

# Knowledge Base Exchange

Marcelo Arenas<sup>1</sup>, Elena Botoeva<sup>2</sup>, and Diego Calvanese<sup>2</sup>

<sup>1</sup> Dept. of Computer Science, PUC Chile  
marenas@ing.puc.cl

<sup>2</sup> KRDB Research Centre, Free Univ. of Bozen-Bolzano, Italy  
lastname@inf.unibz.it

**Abstract.** In this paper, we study the problem of exchanging knowledge between two knowledge bases (KBs) connected through mappings, with a special interest in exchanging implicit knowledge, not only data like in the traditional database exchange setting. As representation formalism we use Description Logics (DL) that exhibit a reasonable tradeoff between expressive power and complexity of reasoning. Thus, we assume that the source and target KBs are given as a DL TBox+ABox, while the mappings have the form of DL TBox assertions. We study the problem of translating the knowledge in the source KB according to these mappings. We define a general framework of KB exchange, and specify the problems of computing and representing different kinds of solutions, i.e., target KBs with specified properties, given a source KB and a mapping. We then develop first results and techniques and study the complexity of KB exchange for the case of *DL-Lite<sub>RDFS</sub>*, a DL that corresponds to the FOL fragment of RDFS.

## 1 Introduction

In data exchange, data structured under one schema (called source schema) must be restructured and translated into an instance of a different schema (called target schema), and the way in which this restructuring should occur is specified by means of a mapping from the source schema to the target schema [5]. Such a problem has been studied extensively in recent years, under various choices for the languages used to specify the source and target schema, and the mappings [2]. While incomplete information in this setting is introduced by the mapping layer (see also [6]), one fundamental assumption in the works on data exchange is that the source is a (completely specified) database.

In this paper, we go beyond this setting, and consider data exchange in the case where implicit knowledge is present in the source, by which new data may be inferred. We follow the line of the work in [1], where a general framework for data exchange is proposed, in which the source data may be incompletely specified, and thus (possibly infinitely) many source instances are implicitly represented. The framework in [1] 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).

We refine that framework to the case where as a representation system we use Description Logics (DL) knowledge bases (KBs) constituted by a TBox and an ABox, and where 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 our case solutions are target DL KBs, derived from the source KB and the mapping. Besides a notion of (universal) solution based on the correspondence between models of source and target KBs, we introduce also the weaker notion of (*universal*) *CQ-solution*, based on the correspondence between answers to conjunctive queries over source and target KBs.

In our setting where we exchange DL KBs, in order to minimize the exchange (and hence transfer and materialization) of explicit (i.e., ABox) information, we are interested in computing universal (CQ-)solutions that contain as much implicit knowledge as possible. This leads us to define a new problem, called *representability*, whose goal is to compute from a source TBox and a mapping, a target TBox that leads to a universal (CQ-)solution when it is combined with a suitable ABox computed from the source ABox, independently of the actual source ABox.

We then develop first results and techniques for KB exchange and for the representability problem in the case of KBs expressed in  $DL-Lite_{RDFS}$ , a DL that corresponds to the FOL fragment of RDFS.  $DL-Lite_{RDFS}$  is a fragment of  $DL-Lite_{\mathcal{R}}$  [4] that does not allow for existential quantification (i.e., concepts of the form  $\exists R$ ) in the right-hand side of concept inclusions, nor for disjointness assertions.

The paper is organized as follows. In Section 2 we give preliminary notions on DLs and conjunctive queries (CQs). In Section 3 we define our framework of KB exchange. In Section 4 we present the techniques for deciding the defined reasoning tasks. Finally, in Section 5 we draw some conclusions and outline issues for future work.

## 2 Preliminaries

We introduce here the necessary notions about the description logic (DL) that we use in this article, and about conjunctive queries, which we adopt as our query formalism.

### 2.1 $DL-Lite_{\mathcal{R}}$ Knowledge Bases

The DLs of the  $DL-Lite$  family [4] of light-weight description logics 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 adopt here  $DL-Lite_{\mathcal{R}}$ , and present now its syntax and semantics.

Let  $N_C, N_R, N_a$  be sets of concept, role, and individual names, respectively, and assume that  $A \in N_C$  and  $P \in N_R$ . In  $DL-Lite_{\mathcal{R}}$ ,  $B$  and  $C$  are used to denote basic and complex concept descriptions, respectively, and  $R$  and  $Q$  are used to denote basic and complex roles, respectively. These concept and roles constructs are defined by the following grammar:

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

In the following, for a basic role  $R$ , we use  $R^-$  to denote  $P^-$  when  $R = P$ , and  $P$  when  $R = P^-$ .

A  $DL\text{-Lite}_{\mathcal{R}}$  TBox is a finite set of concept inclusions  $B \sqsubseteq C$  and role inclusions  $R \sqsubseteq Q$ . A  $DL\text{-Lite}_{\mathcal{R}}$  ABox is a finite set of membership assertions of the form  $A(a)$  and  $P(a, b)$ , where  $a, b \in N_a$ . A  $DL\text{-Lite}_{\mathcal{R}}$  KB  $\mathcal{K}$  is a pair  $\langle \mathcal{T}, \mathcal{A} \rangle$ , where  $\mathcal{T}$  is a  $DL\text{-Lite}_{\mathcal{R}}$  TBox and  $\mathcal{A}$  is a  $DL\text{-Lite}_{\mathcal{R}}$  ABox. As usual, TBoxes represent implicit knowledge, while ABoxes represent the data.

In the following, we will also make use of a restricted form of  $DL\text{-Lite}_{\mathcal{R}}$  TBoxes. A  $DL\text{-Lite}_{\mathcal{R}}$  TBox is said to be *definite* if there are only atomic concepts and atomic roles on the right-hand side of its inclusions. In other words, a definite TBox may not mention in its right-hand side a concept of the form  $\exists R$ , and may not contain concept or role disjointness assertions. We call  $DL\text{-Lite}_{\mathcal{RDFs}}$  the fragment of  $DL\text{-Lite}_{\mathcal{R}}$  obtained by considering only definite  $DL\text{-Lite}_{\mathcal{R}}$  TBoxes. Intuitively,  $DL\text{-Lite}_{\mathcal{RDFs}}$  corresponds to the fragment of RDFS [3] that is embeddable in FOL (and hence in DLs).

The semantics of  $DL\text{-Lite}_{\mathcal{R}}$  is defined in the standard way. We just remark that we use  $\text{MOD}(\mathcal{K})$  to denote the set of all models of KB  $\mathcal{K}$ . From now on, we assume that interpretations satisfy the standard name assumption, that is, we assume given a fixed infinite set  $\mathbf{U}$  of individual names, and we assume that for every interpretation  $\mathcal{I} = \langle \Delta^{\mathcal{I}}, \cdot^{\mathcal{I}} \rangle$ , it holds that  $\Delta^{\mathcal{I}} \subseteq \mathbf{U}$  and  $a^{\mathcal{I}} = a$  for every  $a \in \Delta^{\mathcal{I}}$ . Notice that this implies the Unique Name Assumption (UNA), i.e., different individuals are interpreted as different domain elements.

