Representability in $DL\text{-}Lite_{\mathcal{R}}$ Knowledge Base Exchange

M. Arenas[1], E. Botoeva[2], D. Calvanese[2], V. Ryzhikov[2], and E. Sherkhonov[3]

[1] Dept. of Computer Science, PUC Chile
marenas@ing.puc.cl
[2] KRDB Research Centre, Free Univ. of Bozen-Bolzano, Italy
lastname@inf.unibz.it
[3] ISLA, University of Amsterdam, Netherlands
e.sherkhonov@uva.nl

Abstract. Knowledge base exchange can be considered as a generalization of data exchange in which the aim is to exchange between a source and a target connected through mappings, not only explicit knowledge, i.e., data, but also implicit knowledge in the form of axioms. Such problem has been investigated recently using Description Logics (DLs) as representation formalism, thus assuming that the source and target KBs are given as a DL TBox+ABox, while the mappings have the form of DL TBox assertions. In this paper we are interested in the problem of representing a given source TBox by means of a target TBox that captures at best the intensional information in the source. In previous work, results on representability have been obtained for $DL\text{-}Lite_{RDFS}$, a DL corresponding to the FOL fragment of RDFS. We extend these results to the positive fragment of $DL\text{-}Lite_{\mathcal{R}}$, in which, differently from $DL\text{-}Lite_{RDFS}$, the assertions in the TBox and the mappings may introduce existentially implied individuals. For this we need to overcome the challenge that the chase, a key notion in data and knowledge base exchange, is not guaranteed anymore to be finite.

1 Introduction

Knowledge base exchange is an extension of the data exchange setting, where the source may contain implicit knowledge by which new data may be inferred. The first framework for data exchange with incompletely specified data in the source was proposed in [3]. This framework is based on the general notion of *representation system*, as a mechanism to represent multiple instances of a data schema, and considers the problem of incomplete data exchanges under mappings constituted by a set of tuple generating dependencies (tgds), i.e., mappings between pairs of conjunctive queries. Given that the source data may be incompletely specified, (possibly infinitely) many source instances are implicitly represented. This framework was refined in [1, 2] to the case where as a representation system Description Logics (DL) knowledge bases (KBs) were used: the TBox and the ABox of a DL KB represent implicit and explicit information respectively, and mappings are sets of DL inclusions. While in the traditional data exchange setting, given a source instance and a mapping specification, *(universal) solutions* are target instances derived from the source instance and the mapping, in this case solutions are target DL KBs, derived from the source KB and the mapping.

In such a setting, in order to minimize the exchange (and hence transfer and materialization) of explicit (i.e., ABox) information, one is interested in computing universal solutions that contain as much implicit knowledge as possible. Therefore, the notion of *representability* was defined, which helps us in understanding the capacity of (universal) solutions to transfer implicit knowledge: we say that a source TBox is representable under a mapping if there exists a target TBox that leads to a universal solution when it is combined with a suitable ABox computed from the source ABox, independently of the actual source ABox. *Weak representability* is concerned with representability under a mapping extended with assertions that are implied by the given mapping and the source TBox. (Weak) representability of a source TBox under a mapping implies that the only knowledge that remains to be transferred explicitly via the (extended) mapping is the one in the source ABox. Therefore, checking (weak) representability and computing a representation of a source TBox under a(n extended) mapping turn out to be crucial problems in the context of KB exchange.

In [1, 2] the problems of deciding representability and weak-representability, and of computing a representation for a given mapping and a TBox was tackled for *DL-Lite*$_{RDFS}$, the DL counterpart of RDFS [5] and a member of the *DL-Lite* family of DLs [6]. It has been shown that these problems can be solved in polynomial time for *DL-Lite*$_{RDFS}$ mappings and TBoxes. Moreover, due to the simplicity of the logic, the characterization of representations is concise and simple.

In this paper we extend those results to the case of *DL-Lite*$_{\mathcal{R}}$ without disjointness assertions, a DL that we call *DL-Lite*$_{\mathcal{R}}^{pos}$. The presence of existential quantifiers on the right-hand side of concept inclusions makes the problem considerably more complicated than for *DL-Lite*$_{RDFS}$. However, we show that also in the presence of existentials on the right-hand side we are able to decide representability and weak representability of a *DL-Lite*$_{\mathcal{R}}^{pos}$ TBox under a *DL-Lite*$_{\mathcal{R}}^{pos}$ mapping in polynomial time and to construct a polynomial size representation.

2 Preliminaries

2.1 *DL-Lite*$_{\mathcal{R}}^{pos}$ Knowledge Bases

The DLs of the *DL-Lite* family [6] are characterized by the fact that reasoning can be done in polynomial time, and that data complexity of reasoning and conjunctive query answering is in AC^0. We present now the syntax and semantics of *DL-Lite*$_{\mathcal{R}}^{pos}$, which is the DL that we adopt here, together with a sub-language of it.

In the following, we use A and P to denote concept and role names, respectively, and B and R to denote generic concepts and roles, respectively. The latter are defined by the following grammar:

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

For a role R, we use R^- to denote P^- when $R = P$, and P when $R = P^-$.

A *DL-Lite*$_{\mathcal{R}}^{pos}$ TBox is a finite set of concept inclusions $B \sqsubseteq B'$ and role inclusions $R \sqsubseteq R'$. A *DL-Lite*$_{\mathcal{R}}^{pos}$ ABox is a finite set of membership assertions of the form $A(u)$ and $P(u, v)$, where u and v are *individuals* or *labeled nulls*. We distinguish between the two, since individuals are interpreted under the unique name assumption, while labeled nulls obtain their meaning through assignments (see below). Notice that we include labeled nulls in ABoxes as they are needed in the exchange of KBs. A *DL-Lite*$_{\mathcal{R}}^{pos}$ KB $\mathcal{K}$ is a pair $\langle \mathcal{T}, \mathcal{A} \rangle$, where $\mathcal{T}$ is a *DL-Lite*$_{\mathcal{R}}^{pos}$ TBox and $\mathcal{A}$ is a *DL-Lite*$_{\mathcal{R}}^{pos}$ ABox.

Note that *DL-Lite*$_{\mathcal{R}}^{pos}$ is the fragment of the DL *DL-Lite*$_{\mathcal{R}}$ studied in [6] without disjointness assertions on concepts and roles. In *DL-Lite*$_{\mathcal{R}}$, B and R are called *basic* concepts and *basic* roles, respectively, and for coherence with previous work on the *DL-Lite* family, we adopt here this terminology as well. We call *DL-Lite*$_{\scriptscriptstyle RDFS}$ the fragment of *DL-Lite*$_{\mathcal{R}}^{pos}$ (and hence of *DL-Lite*$_{\mathcal{R}}$) in which there are only atomic concepts and atomic roles on the right-hand side of inclusions.

