition about how the membership problem for UCQ-representations is solved, we give an example showing how one can deal with some of these inconsistency issues.

Example 6.3 *Assume that $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$, where $\Sigma_1 = \{F(\cdot), G(\cdot), H(\cdot)\}$, $\Sigma_2 = \{F'(\cdot), G'(\cdot), H'(\cdot)\}$ and $\mathcal{T}_{12} = \{F \sqsubseteq F', G \sqsubseteq G', H \sqsubseteq H'\}$. Moreover, assume that $\mathcal{T}_1 = \{F \sqsubseteq G\}$ and $\mathcal{T}_2 = \{F' \sqsubseteq G'\}$. In this case, it is clear that $\mathcal{T}_2$ is a* UCQ*-representation of $\mathcal{T}_1$ under $\mathcal{M}$. However, if we add inclusion $H \sqsubseteq \neg G'$ to $\mathcal{T}_{12}$, then $\mathcal{T}_2$ is no longer a* UCQ*-representation of $\mathcal{T}_1$ under $\mathcal{M}$. To see why this is the case, consider an ABox $\mathcal{A}_1 = \{F(a), H(a)\}$, which is consistent with $\mathcal{T}_1$, and a query $q = F'(b)$, where b is a fresh constant. Then we have that $cert(q, \langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle) = \{()\}$ as KB $\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle$ is inconsistent, while $cert(q, \langle \mathcal{T}_2, \mathcal{A}_2 \rangle) = \emptyset$ for* UCQ*-solution $\mathcal{A}_2 = \{F'(a), H'(a)\}$ for $\mathcal{A}_1$ under $\mathcal{M}$. Thus, we conclude that Equation (†) is violated in this case.*

One can deal with the issue raised in the previous example by checking that on every pair (B, B') of $\mathcal{T}_1$-consistent basic concepts over Σ_1,[4] it holds that: (B, B') is $(\mathcal{T}_1 \cup \mathcal{T}_{12})$-consistent if and only if (B, B') is $(\mathcal{T}_{12} \cup \mathcal{T}_2)$-consistent, and likewise for every pair of basic roles over Σ_1. This condition guarantees that for every ABox $\mathcal{A}_1$ over Σ_1 that is consistent with $\mathcal{T}_1$, it holds that: $\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle$ is consistent if and only if there exists an extended ABox $\mathcal{A}_2$ over Σ_2 such that $\mathcal{A}_2$ is a UCQ-solution for $\mathcal{A}_1$ under $\mathcal{M}$ and $\langle \mathcal{T}_2, \mathcal{A}_2 \rangle$ is consistent. Thus, the previous condition ensures that the sets on the left- and right-hand side of Equation (†) coincide whenever the intersection on either of these sides is taken over an empty set.

The following theorem, which requires of a lengthy and non-trivial proof, shows that there exists an efficient algorithm for the membership problem for UCQ-representations that can deal with all the aforementioned issues.

Theorem 6.4 *The membership problem for* UCQ*-representations is* NLOGSPACE*-complete.*

We conclude by pointing out that the non-emptiness problem for UCQ-representations can also be solved efficiently. We give an intuition of how this can be done in the following example, where we say that $\mathcal{T}_1$ is UCQ*-representable under* $\mathcal{M}$ if there exists a UCQ-representation $\mathcal{T}_2$ of $\mathcal{T}_1$ under $\mathcal{M}$.

Example 6.5 *Assume that $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$, where $\Sigma_1 = \{F(\cdot), G(\cdot), H(\cdot)\}$, $\Sigma_2 = \{F'(\cdot), G'(\cdot)\}$ and $\mathcal{T}_{12} = \{F \sqsubseteq F', G \sqsubseteq G', H \sqsubseteq F'\}$. Moreover, assume that $\mathcal{T}_1 = \{F \sqsubseteq G\}$. Then it follows that $\mathcal{T}_1 \cup \mathcal{T}_{12} \models F \sqsubseteq G'$, and in order for $\mathcal{T}_1$ to be* UCQ*-representable under $\mathcal{M}$, the following condition must be satisfied:*

($\star$) *there exists a concept B' over Σ_2 s.t. $\mathcal{T}_{12} \models F \sqsubseteq B'$, and for each concept B over Σ_1 with $\mathcal{T}_1 \cup \mathcal{T}_{12} \models B \sqsubseteq B'$ it follows that $\mathcal{T}_1 \cup \mathcal{T}_{12} \models B \sqsubseteq G'$.*