A *signature*  $\Sigma$  is a set of concept and role names. An interpretation  $\mathcal{I} = \langle \Delta^{\mathcal{I}}, \cdot^{\mathcal{I}} \rangle$  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  $\mathcal{A}$ , a concept inclusions  $B \sqsubseteq C$ , a role inclusions  $R \sqsubseteq Q$ , and membership assertions  $A(a)$  and  $R(a, b)$ ).

## 2.2 Conjunctive Queries and Certain Answers

A *conjunctive query* (CQ) over a signature  $\Sigma$  is a first-order formula of the form  $q(\mathbf{x}) = \exists \mathbf{y}. \text{conj}(\mathbf{x}, \mathbf{y})$ , where  $\mathbf{x}, \mathbf{y}$  are tuples of variables and  $\text{conj}(\mathbf{x}, \mathbf{y})$  is a conjunction of atoms of the form: (1)  $A(t)$ , with  $A$  a concept name in  $\Sigma$  and  $t$  either a constant from  $\mathbf{U}$  or a variable from  $\mathbf{x}$  or  $\mathbf{y}$ , or (2)  $P(t_1, t_2)$ , with  $P$  a role name in  $\Sigma$  and  $t_i$  ( $i = 1, 2$ ) either a constant from  $\mathbf{U}$  or a variable from  $\mathbf{x}$  or  $\mathbf{y}$ . In a conjunctive query  $q(\mathbf{x}) = \exists \mathbf{y}. \text{conj}(\mathbf{x}, \mathbf{y})$ ,  $\mathbf{x}$  is the tuple of free variables of  $q(\mathbf{x})$ . A *union of conjunctive queries* (UCQ) is a formula of the form:  $q(\mathbf{x}) = \bigvee_{i=1}^n \exists \mathbf{y}_i. \text{conj}_i(\mathbf{x}, \mathbf{y}_i)$ , where each  $\text{conj}_i(\mathbf{x}, \mathbf{y}_i)$  is as before. A query  $q$  (either a CQ or a UCQ) is said to be a query over a KB  $\mathcal{K}$  if  $q$  is a query over a signature  $\Sigma$  and  $\Sigma \subseteq \Sigma(\mathcal{K})$ .

Let  $q$  be a CQ  $\exists y_1 \dots \exists y_\ell. \text{conj}(x_1, \dots, x_k, y_1, \dots, y_\ell)$  over a signature  $\Sigma$  and  $\mathcal{I} = \langle \Delta^{\mathcal{I}}, \cdot^{\mathcal{I}} \rangle$  an interpretation of  $\Sigma$ . Then the answer of  $q$  over  $\mathcal{I}$ , denoted by  $q^{\mathcal{I}}$ , is defined as the set of tuples  $(a_1, \dots, a_k)$  of elements from  $\Delta^{\mathcal{I}}$  for which there exist a tuple  $(b_1, \dots, b_\ell)$  of elements from  $\Delta^{\mathcal{I}}$  such that  $\mathcal{I}$  satisfies every conjunct in  $\text{conj}(a_1, \dots, a_k, b_1, \dots, b_\ell)$ . Moreover, given a UCQ  $q = \bigvee_{i=1}^n q_i$ , the answer of  $q$  over an interpretation  $\mathcal{I}$ , denoted by  $q^{\mathcal{I}}$ , is defined as  $\bigcup_{i=1}^n q_i^{\mathcal{I}}$ . Finally, given a query  $q$  (either a CQ or a UCQ) over a KB  $\mathcal{K}$ , the answer to  $q$  over  $\mathcal{K}$ , denoted by  $\text{cert}(q, \mathcal{K})$ , is defined as  $\text{cert}(q, \mathcal{K}) = \bigcap_{\mathcal{I} \in \text{MOD}(\mathcal{K})} q^{\mathcal{I}}$ . Each tuple in  $\text{cert}(q, \mathcal{K})$  is called a *certain answer* for  $q$  over  $\mathcal{K}$ .

Certain answers in  $DL\text{-Lite}_{\mathcal{R}}$  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, or *labeled nulls* interpreted as not necessarily distinct domain elements. For  $DL\text{-Lite}_{\mathcal{R}}$  KBs, we employ the notion of oblivious chase as defined in [4]. For such a KB  $\langle \mathcal{T}, \mathcal{A} \rangle$ , the chase of  $\mathcal{A}$  w.r.t.  $\mathcal{T}$ , denoted  $\text{chase}_{\mathcal{T}}(\mathcal{A})$ , is a chase obtained from  $\mathcal{A}$  by adding facts implied by inclusions in  $\mathcal{T}$ , and introducing labeled nulls whenever required by an inclusion with  $\exists R$  in the right-hand side (see [4] for details).

### 3 Exchanging Knowledge Bases

In this section, we introduce the knowledge exchange framework used in the paper. The starting point to define this framework is the notion of mapping, which has been shown to be of fundamental importance in the context of data exchange [5]. Formally, a  $DL\text{-Lite}_{\mathcal{R}}$ -mapping (or just *mapping*) is a tuple  $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$ , where  $\Sigma_1, \Sigma_2$  are disjoint signatures and  $\mathcal{T}_{12}$  is a  $DL\text{-Lite}_{\mathcal{R}}$  TBox whose inclusions are of the form: (1)  $C_1 \sqsubseteq C_2$ , where  $C_1, C_2$  are complex concepts over  $\Sigma_1$  and  $\Sigma_2$ , respectively, and (2)  $Q_1 \sqsubseteq Q_2$ , where  $Q_1$  and  $Q_2$  are complex roles over  $\Sigma_1$  and  $\Sigma_2$ , respectively.

Let  $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$  be a mapping. Intuitively, mapping  $\mathcal{M}$  specifies how a knowledge base over the vocabulary  $\Sigma_1$  should be translated into a knowledge base over the vocabulary  $\Sigma_2$ . This intuition is formalized in terms of the notion of *solution*, which is defined as follows. Given an interpretation  $\mathcal{I}_1$  of  $\Sigma_1$  and an interpretation  $\mathcal{I}_2$  of  $\Sigma_2$ , pair  $(\mathcal{I}_1, \mathcal{I}_2)$  *satisfies* TBox  $\mathcal{T}_{12}$ , denoted by  $(\mathcal{I}_1, \mathcal{I}_2) \models \mathcal{T}_{12}$ , if for each concept inclusion  $C_1 \sqsubseteq C_2 \in \mathcal{T}_{12}$ , it holds that  $C_1^{\mathcal{I}_1} \subseteq C_2^{\mathcal{I}_2}$ , and for each role inclusion  $Q_1 \sqsubseteq Q_2 \in \mathcal{T}_{12}$ , it holds that  $Q_1^{\mathcal{I}_1} \subseteq Q_2^{\mathcal{I}_2}$ . Moreover, given an interpretation  $\mathcal{I}$  of  $\Sigma_1$ ,  $\text{SAT}_{\mathcal{M}}(\mathcal{I})$  is defined as the set of interpretations  $\mathcal{J}$  of  $\Sigma_2$  such that  $(\mathcal{I}, \mathcal{J}) \models \mathcal{T}_{12}$ , and given a set  $\mathcal{X}$  of interpretations of  $\Sigma_1$ ,  $\text{SAT}_{\mathcal{M}}(\mathcal{X})$  is defined as:

$$\text{SAT}_{\mathcal{M}}(\mathcal{X}) = \bigcup_{\mathcal{I} \in \mathcal{X}} \text{SAT}_{\mathcal{M}}(\mathcal{I}).$$

Then the notion of solution under a mapping is defined as follows, by considering this notion of satisfaction and the knowledge exchange framework proposed in [1].

