Adding Weight to *DL-Lite*

A. Artale,[1] D. Calvanese,[1] R. Kontchakov,[2] and M. Zakharyaschev[2]

[1] KRDB Research Centre
Free University of Bozen-Bolzano
I-39100 Bolzano, Italy
lastname@inf.unibz.it

[2] School of Comp. Science and Inf. Sys.
Birkbeck College
London WC1E 7HX, UK
{roman,michael}@dcs.bbk.ac.uk

1 Introduction

Description logics (DLs) have recently been used to provide access to large amounts of data through a high-level conceptual interface, which is of relevance in several application contexts, notably data integration and ontology-based data access. Besides the traditional reasoning services of knowledge base satisfiability and instance checking, a further important service in that context is that of answering complex database-like queries by fully taking into account the axioms in the TBox and the data stored in the ABox. The key property for such an approach to be viable in practice is the efficiency of query evaluation, in particular for conjunctive queries and, more generally, for positive existential queries (this class of queries includes unions of conjunctive queries) [1]. To address these needs, the *DL-Lite family* of description logics has been proposed and investigated in [6–8, 16], with the aim of identifying a class of DLs that could capture typical conceptual modeling formalisms, such as UML class diagrams and ER models, and for which query answering could be performed efficiently in terms of *data complexity*. The data complexity measure presupposes that only the size of the ABox is considered as variable while the sizes of the TBox and the query are regarded as fixed. Such a measure is important since in the typical application contexts we are interested in here, the size of the data stored in the ABox largely dominates that of the TBox and the query. As shown in [6–8, 16], for the logics of the *DL-Lite* family, a (union of) conjunctive queries posed over a TBox can be answered by rewriting it into a new union of conjunctive queries that has 'compiled in' the assertions in the TBox, and that can simply be evaluated (by a relational engine) over the ABox to produce the correct answer to the original query. In other words, it was shown that such logics enjoy *FO rewritability* [7, 8], and so belong to the complexity class FO in terms of descriptive complexity theory, and to the class AC^0 in terms of circuit complexity [10].

Successive work [3] has shown that some of the nice computational properties of *DL-Lite* logics can be preserved, even when they are extended with additional constructs used in conceptual modeling. In particular, it was proved in [3] that the data complexity of answering positive existential queries stays in AC^0 for the logic $\textit{DL-Lite}^{\mathcal{N}}_{horn}$ which allows conjunctions on the left-hand side of concept inclusions as well as arbitrary number restrictions. Moreover, the same data complexity bound holds also for satisfiability and instance checking in the logic

Language	Combined complexity	Data complexity	
	Satisfiability	Instance checking	Query answering
$DL\text{-}Lite_{core}^{(\mathcal{RN})}$	NLogSpace	in AC^0	in AC^0
$DL\text{-}Lite_{horn}^{(\mathcal{RN})}$	P $\leq$ [Th.1]	in AC^0	in AC^0 $\leq$ [Th.3]
$DL\text{-}Lite_{krom}^{(\mathcal{RN})}$	NLogSpace $\leq$ [Th.1]	in AC^0	coNP $\geq$[18]
$DL\text{-}Lite_{bool}^{(\mathcal{RN})}$	NP $\leq$ [Th.1]	in AC^0 $\leq$ [Th.2]	coNP $\leq$[14, 13, 9]

Table 1. Combined and data complexity.

$DL\text{-}Lite_{bool}^{\mathcal{N}}$ which allows full Booleans as concept constructs. (Note that these results hold only under the unique name assumption. *DL-Lite* logics without this assumption are investigated in [4].)

One aim of this paper is to extend $DL\text{-}Lite_{horn}^{\mathcal{N}}$, $DL\text{-}Lite_{bool}^{\mathcal{N}}$ and their fragments with a number of new constructs without spoiling their computational properties. The resulting logic is called $DL\text{-}Lite_{horn}^{(\mathcal{RN})}$. Another aim is to present explicit (exponential) FO rewritings of positive existential queries over $DL\text{-}Lite_{horn}^{(\mathcal{RN})}$ KBs. The constructs we add to our logics are as follows: (i) role inclusions, (ii) qualified number restrictions, and (iii) role disjointness, symmetry, asymmetry, reflexivity, and irreflexivity constraints. Needless to say that when adding (i) and (ii), we have to restrict the interaction of these constructs with number restrictions (otherwise, even the logic $DL\text{-}Lite_{core}^{\mathcal{R},\mathcal{F}}$ with extremely primitive concept inclusions, but with unrestricted role inclusions and global functionality constraints is ExpTime-complete for combined complexity and P-complete for data complexity [11].) This will be done by generalizing the ideas of [16]. Our main tool for dealing with *DL-Lite* logics is embedding into the one-variable fragment $\mathcal{QL}^1$ of first-order logic without equality and function symbols, which seems to be a natural logic-based characterization of the *DL-Lite* logics.