The semantics of *DL-Lite*$_{\mathcal{R}}^{pos}$ is, as usual in DLs, based on the notion of first-order *interpretation* $\mathcal{I} = \langle \Delta^{\mathcal{I}}, \cdot^{\mathcal{I}} \rangle$, 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; *(2)* $P^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$, for every role name P; *(3)* $a^{\mathcal{I}} \in \Delta^{\mathcal{I}}$, for every individual name a; and *(4)* such that:

$$\begin{array}{rcl} (\exists R)^{\mathcal{I}} & = & \{x \in \Delta^{\mathcal{I}} \mid \text{there exists } y \in \Delta^{\mathcal{I}} \text{ s.t. } (x, y) \in R^{\mathcal{I}}\} \\ (P^{-})^{\mathcal{I}} & = & \{(y, x) \in \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}} \mid (x, y) \in P^{\mathcal{I}}\} \end{array}$$

Moreover, satisfaction of concept and role inclusions is defined as follows: $\mathcal{I} \models B \sqsubseteq B'$ if $B^{\mathcal{I}} \subseteq B'^{\mathcal{I}}$, and $\mathcal{I} \models R \sqsubseteq R'$ if $R^{\mathcal{I}} \subseteq R'^{\mathcal{I}}$. Finally, satisfaction of membership assertions is defined as follows. A *substitution* over an interpretation $\mathcal{I}$ is a function h from individuals and labeled nulls to $\Delta^{\mathcal{I}}$ such that $h(a) = a^{\mathcal{I}}$ for each individual a. Then $(\mathcal{I}, h) \models A(u)$ if $h(u) \in A^{\mathcal{I}}$, and $(\mathcal{I}, h) \models P(u, v)$ if $(h(u), h(v)) \in P^{\mathcal{I}}$.

An interpretation $\mathcal{I}$ is a *model* of a *DL-Lite*$_{\mathcal{R}}^{pos}$ TBox $\mathcal{T}$ if for every $\alpha \in \mathcal{T}$, it holds that $\mathcal{I} \models \alpha$, and it is a *model* of a *DL-Lite*$_{\mathcal{R}}^{pos}$ ABox $\mathcal{A}$ if there exists a substitution h over $\mathcal{I}$ such that for every $\alpha \in \mathcal{A}$, it holds that $(\mathcal{I}, h) \models \alpha$. Finally, $\mathcal{I}$ is a *model* of a *DL-Lite*$_{\mathcal{R}}$ KB $\mathcal{K} = \langle \mathcal{T}, \mathcal{A} \rangle$ if $\mathcal{I}$ is a model of both $\mathcal{T}$ and $\mathcal{A}$. The set of all models of $\mathcal{K}$ is denoted $\textsc{Mod}(\mathcal{K})$, and $\mathcal{K}$ is *consistent* if $\textsc{Mod}(\mathcal{K}) \neq \emptyset$. We observe that in *DL-Lite*$_{\mathcal{R}}^{pos}$ one cannot express any form of negative information, and hence a *DL-Lite*$_{\mathcal{R}}^{pos}$ KB is always consistent.

We assume that interpretations satisfy the standard names assumption, that is, we assume given a fixed infinite set $\mathbf{U}$ of individual names, and we assume that for every interpretation $\mathcal{I}$, it holds that $\Delta^{\mathcal{I}} \subseteq \mathbf{U}$ and $a^{\mathcal{I}} = a$ for every individual name a. This implies that interpretations satisfy the unique name assumption over individual names.

A *signature* Σ is a set of concept and role names. An interpretation $\mathcal{I}$ is said to be an interpretation of Σ if it is defined exactly on the concept and role names in Σ. Given a KB $\mathcal{K}$, the *signature* $\Sigma(\mathcal{K})$ *of* $\mathcal{K}$ is the alphabet of concept and role names occurring in $\mathcal{K}$, and $\mathcal{K}$ is said to be *defined over* (or simply, *over*) a signature Σ if $\Sigma(\mathcal{K}) \subseteq \Sigma$ (and likewise for a TBox $\mathcal{T}$, an ABox $\mathcal{A}$, inclusions $B \sqsubseteq C$ and $R \sqsubseteq Q$, and membership assertions $A(u)$ and $P(u, v)$).

2.2 Queries, Certain Answers, and Chase

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 $q^{\mathcal{I}} \subseteq \Delta^{k}$. In particular, if $k = 0$, then q is called 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). A query q is said to be a query over a KB $\mathcal{K}$ if q is a query over a signature Σ and $\Sigma \subseteq \Sigma(\mathcal{K})$. Moreover, the answer to q over $\mathcal{K}$, denoted by $cert(q, \mathcal{K})$, is defined as $cert(q, \mathcal{K}) = \bigcap_{\mathcal{I} \in \textsc{Mod}(\mathcal{K})} q^{\mathcal{I}}$. Each tuple in $cert(q, \mathcal{K})$ is called a *certain answer* for q over $\mathcal{K}$. 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.

In this paper, we adopt the class of unions of conjunctive queries as our main query formalism. A *conjunctive query (CQ) over a signature* Σ is a first-order formula of the form $q(\boldsymbol{x}) = \exists \boldsymbol{y}.conj(\boldsymbol{x}, \boldsymbol{y})$, where $\boldsymbol{x}$, $\boldsymbol{y}$ are tuples of variables and $conj(\boldsymbol{x}, \boldsymbol{y})$ is a conjunction of atoms of the form: (1) $A(t)$, with A a concept name in Σ and t either an individual from $\mathbf{U}$ or a variable from $\boldsymbol{x}$ or $\boldsymbol{y}$, or (2) $P(t_1, t_2)$, with P a role name in Σ and t_i ($i = 1, 2$) either an individual from $\mathbf{U}$ or a variable from $\boldsymbol{x}$ or $\boldsymbol{y}$. In a CQ $q(\boldsymbol{x}) = \exists \boldsymbol{y}.conj(\boldsymbol{x}, \boldsymbol{y})$ over a signature Σ, $\boldsymbol{x}$ is the tuple of free variables of $q(\boldsymbol{x})$. Moreover, 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 defined as 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 $conj(\boldsymbol{a}, \boldsymbol{b})$. Finally, a union of conjunctive queries (UCQ) over a signature Σ is a finite set of CQs over Σ that have the same free variables. A UCQ $q(\boldsymbol{x})$ is interpreted as the first-order formula $\bigvee_{q_i \in q} q_i(\boldsymbol{x})$, and its semantics is defined as $q^{\mathcal{I}} = \bigcup_{q_i \in q} q_i{}^{\mathcal{I}}$.

Certain answers in *DL-Lite*$_{\mathcal{R}}^{pos}$ can be characterized through the notion of chase. We call a *chase* a (possibly infinite) set of assertions of the form $A(t)$, $P(t, t')$, where t, t' are either individuals from $\mathbf{U}$, or labeled nulls interpreted as not necessarily distinct domain elements (see the definition of the semantics of *DL-Lite*$_{\mathcal{R}}^{pos}$ in Section 2.1). For *DL-Lite*$_{\mathcal{R}}^{pos}$ KBs, we employ the notion of restricted chase as defined in [6]. For such a KB $\langle \mathcal{T}, \mathcal{A} \rangle$, the chase of $\mathcal{A}$ w.r.t. $\mathcal{T}$, denoted *chase*$_{\mathcal{T}}(\mathcal{A})$, is a chase obtained from $\mathcal{A}$ by adding facts implied by inclusions in $\mathcal{T}$, and introducing fresh labeled nulls whenever required by an inclusion with $\exists R$ in the right-hand side (see [6] for details).