**Definition 1.** Let  $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$  be a mapping,  $\mathcal{K}_1$  a KB over  $\Sigma_1$ , and  $\mathcal{K}_2$  a KB over  $\Sigma_2$ . Then  $\mathcal{K}_2$  is said to be a *solution* for  $\mathcal{K}_1$  under  $\mathcal{M}$  if:

$$\text{MOD}(\mathcal{K}_2) \subseteq \text{SAT}_{\mathcal{M}}(\text{MOD}(\mathcal{K}_1)).$$

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

Let  $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$ . A KB  $\mathcal{K}_1$  over  $\Sigma_1$  can have an infinite number of solutions under  $\mathcal{M}$ . Thus, it is natural to ask what is a *good* solution for this knowledge base. Next we introduce the notion of universal solution, which is a simple extension of the concept of solution introduced in Definition 1, and is based on the notion of universal solution introduced in [1].

**Definition 2.** Let  $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$  be a mapping,  $\mathcal{K}_1$  a KB over  $\Sigma_1$ , and  $\mathcal{K}_2$  a KB over  $\Sigma_2$ . Then  $\mathcal{K}_2$  is said to be a *universal solution* for  $\mathcal{K}_1$  under  $\mathcal{M}$  if:

$$\text{MOD}(\mathcal{K}_2) = \text{SAT}_{\mathcal{M}}(\text{MOD}(\mathcal{K}_1)).$$

In the preceding definition, KB  $\mathcal{K}_2$  is considered to be a good solution for KB  $\mathcal{K}_1$  under mapping  $\mathcal{M}$  as the models of  $\mathcal{K}_2$  exactly correspond to the valid translations of the models of  $\mathcal{K}_1$  according to  $\mathcal{M}$ . We illustrate Definitions 1 and 2 in an example.

*Example 1.* Let  $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$ , where  $\Sigma_1 = \{A_1, B_1\}$ ,  $\Sigma_2 = \{A_2, B_2\}$ , and  $\mathcal{T}_{12} = \{A_1 \sqsubseteq A_2, B_1 \sqsubseteq B_2\}$ . Furthermore, assume that  $\mathcal{K}_1 = \langle \mathcal{T}_1, \mathcal{A}_1 \rangle$ , where  $\mathcal{T}_1 = \{B_1 \sqsubseteq A_1\}$  and  $\mathcal{A}_1 = \{B_1(a)\}$ . Then the following knowledge bases over  $\Sigma_2$  are solutions for  $\mathcal{K}_1$  under  $\mathcal{M}$ :

$$\begin{aligned} \mathcal{K}_2 &= \langle \mathcal{T}_2, \mathcal{A}_2 \rangle, & \text{where } \mathcal{T}_2 &= \emptyset, & \mathcal{A}_2 &= \{B_2(a), A_2(a)\} \\ \mathcal{K}'_2 &= \langle \mathcal{T}'_2, \mathcal{A}'_2 \rangle, & \text{where } \mathcal{T}'_2 &= \{B_2 \sqsubseteq A_2\}, & \mathcal{A}'_2 &= \{B_2(a)\} \end{aligned}$$

Moreover,  $\mathcal{K}_2$  is a universal solution for  $\mathcal{K}_1$  under  $\mathcal{M}$ , while  $\mathcal{K}'_2$  is not. In fact, if  $\mathcal{I}_2$  is an interpretation of  $\Sigma_2$  such that:  $\Delta^{\mathcal{I}_2} = \{a, b\}$ ,  $A_2^{\mathcal{I}_2} = \{a\}$ , and  $B_2^{\mathcal{I}_2} = \{a, b\}$ , then we have that  $\mathcal{I}_2 \notin \text{MOD}(\mathcal{K}'_2)$  since  $\mathcal{I}_2$  does not satisfy inclusion  $B_2 \sqsubseteq A_2$ , but  $\mathcal{I}_2 \in \text{SAT}_{\mathcal{M}}(\text{MOD}(\mathcal{K}_1))$  since  $(\mathcal{I}_1, \mathcal{I}_2) \models \mathcal{T}_{12}$  for  $\mathcal{I}_1 \in \text{MOD}(\mathcal{K}_1)$  defined as  $\Delta^{\mathcal{I}_1} = \{a\}$ ,  $A_1^{\mathcal{I}_1} = \{a\}$  and  $B_1^{\mathcal{I}_1} = \{a\}$ . ■

In the previous example,  $\mathcal{K}'_2$  is not a universal solution for  $\mathcal{K}_1$  under  $\mathcal{M}$  since inclusion  $B_2 \sqsubseteq A_2$  cannot be deduced from the information in  $\mathcal{K}_1$  and  $\mathcal{M}$ . Or, more formally,  $\mathcal{K}'_2$  is not a universal solution as  $B_2 \sqsubseteq A_2$  is not implied by  $\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle$ . However,  $\mathcal{K}'_2$  can also be considered as a solution of  $\mathcal{K}_1$  that is desirable to materialize, as the implicit knowledge in  $\mathcal{K}_2$  (i.e., TBox  $\mathcal{T}_2$ ) represents the implicit knowledge in  $\mathcal{K}_1$  (i.e., TBox  $\mathcal{T}_1$ ), given the way that concepts  $A_1$  and  $B_1$  have to be translated according to mapping  $\mathcal{M}$ . In fact, solution  $\mathcal{K}'_2$  is as good as solution  $\mathcal{K}_2$  in terms of the information that can be extracted from them by using some specific queries, but with the advantage that  $\mathcal{K}'_2$  represents knowledge in a more compact way. In what follows, we introduce a new class of good solutions that captures this intuition with respect to the widely used fragment of CQs.

**Definition 3.** 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  $\Sigma_1$ , and  $\mathcal{K}_2$  a KB over  $\Sigma_2$ . Then  $\mathcal{K}_2$  is said to be a CQ-solution for  $\mathcal{K}_1$  under  $\mathcal{M}$  if for every CQ  $q$  over  $\Sigma_2$ , we have that  $\text{cert}(q, \langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle) \subseteq \text{cert}(q, \mathcal{K}_2)$ .