The idea is then to add the inclusion $B' \sqsubseteq G'$ to a UCQ*-representation $\mathcal{T}_2$ so that $\mathcal{T}_{12} \cup \mathcal{T}_2 \models F \sqsubseteq G'$ as well. In our case, concept F' satisfies the condition $\mathcal{T}_{12} \models F \sqsubseteq F'$, but it does not satisfy the second requirement as $\mathcal{T}_1 \cup \mathcal{T}_{12} \models H \sqsubseteq F'$ and $\mathcal{T}_1 \cup \mathcal{T}_{12} \not\models H \sqsubseteq G'$. In fact, $F' \sqsubseteq G'$ cannot be added to $\mathcal{T}_2$ as it would result in $\mathcal{T}_{12} \cup \mathcal{T}_2 \models H \sqsubseteq G'$, hence in Equation (†), the inclusion from right to left would be violated. There is no way to reflect the inclusion $F \sqsubseteq G'$ in the target, so in this case $\mathcal{T}_1$ is not* UCQ*-representable under $\mathcal{M}$.*

The proof of the following result requires of some involved extensions of the techniques used to prove Theorem 6.4.

Theorem 6.6 *The non-emptiness problem for* UCQ*-representations is* NLOGSPACE*-complete.*

The techniques used to prove Theorem 6.6, which is sketched in the example below.

Example 6.7 *Consider $\mathcal{M}$ and $\mathcal{T}_1$ from Example 6.5, but assuming that $\mathcal{T}_{12}$ does not contain the inclusion $H \sqsubseteq F'$. Again, $\mathcal{T}_1 \cup \mathcal{T}_{12} \models F \sqsubseteq G'$, but now condition ($\star$) is satisfied. Then, an algorithm for computing a representation essentially needs to take any B' given by condition ($\star$) and add the inclusion $B' \sqsubseteq F'$ to $\mathcal{T}_2$. In this case, $\mathcal{T}_2 = \{F' \sqsubseteq G'\}$ is a* UCQ*-representation of $\mathcal{T}_1$ under $\mathcal{M}$.*

7 Conclusions

In this paper, we have studied the problem of KB exchange for OWL 2 QL, improving on previously known results with respect to both the expressiveness of the ontology language and the understanding of the computational properties of the problem. Our investigation leaves open several issues, which we intend to address in the future. First, it would be good to have characterizations of classes of source KBs and mappings for which universal (UCQ-)solutions are guaranteed to exist. As for the computation of universal solutions, while we have pinned-down the complexity of membership for extended ABoxes as NP-complete, an exact bound for the other case is still missing. Moreover, it is easy to see that allowing for inequalities between terms (e.g., $a \neq b$ in Example 3.1) and for negated atoms in the (target) ABox would allow one to obtain more universal solutions, but a full understanding of this case is still missing. Finally, we intend to investigate the challenging problem of computing universal UCQ-solutions, adopting also here an automata-based approach.

8 Acknowledgements

This research has been partially supported by the EU IP project Optique (grant FP7-318338), by EU Marie Curie action FP7-PEOPLE-2009-IRSES (grant 24761), by ERC grant agreement DIADEM, no. 246858, and by Fondecyt grant 1131049. Diego Calvanese has been partially supported by the Wolfgang Pauli Institute Vienna. The authors are also grateful to Evgeny Sherkhonov and Roman Kontchakov for helpful comments and discussions.

[4] A pair $(B, B)'$ is $\mathcal{T}$-consistent for a TBox $\mathcal{T}$, if the KB $\langle \mathcal{T}, \{B(a), B'(a)\} \rangle$ is consistent, where a is an arbitrary constant.

References

[Adjiman *et al.*, 2006] Philippe Adjiman, Philippe Chatalic, François Goasdoué, Marie-Christine Rousset, and Laurent Simon. Distributed reasoning in a peer-to-peer setting: Application to the Semantic Web. *J. of Artificial Intelligence Research*, 25:269–314, 2006.

[Arenas *et al.*, 2009] Marcelo Arenas, Jorge Pérez, Juan L. Reutter, and Cristian Riveros. Composition and inversion of schema mappings. *SIGMOD Record*, 38(3):17–28, 2009.

[Arenas *et al.*, 2011] Marcelo Arenas, Jorge Pérez, and Juan L. Reutter. Data exchange beyond complete data. In *PODS 2011*, pages 83–94, 2011.

[Arenas *et al.*, 2012a] Marcelo Arenas, Elena Botoeva, Diego Calvanese, Vladislav Ryzhikov, and Evgeny Sherkhonov. Exchanging description logic knowledge bases. In *KR 2012*, 2012.

[Arenas *et al.*, 2012b] Marcelo Arenas, Elena Botoeva, Diego Calvanese, Vladislav Ryzhikov, and Evgeny Sherkhonov. Representability in *DL-Lite*$_{\mathcal{R}}$ knowledge base exchange. In *Proc. of the 25th Int. Workshop on Description Logic (DL 2012)*, volume 846 of *CEUR Electronic Workshop Proceedings,* `http://ceur-ws.org/`, 2012.