2.3 Knowledge Base Exchange Framework

Assume that Σ_1, Σ_2 are signatures with no concepts or roles in common. Then we say that an inclusion $N_1 \sqsubseteq N_2$ is an *inclusion from* Σ_1 *to* Σ_2, if N_1 is a concept or a role over Σ_1 and N_2 is a concept or a role over Σ_2. For a DL $\mathcal{L}$ (e.g., *DL-Lite*$_{\mathcal{R}}^{pos}$), we define an $\mathcal{L}$-*mapping* (or just *mapping*, when $\mathcal{L}$ is clear from the context) as a tuple $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$, where $\mathcal{T}_{12}$ is a TBox in $\mathcal{L}$ consisting of inclusions from Σ_1 to Σ_2:

(1) $C_1 \sqsubseteq C_2$, where C_1, C_2 are concepts in $\mathcal{L}$ over Σ_1 and Σ_2, respectively, and
(2) $Q_1 \sqsubseteq Q_2$, where Q_1 and Q_2 are roles in $\mathcal{L}$ over Σ_1 and Σ_2, respectively.

If $\mathcal{T}_{12}$ is an $\mathcal{L}$-TBox, for a DL $\mathcal{L}$ (e.g., *DL-Lite*$_{\mathcal{R}}^{pos}$), then $\mathcal{M}$ is called an $\mathcal{L}$-*mapping*.

The semantics of a mapping is defined in terms of the notion of satisfaction. More specifically, given a mapping $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$, an interpretation $\mathcal{I}$ of Σ_1 and an interpretation $\mathcal{J}$ of Σ_2, pair $(\mathcal{I}, \mathcal{J})$ *satisfies* TBox $\mathcal{T}_{12}$, denoted by $(\mathcal{I}, \mathcal{J}) \models \mathcal{T}_{12}$, if for each concept inclusion $C_1 \sqsubseteq C_2 \in \mathcal{T}_{12}$, it holds that $C_1{}^{\mathcal{I}} \subseteq C_2{}^{\mathcal{J}}$, and for each role inclusion $Q_1 \sqsubseteq Q_2 \in \mathcal{T}_{12}$, it holds that $Q_1{}^{\mathcal{I}} \subseteq Q_2{}^{\mathcal{J}}$. Moreover, given an interpretation $\mathcal{I}$ of Σ_1, $\textsc{Sat}_{\mathcal{M}}(\mathcal{I})$ 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: $\textsc{Sat}_{\mathcal{M}}(\mathcal{X}) = \bigcup_{\mathcal{I}\in\mathcal{X}} \textsc{Sat}_{\mathcal{M}}(\mathcal{I})$.

Let $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$ be a mapping, $\mathcal{K}_1$ a KB over Σ_1, and $\mathcal{K}_2$ a KB over Σ_2.

- $\mathcal{K}_2$ is called a *solution* for $\mathcal{K}_1$ under $\mathcal{M}$ if $\textsc{Mod}(\mathcal{K}_2) \subseteq \textsc{Sat}_{\mathcal{M}}(\textsc{Mod}(\mathcal{K}_1))$, and
- $\mathcal{K}_2$ is called a *universal solution* for $\mathcal{K}_1$ under $\mathcal{M}$ if $\textsc{Mod}(\mathcal{K}_2) = \textsc{Sat}_{\mathcal{M}}(\textsc{Mod}(\mathcal{K}_1))$.

Universal solutions present several limitations, as argued in [1, 2]. First, a universal solution (in *DL-Lite*$_{\mathcal{R}}^{pos}$ and *DL-Lite*$_{\mathcal{R}}$) does not always exist. Second, if it exists, then its TBox is trivial (that is, equivalent to the empty TBox). Finally, in the worst case the smallest universal solution is exponential in the size of the mapping and the source KB. A notion of solution parametrized w.r.t. a query language was proposed in [1, 2] in order to overcome these limitations. Such a notion, though weaker, is in line with the objective of (data and) KB exchange of providing in the target sufficient information to answer queries that could also be posed over the source.

Let $\mathcal{Q}$ be a class of queries, $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$ a mapping, $\mathcal{K}_1 = \langle \mathcal{T}_1, \mathcal{A}_1\rangle$ a KB over Σ_1, and $\mathcal{K}_2$ a KB over Σ_2. Then

- $\mathcal{K}_2$ is called a $\mathcal{Q}$*-solution* for $\mathcal{K}_1$ under $\mathcal{M}$ if for every query $q \in \mathcal{Q}$ over Σ_2, $cert(q, \langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1\rangle) \subseteq cert(q, \mathcal{K}_2)$, and
- $\mathcal{K}_2$ is called a *universal* $\mathcal{Q}$*-solution* for $\mathcal{K}_1$ under $\mathcal{M}$ if for every query $q \in \mathcal{Q}$ over Σ_2, $cert(q, \langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1\rangle) = cert(q, \mathcal{K}_2)$.

The definitions of solutions are illustrated in the following example.

Example 1. Assume $\Sigma_1 = \{Painting(\cdot), PaintedBy(\cdot,\cdot), ArtMovement(\cdot,\cdot)\}$ and $\Sigma_2 = \{ArtPiece(\cdot), ArtAuthor(\cdot,\cdot), HasStyle(\cdot,\cdot), HasGenre(\cdot,\cdot)\}$. Consider mapping $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$, where $\mathcal{T}_{12}$ is the following TBox:

$$\begin{array}{ll} Painting \sqsubseteq ArtPiece & PaintedBy \sqsubseteq ArtAuthor \\ Painting \sqsubseteq \exists HasGenre & ArtMovement \sqsubseteq HasStyle \end{array}$$

Further, assume $\mathcal{T}_1 = \{Painting \equiv \exists PaintedBy, Painting \sqsubseteq \exists ArtMovement\}$ and $\mathcal{A}_1 = \{Painting(\mathsf{blacksquare})\}$. Then, a universal solution for the KB $\mathcal{K}_1 = \langle \mathcal{T}_1, \mathcal{A}_1\rangle$ under $\mathcal{M}$ is the KB $\mathcal{K}_2 = \langle \mathcal{T}_2, \mathcal{A}_2\rangle$, where $\mathcal{T}_2 = \emptyset$ and $\mathcal{A}_2$ is the following ABox, where $\_n01$, $\_n02$, and $\_m01$ are labelled nulls:

$$\begin{array}{ll} ArtPiece(\mathsf{blacksquare}) & ArtAuthor(\mathsf{blacksquare}, \_n01) \\ HasGenre(\mathsf{blacksquare}, \_m01) & HasStyle(\mathsf{blacksquare}, \_n02) \end{array}$$

Now, consider KB $\mathcal{K}_2' = \langle \mathcal{T}_2', \mathcal{A}_2'\rangle$ with non-empty TBox, where $\mathcal{T}_2' = \{ArtPiece \equiv \exists ArtAuthor, ArtPiece \sqsubseteq \exists HasStyle, ArtPiece \sqsubseteq \exists HasGenre\}$ and $\mathcal{A}_2' = \{ArtPiece(\mathsf{blacksquare})\}$. Then we have that $\mathcal{K}_2'$ is a solution for $\mathcal{K}_1$ under $\mathcal{M}$. However, we also have that $\mathcal{K}_2'$ is not a universal solution for $\mathcal{K}_1$ under $\mathcal{M}$. Notably, both KB $\mathcal{K}_2$ and KB $\mathcal{K}_2'$ are universal UCQ-solutions for KB $\mathcal{K}_1$ under mapping $\mathcal{M}$. ■

In order to understand the capacity of universal solutions, and also of the query-languages based notions of solutions to transfer implicit knowledge, the notion of *representability* has been introduced in [1, 2]. Here we adapt that definition to the case where the KB is always satisfiable, as in *DL-Lite*$_{\mathcal{R}}^{pos}$. In the definition below we use $chase_{\mathcal{T},\Sigma}(\mathcal{A})$ to denote the projection of $chase_{\mathcal{T}}(\mathcal{A})$ on the signature Σ.

