Computing Solutions in OWL 2 QL Knowledge Base Exchange

M. Arenas[1], E. Botoeva[2], D. Calvanese[2], and V. Ryzhikov[2]

[1] Dept. of Computer Science, PUC Chile & University of Oxford
marenas@ing.puc.cl
[2] KRDB Research Centre, Free Univ. of Bozen-Bolzano, Italy
lastname@inf.unibz.it

Abstract. The problem of exchanging knowledge bases from a source signature to a target signature connected through a mapping has recently attracted attention in knowledge representation. In this paper, we study this problem for knowledge bases and mappings expressed in OWL 2 QL, one of the profiles of the standard Web Ontology Language OWL 2. More specifically, we consider the membership and non-emptiness problems associated with computing universal solutions, which have been identified as one of the most desirable translations to be materialized. We study two settings: when ABoxes are in OWL 2 QL and when null values are allowed in the ABox language. For the former case, we provide a novel technique based on reachability games on graphs to show that the non-emptiness and membership problems are in PTime. For the latter case, we report a range of complexity results from NP to EXPTIME. We also consider the problem of computing universal UCQ-solutions, which provide an alternative notion of translation containing sufficient information to properly answer union of conjunctive queries, reporting a PSPACE lower bound for the membership problem.

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 [17]; in peer-to-peer systems, a set of peers declaratively linked to each other collectively provide access to the information assets they maintain [14,1,13]; in ontology matching, the aim is to understand and derive the correspondences between elements in two ontologies [11,20]; 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 [12,8].

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 [5]. 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 proposed recently in [2,3,4], based on lightweight Description Logics (DLs) of the *DL-Lite* family [9]: 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.

In this paper, we adjust the above mentioned framework to OWL 2 QL [19], one of the profiles of the standard Web Ontology Language OWL 2 [7], and then study the problem of computing *universal solutions*, which have been identified as one of the most desirable translations to be materialized. We investigate both the task of checking *membership*, where a candidate universal solution is given and one needs to check its correctness, and *non-emptiness*, where the aim is to determine the existence of a universal solution. We prove that both problems can be solved in PTIME using a novel reduction to the problem of finding a *winning strategy* in reachability games on graphs [18].

Then, we argue that for certain natural shapes of source KBs and mappings the universal solutions do not exist, unless null values are allowed in ABox languages. So, we consider *extended* ABoxes that may contain nulls, presenting a number of complexity results ranging from NP to EXPTIME for this setting.

Finally, we consider *universal* UCQ-*solutions*, an alternative notion for materialization of knowledge in the target that contains sufficient information to answer unions of conjunctive queries. We show that universal UCQ-solutions exist for certain source KBs and mappings, where universal solutions (even with extended ABoxes) do not. We also report PSPACE-hardness for the membership problem for universal UCQ-solutions.

The paper is organized as follows. We give preliminary notions on DLs and queries in Section 2, and on KB exchange in Section 3. In Section 4, we present the known results and give some intuition about the shape of universal solutions. In Sections 5 and 6, we present the results on the complexity of computing universal solutions for KBs and extended KBs, respectively. In Section 7, we consider universal UCQ-solutions and in Section 8, we present conclusions and outline some future work.

2 Preliminaries

The DLs of the *DL-Lite* family [6] are characterized by the fact that standard reasoning can be done in polynomial time. We adapt here *DL-Lite*$_\mathcal{R}$, 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:

$$R ::= P \mid P^- \qquad B ::= A \mid \exists R$$
$$Q ::= R \mid \neg R \qquad C ::= B \mid \neg B$$

Below, 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 $a, b \in N_a$. Here, we also consider extended ABoxes, 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 $u, v \in (N_a \cup N_\ell)$. We denote by $\mathsf{Ind}(A)$ the set of constants occurring in $\mathcal{A}$. 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). 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: (1) $(\exists R)^\mathcal{I} = \{x \in \Delta^\mathcal{I} \mid \exists y \in \Delta^\mathcal{I}$ such that $(x, y) \in R^\mathcal{I}\}$; (2) $(\neg B)^\mathcal{I} = \Delta^\mathcal{I} \setminus B^\mathcal{I}$; (3) $(P^-)^\mathcal{I} = \{(y, x) \in \Delta^\mathcal{I} \times \Delta^\mathcal{I} \mid (x, y) \in P^\mathcal{I}\}$; and (4) $(\neg R)^\mathcal{I} = (\Delta^\mathcal{I} \times \Delta^\mathcal{I}) \setminus R^\mathcal{I}$. Note that, consistently with the semantics of OWL 2 QL, we do *not* make the unique name assumption, i.e., distinct constants $a, b \in N_a$ may be interpreted as the same object. 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: (1) for every $B(u) \in \mathcal{A}$, it holds that $h(u) \in B^\mathcal{I}$; and (2) 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) ABox $\mathcal{A}$ over a signature Σ and a membership assertion α over Σ, we use notation $\mathcal{A} \models \alpha$ to indicate that for every interpretation $\mathcal{I}$ of Σ, if $\mathcal{I} \models \mathcal{A}$, then $\mathcal{I} \models \alpha$ (and likewise for (extended) KBs).