The complexity results obtained in this paper are summarized in Table 1.

2 $DL\text{-}Lite_{bool}^{(\mathcal{RN})}$ and its fragments

We start by defining the description logic $DL\text{-}Lite_{bool}^{(\mathcal{RN})}$, the most expressive of our logics, which subsumes, in particular, all members of the *DL-Lite* family [6–8].

The language of $DL\text{-}Lite_{bool}^{(\mathcal{RN})}$ contains *object names* $a_0, a_1, \dots$, *concept names* $A_0, A_1, \dots$, and *role names* $P_0, P_1, \dots$. Complex *concepts* C and *roles* R are defined as follows:

$$B \quad ::= \quad \bot \mid A_i \mid \geq q\, R, \qquad\qquad R \quad ::= \quad P_i \mid P_i^-,$$
$$C \quad ::= \quad B \mid \neg C \mid \geq q\, R.C \mid C_1 \sqcap C_2,$$

where q is a positive integer. The concepts of the form B will be called *basic*. A $DL\text{-}Lite_{bool}^{(\mathcal{RN})}$ *TBox*, $\mathcal{T}$, is a finite set of *concept inclusions* (CIs, for short), *role inclusions*, and *role constraints* of the form:

$$C_1 \sqsubseteq C_2, \qquad R_1 \sqsubseteq R_2, \qquad \mathsf{Dis}(R_1, R_2), \qquad \mathsf{Irr}(P_k), \qquad \text{and} \qquad \mathsf{Ref}(P_k).$$

We write $inv(R)$ for P_k^- if $R = P_k$, and for P_k if $R = P_k^-$. Denote by $\sqsubseteq_{\mathcal{T}}^*$ the reflexive and transitive closure of $\{(R, R'), (inv(R), inv(R')) \mid R \sqsubseteq R' \in \mathcal{T}\}$. Say that R' is a *proper sub-role of* R in $\mathcal{T}$ if $R' \sqsubseteq^*_{\mathcal{T}} R$ and $R \not\sqsubseteq^*_{\mathcal{T}} R'$. We impose the following syntactic conditions on $DL\text{-}Lite^{(\mathcal{RN})}_{bool}$ TBoxes $\mathcal{T}$ (cf. $DL\text{-}Lite_{\mathcal{A}}$ [16]):

(inter) if R has a proper sub-role in $\mathcal{T}$ then $\mathcal{T}$ contains no *negative occurrences*[1] of number restrictions $\geq q\, R$ or $\geq q\, inv(R)$ with $q \geq 2$;

(exists) $\mathcal{T}$ may contain only positive occurrences of $\geq q\, R.C$, and if $\geq q\, R.C$ occurs in $\mathcal{T}$ then $\mathcal{T}$ does not contain negative occurrences of $\geq q'\, R$ or $\geq q'\, inv(R)$, for $q' \geq 2$.

It follows that no TBox can contain both a functionality constraint $\geq 2\, R \sqsubseteq \bot$ and an occurrence of $\geq q\, R.C$, for some $q \geq 1$ and some role R.

An *ABox*, $\mathcal{A}$, is a finite set of assertions of the form: $A_k(a_i)$, $P_k(a_i, a_j)$ and $\neg P_k(a_i, a_j)$. Taken together, $\mathcal{T}$ and $\mathcal{A}$ constitute the $DL\text{-}Lite^{(\mathcal{RN})}_{bool}$ *knowledge base* (KB, for short) $\mathcal{K} = (\mathcal{T}, \mathcal{A})$.

As usual in description logic, an *interpretation*, $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$, consists of a nonempty *domain* $\Delta^{\mathcal{I}}$ and an interpretation function $\cdot^{\mathcal{I}}$ that assigns to each object name a_i an element $a_i^{\mathcal{I}} \in \Delta^{\mathcal{I}}$, to each concept name A_i a subset $A_i^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$, and to each role name P_i a binary relation $P_i^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$. In this paper, we adopt the *unique name assumption* (UNA): $a_i^{\mathcal{I}} \neq a_j^{\mathcal{I}}$, for all $i \neq j$, and refer the reader to [4] for results on the *DL-Lite* logics without UNA.