Let $\mathcal{L}$ be a DL, $\mathcal{Q}$ a class of queries, $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$ an $\mathcal{L}$-mapping, and $\mathcal{T}_1$ an $\mathcal{L}$-TBox over Σ_1. Then,

- $\mathcal{T}_1$ is *($\mathcal{Q}$-)representable under* $\mathcal{M}$ if there exists an $\mathcal{L}$-TBox $\mathcal{T}_2$ over Σ_2, called a *($\mathcal{Q}$-)representation of* $\mathcal{T}_1$ *under* $\mathcal{M}$, such that for every ABox $\mathcal{A}_1$ over Σ_1, $\langle \mathcal{T}_2, chase_{\mathcal{T}_{12},\Sigma_2}(\mathcal{A}_1)\rangle$ is a ($\mathcal{Q}$-)universal solution for $\langle \mathcal{T}_1, \mathcal{A}_1\rangle$ under $\mathcal{M}$.
- $\mathcal{T}_1$ is *weakly ($\mathcal{Q}$-)representable under* $\mathcal{M}$ if there exists a mapping $\mathcal{M}^\star = (\Sigma_1, \Sigma_2, \mathcal{T}_{12}^\star)$ such that $\mathcal{T}_{12} \subseteq \mathcal{T}_{12}^\star$, $\mathcal{T}_1 \cup \mathcal{T}_{12} \models \mathcal{T}_{12}^\star$, and $\mathcal{T}_1$ is ($\mathcal{Q}$-)representable under $\mathcal{M}^\star$.

Example 2. Let $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$ and $\mathcal{T}_1$ be as in Example 1. Then we have that $\mathcal{T}_2 = \{ArtPiece \equiv \exists ArtAuthor, ArtPiece \sqsubseteq \exists HasStyle, ArtPiece \sqsubseteq \exists HasGenre\}$ is a UCQ-representation of $\mathcal{T}_1$ under $\mathcal{M}$.

On the other hand, if $\mathcal{M}' = (\Sigma_1, \Sigma_2, \mathcal{T}_{12}')$ with $\mathcal{T}_{12}' = \{PaintedBy \sqsubseteq ArtPiece\}$, then we have that $\mathcal{T}_1$ is not UCQ-representable under $\mathcal{M}'$: take ABox $\mathcal{A}_1 = \{Painting(\mathsf{blacksquare})\}$, then $chase_{\mathcal{T}_{12}',\Sigma_2}(\mathcal{A}_1) = \emptyset$ and for no TBox $\mathcal{T}_2'$, $\langle \mathcal{T}_2', chase_{\mathcal{T}_{12}',\Sigma_2}(\mathcal{A}_1)\rangle$ is a universal UCQ-solution for $\langle \mathcal{T}_1, \mathcal{A}_1\rangle$ under $\mathcal{M}'$. However, if we consider $\mathcal{T}_{12}^\star = \mathcal{T}_{12}' \cup \{Painting \sqsubseteq \exists ArtAuthor\}$, we conclude that $\mathcal{T}_1$ is weakly UCQ-representable under $\mathcal{M}'$ since $\mathcal{T}_{12}' \subseteq \mathcal{T}_{12}^\star$, $\mathcal{T}_1 \cup \mathcal{T}_{12}' \models \mathcal{T}_{12}^\star$ and $\mathcal{T}_1$ is UCQ-representable under $\mathcal{M}^\star = (\Sigma_1, \Sigma_2, \mathcal{T}_{12}^\star)$ (in fact, $\emptyset$ is a UCQ-representation of $\mathcal{T}_1$ under $\mathcal{M}^\star$). ■

3 Solving UCQ-Representability for *DL-Lite*$_{\mathcal{R}}^{pos}$

In this section, we show that the UCQ-representability problem can be solved in polynomial time for the case where TBoxes and mappings are expressed in *DL-Lite*$_{\mathcal{R}}^{pos}$. More specifically, we give a polynomial time algorithm UCQREP^{pos} that, given a *DL-Lite*$_{\mathcal{R}}^{pos}$-mapping $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$ and a *DL-Lite*$_{\mathcal{R}}^{pos}$-TBox $\mathcal{T}_1$, verifies whether $\mathcal{T}_1$ is UCQ-representable under $\mathcal{M}$, and if this is the case computes a UCQ-representation of $\mathcal{T}_1$ under $\mathcal{M}$. Moreover, we also show that this algorithm can be used to solve the UCQ-representability problem for the case of *DL-Lite*$_{RDFS}$. It is important to notice that the algorithm we present can be used to compute universal UCQ-solutions of polynomial size, which make good use of the source implicit knowledge. Thus, this algorithm computes solutions with good properties to be used in practice.

A related problem is that of query inseparability [7], which can be formulated as follows: given TBoxes $\mathcal{T}_1$ and $\mathcal{T}_2$, and a signature Σ, decide whether for each ABox $\mathcal{A}$ over Σ and for each query q over Σ, $cert(q, \langle \mathcal{T}_1, \mathcal{A}\rangle) = cert(q, \langle \mathcal{T}_2, \mathcal{A}\rangle)$. In contrast to our polynomial result for UCQ-representability, query inseparability has been proved to be PSPACE-hard for *DL-Lite*$_{\mathcal{R}}$ TBoxes and CQs [7], and an analysis of the proof shows that the same lower bound holds already for *DL-Lite*$_{\mathcal{R}}^{pos}$.

3.1 Checking Whether a Given Target TBox is a UCQ-Representation

We start by considering the decision problem associated with UCQ-representability: Given a *DL-Lite*$_{\mathcal{R}}$-mapping $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$, a *DL-Lite*$_{\mathcal{R}}$-TBox $\mathcal{T}_1$ over Σ_1, and a *DL-Lite*$_{\mathcal{R}}$-TBox $\mathcal{T}_2$ over Σ_2, check whether $\mathcal{T}_2$ is a UCQ-representation of $\mathcal{T}_1$ under $\mathcal{M}$, i.e., for each ABox $\mathcal{A}_1$ over Σ_1, $\langle \mathcal{T}_2, \mathit{chase}_{\mathcal{T}_{12},\Sigma_2}(\mathcal{A}_1)\rangle$ is a universal UCQ-solution for $\langle \mathcal{T}_1, \mathcal{A}_1\rangle$ under $\mathcal{M}$. This problem can be solved in two steps:

(C1) Check whether for each ABox $\mathcal{A}_1$ over Σ_1, $\langle \mathcal{T}_2, \mathit{chase}_{\mathcal{T}_{12},\Sigma_2}(\mathcal{A}_1)\rangle$ is a UCQ-solution for $\langle \mathcal{T}_1, \mathcal{A}_1\rangle$ under $\mathcal{M}$.
(C2) Check whether for each ABox $\mathcal{A}_1$ over Σ_1 and for each UCQ q over Σ_2, we have that $\mathit{cert}(q, \langle \mathcal{T}_2, \mathit{chase}_{\mathcal{T}_{12},\Sigma_2}(\mathcal{A}_1)\rangle) \subseteq \mathit{cert}(q, \langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1\rangle)$.

