Citation for published version


DOI

Link to record in KAR

https://kar.kent.ac.uk/91308/

Document Version

Publisher pdf
Representability in $DL$-$Lite_R$ 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 $DL$-$Lite_{RDFS}$, a DL corresponding to the FOL fragment of RDFS. We extend these results to the positive fragment of $DL$-$Lite_R$, in which, differently from $DL$-$Lite_{RDFS}$, 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$_{R}$ without disjointness assertions, a DL that we call DL-Lite$_{pos}$. 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$_{pos}^R$ TBox under a DL-Lite$_{pos}^R$ mapping in polynomial time and to construct a polynomial size representation.

2 Preliminaries

2.1 DL-Lite$_{pos}^R$ 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$_{pos}^R$, 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:

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

For a role $R$, we use $R^-$ to denote $P^-$ when $R = P$, and $P$ when $R = P^-$. A DL-Lite$_{pos}^R$ TBox is a finite set of concept inclusions $B \sqsubseteq B'$ and role inclusions $R \sqsubseteq R'$. A DL-Lite$_{pos}^R$ 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
for every role name

\[ (3) \quad a \subseteq \Delta \] for every individual name \( a \); and (4) such that:

\[
(\exists R)^I = \{ x \in \Delta \mid \text{there exists } y \in \Delta \text{ s.t. } (x, y) \in R \}
\]

\[
(P^-)^I = \{ (y, x) \in \Delta \times \Delta \mid (x, y) \in P \}
\]

Moreover, satisfaction of concept and role inclusions is defined as follows: \( \mathcal{I} \models B \subseteq B' \) if \( B \subseteq \Delta \) and \( \mathcal{I} \models R \subseteq R' \) if \( R \subseteq \Delta \times \Delta \). 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 \( \Delta \) such that \( h(a) = a' \) for each individual \( a \). Then \( (\mathcal{I}, h) \models A(u) \) if \( h(u) \in A \), and \( (\mathcal{I}, h) \models P(u, v) \) if \( (h(u), h(v)) \in P \).

An interpretation \( \mathcal{I} \) is a model of a DL-Lite KB \( \mathcal{T} \) if for every \( \alpha \in \mathcal{I} \), it holds that \( \mathcal{I} \models \alpha \), and it is a model of a DL-Lite 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 KB \( \mathcal{K} = (\mathcal{T}, \mathcal{A}) \) if \( \mathcal{I} \) is a model of both \( \mathcal{T} \) and \( \mathcal{A} \). The set of all models of \( \mathcal{K} \) is denoted \( \text{Mod}(\mathcal{K}) \), and \( \mathcal{K} \) is consistent if \( \text{Mod}(\mathcal{K}) \neq \emptyset \). We observe that in DL-Lite one cannot express any form of negative information, and hence a DL-Lite KB is always consistent.

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

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

### 2.2 Queries, Certain Answers, and Chase

A k-ary query \( q \) over a signature \( \Sigma \), with \( k \geq 0 \), is a function that maps every interpretation \( (\Delta, \cdot) \) of \( \Sigma \) into a k-ary relation \( q^\Delta \subseteq \Delta^k \). In particular, if \( k = 0 \), then \( q \) is called a Boolean query, and \( q^\Delta \) 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 \( K \) if \( q \) is a query over a signature \( \Sigma \) and \( \Sigma \subseteq \Sigma(K) \). Moreover, the answer to \( q \) over \( K \), denoted by \( \text{cert}(q, K) \), is defined as \( \text{cert}(q, K) = \bigcap_{I \in \text{Mod}(K)} q^I \).

Each tuple in \( \text{cert}(q, K) \) is called a certain answer for \( q \) over \( K \). Notice that if \( q \) is a Boolean query, then \( \text{cert}(q, K) \) evaluates to true if \( q^I \) evaluates to true for every \( I \in \text{Mod}(K) \), 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 \( \Sigma \) is a first-order formula of the form \( q(x) = \exists y.\text{conj}(x, y) \), where \( x, y \) are tuples of variables and \( \text{conj}(x, y) \) is a conjunction of atoms of the form: (1) \( A(t) \), with \( A \) a concept name in \( \Sigma \) and \( t \) either an individual from \( U \) or a variable from \( x \) or \( y \), or (2) \( P(t_1, t_2) \), with \( P \) a role name in \( \Sigma \) and \( t_i \) (\( i = 1, 2 \)) either an individual from \( U \) or a variable from \( x \) or \( y \). In a CQ \( q(x) = \exists y.\text{conj}(x, y) \) over a signature \( \Sigma \), \( x \) is the tuple of free variables of \( q(x) \). Moreover, given an interpretation \( I = (\Delta^I, \cdot^I) \) of \( \Sigma \), the answer of \( q \) over \( I \), denoted by \( q^I \), is defined as the set of tuples \( a \) of elements from \( \Delta^I \) for which there exist a tuple \( b \) of elements from \( \Delta^I \) such that \( I \) satisfies every conjunct in \( \text{conj}(a, b) \). Finally, a union of conjunctive queries (UCQ) over a signature \( \Sigma \) is a finite set of CQs over \( \Sigma \) that have the same free variables. A UCQ \( q(x) \) is interpreted as the first-order formula \( \bigvee_{q_i \in q} q_i(x) \), and its semantics is defined as \( q^I = \bigcup_{q_i \in q} q_i^I \).