We also need to introduce the notions of Σ-types and Σ-homomorphisms. For an interpretation $\mathcal{I}$ and a signature Σ, the Σ*-types* $\mathbf{t}^\mathcal{I}_\Sigma(x)$ and $\mathbf{r}^\mathcal{I}_\Sigma(x, y)$ for $x, y \in \Delta^\mathcal{I}$ are given by the set of concepts B and roles R over Σ, respectively, such that $x \in B^\mathcal{I}$ and $(x, y) \in R^\mathcal{I}$. We also use $\mathbf{t}^\mathcal{I}(x)$ and $\mathbf{r}^\mathcal{I}(x, y)$ to refer to the types over the signature of all *DL-Lite*$_\mathcal{R}$ concepts and roles. A Σ*-homomorphism* from an interpretation $\mathcal{I}$ to an interpretation $\mathcal{J}$ is a function $h : \Delta^\mathcal{I} \mapsto \Delta^\mathcal{J}$ such that $h(a^\mathcal{I}) = a^\mathcal{J}$, for all individual names a interpreted in $\mathcal{I}$, $\mathbf{t}^\mathcal{I}_\Sigma(x) \subseteq \mathbf{t}^\mathcal{J}_\Sigma(h(x))$ and $\mathbf{r}^\mathcal{I}_\Sigma(x, y) \subseteq \mathbf{r}^\mathcal{J}_\Sigma(h(x), h(y))$ for all $x, y \in \Delta^\mathcal{I}$. We say that $\mathcal{I}$ is Σ*-homomorphically embeddable* into $\mathcal{J}$ if there exists a Σ-homomorphism from $\mathcal{I}$ to $\mathcal{J}$. If Σ is the set of all *DL-Lite*$_\mathcal{R}$ concepts and roles, we call Σ-homomorphism simply *homomorphism*.

DL-Lite$_\mathcal{R}$ enjoys the canonical model property. Let $\mathcal{K} = \langle \mathcal{T}, \mathcal{A} \rangle$ be a (non-extended) KB, and $\sqsubseteq^\mathcal{R}_\mathcal{T}$ the reflexive and transitive closure of the role relation on the set of all basic roles over N_R induced by $\mathcal{T}$ (that is, the reflexive and transitive closure of $\{(R_1, R_2) \mid R_1 \sqsubseteq R_2 \in \mathcal{T}$ or $R_1^- \sqsubseteq R_2^- \in \mathcal{T}\}$). Then define

$[R] = \{S \mid R \sqsubseteq_{\mathcal{T}}^{\mathcal{R}} S \text{ and } S \sqsubseteq_{\mathcal{T}}^{\mathcal{R}} R\}$, $[R] \leq_{\mathcal{T}} [S]$ if $R \sqsubseteq_{\mathcal{T}}^{\mathcal{R}} S$, and a *generating relationship* $\leadsto_{\mathcal{K}}$ as follows:

- $a \leadsto_{\mathcal{K}} w_{[R]}$, if (1) $\mathcal{K} \models \exists R(a)$; (2) $\mathcal{K} \not\models R(a,b)$ for every $b \in N_a$; (3) $[R'] = [R]$ for every $[R']$ such that $[R'] \leq_{\mathcal{T}} [R]$ and $\mathcal{K} \models \exists R'(a)$.
- $w_{[S]} \leadsto_{\mathcal{K}} w_{[R]}$, if (1) $\mathcal{T} \models \exists S^- \sqsubseteq \exists R$; (2) $[S^-] \neq [R]$; (3) $[R'] = [R]$ for every $[R']$ such that $[R'] \leq_{\mathcal{T}} [R]$ and $\mathcal{T} \models \exists S^- \sqsubseteq \exists R'$.

Denote by $\mathsf{path}(\mathcal{K})$ the set of all $\mathcal{K}$-paths, where a *$\mathcal{K}$-path* is a sequence $aw_{[R_1]} \dots w_{[R_n]}$ such that $n \geq 0$, $a \in N_a$, $a \leadsto_{\mathcal{K}} w_{[R_1]}$ and $w_{[R_i]} \leadsto_{\mathcal{K}} w_{[R_{i+1}]}$ for $1 \leq i \leq n-1$. Moreover, for every $\sigma \in \mathsf{path}(\mathcal{K})$, denote by $\mathsf{tail}(\sigma)$ the last element in σ. Finally, the *canonical* (or, *universal*) *model* of $\mathcal{K}$, denoted $\mathcal{U}_{\mathcal{K}}$, is defined as:

$$
\begin{aligned}
\Delta^{\mathcal{U}_{\mathcal{K}}} &= \mathsf{path}(\mathcal{K}),\\
a^{\mathcal{U}_{\mathcal{K}}} &= a, \text{ for } a \in N_a,\\
A^{\mathcal{U}_{\mathcal{K}}} &= \{a \in \mathsf{Ind}(\mathcal{A}) \mid \mathcal{K} \models A(a)\} \cup \{\sigma \cdot w_{[R]} \in \Delta^{\mathcal{U}_{\mathcal{K}}} \mid \mathcal{T} \models \exists R^- \sqsubseteq A\},\\
P^{\mathcal{U}_{\mathcal{K}}} &= \{(a,b) \in \mathsf{Ind}(\mathcal{A}) \times \mathsf{Ind}(\mathcal{A}) \mid \mathcal{K} \models P(a,b)\} \cup\\
&\quad\ \{(\sigma, \sigma \cdot w_{[R]}) \mid \mathsf{tail}(\sigma) \leadsto_{\mathcal{K}} w_{[R]}, [R] \leq_{\mathcal{T}} [P]\} \cup\\
&\quad\ \{(\sigma \cdot w_{[R]}, \sigma) \mid \mathsf{tail}(\sigma) \leadsto_{\mathcal{K}} w_{[R]}, [R^-] \leq_{\mathcal{T}} [P]\}.
\end{aligned}
$$

Theorem 1 ([16]). *If $\mathcal{K}$ is consistent, $\mathcal{U}_{\mathcal{K}}$ is a model of $\mathcal{K}$. For every model $\mathcal{I} \models \mathcal{K}$, there exists a homomorphism from $\mathcal{U}_{\mathcal{K}}$ to $\mathcal{I}$.*