Moreover,  $\mathcal{K}_2$  is said to be a universal CQ-solution for  $\mathcal{K}_1$  under  $\mathcal{M}$  if for every CQ  $q$  over  $\Sigma_2$ , we have that  $\text{cert}(q, \langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle) = \text{cert}(q, \mathcal{K}_2)$ .

Notably, we have in Example 1 that both KB  $\mathcal{K}_2$  and KB  $\mathcal{K}'_2$  are universal CQ-solutions for KB  $\mathcal{K}_1$  under mapping  $\mathcal{M}$ .

A natural question at this point is what is the relationship between the notions of solution presented in this section. The following proposition, which is straightforward to prove, shows that the notion of (universal) CQ-solution is weaker than the notion of (universal) solution.

**Proposition 1.** Let  $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$  be a mapping,  $\mathcal{K}_1$  a KB over  $\Sigma_1$ , and  $\mathcal{K}_2$  a KB over  $\Sigma_2$ . If  $\mathcal{K}_2$  is a (universal) solution for  $\mathcal{K}_1$  under  $\mathcal{M}$ , then  $\mathcal{K}_2$  is a (universal) CQ-solution for  $\mathcal{K}_1$  under  $\mathcal{M}$ .

In this paper, we study several fundamental problems related to the notions of solution presented here. These problems are formally introduced below.

### 3.1 Knowledge Base Exchange: Reasoning Tasks

In the data exchange scenario [5], as well as in the knowledge exchange scenario [1], the problem of materializing solutions has been identified as the fundamental problem to solve. Given a class  $\mathcal{U}$  of mappings (for example, the class of  $DL\text{-Lite}_{\mathcal{R}}$ -mappings) and a DL  $\mathcal{L}$  (for example,  $DL\text{-Lite}_{\mathcal{R}}$ ), the problem of computing solutions is defined as follows for  $\mathcal{U}$  and  $\mathcal{L}$ :

PROBLEM	: $\text{COMPSOL}(\mathcal{U}, \mathcal{L})$
INPUT	: A mapping $\mathcal{M} \in \mathcal{U}$ and an $\mathcal{L}$ -knowledge base $\mathcal{K}_1$ over $\Sigma_1$
TO DO	: Compute an $\mathcal{L}$ -knowledge base $\mathcal{K}_2$ over $\Sigma_2$ such that $\mathcal{K}_2$ is a solution for $\mathcal{K}_1$ under $\mathcal{M}$

Given a class  $\mathcal{U}$  of mappings and a DL  $\mathcal{L}$ , the computation problems  $\text{COMPUNIVSOL}(\mathcal{U}, \mathcal{L})$ ,  $\text{COMPQCSOL}(\mathcal{U}, \mathcal{L})$ , and  $\text{COMPUNIVCQSOL}(\mathcal{U}, \mathcal{L})$  are defined exactly as above, but considering universal solutions, CQ-solutions, and universal CQ-solutions instead of solutions, respectively.

In the rest of this paper, we study these problems, and some other fundamental problems associated to them, for a restriction of the class of mappings introduced in this section. More precisely, a mapping  $\mathcal{M} = \langle \Sigma_1, \Sigma_2, \mathcal{T}_{12} \rangle$  is said to be *definite* if  $\mathcal{T}_{12}$  is a  $DL\text{-Lite}_{\text{RDFS}}$  TBox (recall the definition of definite TBoxes in Section 2.1). We use  $\mathcal{U}_{\text{def}}$  to denote the class of definite mappings. Then, as our first result, we obtained that the chase can be used to compute universal solutions for definite mappings and  $DL\text{-Lite}_{\text{RDFS}}$  TBoxes. In what follows, let  $\text{chase}_{\mathcal{T}, \Sigma}(\mathcal{A})$  denote the projection of  $\text{chase}_{\mathcal{T}}(\mathcal{A})$  on the signature  $\Sigma$ .

**Proposition 2.** *Let  $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$  be a definite mapping and  $\mathcal{K}_1 = \langle \mathcal{T}_1, \mathcal{A}_1 \rangle$  a  $DL\text{-Lite}_{\text{RDFS}}$  KB over  $\Sigma_1$ . Then  $\langle \emptyset, \text{chase}_{\mathcal{T}_{12}, \Sigma_2}(\text{chase}_{\mathcal{T}_1}(\mathcal{A}_1)) \rangle$  is a universal solution for  $\mathcal{K}_1$  under  $\mathcal{M}$ .*

Thus, given that for definite mappings and  $DL\text{-Lite}_{\text{RDFS}}$  TBoxes, the sets  $\text{chase}_{\mathcal{T}_1}(\mathcal{A}_1)$  and  $\text{chase}_{\mathcal{T}_{12}, \Sigma_2}(\text{chase}_{\mathcal{T}_1}(\mathcal{A}_1))$  are always finite and can be computed in polynomial time [4], we obtain as a corollary of Propositions 1 and 2 that the problems of computing solutions defined above can be solved in polynomial time for definite mappings and  $DL\text{-Lite}_{\text{RDFS}}$  TBoxes.

**Corollary 1.**  $\text{COMPSOL}(\mathcal{U}_{\text{def}}, DL\text{-Lite}_{\text{RDFS}})$ ,  $\text{COMPUNIVSOL}(\mathcal{U}_{\text{def}}, DL\text{-Lite}_{\text{RDFS}})$ ,  $\text{COMPQCSOL}(\mathcal{U}_{\text{def}}, DL\text{-Lite}_{\text{RDFS}})$ , and  $\text{COMPUNIVCQSOL}(\mathcal{U}_{\text{def}}, DL\text{-Lite}_{\text{RDFS}})$  can all be solved in polynomial time.

Unfortunately, the solutions obtained by using Proposition 2 are of little interest to us because the generated target ABoxes can be very large. Hence, we turn our attention to the case of non-empty target TBoxes and, more specifically, to the problem of computing universal CQ-solutions that include as much implicit knowledge as possible. This gives rise to the following *separation properties*.

**Definition 4.** *Let  $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$  be a mapping and  $\mathcal{T}_1$  a TBox over  $\Sigma_1$ .*

- $\mathcal{T}_1$  is representable in  $\mathcal{M}$  if there exists a TBox  $\mathcal{T}_2$  over  $\Sigma_2$  such that for every ABox  $\mathcal{A}_1$  over  $\Sigma_1$ , it holds that  $\langle \mathcal{T}_2, \text{chase}_{\mathcal{T}_{12}, \Sigma_2}(\mathcal{A}_1) \rangle$  is a universal CQ-solution for  $\langle \mathcal{T}_1, \mathcal{A}_1 \rangle$  under  $\mathcal{M}$ .  $\mathcal{T}_2$  is called a representation of  $\mathcal{T}_1$  in  $\mathcal{M}$ .