Certain answers in DL-Lite\text{pos} \( _R \) 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 \( U \), or labeled nulls interpreted as not necessarily distinct domain elements (see the definition of the semantics of DL-Lite\text{pos} \( _R \) in Section 2.1). For DL-Lite\text{pos} \( _R \) KBs, we employ the notion of restricted chase as defined in [6]. For such a KB \( \langle T, A \rangle \), the chase of \( A \) w.r.t. \( T \), denoted \( \text{chase}_T(A) \), is a chase obtained from \( A \) by adding facts implied by inclusions in \( T \), 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 \( \Sigma_1, \Sigma_2 \) are signatures with no concepts or roles in common. Then we say that an inclusion \( N_1 \subseteq N_2 \) is an inclusion from \( \Sigma_1 \) to \( \Sigma_2 \), if \( N_1 \) is a concept or a role over \( \Sigma_1 \) and \( N_2 \) is a concept or a role over \( \Sigma_2 \). For a DL \( \mathcal{L} \) (e.g., DL-Lite\text{pos} \( _R \)), we define an \( \mathcal{L} \)-mapping (or just mapping, when \( \mathcal{L} \) is clear from the context) as a tuple \( \mathcal{M} = (\Sigma_1, \Sigma_2, T_{12}) \), where \( T_{12} \) is a TBox in \( \mathcal{L} \) consisting of inclusions from \( \Sigma_1 \) to \( \Sigma_2 \):

1. \( C_1 \subseteq C_2 \), where \( C_1, C_2 \) are concepts in \( \mathcal{L} \) over \( \Sigma_1 \) and \( \Sigma_2 \), respectively, and
2. \( Q_1 \subseteq Q_2 \), where \( Q_1 \) and \( Q_2 \) are roles in \( \mathcal{L} \) over \( \Sigma_1 \) and \( \Sigma_2 \), respectively.

If \( T_{12} \) is an \( \mathcal{L} \)-TBox, for a DL \( \mathcal{L} \) (e.g., DL-Lite\text{pos} \( _R \)), 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 \( \mathcal{M} = (\Sigma_1, \Sigma_2, T_{12}) \), an interpretation \( I \) of \( \Sigma_1 \) and an interpretation \( J \) of \( \Sigma_2 \), pair \( (I, J) \) satisfies TBox \( T_{12} \), denoted by \( (I, J) \models T_{12} \), if for each concept inclusion \( C_1 \subseteq C_2 \), it holds that \( C_1^I \subseteq C_2^I \), and for each role inclusion \( Q_1 \subseteq Q_2 \), it holds that \( Q_1^I \subseteq Q_2^I \). Moreover, given an interpretation \( I \) of \( \Sigma_1 \), \( \text{SAT}_{\mathcal{M}}(I) \) is defined as the set of interpretations \( J \) of \( \Sigma_2 \) such
that \((I, J) \models T_{12}\), and given a set \(\mathcal{X}\) of interpretations of \(\Sigma_1\), \(\text{SAT}_M(\mathcal{X})\) is defined as:
\[
\text{SAT}_M(\mathcal{X}) = \bigcup_{\mathcal{I} \in \mathcal{X}} \text{SAT}_M(\mathcal{I}).
\]
Let \(M = (\Sigma_1, \Sigma_2, T_{12})\) be a mapping, \(K_1\) a KB over \(\Sigma_1\), and \(K_2\) a KB over \(\Sigma_2\).

- \(K_2\) is called a solution for \(K_1\) under \(M\) if \(\text{MOD}(K_2) \subseteq \text{SAT}_M(\text{MOD}(K_1))\), and
- \(K_2\) is called a universal solution for \(K_1\) under \(M\) if \(\text{MOD}(K_2) = \text{SAT}_M(\text{MOD}(K_1))\).