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 $q^{\mathcal{I}} \subseteq (\Delta^{\mathcal{I}})^k$. Given a (non-extended) KB $\mathcal{K}$ over Σ, the set of *certain answers* to q over $\mathcal{K}$, denoted by $cert(q, \mathcal{K})$, is defined as: $\bigcap_{\mathcal{I} \in \textsc{Mod}(\mathcal{K})} \{(a_1, \dots, a_k) \mid \{a_1, \dots, a_k\} \subseteq N_a \text{ and } (a_1^{\mathcal{I}}, \dots, a_k^{\mathcal{I}}) \in q^{\mathcal{I}}\}$. A *conjunctive query* (CQ) *over a signature Σ* is a formula of the form $q(\boldsymbol{x}) = \exists \boldsymbol{y}.\, \varphi(\boldsymbol{x}, \boldsymbol{y})$, where $\boldsymbol{x}$, $\boldsymbol{y}$ are tuples of variables and $\varphi(\boldsymbol{x}, \boldsymbol{y})$ 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 $\boldsymbol{x}$ or $\boldsymbol{y}$. 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 $\boldsymbol{a}$ of elements from $\Delta^{\mathcal{I}}$ for which there exist a tuple $\boldsymbol{b}$ of elements from $\Delta^{\mathcal{I}}$ such that $\mathcal{I}$ satisfies every conjunct in $\varphi(\boldsymbol{a}, \boldsymbol{b})$. A union of conjunctive queries (UCQ) over a signature Σ is a formula of the form $q(\boldsymbol{x}) = \bigvee_{i=1}^{n} q_i(\boldsymbol{x})$, 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 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 [2]. The semantics of such a mapping is defined in [2] in terms of a notion of satisfaction for interpretations, which has to be extended in our case to deal with interpretations not satisfying the unique name assumption (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 (1) for every $a \in N_a$, it holds that $a^{\mathcal{I}} = a^{\mathcal{J}}$, (2) for every concept inclusion $B \sqsubseteq C \in \mathcal{T}_{12}$, it holds that $B^{\mathcal{I}} \subseteq C^{\mathcal{J}}$, and (3) 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. Finally, $\text{SAT}_{\mathcal{M}}(\mathcal{I})$ 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, $\text{SAT}_{\mathcal{M}}(\mathcal{X})$ is defined as $\bigcup_{\mathcal{I} \in \mathcal{X}} \text{SAT}_{\mathcal{M}}(\mathcal{I})$.

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. Let $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$ be a mapping, $\mathcal{K}_1 = \langle \mathcal{T}_1, \mathcal{A}_1 \rangle$ a KB over Σ_1 and $\mathcal{K}_2 = \langle \mathcal{T}_2, \mathcal{A}_2 \rangle$ an extended KB over Σ_2. The first such notion is the concept of solution, which is formalized as follows: $\mathcal{K}_2$ is a *solution* for $\mathcal{K}_1$ under $\mathcal{M}$ if $\text{MOD}(\mathcal{K}_2) \subseteq \text{SAT}_{\mathcal{M}}(\text{MOD}(\mathcal{K}_1))$. 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 $\mathcal{M}$. Then, $\mathcal{K}_2$ is a *universal solution* for $\mathcal{K}_1$ under $\mathcal{M}$ if $\text{MOD}(\mathcal{K}_2) = \text{SAT}_{\mathcal{M}}(\text{MOD}(\mathcal{K}_1))$. Thus, $\mathcal{K}_2$ is designed to exactly represent the space of interpretations obtained by translating the interpretations of $\mathcal{K}_1$ under $\mathcal{M}$ [2].

A second class of translations is obtained in [2] 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. Let $\mathcal{K}_1$, $\mathcal{M}$ as above and $\mathcal{K}_2$ a KB over Σ_2. Then $\mathcal{K}_2$ a UCQ-*solution* for $\mathcal{K}_1$ under $\mathcal{M}$ if for every query $q \in \text{UCQ}$ 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 $q \in \text{UCQ}$ over Σ_2: $cert(q, \langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle) = cert(q, \mathcal{K}_2)$.

Arguably, the most important problem in knowledge exchange [5,2], as well as in data exchange [12,15], is the task of computing a translation of a KB according to a mapping. To study the complexity of this task for the notions of solution just presented, we introduce the following decision problems. The *membership* problem for universal solutions (resp. universal UCQ-solutions) has as input a mapping $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$, a KB $\mathcal{K}_1$ over Σ_1, and an extended KB $\mathcal{K}_2$ (resp. a KB $\mathcal{K}_2$) over Σ_2. Then the question to answer is whether $\mathcal{K}_2$ is a universal solution (resp. universal UCQ-solution) for $\mathcal{K}_1$ under $\mathcal{M}$. In the *non-emptiness* problem for universal solutions the input is the same, except for $\mathcal{K}_2$, and the question to answer is whether a universal solution $\mathcal{K}_2$ exists (analogously for universal UCQ-solutions). The non-emptiness problems are directly related with the problem of computing translations of a KB according to a mapping.

4 The Shape of Universal Solutions

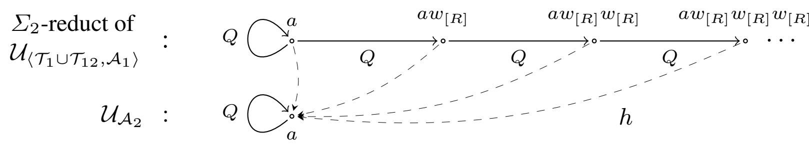
In what follows, we show some known results and examples of universal solutions. First of all, it was shown in [3] that a KB $\mathcal{K}_2 = \langle \mathcal{T}_2, \mathcal{A}_2 \rangle$ (extended or not) over Σ_2 is a universal solution for $\mathcal{K}_1 = \langle \mathcal{T}_2, \mathcal{A}_2 \rangle$ over Σ_1 under $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$ only if TBox

$\mathcal{T}_2$ is trivial (we call a TBox $\mathcal{T}$ trivial if $\mathcal{T}$ is equivalent to the empty set of formulas). Therefore, in the context of universal solutions, we only consider target KBs of the form $\langle \emptyset, \mathcal{A}_2 \rangle$, and we treat ABoxes $\mathcal{A}_2$ as such KBs.