The role and concept constructs are interpreted in $\mathcal{I}$ in the standard way. We will also use standard abbreviations such as $\top = \neg\bot$, $\exists R = (\geq 1\, R)$ and $\leq q\, R = \neg(\geq q+1\, R)$. The *satisfaction relation* $\models$ is also standard; we only mention here that $\mathcal{I} \models \mathsf{Dis}(R_1, R_2)$ iff $R_1^{\mathcal{I}} \cap R_2^{\mathcal{I}} = \emptyset$ (R_1 and R_2 are *disjoint*), $\mathcal{I} \models \mathsf{Irr}(P_k)$ iff $(x, x) \notin P_k^{\mathcal{I}}$ for all $x \in \Delta^{\mathcal{I}}$ (P_k is *irreflexive*), $\mathcal{I} \models \mathsf{Ref}(P_k)$ iff $(x, x) \in P_k^{\mathcal{I}}$ for all $x \in \Delta^{\mathcal{I}}$ (P_k is *reflexive*). Note that symmetric and asymmetric role constraints can be regarded as syntactic sugar in this language: $\mathsf{Sym}(P_k)$ and $\mathsf{Asym}(P_k)$ can be equivalently replaced with $P_k^- \sqsubseteq P_k$ and $\mathsf{Dis}(P_k, P_k^-)$, respectively (extending a TBox with $P_k^- \sqsubseteq P_k$ cannot violate **(inter)** as P_k^- is not a *proper* sub-role of P_k). A KB $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ is said to be *satisfiable* (or *consistent*) if there is an interpretation, $\mathcal{I}$, satisfying all the members of $\mathcal{T}$ and $\mathcal{A}$. In this case we write $\mathcal{I} \models \mathcal{K}$ (as well as $\mathcal{I} \models \mathcal{T}$ and $\mathcal{I} \models \mathcal{A}$) and say that $\mathcal{I}$ is a *model of* $\mathcal{K}$.

It is to be emphasized that such constructs as role constraints and qualified number restrictions are used in conceptual modeling and also belong to the OWL 2 proposal; moreover, as we show, adding them does not affect the computational complexity of our logics.

Similarly to classical logic, we adopt the following definitions: a TBox $\mathcal{T}$ is

- a $DL\text{-}Lite^{(\mathcal{RN})}_{horn}$ TBox if its CIs are of the form $B_1 \sqcap \cdots \sqcap B_k \sqsubseteq B$ (the B_i and B are basic concepts and, by definition, the empty conjunction is $\top$);
- a $DL\text{-}Lite^{(\mathcal{RN})}_{krom}$ TBox[2] if its CIs are of the form $B_1 \sqsubseteq B_2$, $B_1 \sqsubseteq \neg B_2$ or $\neg B_1 \sqsubseteq B_2$;

[1] An occurrence of a concept on the right-hand (resp., left-hand) side of a concept inclusion is called *negative* if it is in the scope of an odd (resp., even) number of negations $\neg$; otherwise the occurrence is called *positive*.

[2] The Krom fragment of first-order logic consists of all formulas in prenex normal form whose quantifier-free part is a conjunction of binary clauses.

– a $DL\text{-}Lite^{(\mathcal{RN})}_{core}$ TBox if its CIs are of the form $B_1 \sqsubseteq B_2$ or $B_1 \sqsubseteq \neg B_2$.

As $B_1 \sqsubseteq \neg B_2$ is equivalent to $B_1 \sqcap B_2 \sqsubseteq \bot$, core TBoxes can be regarded as both Krom and Horn TBoxes. We note here that a concept C occurring in $\mathcal{T}$ in some $\geq q\, R.C$ can be a conjunction of any concepts allowed on the right-hand side of concept inclusions in the respective language.

3 *DL-Lite* in the Light of First-Order Logic

Our main aim in this section is to prove the upper combined complexity bounds for reasoning in $DL\text{-}Lite^{(\mathcal{RN})}_{bool}$ and its fragments and develop the technical tools we need to investigate the data complexity of query answering in $DL\text{-}Lite^{(\mathcal{RN})}_{horn}$.

For a $DL\text{-}Lite^{(\mathcal{RN})}_{bool}$ KB $\mathcal{K} = (\mathcal{T}, \mathcal{A})$, denote by $role^{\pm}(\mathcal{K})$ the set of role names occurring in $\mathcal{K}$ and their inverses and by $ob(\mathcal{A})$ the set of object names occurring in $\mathcal{A}$. Let $Q^R_{\mathcal{T}}$ be the set of natural numbers containing 1 and all the numerical parameters q such that $\geq q\, R$ or $\geq q\, R.C$ occurs in $\mathcal{T}$. Note that $|Q^R_{\mathcal{T}}| \geq 2$ if $\mathcal{T}$ contains a functionality constraint for R. Our main result in this section is:

Theorem 1. (i) *Satisfiability of* $DL\text{-}Lite^{(\mathcal{RN})}_{bool}$ *KBs is* NP*-complete*; (ii) *satisfiability of* $DL\text{-}Lite^{(\mathcal{RN})}_{horn}$ *KBs is* P*-complete*; *and* (iii) *satisfiability of* $DL\text{-}Lite^{(\mathcal{RN})}_{krom}$ *and* $DL\text{-}Lite^{(\mathcal{RN})}_{core}$ *KBs is* NLogSpace*-complete.*

Let us consider first the sub-language of $DL\text{-}Lite^{(\mathcal{RN})}_{bool}$ without qualified number restrictions and role constraints, which will be required for purely technical reasons; we denote it by $DL\text{-}Lite^{(\mathcal{RN})^-}_{bool}$. In Section 4, we will also use $DL\text{-}Lite^{(\mathcal{RN})^-}_{horn}$.