Universal solutions present several limitations, as argued in [1, 2]. First, a universal solution (in DL-Lite\(_R^\text{mu}\) and DL-Lite\(_R\)) 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 \(Q\) be a class of queries, \(M = (\Sigma_1, \Sigma_2, T_{12})\) a mapping, \(K_1 = \langle T_1, A_1 \rangle\) a KB over \(\Sigma_1\), and \(K_2\) a KB over \(\Sigma_2\). Then

- \(K_2\) is called a \(Q\)-solution for \(K_1\) under \(M\) if for every query \(q \in Q\) over \(\Sigma_2\),
  \[
  \text{cert}(q, \langle T_1 \cup T_{12}, A_1 \rangle) \subseteq \text{cert}(q, K_2),
  \]
- \(K_2\) is called a universal \(Q\)-solution for \(K_1\) under \(M\) if for every query \(q \in Q\) over \(\Sigma_2\),
  \[
  \text{cert}(q, \langle T_1 \cup T_{12}, A_1 \rangle) = \text{cert}(q, K_2).
  \]

The definitions of solutions are illustrated in the following example.

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

\[
\begin{align*}
\text{Painting} & \sqsubseteq \text{ArtPiece} & \text{PaintedBy} & \sqsubseteq \text{ArtAuthor} \\
\text{Painting} & \sqsubseteq \exists \text{HasGenre} & \text{ArtMovement} & \sqsubseteq \text{HasStyle}
\end{align*}
\]

Further, assume \(T_1 = \{\text{Painting} \equiv \exists \text{PaintedBy}, \text{Painting} \sqsubseteq \exists \text{ArtMovement}\}\) and \(A_1 = \{\text{Painting(blacksquare)}\}\). Then, a universal solution for the KB \(K_1 = \langle T_1, A_1 \rangle\) under \(M\) is the KB \(K_2 = \langle T_2, A_2 \rangle\), where \(T_2 = \emptyset\) and \(A_2\) is the following ABox, where \(\_n01, \_n02\), and \(\_m01\) are labelled nulls:

\[
\begin{align*}
\text{ArtPiece(blacksquare)} & & \text{ArtAuthor(blacksquare, \_n01)} \\
\text{HasGenre(blacksquare, \_m01)} & & \text{HasStyle(blacksquare, \_n02)}
\end{align*}
\]

Now, consider KB \(K'_2 = \langle T'_2, A'_2 \rangle\) with non-empty TBox, where \(T'_2 = \{\text{ArtPiece} \equiv \exists \text{ArtAuthor}, \text{ArtPiece} \sqsubseteq \exists \text{HasStyle}, \text{ArtPiece} \sqsubseteq \exists \text{HasGenre}\}\) and \(A'_2 = \{\text{ArtPiece(blacksquare)}\}\). Then we have that \(K'_2\) is a solution for \(K_1\) under \(M\). However, we also have that \(K'_2\) is not a universal solution for \(K_1\) under \(M\). Notably, both KB \(K_2\) and KB \(K'_2\) are universal UCQ-solutions for KB \(K_1\) under mapping \(M\). ■

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\textsuperscript{pos}_R. In the definition below we use \(\text{chase}_\Sigma(T)(A)\) to denote the projection of \(\text{chase}_T(A)\) on the signature \(\Sigma\).

Let \(\mathcal{L}\) be a DL, \(Q\) a class of queries, \(M = (\Sigma_1, \Sigma_2, T_{12})\) an \(\mathcal{L}\)-mapping, and \(T_1\) an \(\mathcal{L}\)-TBox over \(\Sigma_1\). Then,

- \(T_1\) is \((Q-)\text{-representable under} M\) if there exists an \(\mathcal{L}\)-TBox \(T_2\) over \(\Sigma_2\), called a \((Q-)\text{-representation of} T_1 \text{under} M\), such that for every ABox \(A_1\) over \(\Sigma_1\), \(\langle T_2, \text{chase}_{T_{12}}(\Sigma_2)(A_1)\rangle\) is a \((Q-)\text{-universal solution for} \langle T_1, A_1 \rangle \text{under} M\).

- \(T_1\) is weakly \((Q-)\text{-representable under} M\) if there exists a mapping \(M^* = (\Sigma_1, \Sigma_2, T^*_1)\) such that \(T_{12} \subseteq T^*_1\), \(T_1 \cup T_{12} \models T^*_1\), and \(T_1\) is \((Q-)\text{-representable under} M^*\).