If both checks succeed, then $\mathcal{T}_2$ is a UCQ-representation of $\mathcal{T}_1$ under $\mathcal{M}$, otherwise not. We develop now techniques to perform these two checks in polynomial time.

Checking Condition (C1). For a *DL-Lite*$_{\mathcal{R}}^{pos}$ TBox $\mathcal{T}$ and a concept or role N, we define the *upward closure of* N *w.r.t.* $\mathcal{T}$ as the set $\mathbb{U}_{\mathcal{T}}(N) = \{N' \mid N' \text{ is concept or role and } \mathcal{T} \models N \sqsubseteq N'\}$, and the *strict upward closure* $\mathbb{S}_{\mathcal{T}}(N)$ as $\mathbb{U}_{\mathcal{T}}(N) \setminus \{N\}$. Then, for a set $\mathbf{N}$ of concepts and roles we define $\mathbb{U}_{\mathcal{T}}(\mathbf{N}) = \bigcup_{N \in \mathbf{N}} \mathbb{U}_{\mathcal{T}}(N)$, and its strict version $\mathbb{S}_{\mathcal{T}}(\mathbf{N})$. Notice that both $\mathbb{S}_{\mathcal{T}_{12}}(\mathbb{U}_{\mathcal{T}_1}(N))$ and $\mathbb{U}_{\mathcal{T}_2}(\mathbb{S}_{\mathcal{M}}(N))$ are sets over Σ_2, for each concept or role N over Σ_1.

With these notions in place, we can provide a necessary and sufficient condition for the satisfaction of condition (C1).

Proposition 1. *Let* $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$ *be a DL-Lite*$_{\mathcal{R}}^{pos}$*-mapping,* $\mathcal{T}_1$ *a DL-Lite*$_{\mathcal{R}}^{pos}$*-TBox over* Σ_1*, and* $\mathcal{T}_2$ *a DL-Lite*$_{\mathcal{R}}^{pos}$*-TBox over* Σ_2*. Then* $\mathcal{T}_1$*,* $\mathcal{T}_2$*, and* $\mathcal{M}$ *satisfy condition (C1) iff the following conditions are satisfied:*

(A) $\mathbb{S}_{\mathcal{M}}(\mathbb{U}_{\mathcal{T}_1}(B)) \subseteq \mathbb{U}_{\mathcal{T}_2}(\mathbb{S}_{\mathcal{M}}(B))$*, for each basic concept* B *over* Σ_1*;*
(B) $\mathbb{S}_{\mathcal{M}}(\mathbb{U}_{\mathcal{T}_1}(R)) \subseteq \mathbb{U}_{\mathcal{T}_2}(\mathbb{S}_{\mathcal{M}}(R))$*, for each basic role* R *over* Σ_1*;*
(C) *for each basic concept* B *and each basic role* R *over* Σ_1 *such that* $\exists R \in \mathbb{U}_{\mathcal{T}_1}(B)$ *and* $\mathbb{S}_{\mathcal{M}}(\mathbb{U}_{\mathcal{T}_1}(\exists R^-)) \neq \emptyset$*, we have that*
if $\mathbb{S}_{\mathcal{M}}(\mathbb{U}_{\mathcal{T}_1}(R)) \neq \emptyset$*, then there exists a role* $Q_{R,B}$ *over* Σ_2 *such that*
(CA) $\exists Q_{R,B} \in \mathbb{U}_{\mathcal{T}_2}(\mathbb{S}_{\mathcal{M}}(B))$*,*
(CB) $\mathbb{S}_{\mathcal{M}}(\mathbb{U}_{\mathcal{T}_1}(R)) \subseteq \mathbb{U}_{\mathcal{T}_2}(Q_{R,B})$*, and*
(CC) $\mathbb{S}_{\mathcal{M}}(\mathbb{U}_{\mathcal{T}_1}(\exists R^-)) \subseteq \mathbb{U}_{\mathcal{T}_2}(\exists Q_{R,B}^-)$*,*
and if $\mathbb{S}_{\mathcal{M}}(\mathbb{U}_{\mathcal{T}_1}(R)) = \emptyset$*, either*
(CD) $\mathbb{S}_{\mathcal{M}}(\mathbb{U}_{\mathcal{T}_1}(\exists R^-)) \subseteq \mathbb{U}_{\mathcal{T}_2}(\mathbb{S}_{\mathcal{M}}(B))$*,*
or there exist roles $Q_{R,B}^1, \ldots, Q_{R,B}^n$ *over* Σ_2 *such that*
(CE) $\exists Q_{R,B}^1 \in \mathbb{U}_{\mathcal{T}_2}(\mathbb{S}_{\mathcal{M}}(B))$*,*
(CF) $\mathcal{T}_2 \models \exists(Q_{R,B}^1)^- \sqsubseteq \exists Q_{R,B}^2, \ldots, \exists(Q_{R,B}^{n-1})^- \sqsubseteq \exists Q_{R,B}^n$*, and*
(CG) $\mathbb{S}_{\mathcal{M}}(\mathbb{U}_{\mathcal{T}_1}(\exists R^-)) \subseteq \mathbb{U}_{\mathcal{T}_2}(\exists(Q_{R,B}^n)^-)$*.*

It is important to notice that the necessary and sufficient condition in Proposition 1 can be checked in polynomial time, as the implication problem for *DL-Lite*$_{\mathcal{R}}^{pos}$ can be solved in polynomial time. In particular, for a basic concept B and a basic role R over Σ_1 such that $\exists R \in \mathbb{U}_{\mathcal{T}_1}(B)$, $\mathbb{S}_{\mathcal{M}}(\mathbb{U}_{\mathcal{T}_1}(\exists R^-)) \neq \emptyset$, $\mathbb{S}_{\mathcal{M}}(\mathbb{U}_{\mathcal{T}_1}(R)) = \emptyset$, and $\mathbb{S}_{\mathcal{M}}(\mathbb{U}_{\mathcal{T}_1}(\exists R^-)) \not\subseteq \mathbb{U}_{\mathcal{T}_2}(\mathbb{S}_{\mathcal{M}}(B))$, checking the existence of roles $Q_{R,B}^1, \ldots, Q_{R,B}^n$ over Σ_2 satisfying conditions **(CE)**, **(CF)**, and **(CG)** can be reduced to checking reachability in a directed graph. Indeed, for each pair of basic concepts B_2, B_2' over Σ_2