[Artale *et al.*, 2009] Alessandro Artale, Diego Calvanese, Roman Kontchakov, and Michael Zakharyaschev. The *DL-Lite* family and relations. *J. of Artificial Intelligence Research*, 36:1–69, 2009.

[Bao *et al.*, 2012] Jie Bao et al. OWL 2 Web Ontology Language document overview (second edition). W3C Recommendation, World Wide Web Consortium, December 2012. `http://www.w3.org/TR/owl2-overview/`.

[Barceló, 2009] Pablo Barceló. Logical foundations of relational data exchange. *SIGMOD Record*, 38(1):49–58, 2009.

[Calvanese *et al.*, 2007] Diego Calvanese, Giuseppe De Giacomo, Domenico Lembo, Maurizio Lenzerini, and Riccardo Rosati. Tractable reasoning and efficient query answering in description logics: The *DL-Lite* family. *J. of Automated Reasoning*, 39(3):385–429, 2007.

[Euzenat and Shvaiko, 2007] Jérôme Euzenat and Pavel Shvaiko. *Ontology Matching*. Springer, 2007.

[Fagin and Kolaitis, 2012] Ronald Fagin and Phokion G. Kolaitis. Local transformations and conjunctive-query equivalence. In *PODS 2012*, pages 179–190, 2012.

[Fagin *et al.*, 2005] Ronald Fagin, Phokion G. Kolaitis, Renée J. Miller, and Lucian Popa. Data exchange: Semantics and query answering. *Theoretical Computer Science*, 336(1):89–124, 2005.

[Fagin *et al.*, 2008] Ronald Fagin, Phokion G. Kolaitis, Alan Nash, and Lucian Popa. Towards a theory of schema-mapping optimization. In *PODS 2008*, pages 33–42, 2008.

[Fagin *et al.*, 2009] Ronald Fagin, Laura M. Haas, Mauricio A. Hernández, Renée J. Miller, Lucian Popa, and Yannis Velegrakis. Clio: Schema mapping creation and data exchange. In *Conceptual Modeling: Foundations and Applications – Essays in Honor of John Mylopoulos*, volume 5600 of *Lecture Notes in Computer Science*, pages 198–236, 2009.

[Fuxman *et al.*, 2006] Ariel Fuxman, Phokion G. Kolaitis, Renée J. Miller, and Wang Chiew Tan. Peer data exchange. *ACM Trans. on Database Systems*, 31(4):1454–1498, 2006.

[Kementsietsidis *et al.*, 2003] Anastasios Kementsietsidis, Marcelo Arenas, and Renée J. Miller. Mapping data in peer-to-peer systems: Semantics and algorithmic issues. In *Proc. of the ACM SIGMOD Int. Conf. on Management of Data*, pages 325–336, 2003.

[Kolaitis, 2005] Phokion G. Kolaitis. Schema mappings, data exchange, and metadata management. In *PODS 2005*, pages 61–75, 2005.

[Konev *et al.*, 2011] Boris Konev, Roman Kontchakov, Michel Ludwig, Thomas Schneider, Frank Wolter, and Michael Zakharyaschev. Conjunctive query inseparability of OWL 2 QL TBoxes. In *Proc. of the 25th AAAI Conference on Artificial Intelligence, (AAAI 2011)*, 2011.

[Lenzerini, 2002] Maurizio Lenzerini. Data integration: A theoretical perspective. In *Proc. of the 21st ACM SIGACT SIGMOD SIGART Symp. on Principles of Database Systems (PODS 2002)*, pages 233–246, 2002.

[Madhavan and Halevy, 2003] Jayant Madhavan and Alon Y. Halevy. Composing mappings among data sources. In *VLDB 2003*, pages 572–583, 2003.

[Motik *et al.*, 2012] Boris Motik, Bernardo Cuenca Grau, Ian Horrocks, Zhe Wu, Achille Fokoue, and Carsten Lutz. OWL 2 Web Ontology Language profiles (second edition). W3C Recommendation, World Wide Web Consortium, December 2012. `http://www.w3.org/TR/owl2-profiles/`.

[Pichler *et al.*, 2013] Reinhard Pichler, Emanuel Sallinger, and Vadim Savenkov. Relaxed notions of schema mapping equivalence revisited. *Theory of Computing Systems*, 52(3):483–541, 2013.

[Shvaiko and Euzenat, 2013] Pavel Shvaiko and Jérôme Euzenat. Ontology matching: State of the art and future challenges. *IEEE Trans. on Knowledge and Data Engineering*, 25(1):158–176, 2013.

[Vardi, 1998] Moshe Y. Vardi. Reasoning about the past with two-way automata. In *Proc. of the 25th Int. Coll. on Automata, Languages and Programming (ICALP'98)*, volume 1443 of *Lecture Notes in Computer Science*, pages 628–641. Springer, 1998.