**Example 2.** Let \(M = (\Sigma_1, \Sigma_2, T_{12})\) and \(T_1\) be as in Example 1. Then we have that \(T_2 = \{\text{ArtPiece} \equiv \exists \text{ArtAuthor}, \text{ArtPiece} \subseteq \exists \text{HasStyle}, \text{ArtPiece} \subseteq \exists \text{HasGenre}\}\) is a UCQ-representation of \(T_1\) under \(M\).

On the other hand, if \(M' = (\Sigma_1, \Sigma_2, T'_{12})\) with \(T'_{12} = \{\text{PaintedBy} \subseteq \text{ArtPiece}\}\), then we have that \(T_1\) is not UCQ-representable under \(M'\): take ABox \(A_1 = \{\text{Painting}(\text{blacksquare})\}\), then \(\text{chase}_{T'_{12}}(\Sigma_2)(A_1) = \emptyset\) and for no TBox \(T_2\), \(\langle T_2, \text{chase}_{T'_{12}}(\Sigma_2)(A_1)\rangle\) is a universal UCQ-solution for \(\langle T_1, A_1 \rangle\) under \(M'\). However, if we consider \(T^*_1 = T_{12} \cup \{\text{Painting} \subseteq \exists \text{ArtAuthor}\}\), we conclude that \(T_1\) is weakly UCQ-representable under \(M'\) since \(T^*_1 \subseteq T_{12}, T_1 \cup T_{12} \models T^*_1\) and \(T_1\) is UCQ-representable under \(M^* = (\Sigma_1, \Sigma_2, T^*_1)\) (in fact, \(\emptyset\) is a UCQ-representation of \(T_1\) under \(M^*\)). □

### 3 Solving UCQ-Representability for DL-Lite\textsuperscript{pos}_R

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\textsuperscript{pos}_R. More specifically, we give a polynomial time algorithm UCQREP\textsuperscript{pos} that, given a DL-Lite\textsuperscript{pos}_R-mapping \(M = (\Sigma_1, \Sigma_2, T_{12})\) and a DL-Lite\textsuperscript{pos}_R-TBox \(T_1\), verifies whether \(T_1\) is UCQ-representable under \(M\), and if this is the case computes a UCQ-representation of \(T_1\) under \(M\). Moreover, we also show that this algorithm can be used to solve the UCQ-representability problem for the case of DL-Lite\textsuperscript{R$_{pos}$}. 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 \(T_1\) and \(T_2\), and a signature \(\Sigma\), decide whether for each ABox \(A\) over \(\Sigma\) and for each query \(q\) over \(\Sigma\), \(\text{cert}(q, \langle T_1, A \rangle) = \text{cert}(q, \langle T_2, A \rangle)\). In contrast to our polynomial result for UCQ-representability, query inseparability has been proved to be PSPACE-hard for DL-Lite\textsuperscript{R$_{pos}$} TBoxes and CQs [7], and an analysis of the proof shows that the same lower bound holds already for DL-Lite\textsuperscript{pos}_R.

#### 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\textsuperscript{R}-mapping \(M = (\Sigma_1, \Sigma_2, T_{12})\), a DL-Lite\textsuperscript{R}-TBox \(T_1\) over \(\Sigma_1\), and
a DL-Lite\textsubscript{R} TBox $T_2$ over $\Sigma_2$, check whether $T_2$ is a UCQ-representation of $T_1$ under $\mathcal{M}$, i.e., for each ABox $A_1$ over $\Sigma_1$, $\langle T_2, chase_{T_{12}, \Sigma_2}(A_1) \rangle$ is a universal UCQ-solution for $\langle T_1, A_1 \rangle$ under $\mathcal{M}$. This problem can be solved in two steps:

(C1) Check whether for each ABox $A_1$ over $\Sigma_1$, $\langle T_2, chase_{T_{12}, \Sigma_2}(A_1) \rangle$ is a UCQ-solution for $\langle T_1, A_1 \rangle$ under $\mathcal{M}$.

(C2) Check whether for each ABox $A_1$ over $\Sigma_1$ and for each UCQ $q$ over $\Sigma_2$, we have that $cert(q, \langle T_2, chase_{T_{12}, \Sigma_2}(A_1) \rangle) \subseteq cert(q, \langle T_1 \cup T_{12}, A_1 \rangle)$.