- $\mathcal{T}_1$  is weakly representable in  $\mathcal{M}$  if there exists a mapping  $\mathcal{M}^* = (\Sigma_1, \Sigma_2, \mathcal{T}_{12}^*)$  such that  $\mathcal{T}_{12} \subseteq \mathcal{T}_{12}^*$ ,  $\mathcal{T}_1 \cup \mathcal{T}_{12} \models \mathcal{T}_{12}^*$  and  $\mathcal{T}_1$  is representable in  $\mathcal{M}^*$ .

The separation problems are interesting to us because a positive answer would mean that we can construct the TBox of a solution by considering only the input TBox and the mapping, independently of the input ABox. On the other hand, the ABox of the solution can be found simply by translating the input ABox according to the rules in the mapping (and the input TBox). We illustrate the notions just defined in the following example.

*Example 2.* Let  $\Sigma_1 = \{A_1, B_1\}$ ,  $\Sigma_2 = \{A_2, B_2\}$ , and  $\mathcal{T}_1 = \{B_1 \sqsubseteq A_1\}$ . If  $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$  with  $\mathcal{T}_{12} = \{A_1 \sqsubseteq A_2, B_1 \sqsubseteq B_2\}$ , then we have that  $\mathcal{T}_2 = \{B_2 \sqsubseteq A_2\}$  is a representation of  $\mathcal{T}_1$  in  $\mathcal{M}$ .

On the other hand, if  $\mathcal{M}' = (\Sigma_1, \Sigma_2, \mathcal{T}'_{12})$  with  $\mathcal{T}'_{12} = \{A_1 \sqsubseteq A_2\}$ , then we have that  $\mathcal{T}_1$  is not representable in  $\mathcal{M}'$ : let  $\mathcal{A}'_1 = \{B_1(a)\}$ , then  $\text{chase}_{\mathcal{T}'_{12}, \Sigma_2}(\mathcal{A}'_1) = \emptyset$  and for no TBox  $\mathcal{T}'_2$ ,  $\langle \mathcal{T}'_2, \text{chase}_{\mathcal{T}'_{12}, \Sigma_2}(\mathcal{A}'_1) \rangle$  is a universal CQ-solution to  $\langle \mathcal{T}_1, \mathcal{A}'_1 \rangle$  under  $\mathcal{M}'$ . However, if we consider  $\mathcal{T}^*_{12} = \mathcal{T}'_{12} \cup \{B_1 \sqsubseteq A_2\}$ , we conclude that  $\mathcal{T}_1$  is weakly representable in  $\mathcal{M}'$  since  $\mathcal{T}^*_{12} \subseteq \mathcal{T}_{12}$ ,  $\mathcal{T}_1 \cup \mathcal{T}^*_{12} \models \mathcal{T}^*_{12}$  and  $\mathcal{T}_1$  is representable in  $\mathcal{M}^* = (\Sigma_1, \Sigma_2, \mathcal{T}^*_{12})$  (in fact,  $\emptyset$  is a representation of  $\mathcal{T}_1$  in  $\mathcal{M}^*$ ). Note, that  $\mathcal{T}'_{12} \subset \mathcal{T}_{12}$ .

Now, if  $\mathcal{M}'' = (\Sigma_1, \Sigma_2, \mathcal{T}''_{12})$  with  $\mathcal{T}''_{12} = \{A_1 \sqsubseteq A_2, B_1 \sqsubseteq B_2, C_1 \sqsubseteq B_2\}$ , then again we have that  $\mathcal{T}_1$  is not representable in  $\mathcal{M}''$ : let  $\mathcal{A}''_1 = \{B_1(a), C_1(c)\}$ , then  $\text{chase}_{\mathcal{T}''_{12}, \Sigma_2}(\mathcal{A}''_1) = \{B_2(a), B_2(c)\}$ . The query  $q() \leftarrow A_2(a)$  evaluates to true in  $\langle \mathcal{T}_1 \cup \mathcal{T}''_{12}, \mathcal{A}''_1 \rangle$ , hence in order for a TBox  $\mathcal{T}''_2$  to be a representation of  $\mathcal{T}_1$  in  $\mathcal{M}''$ , it must contain the inclusion  $B_2 \sqsubseteq A_2$ . On the other hand, it implies that the query  $q'() \leftarrow A_2(c)$  evaluates to true in  $\langle \mathcal{T}''_2, \text{chase}_{\mathcal{T}''_{12}, \Sigma_2}(\mathcal{A}''_1) \rangle$ , whereas it evaluates to false in  $\langle \mathcal{T}_1 \cup \mathcal{T}''_{12}, \mathcal{A}''_1 \rangle$ . However, again if we consider  $\mathcal{T}^{**}_{12} = \mathcal{T}''_{12} \cup \{B_1 \sqsubseteq A_2\}$ , we conclude that  $\mathcal{T}_1$  is weakly representable in  $\mathcal{M}''$ . ■

## 4 Techniques for Deciding Representability

We address now the problem of deciding (weak) representability of a TBox in a mapping, for the case where TBoxes are expressed in *DL-Lite<sub>RDFS</sub>* and mappings are definite. We start by showing some properties that characterize solutions in terms of chases.

In the following for two chases  $\mathcal{C}_1$  and  $\mathcal{C}_2$ , a *homomorphism* from  $\mathcal{C}_1$  to  $\mathcal{C}_2$  is a mapping  $h$  from the individuals and labeled nulls in  $\mathcal{C}_1$  to those in  $\mathcal{C}_2$  such that (1)  $h$  is the identity on the individuals, (2) if  $A(t) \in \mathcal{C}_1$  then  $A(h(t)) \in \mathcal{C}_2$ , and (3) if  $P(t, t') \in \mathcal{C}_1$  then  $P(h(t), h(t')) \in \mathcal{C}_2$ . We write  $\mathcal{C}_1 \rightarrow \mathcal{C}_2$  if there is a homomorphism from  $\mathcal{C}_1$  to  $\mathcal{C}_2$ , and  $\mathcal{C}_1 \rightleftharpoons \mathcal{C}_2$  if  $\mathcal{C}_1 \rightarrow \mathcal{C}_2$  and  $\mathcal{C}_2 \rightarrow \mathcal{C}_1$ .

From now on, we assume  $\Sigma_1$  and  $\Sigma_2$  as given, and for a mapping  $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$ , we use simply  $\mathcal{M}$  to denote also  $\mathcal{T}_{12}$ . Then we have the following characterizations of solutions in terms of chases, which are similar to the characterizations in [1].