such that $B_2 \in \mathbb{S}_{\mathcal{M}}(B)$ and $\mathbb{S}_{\mathcal{M}}(\mathbb{U}_{\mathcal{T}_1}(\exists R^-)) \subseteq \mathbb{U}_{\mathcal{T}_2}(B'_2)$, we use the following approach to check for the existence of the roles $Q^1_{R,B}, \ldots, Q^n_{R,B}$ over Σ_2 such that $\mathcal{T}_2 \models B_2 \sqsubseteq \exists Q^1_{R,B}, \mathcal{T}_2 \models \exists (Q^1_{R,B})^- \sqsubseteq \exists Q^2_{R,B}, \ldots, \mathcal{T}_2 \models \exists (Q^n_{R,B})^- \sqsubseteq B'_2$. Let $G = (V, E)$ be the directed graph defined as:

$$
\begin{aligned}
V &= \{B_2, B'_2\} \cup \{Q_2 \mid Q_2 \text{ is a role in } \Sigma_2\} \\
E &= \{(B_2, B'_2) \mid \mathcal{T}_2 \models B_2 \sqsubseteq B'_2\} \cup \{(B_2, Q_2) \mid \mathcal{T}_2 \models B_2 \sqsubseteq \exists Q_2\} \cup \\
&\quad\ \{(Q_2, B'_2) \mid \mathcal{T}_2 \models \exists Q_2^- \sqsubseteq B'_2\} \cup \{(Q_2, Q'_2) \mid \mathcal{T}_2 \models \exists Q_2^- \sqsubseteq \exists Q'_2\}
\end{aligned}
$$

Then we test for the existence of the roles $Q^1_{R,B}, \ldots, Q^n_{R,B}$ by verifying whether B'_2 is reachable from B_2 in G. If for some pair B_2, B'_2 the aforementioned two-step test succeed, then we have that there exist roles $Q^1_{R,B}, \ldots, Q^n_{R,B}$ that satisfy conditions **(CE)**, **(CF)**, and **(CG)**. Otherwise, we know that such roles do not exist.

Checking Condition (C2). We rely on the following result:

Proposition 2. *Let $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$ be a DL-Lite$^{pos}_{\mathcal{R}}$-mapping, $\mathcal{T}_1$ a DL-Lite$^{pos}_{\mathcal{R}}$-TBox over Σ_1, and $\mathcal{T}_2$ a DL-Lite$^{pos}_{\mathcal{R}}$-TBox over Σ_2. Then $\mathcal{T}_1$, $\mathcal{T}_2$, and $\mathcal{M}$ satisfy condition (C2) iff the following conditions are satisfied:*

(A) $\mathbb{U}_{\mathcal{T}_2}(\mathbb{S}_{\mathcal{M}}(B)) \subseteq \mathbb{S}_{\mathcal{M}}(\mathbb{U}_{\mathcal{T}_1}(B))$ *for each basic concept B over Σ_1;*
(B) $\mathbb{U}_{\mathcal{T}_2}(\mathbb{S}_{\mathcal{M}}(R)) \subseteq \mathbb{S}_{\mathcal{M}}(\mathbb{U}_{\mathcal{T}_1}(R))$ *for each role $R \in \Sigma_1$;*
(C) *for each basic role Q over Σ_2 and each basic concept B over Σ_1 such that $\exists Q \in \mathbb{U}_{\mathcal{T}_2}(\mathbb{S}_{\mathcal{M}}(B))$ and $\mathbb{U}_{\mathcal{T}_2}(\exists Q^-) \neq \{\exists Q^-\}$, there exists a role $R_{Q,B}$ over Σ_1 s.t.*
  (CA) $\exists R_{Q,B} \in \mathbb{U}_{\mathcal{T}_1}(B)$ *and*
  (CB) $Q \in \mathbb{S}_{\mathcal{M}}(R_{Q,B})$.

The necessary and sufficient condition in Proposition 2 can be checked in polynomial time, as the implication problem can be solved in polynomial time for *DL-Lite*$^{pos}_{\mathcal{R}}$.

Thus, given that, by Propositions 1 and 2, both conditions (C1) and (C2) can be tested in polynomial time, we obtain the following result.

Theorem 1. *The problem of verifying, given a DL-Lite$^{pos}_{\mathcal{R}}$-mapping $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$, a DL-Lite$^{pos}_{\mathcal{R}}$-TBox $\mathcal{T}_1$ over Σ_1, and a DL-Lite$^{pos}_{\mathcal{R}}$-TBox $\mathcal{T}_2$ over Σ_2, whether $\mathcal{T}_2$ is a UCQ-representation of $\mathcal{T}_1$ under $\mathcal{M}$ can be solved in polynomial time.*

3.2 The Algorithm for Computing a UCQ-Representation

In what follows, we present the algorithm UCQREP^{pos}, which verifies whether a source TBox $\mathcal{T}_1$ is UCQ-representable under a mapping $\mathcal{M}$, and if this is the case returns a UCQ-representation of $\mathcal{T}_1$ under $\mathcal{M}$.

Intuitively, given a source TBox $\mathcal{T}_1$ and a mapping $\mathcal{M}$, algorithm UCQREP^{pos} constructs the best possible candidate $\mathcal{T}_2$ for a UCQ-representation of $\mathcal{T}_1$ under $\mathcal{M}$ (given the conditions in Propositions 1 and 2), and then checks whether $\mathcal{T}_2$ effectively satisfies the properties required for a UCQ-representation (note, that $\mathcal{T}_2$ is indeed a *DL-Lite*$^{pos}_{\mathcal{R}}$ TBox). To prove the correctness of this algorithm, we need to show that $\mathcal{T}_1$ is UCQ-representable under $\mathcal{M}$ if and only if $\mathcal{T}_2$ is a UCQ-representation of $\mathcal{T}_1$ under $\mathcal{M}$. This

Algorithm: UCQREP$^{pos}(\mathcal{T}_1, \mathcal{M})$
Input: A *DL-Lite*$_{\mathcal{R}}^{pos}$-mapping $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$ and a *DL-Lite*$_{\mathcal{R}}^{pos}$-TBox $\mathcal{T}_1$ over Σ_1.
Output: A *DL-Lite*$_{\mathcal{R}}^{pos}$-TBox $\mathcal{T}_2$ over Σ_2 that is a UCQ-representation of $\mathcal{T}_1$ under $\mathcal{M}$, if $\mathcal{T}_1$ is UCQ-representable under $\mathcal{M}$. The keyword *false* otherwise.