If both checks succeed, then $T_2$ is a UCQ-representation of $T_1$ under $\mathcal{M}$, otherwise not. We develop now techniques to perform these two checks in polynomial time.

Checking Condition (C1). For a DL-Lite\textsubscript{R} TBox $T$ and a concept or role $N$, we define the upward closure of $N$ w.r.t. $T$ as the set $U_T(N) = \{N' | N' \text{ is concept or role and } T \models N \subseteq N'\}$, and the strict upward closure $S_T(N)$ as $U_T(N) \setminus \{N\}$. Then, for a set $\Sigma$ of concepts and roles we define $U_T(\Sigma) = \bigcup_{N \in \Sigma} U_T(N)$, and its strict version $S_T(\Sigma)$.

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

**Proposition 1.** Let $\mathcal{M} = (\Sigma_1, \Sigma_2, T_{12})$ be a DL-Lite\textsubscript{R} TBox over $\Sigma_1$, and $T_2$ a DL-Lite\textsubscript{R} TBox over $\Sigma_2$. Then $T_1$, $T_2$, and $\mathcal{M}$ satisfy condition (C1) iff the following conditions are satisfied:

(A) $S_M(U_{T_1}(B)) \subseteq U_{T_2}(S_M(B))$, for each basic concept $B$ over $\Sigma_1$;

(B) $S_M(U_{T_1}(R)) \subseteq U_{T_2}(S_M(R))$, for each basic role $R$ over $\Sigma_1$;

(C) for each basic concept $B$ and each basic role $R$ over $\Sigma_1$ such that $\exists R \in U_{T_1}(B)$ and $S_M(U_{T_1}(\exists R^-)) \neq \emptyset$, we have that

(CA) $\exists Q_{R,B} \in U_{T_1}(S_M(B))$,

(CB) $S_M(U_{T_1}(R)) \subseteq U_{T_2}(Q_{R,B})$, and

(CC) $S_M(U_{T_1}(\exists R^-)) \subseteq U_{T_1}(\exists Q_{R,B})$.

and if $S_M(U_{T_1}(R)) = \emptyset$, either

(CD) $S_M(U_{T_1}(\exists R^-)) \subseteq U_{T_2}(S_M(B))$,

or there exist roles $Q_{R,B}^1, \ldots, Q_{R,B}^n$ over $\Sigma_2$ such that

(CE) $\exists Q_{R,B}^1 \in U_{T_2}(S_M(B))$,

(CF) $T_2 \models \exists(Q_{R,B}^1)^- \subseteq \exists Q_{R,B}^2, \ldots, \exists(Q_{R,B}^{n-1})^- \subseteq \exists Q_{R,B}^n$, and

(CG) $S_M(U_{T_1}(\exists R^-)) \subseteq U_{T_2}(\exists Q_{R,B}^-)$.

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\textsubscript{R}, can be solved in polynomial time. In particular, for a basic concept $B$ and a basic role $R$ over $\Sigma_1$ such that $\exists R \in U_{T_1}(B)$, $S_M(U_{T_1}(\exists R^-)) \neq \emptyset$, $S_M(U_{T_1}(R)) = \emptyset$, and $S_M(U_{T_1}(\exists R^-)) \not\subseteq U_{T_2}(S_M(B))$, checking the existence of roles $Q_{R,B}^1, \ldots, Q_{R,B}^n$ over $\Sigma_2$ 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 $\Sigma_2$, checking the existence of a role $Q_{R,B}^i$ in $U_{T_2}(S_M(B))$ is equivalent to checking whether there exists a path of type $R$ from $B$ to $B'$ in a directed graph $G = (V, E)$, where $V$ is the set of all basic concepts and roles in $\Sigma_2$, and $E$ is the set of all directed edges $(B, R, B')$ in $G$. This problem can be solved in polynomial time, as the reachability problem in a directed graph can be solved in polynomial time using a depth-first search algorithm.
such that $B_2 \in S_M(B)$ and $S_M(\cup_{T_1} (\exists R^\rightarrow)) \subseteq \cup_{T_2}(B'_2)$, we use the following approach to check for the existence of the roles $Q_{R,B}^1, \ldots, Q_{R,B}^n$ over $\Sigma_2$ such that $T_2 \models B_2 \subseteq \exists Q_{R,B}^1, T_2 \models \exists Q_{R,B}^2 \subseteq \exists Q_{R,B}^2, \ldots, T_2 \models \exists Q_{R,B}^n \subseteq B'_2$. Let $G = (V, E)$ be the directed graph defined as:

$$V = \{B_2, B'_2\} \cup \{Q_2 | Q_2 \text{ is a role in } \Sigma_2\}$$

$$E = \{(B_2, B'_2) | T_2 \models B_2 \subseteq B'_2\} \cup \{(B_2, Q_2) | T_2 \models B_2 \subseteq \exists Q_2\} \cup \{(Q_2, B'_2) | T_2 \models \exists Q'_2 \subseteq B'_2\} \cup \{(Q_2, Q'_2) | T_2 \models \exists Q'_2 \subseteq \exists Q'_2\}$$