**Proposition 3.** *Let  $\mathcal{M}$  be a definite mapping,  $\mathcal{K}_1 = \langle \mathcal{T}_1, \mathcal{A}_1 \rangle$  a *DL-Lite<sub>RDFS</sub>* KB over  $\Sigma_1$ , and  $\mathcal{K}_2 = \langle \mathcal{T}_2, \mathcal{A}_2 \rangle$  a *DL-Lite<sub>RDFS</sub>* KB over  $\Sigma_2$ . If  $\text{chase}_{\mathcal{M}, \Sigma_2}(\text{chase}_{\mathcal{T}_1}(\mathcal{A}_1)) \rightarrow \text{chase}_{\mathcal{T}_2}(\mathcal{A}_2)$ , then  $\mathcal{K}_2$  is a CQ-solution for  $\mathcal{K}_1$  under  $\mathcal{M}$ .*

**Corollary 2.** *Let  $\mathcal{M}$  be a definite mapping,  $\mathcal{K}_1 = \langle \mathcal{T}_1, \mathcal{A}_1 \rangle$  a  $DL\text{-Lite}_{\text{RDFS}}$  KB over  $\Sigma_1$ , and  $\mathcal{K}_2 = \langle \mathcal{T}_2, \mathcal{A}_2 \rangle$  a  $DL\text{-Lite}_{\text{RDFS}}$  KB over  $\Sigma_2$ . Then  $\mathcal{K}_2$  is a universal CQ-solution for  $\mathcal{K}_1$  under  $\mathcal{M}$  if  $\text{chase}_{\mathcal{M}, \Sigma_2}(\text{chase}_{\mathcal{T}_1}(\mathcal{A}_1)) \simeq \text{chase}_{\mathcal{T}_2}(\mathcal{A}_2)$ .*

#### 4.1 Checking Representability of a TBox in a Mapping

We address now the *representability* problem, which is to decide, given a TBox  $\mathcal{T}_1$  over  $\Sigma_1$  and a mapping  $\mathcal{M}$ , whether  $\mathcal{T}_1$  is representable in  $\mathcal{M}$  (cf. Definition 4).

We start by considering the decision problem associated with representability:

Given a mapping  $\mathcal{M}$ , a TBox  $\mathcal{T}_1$  over  $\Sigma_1$ , and a TBox  $\mathcal{T}_2$  over  $\Sigma_2$ , check whether  $\mathcal{T}_2$  is a representation of  $\mathcal{T}_1$  in  $\mathcal{M}$ , i.e., for each ABox  $\mathcal{A}_1$  over  $\Sigma_1$ ,  $\langle \mathcal{T}_2, \text{chase}_{\mathcal{M}, \Sigma_2}(\mathcal{A}_1) \rangle$  is a universal CQ-solution for  $\langle \mathcal{T}_1, \mathcal{A}_1 \rangle$  under  $\mathcal{M}$ .

For definite mappings and  $DL\text{-Lite}_{\text{RDFS}}$  TBoxes, the decision problem associated with representability can be solved in two steps:

1. Check whether for each ABox  $\mathcal{A}_1$  over  $\Sigma_1$ ,  $\langle \mathcal{T}_2, \text{chase}_{\mathcal{M}, \Sigma_2}(\mathcal{A}_1) \rangle$  is a CQ-solution for  $\langle \mathcal{T}_1, \mathcal{A}_1 \rangle$  under  $\mathcal{M}$ .
2. Check whether for each ABox  $\mathcal{A}_1$  over  $\Sigma_1$  and for each CQ  $q$  over  $\Sigma_2$ , we have that  $\text{cert}(q, \langle \mathcal{T}_2, \text{chase}_{\mathcal{M}, \Sigma_2}(\mathcal{A}_1) \rangle) \subseteq \text{cert}(q, \langle \mathcal{T}_1 \cup \mathcal{M}, \mathcal{A}_1 \rangle)$ .

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

**Step 1:** *Checking whether for each ABox  $\mathcal{A}_1$  over  $\Sigma_1$ ,  $\langle \mathcal{T}_2, \text{chase}_{\mathcal{M}, \Sigma_2}(\mathcal{A}_1) \rangle$  is a CQ-solution for  $\langle \mathcal{T}_1, \mathcal{A}_1 \rangle$  under  $\mathcal{M}$ .*

We introduce now some notions that help us to characterize when a mapping is able to “translate” the inclusions in  $\mathcal{T}_1$  to the target TBox.

We use  $pn(X)$  to denote  $A$  if  $X$  is  $A$ , and  $P$  if  $X$  is  $\exists P$ ,  $\exists P^-$ ,  $P$ , or  $P^-$ . For  $X, Y$   $DL\text{-Lite}_{\text{RDFS}}$  expressions, we say that  $X$  is *compatible with*  $Y$  if (i)  $pn(X) = pn(Y)$ , and (ii) if  $X$  is  $\exists R$ , then  $Y$  is  $\exists R$ ,  $R$ , or  $R^-$ . For a  $DL\text{-Lite}_{\text{RDFS}}$  inclusion  $\alpha = X_1 \sqsubseteq X_2$ , we use  $lhs(\alpha)$  to denote  $X_1$ , and  $rhs(\alpha)$  to denote  $X_2$ . We say that  $\alpha$  is *left-compatible* (resp., *right-compatible*) with  $X$ , if  $X$  is compatible with  $lhs(\alpha)$  (resp.,  $rhs(\alpha)$ ).

Let  $\mathcal{M}$  be a definite mapping,  $\alpha = N_1 \sqsubseteq M_1$  a  $DL\text{-Lite}_{\text{RDFS}}$  inclusion over  $\Sigma_1$ , and  $\mu \in \mathcal{M}$  left-compatible with  $M_1$ . Then the *translation set* of  $\alpha$  and  $\mu$  in  $\mathcal{M}$ , denoted  $M(\alpha, \mu, \mathcal{M})$ , is the set of  $DL\text{-Lite}_{\text{RDFS}}$  inclusions over  $\Sigma_2$  such that, if there

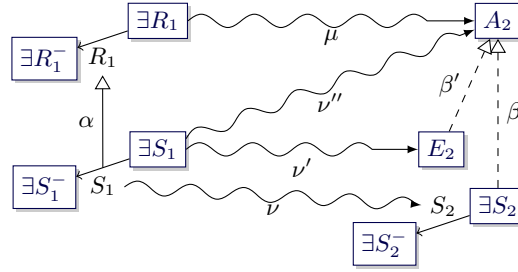
**Table 1.** Definition of  $M(\alpha, \mu, \mathcal{M})$ .  $A_i, E_i$  are atomic concepts,  $R_i, S_i$  are basic roles.