1. Let $\mathcal{T}_2$ be a TBox over Σ_2 defined as:
$$\mathcal{T}_2 = \{N_2 \sqsubseteq M_2 \mid N_1 \text{ a basic concept or role over } \Sigma_1,\ N_2 \in \mathbb{S}_{\mathcal{M}}(N_1), M_2 \in \mathbb{S}_{\mathcal{M}}(\mathbb{U}_{\mathcal{T}_1}(N_1))\}$$
2. Remove from $\mathcal{T}_2$ every inclusion $N_2 \sqsubseteq M_2$ such that *(i)* $N_2 \in \mathbb{S}_{\mathcal{M}}(N_1)$ for some N_1 over Σ_1, and *(ii)* for every M_1 over Σ_1 such that $M_2 \in \mathbb{S}_{\mathcal{M}}(M_1)$, it holds that $\mathcal{T}_1 \not\models N_1 \sqsubseteq M_1$. Moreover, if $N_2 = \exists R_2$ and $M_2 = \exists R_2'$, then also remove inclusions $R_2 \sqsubseteq R_2'$ and $R_2^- \sqsubseteq {R_2'}^-$ from $\mathcal{T}_2$.
3. Remove from $\mathcal{T}_2$ every inclusion of the form either $\exists R_2^- \sqsubseteq B_2$ or $R_2 \sqsubseteq R_2'$ or $R_2^- \sqsubseteq R_2'$ for roles R_2, R_2' and a concept B_2 over Σ_2, if there exists a concept B_1 over Σ_1 such that *(i)* $\exists R_2 \in \mathbb{S}_{\mathcal{M}}(B_1)$, and *(ii)* for every role R_1 over Σ_1 such that $\exists R_1 \in \mathbb{U}_{\mathcal{T}_1}(B_1)$ and $R_2 \in \mathbb{S}_{\mathcal{M}}(R_1)$, it holds that $\mathcal{T}_1 \not\models B_1 \sqsubseteq \exists R_1$.
4. Verify whether $\mathcal{T}_2$ is a UCQ-representation of $\mathcal{T}_1$ under $\mathcal{M}$. If the test succeeds, return $\mathcal{T}_2$, otherwise return *false*.

Fig. 1. Algorithm to compute the UCQ-representation of a *DL-Lite*$_{\mathcal{R}}^{pos}$ TBox $\mathcal{T}_1$ under a *DL-Lite*$_{\mathcal{R}}^{pos}$ mapping $\mathcal{M}$.

is done in the following theorem, where it is also proved that the algorithm works in polynomial time. The latter is a consequence of the fact that $\mathcal{T}_2$ is of polynomial size in the sizes of $\mathcal{T}_1$ and $\mathcal{M}$, and that, by Theorem 1, it is possible to check in polynomial time whether $\mathcal{T}_2$ is a UCQ-representation of $\mathcal{T}_1$ under $\mathcal{M}$.

Theorem 2. *Algorithm* UCQREP^{pos} *is correct and runs in polynomial time.*

The following examples illustrate how algorithm UCQREP^{pos} works.

Example 3. Let $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$, where $\Sigma_1 = \{A_1(\cdot), B_1(\cdot), C_1(\cdot)\}$, $\Sigma_2 = \{A_2(\cdot), B_2(\cdot)\}$, and $\mathcal{T}_{12} = \{A_1 \sqsubseteq A_2, B_1 \sqsubseteq B_2, C_1 \sqsubseteq B_2\}$. Furthermore, assume that $\mathcal{T}_1 = \{B_1 \sqsubseteq A_1\}$. Then, in step 1, the algorithm constructs the TBox $\mathcal{T}_2 = \{B_2 \sqsubseteq A_2\}$. In step 2, it removes the only axiom from $\mathcal{T}_2$ as $B_2 \in \mathbb{S}_{\mathcal{M}}(C_1)$ and $\mathcal{T}_1 \not\models C_1 \sqsubseteq A_1$. In step 3, it does nothing, and finally, at the last step it checks whether the empty TBox $\mathcal{T}_2$ is a UCQ-representation of $\mathcal{T}_1$ under $\mathcal{M}$. Since $A_2 \in \mathbb{S}_{\mathcal{M}}(\mathbb{U}_{\mathcal{T}_1}(B_1))$ and $A_2 \notin \mathbb{U}_{\mathcal{T}_2}(\mathbb{S}_{\mathcal{M}}(B_1))$, the algorithm returns *false*. ■

Example 4. Let $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$, where $\Sigma_1 = \{B_1(\cdot), P_1(\cdot,\cdot), R_1(\cdot,\cdot)\}$, $\Sigma_2 = \{A_2(\cdot), B_2(\cdot), R_2(\cdot,\cdot)\}$, and $\mathcal{T}_{12} = \{\exists P_1^- \sqsubseteq A_2, B_1 \sqsubseteq B_2, R_1 \sqsubseteq R_2\}$. Furthermore, assume that $\mathcal{T}_1 = \{B_1 \sqsubseteq \exists P_1, B_1 \sqsubseteq \exists R_1, \exists R_1^- \sqsubseteq \exists P_1^-\}$. Then, in step 1, the algorithm constructs the TBox $\mathcal{T}_2 = \{B_2 \sqsubseteq \exists R_2, \exists R_2^- \sqsubseteq A_2\}$. It does not remove anything in steps 2 and 3. Finally, at the last step it successfully checks that $\mathcal{T}_2$ is a UCQ-representation of $\mathcal{T}_1$ under $\mathcal{M}$ and returns $\mathcal{T}_2$. ■

3.3 Solving UCQ-Representability for *DL-Lite*$_{RDFS}$

It is not difficult to see that if the input of algorithm UCQREP^{pos} is a *DL-Lite*$_{RDFS}$-mapping $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$ and a *DL-Lite*$_{RDFS}$-TBox $\mathcal{T}_1$ over Σ_1, then the set $\mathcal{T}_2$ computed by this algorithm is a *DL-Lite*$_{\mathcal{R}}^{pos}$-TBox over Σ_2 that can be easily transformed into an equivalent *DL-Lite*$_{RDFS}$-TBox. Indeed, $\mathcal{T}_2$ might contain inclusions between basic concepts of the form $\exists R_2 \sqsubseteq \exists R'_2$, but this occurs only if $\mathcal{T}_2$ implies also the role inclusion $R_2 \sqsubseteq R'_2$. Hence, all concept inclusions that would fall outside *DL-Lite*$_{RDFS}$ are implied by the *DL-Lite*$_{RDFS}$ fragment of $\mathcal{T}_2$ and can be removed from $\mathcal{T}_2$ without affecting its semantics. Thus, we conclude that algorithm UCQREP^{pos} can also be used to solve in polynomial time the UCQ-representability problem for *DL-Lite*$_{RDFS}$ mappings and TBoxes.

4 Solving Weak UCQ-Representability for *DL-Lite*$_{\mathcal{R}}^{pos}$

In this section, we show that also the weak UCQ-representability problem can be solved in polynomial time when TBoxes and mappings are expressed *DL-Lite*$_{\mathcal{R}}^{pos}$. We first need to introduce some terminology. Given a *DL-Lite*$_{\mathcal{R}}$-TBox $\mathcal{T}$ over a signature Σ and a UCQ q over Σ, a UCQ q_r over Σ is said to be a *perfect reformulation* of q w.r.t. $\mathcal{T}$ if for every ABox $\mathcal{A}$ over Σ, it holds that [6]: $cert(q, \langle \mathcal{T}, \mathcal{A} \rangle) = cert(q_r, \langle \emptyset, \mathcal{A} \rangle)$. That is, the certain answers to the UCQ q over a KB $\langle \mathcal{T}, \mathcal{A} \rangle$ can be computed by posing the UCQ q_r over the ABox $\mathcal{A}$. It is well-known that every UCQ q admits a perfect reformulation w.r.t. a *DL-Lite*$_{\mathcal{R}}$-TBox $\mathcal{T}$, which can be computed in polynomial time [6].