Example 1. Assume that $\mathcal{M} = \big(\{A(\cdot), B(\cdot)\}, \{A'(\cdot), B'(\cdot)\}, \{A \sqsubseteq A', B \sqsubseteq B'\}\big)$, and let $\mathcal{K}_1 = \langle \mathcal{T}_1, \mathcal{A}_1 \rangle$, where $\mathcal{T}_1 = \{\}$ and $\mathcal{A}_1 = \{A(a), B(b)\}$. Then the ABox $\mathcal{A}_2 = \{A'(a), B'(b)\}$ is a straightforward universal solution for $\mathcal{K}_1$ under $\mathcal{M}$.

It can be shown (c.f. Lemma 1) that in the language without disjointness assertions, an ABox $\mathcal{A}_2$ is a universal solution for $\langle \mathcal{T}_1, \mathcal{A}_1 \rangle$ under $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$ if and only if $\mathcal{U}_{\mathcal{A}_2}$ is Σ_2-homomorphically equivalent to $\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}$. This fact is used in the following more involved example:

Example 2. Assume $\mathcal{M} = \big(\{A(\cdot), R(\cdot,\cdot), S(\cdot,\cdot)\}, \Sigma_2, \mathcal{T}_{12}\big)$ where $\Sigma_2 = \{Q(\cdot,\cdot)\}$, $\mathcal{T}_{12} = \{R \sqsubseteq Q, S \sqsubseteq Q\}$, and that $\mathcal{K}_1 = \langle \mathcal{T}_1, \mathcal{A}_1 \rangle$, where $\mathcal{T}_1 = \{A \sqsubseteq \exists R, \exists R^- \sqsubseteq \exists R\}$ and $\mathcal{A}_1 = \{A(a), S(a,a)\}$. Let $\mathcal{A}_2 = \{Q(a,a)\}$. In the following picture, it is easy to see h is a Σ_2-homomorphism from $\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}$ to $\mathcal{U}_{\mathcal{A}_2}$. The existence of a Σ_2-homomorphism in the other direction is trivial and, hence, $\mathcal{A}_2$ is a universal solution for $\mathcal{K}_1$ under $\mathcal{M}$.



■

The following example shows that extended ABoxes are necessary to guarantee the existence of universal solutions in certain cases.

Example 3. Assume that $\mathcal{M} = \big(\{A(\cdot), R(\cdot,\cdot)\}, \{B(\cdot)\}, \{\exists R^- \sqsubseteq B\}\big)$, and let $\mathcal{K}_1 = \langle \mathcal{T}_1, \mathcal{A}_1 \rangle$, where $\mathcal{T}_1 = \{A \sqsubseteq \exists R\}$ and $\mathcal{A}_1 = \{A(a)\}$. Then the ABox $\mathcal{A}_2 = \{B(n)\}$, 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. ■

Finally, we discuss the impact of disjointness assertions on the universal solutions.

Example 4. Consider Example 1 with $\mathcal{T}_1 = \{A \sqsubseteq \neg B\}$. With this seemingly harmless disjointness assertion $\mathcal{A}_2$ 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 unique name assumption on the one hand, and the presence of the disjointness assertion in $\mathcal{T}_1$ that forces a and b to be interpreted differently in the source, on the other hand. Thus, for a model $\mathcal{J}$ of $\mathcal{A}_2$ such that $a^{\mathcal{J}} = b^{\mathcal{J}}$, $A'^{\mathcal{J}} = B'^{\mathcal{J}} = \{a^{\mathcal{J}}\}$, there is no model $\mathcal{I}$ of $\mathcal{K}_1$ such that $(\mathcal{I}, \mathcal{J}) \models \mathcal{T}_{12}$ (hence, $a^{\mathcal{I}} = a^{\mathcal{J}}$ and $b^{\mathcal{I}} = b^{\mathcal{J}}$). In general, there is no universal solution for $\mathcal{K}_1$ under $\mathcal{M}$, even though $\mathcal{K}_1$ and $\mathcal{T}_{12}$ are consistent with each other. ■

The problem raised by the latter example could be solved by allowing for inequalities between constants in the ABoxes. A similar problem appears with disjointness assertions in the mapping, but it requires negative facts to be present for a universal solution to exist (i.e., facts of the form $\neg A(a)$, $\neg P(a,b)$), which are not part of OWL 2 QL.

On the other hand, having disjointness assertion in the source or the mapping does not exclude the existence of universal UCQ-solutions, which is explained in Example 6.

5 Computing Universal Solutions: The Case of Knowledge Bases

In this section, we show that both the membership and the non-emptiness problems for universal solutions without null values are in PTIME.