$\alpha$	$\mu$	$\nu$	$\beta$
$A_1 \sqsubseteq E_1$	$E_1 \sqsubseteq E_2$	$A_1 \sqsubseteq A_2$	$A_2 \sqsubseteq E_2$
$\exists R_1 \sqsubseteq E_1$	$E_1 \sqsubseteq E_2$	$\exists R_1 \sqsubseteq A_2$	$A_2 \sqsubseteq E_2$
		$R_1 \sqsubseteq R_2$	$\exists R_2 \sqsubseteq E_2$
$R_1 \sqsubseteq S_1$	$S_1 \sqsubseteq S_2$	$R_1 \sqsubseteq R_2$	$R_2 \sqsubseteq S_2$
	$\exists S_1 \sqsubseteq E_2$	$\exists R_1 \sqsubseteq A_2$	$A_2 \sqsubseteq E_2$
		$R_1 \sqsubseteq R_2$	$\exists R_2 \sqsubseteq E_2$
	$\exists S_1^- \sqsubseteq E_2$	$\exists R_1^- \sqsubseteq A_2$	$A_2 \sqsubseteq E_2$
$R_1 \sqsubseteq R_2$		$\exists R_2^- \sqsubseteq E_2$	



exists an inclusion  $\nu$  in  $\mathcal{M}$  left-compatible with  $N_1$ , then  $\beta \in \mathcal{M}(\alpha, \mu, \mathcal{M})$ , where  $\beta$  is uniquely defined by  $\alpha, \mu, \nu$  as shown in Table 1. When the mapping  $\mathcal{M}$  is clear from the context, we write  $\mathcal{M}(\alpha, \mu)$ . We say that  $\alpha$  is *redundant* w.r.t.  $\mu = M'_1 \sqsubseteq M_2$  (in  $\mathcal{M}$ ) if  $M_2 \sqsubseteq M_2 \in \mathcal{M}(\alpha, \mu)$ . We explain these definitions in an example.

*Example 3.* Let  $\alpha = S_1 \sqsubseteq R_1$  and  $\mu = \exists R_1 \sqsubseteq A_2$ .



Let  $\nu = S_1 \sqsubseteq S_2$  and  $\nu' = \exists S_1 \sqsubseteq E_2$  be in  $\mathcal{M}$ . Then  $\{\exists S_2 \sqsubseteq A_2, E_2 \sqsubseteq A_2\} \subseteq \mathcal{M}(\alpha, \mu, \mathcal{M})$ . If  $\nu'' = \exists S_1 \sqsubseteq A_2 \in \mathcal{M}$ , then  $\alpha$  is redundant w.r.t.  $\mu$ .

If none of  $\nu, \nu'$  and  $\nu''$  is in  $\mathcal{M}$ , then  $\mathcal{M}(\alpha, \mu, \mathcal{M})$  is empty. ■

With these notions in place, we can characterize CQ-solutions in the context of the representability problem.

**Proposition 4.** *Let  $\mathcal{M}$  be a definite mapping,  $\mathcal{T}_1$  a  $DL\text{-Lite}_{\text{RDFS}}$  TBox over  $\Sigma_1$ , and  $\mathcal{T}_2$  a  $DL\text{-Lite}_{\text{RDFS}}$  TBox over  $\Sigma_2$ . Then for each ABox  $\mathcal{A}_1$ , the KB  $\langle \mathcal{T}_2, \text{chase}_{\mathcal{M}, \Sigma_2}(\mathcal{A}_1) \rangle$  is a CQ-solution for  $\langle \mathcal{T}_1, \mathcal{A}_1 \rangle$  under  $\mathcal{M}$  if and only if for each inclusion  $\alpha$ , s.t.  $\mathcal{T}_1 \models \alpha$ , and for each inclusion  $\mu \in \mathcal{M}$  left-compatible with  $\text{rhs}(\alpha)$ , there exists  $\beta \in \mathcal{M}(\alpha, \mu)$  such that  $\mathcal{T}_2 \models \beta$ .*

**Step 2:** *Checking whether for each ABox  $\mathcal{A}_1$  over  $\Sigma_1$ , and for each CQ  $q$  over  $\Sigma_2$ , we have that  $\text{cert}(q, \langle \mathcal{T}_2, \text{chase}_{\mathcal{M}, \Sigma_2}(\mathcal{A}_1) \rangle) \subseteq \text{cert}(q, \langle \mathcal{T}_1 \cup \mathcal{M}, \mathcal{A}_1 \rangle)$ .*

Recall the definition of the translation set. Now, let  $\beta = N_2 \sqsubseteq M_2$  be a  $DL\text{-Lite}_{\text{RDFS}}$  inclusion over  $\Sigma_2$ , and  $\nu \in \mathcal{M}$  right-compatible with  $N_2$ . Then the *reverse-translation set* of  $\beta$  and  $\nu$  in  $\mathcal{M}$ , denoted  $\mathcal{M}^-(\beta, \nu, \mathcal{M})$ , is the set of  $DL\text{-Lite}_{\text{RDFS}}$  inclusions over  $\Sigma_1$  such that, if there exists an inclusion  $\mu$  in  $\mathcal{M}$  right-compatible with  $M_2$ , then  $\alpha \in \mathcal{M}^-(\beta, \nu, \mathcal{M})$ , where  $\alpha$  is uniquely defined by  $\beta, \nu$ , and  $\mu$  as shown in Table 1.

**Proposition 5.** *Let  $\mathcal{M}$  be a definite mapping,  $\mathcal{T}_1$  a  $DL\text{-Lite}_{\text{RDFS}}$  TBox over  $\Sigma_1$ , and  $\mathcal{T}_2$  a  $DL\text{-Lite}_{\text{RDFS}}$  TBox over  $\Sigma_2$ . Then for each ABox  $\mathcal{A}_1$  over  $\Sigma_1$  and for each CQ  $q$  over  $\Sigma_2$ ,  $\text{cert}(q, \langle \mathcal{T}_2, \text{chase}_{\mathcal{M}, \Sigma_2}(\mathcal{A}_1) \rangle) \subseteq \text{cert}(q, \langle \mathcal{T}_1 \cup \mathcal{M}, \mathcal{A}_1 \rangle)$  if and only if for each inclusion  $\beta$ , s.t.  $\mathcal{T}_2 \models \beta$ , and for each inclusion  $\nu \in \mathcal{M}$  right-compatible with  $\text{lhs}(\beta)$ , there exists  $\alpha \in \mathcal{M}^-(\beta, \nu)$ , s.t.  $\mathcal{T}_1 \models \alpha$ .*

With the techniques for Steps 1 and 2 at hand, we are ready to characterize representations.

**Corollary 3.** *Let  $\mathcal{M}$  be a definite mapping,  $\mathcal{T}_1$  a  $DL\text{-Lite}_{\text{RDFS}}$  TBox over  $\Sigma_1$ , and  $\mathcal{T}_2$  a  $DL\text{-Lite}_{\text{RDFS}}$  TBox over  $\Sigma_2$ . Then for each ABox  $\mathcal{A}_1$  over  $\Sigma_1$ ,  $\langle \mathcal{T}_2, \text{chase}_{\mathcal{M}, \Sigma_2}(\mathcal{A}_1) \rangle$  is a universal CQ-solution for  $\langle \mathcal{T}_1, \mathcal{A}_1 \rangle$  under  $\mathcal{M}$  iff*

- for each inclusion  $\alpha$ , s.t.  $\mathcal{T}_1 \models \alpha$ , and for each inclusion  $\mu \in \mathcal{M}$  left-compatible with  $\text{rhs}(\alpha)$ , there exists  $\beta \in \text{M}(\alpha, \mu)$  s.t.  $\mathcal{T}_2 \models \beta$ , and
- for each inclusion  $\beta$ , s.t.  $\mathcal{T}_2 \models \beta$ , and for each inclusion  $\nu \in \mathcal{M}$  right-compatible with  $\text{lhs}(\beta)$ , there exists  $\alpha \in \text{M}^-(\beta, \nu)$ , s.t.  $\mathcal{T}_1 \models \alpha$ .