Let $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ be a $DL\text{-}Lite^{(\mathcal{RN})^-}_{bool}$ KB and let Id be a distinguished role name, which will be used to simulate the *identity relation* required for encoding the role constraints. We assume that either $\mathcal{K}$ does not contain Id at all or satisfies the following conditions:

(Id$_1$) $Id(a_i, a_j) \in \mathcal{A}$ iff $i = j$, for all $a_i, a_j \in ob(\mathcal{A})$,
(Id$_2$) $\{\top \sqsubseteq \exists Id,\ \ Id^- \sqsubseteq Id\} \subseteq \mathcal{T}$, and $Q^{Id}_{\mathcal{T}} = Q^{Id^-}_{\mathcal{T}} = \{1\}$,
(Id$_3$) Id is only allowed in role inclusions of the form $Id^- \sqsubseteq Id$ and $Id \sqsubseteq R$.

We assume, without loss of generality, that $Q^R_{\mathcal{T}} \subseteq Q^{R'}_{\mathcal{T}}$ whenever $R \sqsubseteq^*_{\mathcal{T}} R'$ (for if this is not the case we can always introduce the missing numbers in $Q^{R'}_{\mathcal{T}}$, e.g., by adding $\bot \sqsubseteq\ \geq q\, R'$ to the TBox).

We present now a reduction of the satisfiability problem for $DL\text{-}Lite^{(\mathcal{RN})^-}_{bool}$ KBs to satisfiability of first-order formulas with one variable, or $\mathcal{QL}^1$-formulas. With every object name $a_i \in ob(\mathcal{A})$ we associate the individual constant a_i of $\mathcal{QL}^1$ and with every concept name A_i the unary predicate $A_i(x)$ from the signature of $\mathcal{QL}^1$. For each role $R \in role^{\pm}(\mathcal{K})$, we introduce $|Q^R_{\mathcal{T}}|$-many fresh unary predicates $E_q R(x)$, for $q \in Q^R_{\mathcal{T}}$. The intended meaning of these predicates is as follows: for a role name P_k, $E_1 P_k(x)$ and $E_1 P_k^-(x)$ represent the domain and range of P_k, respectively; more generally, for each $q \in Q^R_{\mathcal{T}}$, $E_q P_k(x)$ and $E_q P_k^-(x)$ represent the sets of points with *at least* q distinct P_k-successors and *at* *least* q distinct P_k-predecessors, respectively. We write $\mathit{inv}(E_qR)(x)$ for $E_qP_k^-(x)$ if $R = P_k$, and for $E_qP_k(x)$ if $R = P_k^-$. Additionally, for every pair of roles $P_k, P_k^- \in \mathit{role}^\pm(\mathcal{K})$, we take two fresh individual constants dp_k and dp_k^- of $\mathcal{QL}^1$, which will serve as 'representatives' of the points from the domain and range of P_k, respectively (provided that it is not empty). Denote the set of all those dp_k and dp_k^- by $dr(\mathcal{K})$ and write $\mathit{inv}(dr)$ for dp_k^- if $R = P_k$, and for dp_k if $R = P_k^-$. By induction on the construction of concept C we define the $\mathcal{QL}^1$-formula C^*:

$$\bot^* = \bot, \qquad (A_i)^* = A_i(x), \qquad (\geq q\, R)^* = E_qR(x),$$
$$(\neg C)^* = \neg C^*(x), \qquad (C_1 \sqcap C_2)^* = C_1^*(x) \wedge C_2^*(x).$$

For every role $R \in \mathit{role}^\pm(\mathcal{K})$, we need two $\mathcal{QL}^1$-formulas:

$$\varepsilon_R(x) = E_1R(x) \to \mathit{inv}(E_1R)(\mathit{inv}(dr)),$$
$$\delta_R(x) = \bigwedge_{\substack{q,q' \in Q_{\mathcal{T}}^R,\ \ q' > q \\ q' > q'' > q \text{ for no } q'' \in Q_{\mathcal{T}}^R}} \big(E_{q'}R(x) \to E_qR(x)\big).$$