Assume that Σ_1, Σ_2 are disjoint signatures, and that $\mathcal{K} = \langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle$ is a KB such that $\mathcal{T}_1$ is defined over Σ_1 and $\mathcal{T}_{12}$ is a set of inclusions from Σ_1 to Σ_2. Moreover, let $\mathcal{U}_{\mathcal{K}}$ be the canonical model of $\mathcal{K}$. Then a basic concept B over Σ_1 is said to be *safe* in $\mathcal{U}_{\mathcal{K}}$ if $d \not\in N_a$ and $\mathbf{t}_{\Sigma_2}^{\mathcal{U}_{\mathcal{K}}}(d) = \emptyset$ for every $d \in B^{\mathcal{U}_{\mathcal{K}}}$. Intuitively, safeness for B means no constant "associated" with B and no target concept that is "associated" with B via $\mathcal{T}_1$ and $\mathcal{T}_{12}$ will be mentioned in the target; in Example 4 neither A nor B is safe in $\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}$. Furthermore, a pair of basic concepts (B, C) is is said to be *safe* if B or C is safe. Intuitively, if a pair (B, C) is not safe and $(B \sqsubseteq \neg C) \in \mathcal{T}_1$, then universal solution cannot exist, as explained in Example 4. Similarly, we say a basic role R over Σ_1 is safe if either $d \not\in N_a$ and $\mathbf{t}_{\Sigma_2}^{\mathcal{U}_{\mathcal{K}}}(d) = \emptyset$, or $d' \not\in N_a$ and $\mathbf{t}_{\Sigma_2}^{\mathcal{U}_{\mathcal{K}}}(d') = \emptyset$, for every $(d, d') \in R^{\mathcal{U}_{\mathcal{K}}}$. A pair of roles (R, Q) is safe if 1) R or Q is safe, and 2) $\mathbf{t}_{\Sigma_2}^{\mathcal{U}_{\mathcal{K}}}(d') = \emptyset$ or $\mathbf{t}_{\Sigma_2}^{\mathcal{U}_{\mathcal{K}}}(d'') = \emptyset$ for every $d, d', d'' \in \Delta^{\mathcal{U}_{\mathcal{K}}}$ such that $(d, d') \in R^{\mathcal{U}_{\mathcal{K}}}$ and $(d, d'') \in Q^{\mathcal{U}_{\mathcal{K}}}$.

Lemma 1. *A KB $\mathcal{K}_2 = \langle \mathcal{T}_2, \mathcal{A}_2 \rangle$ over Σ_2 is a universal solution for a KB $\mathcal{K}_1 = \langle \mathcal{T}_1, \mathcal{A}_1 \rangle$ under a mapping $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$ iff the following conditions hold:*

(tr) *$\mathcal{T}_2$ is a trivial TBox,*
(hom) *$\mathcal{U}_{\mathcal{A}_2}$ is Σ_2- homomorphically equivalent to $\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}$,*
(ps1) *each pair of concepts (B, C) is safe in $\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}$ whenever $(B \sqsubseteq \neg C) \in \mathcal{T}_1$,*
(pm1) *$B^{\mathcal{U}_{\mathcal{K}}} = \emptyset$ for each basic concept B such that $(B \sqsubseteq \neg B') \in \mathcal{T}_{12}$,*
(ps2) *each pair of roles (R, Q) is safe in $\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}$ whenever $(R \sqsubseteq \neg Q) \in \mathcal{T}_1$,*
(pm2) *$R^{\mathcal{U}_{\mathcal{K}}} = \emptyset$ for each basic role R such that $(R \sqsubseteq \neg R') \in \mathcal{T}_{12}$.*

It can be readily verified that conditions **(tr)**, **(ps1)**, **(ps2)**, **(pm1)**, **(pm2)** and the existence of a Σ_2-homomorphism from $\mathcal{U}_{\mathcal{A}_2}$ to $\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}$ required by **(hom)**, are solvable in polynomial time. To solve the problem of existence of a Σ_2-homomorphism in the opposite direction, we are going to employ the technique of reachability games on graphs. Below we present the required basic notions.

5.1 Reachability games on graphs

A game is defined by a game graph (a playground) and a winning condition. A *game graph* is a triple $G = (S_0, S_1, T)$, where $S = S_0 \cup S_1$ is a finite set of states, $S_0 \cap S_1 = \emptyset$ and $T \subseteq S \times S$ is a transition relation. The game starts in some state $s_0 \in S$, and it is played in turns. In each turn, if the current state s is in S_i ($i = 0, 1$), then Player i chooses some state $s' \in S$ such that $(s, s') \in T$. Thus, each play in the game is viewed as a path π, which can be infinite ($\pi = s_0, s_1, s_2, \ldots$, where $s_i \in S$ and $(s_i, s_{i+1}) \in T$ for every $i \geq 0$) or finite ($\pi = s_0, s_1, s_2, \ldots, s_k \in S^{k+1}$, where $(s_i, s_{i+1}) \in T$ for every $i \in \{0, \ldots, k-1\}$ and $\{s \mid (s_k, s) \in T\} = \emptyset$).

The winning condition defines what are the plays won by Player 0. We will consider a *reachability acceptance condition* specified as follows: given a set of accepting states $F \subseteq S$, a play π is a *win* for Player 0 iff some vertex from F occurs in π. Finally, a *reachability game* is a pair $\mathcal{G} = (G, F)$ where G is a game graph and F is a set of accepting states.

A *strategy for Player 0 from* state s is a (partial) function $f_0 : S^* S_0 \to S$ such that it assigns to each sequence of states $s_0, s_1, \dots, s_k$ with $s_0 = s$ and $s_k \in S_0$, a successor state s_{k+1} such that $(s_k, s_{k+1}) \in T$. A play $\pi = s_0 s_1 \cdots$ is said to *conform* with strategy f_0 if $s_{i+1} = f_0(s_0 s_1 \dots s_i)$ for every $s_i \in S_0$. Then, a strategy f_0 is a *winning strategy* for Player 0 *from* $s \in S$ if every play that conforms with f_0 and starts in s is a win for Player 0.

Proposition 1 ([18], [10]). *Given a reachability game $\mathcal{G} = (G, F)$ and a state s in G, it can be checked in* PTIME *whether Player 0 has a winning strategy from s.*

5.2 The reduction

Assume 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 $\mathcal{K}_2 = \langle \emptyset, \mathcal{A}_2 \rangle$ over Σ_2 (w.l.o.g., we can assume that the TBox of $\mathcal{K}_2$ is empty). Denote $\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle$ by $\mathcal{K}$. We show how the problem of checking whether there exists a Σ_2-homomorphism h from $\mathcal{U}_\mathcal{K}$ to $\mathcal{U}_{\mathcal{A}_2}$ can be reduced to the problem of existence of a winning strategy for Player 0 in a reachability game.