Then we test for the existence of the roles $Q_{R,B}^1, \ldots, Q_{R,B}^n$ 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 $Q_{R,B}^1, \ldots, Q_{R,B}^n$ 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 $\mathcal{M} = (\Sigma_1, \Sigma_2, T_{12})$ be a DL-Lite$^{pos}_R$-mapping, $T_1$ a DL-Lite$^{pos}_R$-TBox over $\Sigma_1$, and $T_2$ a DL-Lite$^{pos}_R$-TBox over $\Sigma_2$. Then $T_1$, $T_2$, and $\mathcal{M}$ satisfy condition (C2) iff the following conditions are satisfied:

(A) $\cup_{T_2}(S_M(B)) \subseteq S_M(\cup_{T_1}(B))$ for each basic concept $B$ over $\Sigma_1$;

(B) $\cup_{T_2}(S_M(R)) \subseteq S_M(\cup_{T_1}(R))$ for each role $R \in \Sigma_1$;

(C) for each basic role $Q$ over $\Sigma_2$ and each basic concept $B$ over $\Sigma_1$ such that $\exists Q \in \cup_{T_2}(S_M(B))$ and $\cup_{T_2}(\exists Q^\rightarrow) \neq \{\exists Q^\rightarrow\}$, there exists a role $R_{Q,B}$ over $\Sigma_1$ s.t.

(CA) $\exists R_{Q,B} \in \cup_{T_1}(B)$ and

(CB) $Q \in S_M(R_{Q,B})$.

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$^{pos}_R$.

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$^{pos}_R$-mapping $\mathcal{M} = (\Sigma_1, \Sigma_2, T_{12})$, a DL-Lite$^{pos}_R$-TBox $T_1$ over $\Sigma_1$, and a DL-Lite$^{pos}_R$-TBox $T_2$ over $\Sigma_2$, whether $T_2$ is a UCQ-representation of $T_1$ 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 UCQRREP$^{pos}_R$, which verifies whether a source TBox $T_1$ is UCQ-representable under a mapping $\mathcal{M}$, and if this is the case returns a UCQ-representation of $T_1$ under $\mathcal{M}$.

Intuitively, given a source TBox $T_1$ and a mapping $\mathcal{M}$, algorithm UCQRREP$^{pos}_R$ constructs the best possible candidate $T_2$ for a UCQ-representation of $T_1$ under $\mathcal{M}$ (given the conditions in Propositions 1 and 2), and then checks whether $T_2$ effectively satisfies the properties required for a UCQ-representation (note, that $T_2$ is indeed a DL-Lite$^{pos}_R$ TBox). To prove the correctness of this algorithm, we need to show that $T_1$ is UCQ-representable under $\mathcal{M}$ if and only if $T_2$ is a UCQ-representation of $T_1$ under $\mathcal{M}$.
Algorithm: UCQRep\textsuperscript{pos}(T_1,\mathcal{M})
Input: A DL-Lite\textsuperscript{pos}_R\textsuperscript{pos}-mapping \mathcal{M} = (\Sigma_1, \Sigma_2, T_{12}) and a DL-Lite\textsuperscript{pos}_R\textsuperscript{pos}-TBox T_1 over \Sigma_1.
Output: A DL-Lite\textsuperscript{pos}_R\textsuperscript{pos}-TBox T_2 over \Sigma_2 that is a UCQ-representation of T_1 under \mathcal{M}, if T_1 is UCQ-representable under \mathcal{M}. The keyword false otherwise.

1. Let T_2 be a TBox over \Sigma_2 defined as:
   \[ T_2 = \{ N_2 \subseteq M_2 \mid N_1 \text{ a basic concept or role over } \Sigma_1, \]
   \[ N_2 \in \mathcal{S}_M(N_1), M_2 \in \mathcal{S}_M(\cup T_1(N_1)) \}

2. Remove from T_2 every inclusion N_2 \subseteq M_2 such that (i) N_2 \subseteq \mathcal{S}_M(N_1) for some \Sigma_1, and (ii) for every M_1 over \Sigma_1 such that M_2 \subseteq \mathcal{S}_M(M_1), it holds that T_1 \not\sqsubseteq N_1 \subseteq M_1. Moreover, if N_2 = \exists R_2 and M_2 = \exists R_2', then also remove inclusions R_2 \subseteq R_2' and R_2' \subseteq R_2'' from T_2.