Formula $\varepsilon_R(x)$ says that if the domain of R is not empty then its range is not empty either: it contains the constant $\mathit{inv}(dr)$, the 'representative' of the domain of $\mathit{inv}(R)$. The meaning of $\delta_R(x)$ should be obvious. For a KB $\mathcal{K}$, we define

$$\mathcal{K}^{\ddagger_{\mathsf{e}}} = \forall x \Big[\mathcal{T}^{*\mathcal{R}}(x) \ \wedge \bigwedge_{R \in \mathit{role}^\pm(\mathcal{K})} \big(\varepsilon_R(x) \wedge \delta_R(x)\big)\Big] \ \wedge \ \mathcal{A}^{\ddagger_{\mathsf{e}}}, \qquad \text{where}$$

$$\mathcal{T}^{*\mathcal{R}}(x) = \bigwedge_{C_1 \sqsubseteq C_2 \in \mathcal{T}} \big(C_1^*(x) \to C_2^*(x)\big) \ \wedge \bigwedge_{\substack{R \sqsubseteq R' \in \mathcal{T} \text{ or} \\ \mathit{inv}(R) \sqsubseteq \mathit{inv}(R') \in \mathcal{T}}} \ \bigwedge_{q \in Q_{\mathcal{T}}^R} \big(E_qR(x) \to E_qR'(x)\big),$$

$$\mathcal{A}^{\ddagger_{\mathsf{e}}} = \bigwedge_{A_k(a_i) \in \mathcal{A}} A_k(a_i) \ \wedge \bigwedge_{R(a,a') \in \mathsf{Cl}_{\mathcal{T}}^{\mathsf{e}}(\mathcal{A})} E_{q^{\mathsf{e}}_{R,a}}R(a) \ \wedge \bigwedge_{\neg P_k(a_i,a_j) \in \mathcal{A}} (\neg P_k(a_i,a_j))^{\perp_{\mathsf{e}}},$$

$\mathsf{Cl}_{\mathcal{T}}^{\mathsf{e}}(\mathcal{A}) = \{R'(a_i,a_j) \mid R(a_i,a_j) \in \mathcal{A},\ R \sqsubseteq_{\mathcal{T}}^* R'\}$;[3] $q^{\mathsf{e}}_{R,a}$ is the maximum number in $Q_{\mathcal{T}}^R$ such that there are $q^{\mathsf{e}}_{R,a}$ many distinct a_i with $R(a,a_i) \in \mathsf{Cl}_{\mathcal{T}}^{\mathsf{e}}(\mathcal{A})$, and $(\neg P_k(a_i,a_j))^{\perp_{\mathsf{e}}} = \bot$ if $P_k(a_i,a_j) \in \mathsf{Cl}_{\mathcal{T}}^{\mathsf{e}}(\mathcal{A})$ and $\top$ otherwise. Note that the size of $\mathcal{K}^{\ddagger_{\mathsf{e}}}$ is linear in the size of $\mathcal{K}$, *no matter whether the numerical parameters are coded in unary or in binary.* The following lemma is an analogue of [3, Theorem 1] (for the proof see [4]):

Lemma 1. *A DL-Lite*$_{bool}^{(\mathcal{RN})^-}$ *KB* $\mathcal{K}$ *is satisfiable iff the* $\mathcal{QL}^1$*-sentence* $\mathcal{K}^{\ddagger_{\mathsf{e}}}$ *is satisfiable.*

It should be clear that the translation $\cdot^{\ddagger_{\mathsf{e}}}$ can be computed in NLogSpace for combined complexity. Indeed, this is trivial for the first conjunct of $\mathcal{K}^{\ddagger_{\mathsf{e}}}$. To compute $\mathcal{A}^{\ddagger_{\mathsf{e}}}$, we first need to be able to check, given a role R and a pair of objects a_i, a_j, whether $R(a_i,a_j) \in \mathsf{Cl}_{\mathcal{T}}^{\mathsf{e}}(\mathcal{A})$ and second, given $R(a,a') \in \mathsf{Cl}_{\mathcal{T}}^{\mathsf{e}}(\mathcal{A})$, to

[3] We slightly abuse notation and write $R(a_i,a_j) \in \mathcal{A}$ to indicate that $P_k(a_i,a_j) \in \mathcal{A}$ if $R = P_k$, or $P_k(a_j,a_i) \in \mathcal{A}$ if $R = P_k^-$.

compute $q^{\mathbf{e}}_{R,a}$. The $R(a_i,a_j) \in \mathsf{Cl}^{\mathbf{e}}_{\mathcal{T}}(\mathcal{A})$ test can be done by a *non-deterministic* algorithm using space *logarithmic* in $|role^{\pm}(\mathcal{K})|$ (see, e.g., the NLogSpace directed graph reachability problem [12]). The following algorithm computes $q^{\mathbf{e}}_{R,a}$: set $q = 0$ and then enumerate all object names a_i in $\mathcal{A}$ incrementing q each time $R(a,a_i) \in \mathsf{Cl}^{\mathbf{e}}_{\mathcal{T}}(\mathcal{A})$; stop if $q = \max Q^R_{\mathcal{T}}$ or the end of the object name list is reached. The resulting $q^{\mathbf{e}}_{R,a}$ is the maximum number in $Q^R_{\mathcal{T}}$ not exceeding q.