First, for such a homomorphism h to exist, it should be the case that

$$\mathbf{t}^{\mathcal{U}_\mathcal{K}}_{\Sigma_2}(a) \subseteq \mathbf{t}^{\mathcal{U}_{\mathcal{A}_2}}(a) \text{ and } \mathbf{r}^{\mathcal{U}_\mathcal{K}}_{\Sigma_2}(a,b) \subseteq \mathbf{t}^{\mathcal{U}_{\mathcal{A}_2}}(a,b) \text{ for all } a, b \in \mathsf{Ind}(\mathcal{A}_1). \tag{1}$$

These conditions can be clearly checked in PTIME.

Now, to check how the elements $a \in \Delta^{\mathcal{U}_\mathcal{K}}$ with $a \in \mathsf{Ind}(\mathcal{A}_1)$ can be mapped on $\Delta^{\mathcal{U}_{\mathcal{A}_2}}$, we construct a game $\mathcal{G}_a = (G_a, F_a)$. The game graph $G_a = (S_0, S_1, T)$ has the set of states of the kind $(x, y, \mathfrak{p})$, where $x \in \Delta^{\mathcal{U}_{\mathcal{A}_2}}$, $y \in \{\mathsf{tail}(a\sigma) \mid a\sigma \in \Delta^{\mathcal{U}_\mathcal{K}}\}$, and $\mathfrak{p} \in \{\mathfrak{s}, \mathfrak{d}\}$. The states $(x, y, \mathfrak{s})$ form S_0 and will be called *spoiling*; intuitively, the moves going out of such states represent various edges of the tree $\mathcal{U}_\mathcal{K}$ accessible from the end of the current edge. On the other hand, the states $(x, y, \mathfrak{d})$ form S_1 and will be called *duplicating*; the moves from them "show" how the "challenged" edge of the tree $\mathcal{U}_\mathcal{K}$ can be "mapped" on $\mathcal{U}_{\mathcal{A}_2}$. Notice that the size of G_a is $O(|\mathcal{T}_1 \cup \mathcal{T}_{12}| \times |\mathcal{A}_2|)$. The transition relation T is defined as follows:

$$\begin{aligned} T = \{ \big((x,y,\mathfrak{s}), (x',y',\mathfrak{d})\big) \mid y \rightsquigarrow_\mathcal{K} y' \text{ and } x' = x \} \cup \\ \{ \big((x,y,\mathfrak{d}), (x',y',\mathfrak{s})\big) \mid y = w_{[R]},\ cl^{\mathcal{T}_1 \cup \mathcal{T}_{12}}_{\Sigma_2}(\exists R^-) \subseteq \mathbf{t}^{\mathcal{U}_{\mathcal{A}_2}}(x'), \\ cl^{\mathcal{T}_1 \cup \mathcal{T}_{12}}_{\Sigma_2}(R) \subseteq \mathbf{r}^{\mathcal{U}_{\mathcal{A}_2}}(x,x'),\ \text{and } y' = y \}, \end{aligned}$$

where for a TBox $\mathcal{T}$ and a concept B, $cl^\mathcal{T}_\Sigma(B)$ is the set of all concepts B' over Σ such that $\mathcal{T} \models B \sqsubseteq B'$, and for a role R, $cl^\mathcal{T}_\Sigma(R)$ is defined analogously.

The set F_a in the definition of the game is given by the duplicating states that are "dead ends", i.e.,

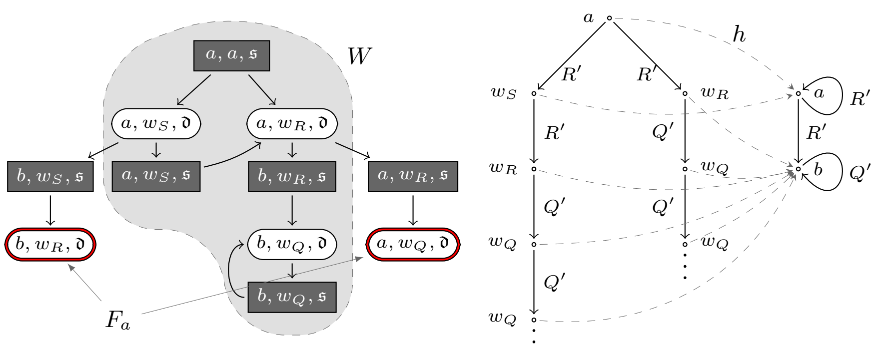
$$\begin{aligned} F_a = \{ (x,y,\mathfrak{d}) \mid y = w_{[R]} \text{ and for all } x' \in \Delta^{\mathcal{U}_{\mathcal{A}_2}},\ cl^{\mathcal{T}_1 \cup \mathcal{T}_{12}}_{\Sigma_2}(\exists R^-) \not\subseteq \mathbf{t}^{\mathcal{U}_{\mathcal{A}_2}}(x') \text{ or} \\ cl^{\mathcal{T}_1 \cup \mathcal{T}_{12}}_{\Sigma_2}(R) \not\subseteq \mathbf{r}^{\mathcal{U}_{\mathcal{A}_2}}(x,x') \}. \end{aligned}$$

Having constructed the game $\mathcal{G}_a = (G_a, F_a)$, we prove that verifying whether the elements $a \in \Delta^{\mathcal{U}_\mathcal{K}}$ can be Σ_2-homomorphically mapped on $\Delta^{\mathcal{U}_{\mathcal{A}_2}}$ reduces to checking whether Player 0 has a winning strategy in $\mathcal{G}_a$ from the state $(a, a, \mathfrak{s})$.

Lemma 2. *There exists a Σ_2-homomorphism from $\mathcal{U}_\mathcal{K}$ to $\mathcal{U}_{\mathcal{A}_2}$ iff* (1) *holds and for each $a \in \mathsf{Ind}(\mathcal{A}_1)$, Player 0 does not have a winning strategy in $\mathcal{G}_a$ from $(a, a, \mathfrak{s})$.*

The example below illustrates the presented reduction.