3. Remove from T_2 every inclusion of the form either \exists R_2 \subseteq B_2 or R_2 \subseteq R_2' or R_2' \subseteq R_2'' for roles R_2, R_2' and a concept B_2 over \Sigma_2, if there exists a concept B_1 over \Sigma_1 such that (i) \exists R_2 \subseteq \mathcal{S}_M(B_1), and (ii) for every role R_1 over \Sigma_1 such that \exists R_1 \subseteq \cup T_1(B_1) and R_2 \subseteq \mathcal{S}_M(R_1), it holds that T_1 \not\sqsubseteq B_1 \subseteq R_1.

4. Verify whether T_2 is a UCQ-representation of T_1 under \mathcal{M}. If the test succeeds, return T_2, otherwise return false.

Fig. 1. Algorithm to compute the UCQ-representation of a DL-Lite\textsuperscript{pos}_R\textsuperscript{pos}-TBox T_1 under a DL-Lite\textsuperscript{pos}_R\textsuperscript{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 T_2 is of polynomial size in the sizes of T_1 and \mathcal{M}, and that, by Theorem 1, it is possible to check in polynomial time whether T_2 is a UCQ-representation of T_1 under \mathcal{M}.

Theorem 2. Algorithm UCQRep\textsuperscript{pos} is correct and runs in polynomial time.

The following examples illustrate how algorithm UCQRep\textsuperscript{pos} works.

Example 3. Let \mathcal{M} = (\Sigma_1, \Sigma_2, T_{12}), where \Sigma_1 = \{A_1(\cdot), B_1(\cdot), C_1(\cdot)\}, \Sigma_2 = \{A_2(\cdot), B_2(\cdot)\}, and T_{12} = \{A_1 \sqsubseteq A_2, B_1 \sqsubseteq B_2, C_1 \sqsubseteq B_2\}. Furthermore, assume that T_1 = \{B_1 \sqsubseteq A_1\}. Then, in step 1, the algorithm constructs the TBox T_2 = \{B_2 \sqsubseteq A_2\}.

In step 2, it removes the only axiom from T_2 as B_2 \in \mathcal{S}_M(C_1) and T_1 \not\sqsubseteq C_1 \sqsubseteq A_1. In step 3, it does nothing, and finally, at the last step it checks whether the empty TBox T_2 is a UCQ-representation of T_1 under \mathcal{M}. Since A_2 \in \mathcal{S}_M(\cup T_1(B_1)) and A_2 \not\in \cup T_2(\mathcal{S}_M(B_1)), the algorithm returns false.

Example 4. Let \mathcal{M} = (\Sigma_1, \Sigma_2, 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 T_{12} = \{\exists P_1 \sqsubseteq A_2, B_1 \sqsubseteq B_2, R_1 \sqsubseteq R_2\}. Furthermore, assume that 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 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 T_2 is a UCQ-representation of T_1 under \mathcal{M} and returns T_2.
3.3 Solving UCQ-Representability for $DL$-Lite$_{RDFS}$

It is not difficult to see that if the input of algorithm UCQR$^{pos}$ is a $DL$-Lite$_{RDFS}$-mapping $\mathcal{M} = (\Sigma_1, \Sigma_2, T_{12})$ and a $DL$-Lite$_{RDFS}$-TBox $T_1$ over $\Sigma_1$, then the set $T_2$ computed by this algorithm is a $DL$-Lite$_{RDFS}^{pos}$-TBox over $\Sigma_2$ that can be easily transformed into an equivalent $DL$-Lite$_{RDFS}$-TBox. Indeed, $T_2$ might contain inclusions between basic concepts of the form $\exists R_2 \sqsubseteq \exists R_2'$, but this occurs only if $T_2$ 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 $T_2$ and can be removed from $T_2$ without affecting its semantics. Thus, we conclude that algorithm UCQR$^{pos}$ 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$_{R}^{pos}$