Now, we can check whether a given  $\mathcal{T}_1$  is representable in a given  $\mathcal{M}$ .

**Theorem 1.** *Let  $\mathcal{M}$  be a definite mapping and  $\mathcal{T}_1$  a  $DL\text{-Lite}_{\text{RDFS}}$  TBox over  $\Sigma_1$ . Then we can check whether  $\mathcal{T}_1$  is representable in  $\mathcal{M}$  in polynomial time.*

## 4.2 Checking Weak Representability

We can easily solve the weak representability problem for  $DL\text{-Lite}_{\text{RDFS}}$  KBs even if the mappings are arbitrary tgds, i.e., assertions of the form  $q_1 \rightarrow q_2$ , mapping a CQ  $q_1$  over  $\Sigma_1$  to a CQ  $q_2$  over  $\Sigma_2$  of the same arity as  $q_1$ . We call such a mapping a tgd-mapping. Let  $\mathcal{T}_1$  be a  $DL\text{-Lite}_{\text{RDFS}}$  TBox and  $\mathcal{M}$  a tgd-mapping. We can enrich  $\mathcal{M}$  by compiling knowledge from  $\mathcal{T}_1$  into it:

- Let  $q_1 \rightarrow q_2$  be a tgd in  $\mathcal{M}$ , with  $q_1$  a CQ over  $\Sigma_1$  and  $q_2$  a CQ over  $\Sigma_2$ . Let the UCQ  $Q_1 = \{q_1^1, \dots, q_1^k\}$  be the perfect reformulation of  $q_1$  w.r.t.  $\mathcal{T}_1$  (see, e.g., [4]). Then we extend  $\mathcal{M}$  with a tgd  $q_1^i \rightarrow q_2$  for each  $q_1^i \in Q_1$ .
- We perform this for each tgd in  $\mathcal{M}$ . The resulting mapping is denoted with  $\mathcal{M}^*$ .

**Proposition 6.** *Let  $\mathcal{M}$  be a tgd-mapping,  $\mathcal{T}_1$  a  $DL\text{-Lite}_{\text{RDFS}}$  TBox over  $\Sigma_1$ , and  $\mathcal{M}^*$  the mapping constructed as described above. Then for each ABox  $\mathcal{A}_1$  over  $\Sigma_1$ , the KB  $\langle \emptyset, \text{chase}_{\mathcal{M}^*, \Sigma_2}(\mathcal{A}_1) \rangle$  is a universal CQ-solution for  $\langle \mathcal{T}_1, \mathcal{A}_1 \rangle$  under  $\mathcal{M}$ .*

From this we immediately get:

**Theorem 2.** *Let  $\mathcal{M}$  be a definite mapping and  $\mathcal{T}_1$  a  $DL\text{-Lite}_{\text{RDFS}}$  TBox over  $\Sigma_1$ . Then  $\mathcal{T}_1$  is weakly representable in  $\mathcal{M}$ .*

Thus, when the source KB is expressed in  $DL\text{-Lite}_{\text{RDFS}}$  it is always possible to separate data by enriching mappings, even if mappings are tgds. When  $\mathcal{M}$  is a set of tgds, the size of  $\mathcal{M}^*$  might be exponential in the size of  $\mathcal{M}$ . If  $\mathcal{M}$  is a  $DL\text{-Lite}_{\text{RDFS}}$  mapping, the size of  $\mathcal{M}^*$  is always polynomial in the size of  $\mathcal{M}$ .

## 5 Conclusions

We have specialized the framework for KB exchange proposed in [1] to the case of DLs, i.e., the source and target KBs are given as DL KBs and the mappings have the form of DL TBox assertions. Moreover, we have defined a new reasoning problem: representability of a TBox in a mapping, whose goal is to compute from a source TBox and a mapping, a target TBox that leads to a universal (CQ-)solution when it is combined with a suitable ABox computed from the source ABox, independently of the

actual source ABox. A variation of this problem, called weak representability, allows for modification of the mapping, so that the source TBox becomes representable.

Then, we have developed first results and techniques for KB exchange and for the representability problem in the case of mappings and KBs expressed in  $DL\text{-}Lite_{RDFS}$  (such mappings are called definite).  $DL\text{-}Lite_{RDFS}$  is a fragment of  $DL\text{-}Lite_{\mathcal{R}}$  that does not allow for existential quantification (i.e., concepts of the form  $\exists R$ ) in the right-hand side of concept inclusions, nor for disjointness assertions. It implies, first, that the chase is always finite in  $DL\text{-}Lite_{RDFS}$ , and, second, that KBs are always consistent. We have shown the following results for definite mappings and  $DL\text{-}Lite_{RDFS}$  KBs: (i) the problems of computing (universal) (CQ-)solutions can be solved in polynomial time; (ii) the problem of representability of a TBox in a mapping is decidable in polynomial time; (iii) every  $DL\text{-}Lite_{RDFS}$  TBox is weakly representable in a definite mapping.

The main direction for future work is to extend the results to the case of full  $DL\text{-}Lite_{\mathcal{R}}$ . This brings up the problem of labelled nulls in the chase, which becomes infinite in general. Moreover, due to the possible presence of disjointness assertions in TBoxes, consistency of the source and target KBs has to be checked.

### Acknowledgments

The authors are grateful to Vladislav Ryzhikov and Evgeny Sherkhonov for helpful discussions, and to the anonymous referees for many helpful comments. The authors were supported by Marie Curie action “International Research Staff Exchange Scheme (IRSES)”, FP7-PEOPLE-2009-IRSES, under grant number 24761. M. Arenas was also supported by Fondecyt grant 1090565. E. Botoeva and D. Calvanese were also supported by the EU under the ICT Collaborative Project ACSI (Artifact-Centric Service Interoperation), grant agreement number FP7-257593, and under the large-scale integrating project (IP) OntoRule (ONTOlogies meet Business RULEs), grant agreement number FP7-231875.

### References

1. M. Arenas, J. Pérez, and J. 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.
2. B. Barcelò. Logical foundations of relational data exchange. *SIGMOD Record*, 38(1):49–58, 2009.
3. D. Brickley and R. V. Guha. 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/>.
4. D. Calvanese, G. De Giacomo, D. Lembo, M. Lenzerini, and R. Rosati. Tractable reasoning and efficient query answering in description logics: The  $DL\text{-}Lite$  family. *J. of Automated Reasoning*, 39(3):385–429, 2007.
5. R. Fagin, P. G. Kolaitis, R. J. Miller, and L. Popa. Data exchange: Semantics and query answering. *Theoretical Computer Science*, 336(1):89–124, 2005.
6. L. Libkin and C. Sirangelo. Data exchange and schema mappings in open and closed worlds. *J. of Computer and System Sciences*, 77(3):542–571, 2011.