Example 5. Assume $\mathcal{M} = \big(\{R(\cdot,\cdot), S(\cdot,\cdot), Q(\cdot,\cdot)\}, \Sigma_2, \mathcal{T}_{12}\big)$, where $\Sigma_2 = \{R'(\cdot,\cdot), Q'(\cdot,\cdot)\}$ and $\mathcal{T}_{12} = \{R \sqsubseteq R', S \sqsubseteq R', Q \sqsubseteq Q'\}\big)$, $\mathcal{K}_1 = \langle \mathcal{T}_1, \mathcal{A}_1 \rangle$, where $\mathcal{T}_1 = \{\exists S^- \sqsubseteq \exists R, \exists R^- \sqsubseteq \exists Q, \exists Q^- \sqsubseteq \exists Q\}$ and $\mathcal{A}_1 = \{\exists R(a), \exists S(a)\}$, and $\mathcal{A}_2 = \{R'(a,a), R'(a,b), Q'(b,b)\}$. Then $F_a = \{(b, w_R, \mathfrak{d}), (a, w_Q, \mathfrak{d})\}$, and the game graph G_a is depicted in a) below (we ignore the states that are not reachable from $(a, a, \mathfrak{s})$; the duplicating states forming S_1 are shown as ovals and the spoiling states forming S_0 are shown as boxes). In b) below we show $\mathcal{U}_\mathcal{K}$ (the domain elements $d \in \Delta^{\mathcal{U}_\mathcal{K}}$ are shown as dots, the labels next to them represent $\mathsf{tail}(d)$, and the labels on the edges (d, d') show the Σ_2-role names P, such that $(d, d') \in P^{\mathcal{U}_\mathcal{K}}$), in c) we show $\mathcal{U}_{\mathcal{A}_2}$, and the dashed arrows from $\mathcal{U}_\mathcal{K}$ to $\mathcal{U}_{\mathcal{A}_2}$ show the homomorphism h.



a) The game graph G_a b) Σ_2-reduct of $\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}$ c) $\mathcal{U}_{\mathcal{A}_2}$

Observe that in the game $\mathcal{G}_a$ Player 0 does not have a winning strategy from $(a, a, \mathfrak{s})$, because there is a way for Player 1 to play (infinitely) so that the game never goes out of the region W shown in a). It is not difficult to see that such strategy of Player 1 can be used to define the homomorphism h, and vice versa. ■

Finally, combining Lemma 2, Lemma 1 and Proposition 1 one obtains:

Theorem 2. *The membership problem for universal solutions with (non-extended) KBs is in* PTIME.

We conclude this section by addressing the non-emptiness problem. It follows from Conditions **(tr)**, **(hom)** that there exists a universal solution for $\mathcal{K}_1$ under $\mathcal{M}$ iff $\mathcal{K}_2 = \langle \emptyset, \mathcal{A}_2 \rangle$ over Σ_2 is a universal solution for $\mathcal{K}_1$ under $\mathcal{M}$, where $\mathcal{A}_2$ satisfies *(i)* $\mathcal{A}_2 \models B(a)$ iff $\mathcal{K} \models B(a)$ and *(ii)* $\mathcal{A}_2 \models R(a,b)$ iff $\mathcal{K} \models R(a,b)$ for all $a, b \in \mathsf{Ind}(\mathcal{A}_1 \cup \mathcal{A}_2)$, Σ_2-concept B, and Σ_2-role R. Obviously, we can construct the required $\mathcal{A}_2$ in PTIME, then it remains to check if $\mathcal{K}_2$ above is a universal solution. It follows:

Theorem 3. *The non-emptiness problem for universal solutions with (non-extended) KBs is in* PTIME. *Moreover, there is an effective algorithm to compute a universal solution in polynomial time (if such a solution exists).*

6 Computing Universal Solutions: The Case of Extended Knowledge Bases

We start with the membership problem for extended KBs. Assume 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 $\mathcal{K}_2 = \langle \emptyset, \mathcal{A}_2 \rangle$ over Σ_2, where $\mathcal{A}_2$ is an extended ABox. It can be shown that an analogue of Lemma 1 holds, provided that the definition of $\mathcal{U}_{\mathcal{A}_2}$ is adjusted in an obvious way to account for nulls, and so is the definition of homomorphism. In this setting, Σ_2-homomorphism from $\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}$ to $\mathcal{U}_{\mathcal{A}_2}$ can be still checked in PTIME, however, the opposite direction cannot be checked efficiently due to nulls in $\mathcal{A}_2$. In fact, it can be shown by reduction from the graph 3-colorability problem that the membership problem for universal solutions with null values is NP-hard. To decide in NP whether there exists a homomorphism h from $\mathcal{U}_{\mathcal{A}_2}$ to $\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}$, we can use the fact that the image $W \subseteq \Delta^{\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}}$ of such a function h on $\Delta^{\mathcal{U}_{\mathcal{A}_2}}$ is of polynomial size. Therefore, for each constant and null in $\mathcal{A}_2$, one needs to guess its homomorphic image, and then check whether the resulting function is a homomorphism. Thus, we obtain:

Theorem 4. *The membership problem for universal solutions with extended KBs is* NP*-complete.*