As follows from the proof of Lemma 1, for a $\mathit{DL\text{-}Lite}^{(\mathcal{RN})^-}_{bool}$ KB $\mathcal{K} = (\mathcal{T},\mathcal{A})$, every model $\mathfrak{M}$ of $\mathcal{K}^{\ddagger_{\mathbf{e}}}$ induces a model $\mathcal{I}_{\mathfrak{M}}$ of $\mathcal{K}$ with the following properties:

(abox) For all $a_i, a_j \in ob(\mathcal{A})$, $(a_i^{\mathcal{I}_{\mathfrak{M}}}, a_j^{\mathcal{I}_{\mathfrak{M}}}) \in R^{\mathcal{I}_{\mathfrak{M}}}$ iff $R(a_i,a_j) \in \mathsf{Cl}^{\mathbf{e}}_{\mathcal{T}}(\mathcal{A})$.

(uniq) The object names $a \in ob(\mathcal{A})$ induce a partitioning of $\Delta^{\mathcal{I}_{\mathfrak{M}}}$ into disjoint labeled trees $\mathfrak{T}_a = (T_a, E_a, \ell_a)$ with nodes T_a, edges E_a, root $a^{\mathcal{I}_{\mathfrak{M}}}$, and a labeling function $\ell_a \colon E_a \to role^{\pm}(\mathcal{K}) \setminus \{Id, Id^-\}$.

(cp) There is a function $cp\colon \Delta^{\mathcal{I}_{\mathfrak{M}}} \to ob(\mathcal{A}) \cup dr(\mathcal{K})$ such that $cp(a^{\mathcal{I}_{\mathfrak{M}}}) = a$ for $a \in ob(\mathcal{A})$, and $cp(w) = dr$, for role R such that $w' \in T_a$, $(w',w) \in E_a$ and $\ell_a(w',w) = inv(R)$, for some $a \in ob(\mathcal{A})$.

(iso) For each $R \in role^{\pm}(\mathcal{K})$, all labeled subtrees generated by $w \in \Delta^{\mathcal{I}_{\mathfrak{M}}}$ with $cp(w) = dr$ are isomorphic.

(con) For all basic concepts B in $\mathcal{K}$ and $w \in \Delta^{\mathcal{I}_{\mathfrak{M}}}$, $w \in B^{\mathcal{I}_{\mathfrak{M}}}$ iff $\mathfrak{M} \models B^*[cp(w)]$.

(role) For every role name P_k, including Id,

$$P_k^{\mathcal{I}_{\mathfrak{M}}} = \big\{(a_i^{\mathcal{I}_{\mathfrak{M}}}, a_j^{\mathcal{I}_{\mathfrak{M}}}) \mid R(a_i,a_j) \in \mathcal{A},\ R \sqsubseteq^*_{\mathcal{T}} P_k\big\} \ \cup\ \big\{(w,w) \mid Id \sqsubseteq^*_{\mathcal{T}} P_k\big\} \ \cup \\ \big\{(w,w') \in E_a \mid a \in ob(\mathcal{A}),\ \ell_a(w,w') = R,\ R \sqsubseteq^*_{\mathcal{T}} P_k\big\}.$$

Such a model will be called an *untangled model of* $\mathcal{K}$ (*the untangled model of* $\mathcal{K}$ *induced by* $\mathfrak{M}$, to be more precise). It should be pointed out that there are two main distinguishing features of untangled models for $\mathit{DL\text{-}Lite}^{(\mathcal{RN})}_{bool}$ KBs: (i) there are at most $|ob(\mathcal{A})| + |role^{\pm}(\mathcal{K})|$ different types of points in them, and (ii) although two points may be connected by a set of roles Ω, one can always select $R \in \Omega$ such that Ω is an upward closure of $\{R\}$ under $\sqsubseteq^*_{\mathcal{T}}$, provided that one of the points is not from $ob(\mathcal{A})$.

The following lemma reduces satisfiability of $\mathit{DL\text{-}Lite}^{(\mathcal{RN})}_{bool}$ KBs to satisfiability of $\mathit{DL\text{-}Lite}^{(\mathcal{RN})^-}_{bool}$ KBs (for the proof see [4, Lemma 5.17]):

Lemma 2. *For every* $\mathit{DL\text{-}Lite}^{(\mathcal{RN})}_{bool}$ *KB* $\mathcal{K}' = (\mathcal{T}',\mathcal{A}')$, *one can construct* (*in linear time and logarithmic space*) *a* $\mathit{DL\text{-}Lite}^{(\mathcal{RN})^-}_{bool}$ *KB* $\mathcal{K} = (\mathcal{T},\mathcal{A})$ *such that*

- *every untangled model* $\mathcal{I}_{\mathfrak{M}}$ *of* $\mathcal{K}$ *is a model of* $\mathcal{K}'$, *provided that*
  there are no $R_1(a_i,a_j), R_2(a_i,a_j) \in \mathsf{Cl}^{\mathbf{e}}_{\mathcal{T}}(\mathcal{A})$ *with* $\mathsf{Dis}(R_1,R_2) \in \mathcal{T}'$, *and*
  there is no $R(a_i,a_i) \in \mathsf{Cl}^{\mathbf{e}}_{\mathcal{T}}(\mathcal{A})$ *with* $\mathsf{Irr}(R) \in \mathcal{T}'$;
- *every model* $\mathcal{I}'$ *of* $\mathcal{K}'$ *gives rise to a model* $\mathcal{I}$ *of* $\mathcal{K}$ *based on the same domain as* $\mathcal{I}'$ *and such that* $\mathcal{I}$ *agrees with* $\mathcal{I}'$ *on all symbols from* $\mathcal{K}'$.