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$_{R}^{pos}$. We first need to introduce some terminology. Given a $DL$-Lite$_{R}$-TBox $\mathcal{T}$ over a signature $\Sigma$ and a UCQ $q$ over $\Sigma$, a UCQ $q_r$ over $\Sigma$ is said to be a perfect reformulation of $q$ w.r.t. $\mathcal{T}$ if for every ABox $\mathcal{A}$ over $\Sigma$, 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$_{R}$-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$_{R}^{pos}$. More precisely, let $\mathcal{M} = (\Sigma_1, \Sigma_2, T_{12})$ be a $DL$-Lite$_{R}$-mapping and $T_1$ a $DL$-Lite$_{R}$-TBox over $\Sigma_1$. Then define a mapping $\text{Comp}(\mathcal{M}, T_1) = (\Sigma_1, \Sigma_2, T_{12})$ that extends $\mathcal{M}$ by compiling the knowledge from $T_1$ into $T_{12}$. Formally, for a basic concept or role $N$ over $\Sigma_1$, let $bq_{N}$ be the UCQ defined as follows: $bq_{N}(x) = A(x)$, $bq_{x,y}(x) = \exists y.P(x, y)$, $bq_{x,y}(y) = \exists x.P(y, x)$, and $bq_{y,x} = P(x, y)$. Then, for every concept inclusion $B \sqsubseteq C$ in $T_{12}$ and for every CQ $q$ in the perfect reformulation of $bq_{B}$ w.r.t. $T_1$, include $C_q \sqsubseteq C$ into $T_{12}^*$, where $C_q$ is the (unique) basic concept such that $bq_{C_q} = q$. Also, for every role inclusion $R \sqsubseteq Q \in T_{12}$ and for every UCQ $q$ in the perfect reformulation of $bq_{R}$ w.r.t. $T_1$, include $R_q \sqsubseteq Q$ into $T_{12}^*$, where $R_q$ is the basic role such that $bq_{R_q} = q$.

It is important to notice that if $\mathcal{M} = (\Sigma_1, \Sigma_2, T_{12})$ is a $DL$-Lite$_{R}^{pos}$-mapping and $T_1$ a $DL$-Lite$_{R}^{pos}$-TBox over $\Sigma_1$, then $\text{Comp}(\mathcal{M}, T_1) = (\Sigma_1, \Sigma_2, T_{12})$ is a $DL$-Lite$_{R}^{pos}$-mapping that can be computed in polynomial time in the sizes of $\mathcal{M}$ and $T_1$. Therefore, given that the set of inclusions defining $\text{Comp}(\mathcal{M}, T_1)$ contains the set of inclusions defining $\mathcal{M}$ and $T_1 \cup T_{12} = T_{12}^*$, we conclude that $\text{Comp}(\mathcal{M}, T_1)$ can be used to check in polynomial time whether $T_1$ is weakly UCQ-representable under $\mathcal{M}$.

**Theorem 3.** Let $\mathcal{M} = (\Sigma_1, \Sigma_2, T_{12})$ be a $DL$-Lite$_{R}^{pos}$-mapping and $T_1$ a $DL$-Lite$_{R}^{pos}$-TBox over $\Sigma_1$. Then $T_1$ is weakly UCQ-representable under $\mathcal{M}$ if and only if $T_1$ is UCQ-representable under $\text{Comp}(\mathcal{M}, T_1)$.

From this result and Theorem 2 we obtain a polynomial time algorithm for solving the weak UCQ-representability problem for $DL$-Lite$_{R}^{pos}$ mappings and TBoxes.
The example below shows a DL-Lite\textsubscript{pos} TBox $\mathcal{T}_1$ and a DL-Lite\textsubscript{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 $\Sigma_1 = \{P_1(\cdot, \cdot), B_2(\cdot)\}$, $\Sigma_2 = \{A_2(\cdot), \exists P_1\}$, and $\mathcal{T}_{12} = \{B_1 \subseteq B_2, \exists P_1 \subseteq A_2\}$. Furthermore, assume that $\mathcal{T}_1 = \{B_1 \subseteq \exists P_1\}$. 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 $\mathcal{M} = \text{COMP}(\mathcal{M}, \mathcal{T}_1)$ and, thus, $\mathcal{T}_1$ is not weakly UCQ-representable under $\mathcal{M}$, as $\mathcal{T}_1$ is not UCQ-representable under $\mathcal{M}$. \hfill \IEEEQEDclosed

Instead, as shown in [1, 2], for each DL-Lite\textsubscript{RDFS} TBox $\mathcal{T}_1$ and DL-Lite\textsubscript{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\textsubscript{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\textsubscript{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\textsubscript{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