Consider now the problem of checking whether there exists a universal solution $\mathcal{K}_2 = \langle \emptyset, \mathcal{A}_2 \rangle$ for $\mathcal{K}_1$ under $\mathcal{M}$. This problem turns out to be harder than the membership problem as now candidate solutions are not part of the input. In fact, we show by reduction from the validity problem for quantified Boolean formulas that checking the existence of a universal solution is PSPACE-hard. As for the upper bound, first, it can be shown that such an $\mathcal{A}_2$ exists if and only if $\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}$ is Σ_2-homomorphically embeddable into a finite part of itself. Then, such a finite subset of $\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}$ projected on Σ_2 can be taken as a universal solution for $\mathcal{K}_1$ under $\mathcal{M}$. As the inclusion of inverse roles is one of the main sources of complexity, we use *two-way alternating automata on infinite trees (2ATA)*, which are a generalization of nondeterministic automata on infinite trees [21] 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: (1) $\mathbb{A}^{can}_{\mathcal{K}}$ is a 2ATA that accepts trees corresponding to the canonical model of $\mathcal{K}$ with nodes arbitrary labeled with a special symbol G; (2) $\mathbb{A}^{mod}_{\mathcal{K}}$ 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 (3) $\mathbb{A}_{fin}$ 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 $\pi_{\Gamma_{\mathcal{K}}}(\mathbb{A}^{can}_{\mathcal{K}})$, $\pi_{\Gamma_{\mathcal{K}}}(\mathbb{A}^{mod}_{\mathcal{K}})$ and $\mathbb{A}_{fin}$, where $\mathcal{K} = \langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle$, $\pi_{\Gamma_{\mathcal{K}}}(\mathbb{A}^{can}_{\mathcal{K}})$ is the projection of $\mathbb{A}^{can}_{\mathcal{K}}$ on a vocabulary $\Gamma_{\mathcal{K}}$ not mentioning symbols from Σ_1, and likewise for $\pi_{\Gamma_{\mathcal{K}}}(\mathbb{A}^{mod}_{\mathcal{K}})$. If the language accepted

Universal solutions	ABoxes	Extended ABoxes
Membership	PTIME	NP-complete
Non-emptiness	PTIME	PSPACE-hard, in EXPTIME

Fig. 1. Complexity results for universal solutions

by $\mathbb{B}$ is empty, then there is no universal solution for $\mathcal{K}_1$ under $\mathcal{M}$, otherwise a universal solution (of exponential size) exists and it can be extracted from the tree accepted by $\mathbb{B}$.

Theorem 5. *The non-emptiness problem for universal solutions with extended KBs is* PSPACE*-hard and in* EXPTIME*. Moreover, there is an effective algorithm to compute a universal solution in exponential time (if such a solution exists).*

7 Universal UCQ-solutions

We start by arguing that universal UCQ-solutions exhibit more robust behavior in the presence of disjointness assertions than universal solutions.

Example 6. Consider $\mathcal{M}$, $\mathcal{K}_1$, and $\mathcal{A}_2$ from Example 4. Recall that $\mathcal{A}_2$ is not a universal solution for $\mathcal{K}_1$ under $\mathcal{M}$. However, $\mathcal{A}_2$ is a universal UCQ-solution for $\mathcal{K}_1$ under $\mathcal{M}$. Moreover, $\mathcal{A}_2$ remains a universal UCQ-solution for $\mathcal{K}_1$ under $\mathcal{M}$ independently of whether the unique name assumption is employed. ■

Unfortunately, universal UCQ-solutions are also harder to compute, which can be explained by the fact that TBoxes have 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 [16], we are able to prove the following:

Theorem 6. *The membership problem for universal* UCQ*-solutions is* PSPACE*-hard.*

8 Conclusions

A summary of our results for universal solutions is presented in Figure 1.

In this paper, we have studied the problem of KB exchange for OWL 2 QL, improving on previously known results both w.r.t. the expressiveness of the ontology language and w.r.t. the understanding of the computational properties of the problem. Our main contribution is a novel PTIME algorithm for the membership and non-emptiness problems for universal solutions when OWL 2 QL ABoxes are considered. Our investigation leaves open several issues, which we intend to address in the future. For the computation of universal solutions when extended ABoxes are allowed, while we have pinned-down the complexity of the membership problem as NP-complete, an exact characterization for the non-emptiness problem is still missing. Moreover, it is easy to see that allowing for inequalities between terms 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.

References

1. 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.
2. Marcelo Arenas, Elena Botoeva, and Diego Calvanese. Knowledge base exchange. In *Proc. of the 24th Int. Workshop on Description Logic (DL 2011)*, volume 745 of *CEUR Electronic Workshop Proceedings,* `http://ceur-ws.org/`, 2011.
3. Marcelo Arenas, Elena Botoeva, Diego Calvanese, Vladislav Ryzhikov, and Evgeny Sherkhonov. Exchanging description logic knowledge bases. In *Proc. of the 13th Int. Conf. on the Principles of Knowledge Representation and Reasoning (KR 2012)*, 2012.
4. 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.
5. Marcelo Arenas, Jorge Pérez, and Juan L. Reutter. Data exchange beyond complete data. In *Proc. of the 30th ACM SIGACT SIGMOD SIGART Symp. on Principles of Database Systems (PODS 2011)*, pages 83–94, 2011.
6. Alessandro Artale, Diego Calvanese, Roman Kontchakov, and Michael Zakharyaschev. The *DL-Lite* family and relations. *J. of Artificial Intelligence Research*, 36:1–69, 2009.
7. 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/`.
8. Pablo Barceló. Logical foundations of relational data exchange. *SIGMOD Record*, 38(1):49–58, 2009.
9. 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.
10. Krishnendu Chatterjee and Monika Henzinger. An o(n2) time algorithm for alternating Büchi games. In *Proc. of the 23rd Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 2012)*, pages 1386–1399, 2012.
11. Jérôme Euzenat and Pavel Schwaiko. *Ontology Matching*. Springer, 2007.
12. 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.
13. 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.
14. 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.
15. Phokion G. Kolaitis. Schema mappings, data exchange, and metadata management. In *Proc. of the 24th ACM SIGACT SIGMOD SIGART Symp. on Principles of Database Systems (PODS 2005)*, pages 61–75, 2005.
16. 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 Conf. on Artificial Intelligence (AAAI 2011)*, 2011.
17. 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.
18. René Mazala. Infinite games. In *Automata, Logics, and Infinite Games*, pages 23–42, 2001.

19. 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/`.
20. 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.
21. 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.