Interestingly, the fundamental notion of perfect reformulation can be used to solve the UCQ-representability problem for *DL-Lite*$_{\mathcal{R}}^{pos}$. More precisely, let $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$ be a *DL-Lite*$_{\mathcal{R}}$-mapping and $\mathcal{T}_1$ a *DL-Lite*$_{\mathcal{R}}$-TBox over Σ_1. Then define a mapping COMP$(\mathcal{M}, \mathcal{T}_1) = (\Sigma_1, \Sigma_2, \mathcal{T}_{12}^\star)$ that extends $\mathcal{M}$ by compiling the knowledge from $\mathcal{T}_1$ into $\mathcal{T}_{12}$. Formally, for a basic concept or role N over Σ_1, let bq_N be the CQ defined as follows: $bq_A(x) = A(x)$, $bq_{\exists P}(x) = \exists y.P(x,y)$, $bq_{\exists P^-}(x) = \exists y.P(y,x)$, $bq_P(x,y) = P(x,y)$, and $bq_{P^-}(x,y) = P(y,x)$. Then, for every concept inclusion $B \sqsubseteq C \in \mathcal{T}_{12}$ and for every CQ q in the perfect reformulation of bq_B w.r.t. $\mathcal{T}_1$, include $C_q \sqsubseteq C$ into $\mathcal{T}_{12}^\star$, where C_q is the (unique) basic concept such that $bq_{C_q} = q$. Also, for every role inclusion $R \sqsubseteq Q \in \mathcal{T}_{12}$ and for every CQ q in the perfect reformulation of bq_R w.r.t. $\mathcal{T}_1$, include $R_q \sqsubseteq Q$ into $\mathcal{T}_{12}^\star$, where R_q is the basic role such that $bq_{R_q} = q$.

It is important to notice that if $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$ is a *DL-Lite*$_{\mathcal{R}}^{pos}$-mapping and $\mathcal{T}_1$ a *DL-Lite*$_{\mathcal{R}}^{pos}$-TBox over Σ_1, then COMP$(\mathcal{M}, \mathcal{T}_1) = (\Sigma_1, \Sigma_2, \mathcal{T}_{12}^\star)$ is a *DL-Lite*$_{\mathcal{R}}^{pos}$-mapping that can be computed in polynomial time in the sizes of $\mathcal{M}$ and $\mathcal{T}_1$. Therefore, given that the set of inclusions defining COMP$(\mathcal{M}, \mathcal{T}_1)$ contains the set of inclusions defining $\mathcal{M}$ and $\mathcal{T}_1 \cup \mathcal{T}_{12} \models \mathcal{T}_{12}^\star$, we conclude that COMP$(\mathcal{M}, \mathcal{T}_1)$ can be used to check in polynomial time whether $\mathcal{T}_1$ is weakly UCQ-representable under $\mathcal{M}$.

Theorem 3. *Let $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$ be a DL-Lite$_{\mathcal{R}}^{pos}$-mapping and $\mathcal{T}_1$ a DL-Lite$_{\mathcal{R}}^{pos}$-TBox over Σ_1. Then $\mathcal{T}_1$ is weakly UCQ-representable under $\mathcal{M}$ if and only if $\mathcal{T}_1$ is UCQ-representable under* COMP$(\mathcal{M}, \mathcal{T}_1)$.

From this result and Theorem 2 we obtain a polynomial time algorithm for solving the weak UCQ-representability problem for *DL-Lite*$_{\mathcal{R}}^{pos}$ mappings and TBoxes.

The example below shows a *DL-Lite*$_{\mathcal{R}}^{pos}$ TBox $\mathcal{T}_1$ and a *DL-Lite*$_{\mathcal{R}}^{pos}$ mapping $\mathcal{M}$ such that $\mathcal{T}_1$ is not weakly UCQ-representable under $\mathcal{M}$.

Example 5. Let $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$, where $\Sigma_1 = \{P_1(\cdot,\cdot), B_1(\cdot)\}$, $\Sigma_2 = \{A_2(\cdot), B_2(\cdot)\}$, and $\mathcal{T}_{12} = \{B_1 \sqsubseteq B_2, \exists P_1^- \sqsubseteq A_2\}$. Furthermore, assume that $\mathcal{T}_1 = \{B_1 \sqsubseteq \exists P_1\}$. Then $\mathcal{T}_1$ is not weakly UCQ-representable under $\mathcal{M}$. In fact, given that the perfect reformulation of $\exists P_1^-$ w.r.t. TBox $\mathcal{T}_1$ is $\exists P_1^-$ itself, and likewise for concept B_1, we have that $\mathcal{M} = \textsc{Comp}(\mathcal{M}, \mathcal{T}_1)$ and, thus, $\mathcal{T}_1$ is not weakly UCQ-representable under $\mathcal{M}$, as $\mathcal{T}_1$ is not UCQ-representable under $\mathcal{M}$. ∎

Instead, as shown in [1,2], for each *DL-Lite*$_{RDFS}$ TBox $\mathcal{T}_1$ and *DL-Lite*$_{RDFS}$ mapping $\mathcal{M}$, $\mathcal{T}_1$ is weakly UCQ-representable under $\mathcal{M}$.

5 Conclusions

In this paper, we have extended previous results on representability in the knowledge exchange framework to *DL-Lite*$_{\mathcal{R}}^{pos}$, a DL of the *DL-Lite* family that allows for existentials in the right-hand side of inclusion assertions, both in the source TBox and in the mapping. We are currently working on extending our results to *DL-Lite*$_{\mathcal{R}}$, which includes disjointness assertions, and to the other DLs in the extended *DL-Lite* family [4]. A further interesting problem that we are investigating is that of checking the existence of universal solutions for *DL-Lite*$_{\mathcal{R}}^{pos}$ and other more expressive DLs.

Acknowledgements. The authors were supported by the EU Marie Curie action FP7-PEOPLE-2009-IRSES (grant 24761), by Fondecyt (grant 1090565), by the EU ICT projects ACSI (grant FP7-257593) and FOX (grant FP7-233599), and by the NWO project DEX (612.001.012).

References

1. Arenas, M., Botoeva, E., Calvanese, D.: Knowledge base exchange. In: Proc. of DL 2011. CEUR, ceur-ws.org, vol. 745 (2011)
2. Arenas, M., Botoeva, E., Calvanese, D., Ryzhikov, V., Sherkhonov, E.: Exchanging description logic knowledge bases. In: Proc. of KR 2012 (2012)
3. Arenas, M., Pérez, J., Reutter, J.L.: Data exchange beyond complete data. In: Proc. of PODS 2011. pp. 83–94 (2011)
4. Artale, A., Calvanese, D., Kontchakov, R., Zakharyaschev, M.: The *DL-Lite* family and relations. J. of Artificial Intelligence Research 36, 1–69 (2009)
5. Brickley, D., Guha, R.V.: RDF vocabulary description language 1.0: RDF Schema. W3C Recommendation, World Wide Web Consortium (Feb 2004), available at http://www.w3.org/TR/rdf-schema/
6. Calvanese, D., De Giacomo, G., Lembo, D., Lenzerini, M., Rosati, R.: Tractable reasoning and efficient query answering in description logics: The *DL-Lite* family. J. of Automated Reasoning 39(3), 385–429 (2007)
7. Konev, B., Kontchakov, R., Ludwig, M., Schneider, T., Wolter, F., Zakharyaschev, M.: Conjunctive query inseparability of OWL 2 QL TBoxes. In: Proc. of AAAI 2011. pp. 221–226 (2011)