Theorem 1 follows from Lemmas 2 and 1, the observation that $\mathcal{K}^{\ddagger_{\mathbf{e}}}$ is a $\mathcal{QL}^1$-formula, for a $\mathit{DL\text{-}Lite}^{(\mathcal{RN})}_{bool}$ KB, a universal Horn $\mathcal{QL}^1$-formula, for a $\mathit{DL\text{-}Lite}^{(\mathcal{RN})}_{horn}$ KB, and a universal Krom $\mathcal{QL}^1$-formula, for a $\mathit{DL\text{-}Lite}^{(\mathcal{RN})}_{krom}$ KB, and the complexity results for the respective fragments of $\mathcal{QL}^1$ [15, 5].

For the data complexity the following result is proved in [4, Section 6]:

Theorem 2. *The satisfiability and instance checking problems for $DL\text{-}Lite_{bool}^{(\mathcal{RN})}$ KBs are in* AC^0 *for data complexity.*

4 FO Rewritability of Query Answering

In this section we study the data complexity of query answering over $DL\text{-}Lite_{horn}^{(\mathcal{RN})}$ KBs. We assume that all concept and role names of a KB occur in its TBox and write $role^{\pm}(\mathcal{T})$ and $dr(\mathcal{T})$ instead of $role^{\pm}(\mathcal{K})$ and $dr(\mathcal{K})$, respectively. Denote by $Bcon(\mathcal{T})$ the set of basic concepts occurring in $\mathcal{T}$ (i.e., concepts of the form A and $\geq q\, R$, for a concept name A occurring in $\mathcal{T}$, $R \in role^{\pm}(\mathcal{T})$ and $q \in Q_{\mathcal{T}}^{R}$).

A *positive existential query* $\mathsf{q}(\boldsymbol{x})$ is a first-order formula $\varphi(\boldsymbol{x})$ constructed by means of conjunction, disjunction and existential quantification from atoms of the from $A(t_1)$ and $P(t_1, t_2)$, where t_1, t_2 are *terms* taken from the list of variables $y_0, y_1, \dots$ and object names $a_0, a_1, \dots$. The free variables of φ are called *distinguished variables* of q. An *assignment* $\mathfrak{a}$ in $\Delta^{\mathcal{I}}$ is a function associating with each variable y an element $\mathfrak{a}(y)$ of $\Delta^{\mathcal{I}}$. We write $a_i^{\mathcal{I},\mathfrak{a}} = a_i^{\mathcal{I}}$ and $y^{\mathcal{I},\mathfrak{a}} = \mathfrak{a}(y)$. For $\mathcal{K} = (\mathcal{T}, \mathcal{A})$, say that a tuple $\boldsymbol{a}$ of object names from $\mathcal{A}$ is a *certain answer* to $\mathsf{q}(\boldsymbol{x})$ w.r.t. $\mathcal{K}$, and write $\mathcal{K} \models \mathsf{q}(\boldsymbol{a})$, if $\mathcal{I} \models \mathsf{q}(\boldsymbol{a})$ whenever $\mathcal{I} \models \mathcal{K}$. The *query answering problem* is, given $\mathcal{K}$, a query $\mathsf{q}(\boldsymbol{x})$ and $\boldsymbol{a} \subseteq ob(\mathcal{A})$, decide whether $\mathcal{K} \models \mathsf{q}(\boldsymbol{a})$. Our main result in this section is the following:

Theorem 3. *The positive existential query answering problem for $DL\text{-}Lite_{horn}^{(\mathcal{RN})}$ KBs is in* AC^0 *for data complexity.*

Proof. Suppose that we are given a *consistent* $DL\text{-}Lite_{horn}^{(\mathcal{RN})}$ KB $\mathcal{K}' = (\mathcal{T}', \mathcal{A}')$ and a positive existential query in prenex form $\mathsf{q}(\boldsymbol{x}) = \exists \boldsymbol{y}\, \varphi(\boldsymbol{x}, \boldsymbol{y})$, $\boldsymbol{y} = y_1, \dots, y_k$, in the signature of $\mathcal{K}'$. Consider the $DL\text{-}Lite_{horn}^{(\mathcal{RN})^-}$ KB $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ provided by Lemma 2. The untangled models of $\mathcal{K}$ produce exactly the same answers as $\mathcal{K}'$:

Lemma 3. *For every tuple $\boldsymbol{a}$ of object names in $\mathcal{K}'$, we have $\mathcal{K}' \models \mathsf{q}(\boldsymbol{a})$ iff $\mathcal{I} \models \mathsf{q}(\boldsymbol{a})$ for all untangled models $\mathcal{I}$ of $\mathcal{K}$.*

Next we show that, as $\mathcal{K}^{\ddagger_e}$ is a universal Horn sentence, it is enough to consider just one special untangled model $\mathcal{I}_0$ of $\mathcal{K}$. Let $\mathfrak{M}_0$ be the *minimal Herbrand model* of $\mathcal{K}^{\ddagger_e}$. We remind the reader (see, e.g., [2, 17]) that $\mathfrak{M}_0$ can be constructed by taking the intersection of all Herbrand models for $\mathcal{K}^{\ddagger_e}$, that is, of all models based on the domain $\Lambda = ob(\mathcal{A}) \cup dr(\mathcal{T})$. It follows that

$$\mathfrak{M}_0 \models B^*[c] \quad \text{iff} \quad \mathcal{K}^{\ddagger_e} \models B^*(c), \qquad \text{for all } c \in \Lambda \text{ and } B \in Bcon(\mathcal{T}).$$

Denote by $\mathcal{I}_0$ the untangled model of $\mathcal{K}$ induced by $\mathfrak{M}_0$ and its domain by $\Delta^{\mathcal{I}_0}$. By Lemma 1 with **(con)** and **(cp)**,

$$a_i^{\mathcal{I}_0} \in B^{\mathcal{I}_0} \quad \text{iff} \quad \mathcal{K} \models B(a_i), \qquad \text{for all } a_i \in ob(\mathcal{A}) \text{ and } B \in Bcon(\mathcal{T}). \tag{1}$$

For each $R \in role^{\pm}(\mathcal{T})$, by Lemma 1, if $R^{\mathcal{I}_0} \neq \emptyset$ then $\mathfrak{M}_0 \models (\exists R)^*[dr]$ and thus, $(\mathcal{T} \cup \{\exists R \sqsubseteq \bot\}, \mathcal{A})$ is not satisfiable, whence $R^{\mathcal{I}} \neq \emptyset$, for all models $\mathcal{I}$ of $\mathcal{K}$. Moreover, if $R^{\mathcal{I}_0} \neq \emptyset$ then

$$w \in B^{\mathcal{I}_0} \quad \text{iff} \quad \mathcal{K} \models \exists R \sqsubseteq B, \qquad \text{for all } w \in \Delta^{\mathcal{I}_0} \text{ with } cp(w) = dr, \tag{2}$$

where $cp\colon \Delta^{\mathcal{I}_0} \to \Lambda$ is the function provided by **(cp)**.

Lemma 4. *If $\mathcal{I}_0 \models \mathsf{q}(\boldsymbol{a})$ then $\mathcal{I} \models \mathsf{q}(\boldsymbol{a})$ for all untangled models $\mathcal{I}$ of $\mathcal{K}$.*

Proof. Suppose $\mathcal{I} \models \mathcal{K}$. As $\mathsf{q}(\boldsymbol{a})$ is a positive existential sentence, it is enough to construct a homomorphism $h\colon \mathcal{I}_0 \to \mathcal{I}$. By **(uniq)**, $\Delta^{\mathcal{I}_0}$ is partitioned into trees $\mathfrak{T}_a$, for $a \in ob(\mathcal{A})$. Define the *depth* of $w \in \Delta^{\mathcal{I}_0}$ to be the length of the path in the respective tree from its root to w. Denote by W_m the set of points of depth $\leq m$; in particular, $W_0 = \{a^{\mathcal{I}_0} \mid a \in ob(\mathcal{A})\}$. We construct h as the union of homomorphisms $h_m\colon W_m \to \mathcal{I}$, $m \geq 0$, such that $h_{m+1}(w) = h_m(w)$, for all $w \in W_m$. For the basis of induction, set $h_0(a_i^{\mathcal{I}_0}) = a_i^{\mathcal{I}}$, for $a_i \in ob(\mathcal{A})$; h_0 is a homomorphism by (1) and **(abox)**. For the induction step, suppose h_m has been defined. If $v \in W_m$, set $h_{m+1}(v) = h_m(v)$. Otherwise, $v \in W_{m+1} \setminus W_m$. By **(uniq)**, there is a unique $u \in W_m$ with $(u, v) \in E_a$, for some $a \in ob(\mathcal{A})$. Let $\ell_a(u, v) = S$. By **(cp)**, $cp(v) = inv(ds)$ and, by **(role)**, $u \in (\exists S)^{\mathcal{I}_0}$. As h_m is a homomorphism, $h_m(u) \in (\exists S)^{\mathcal{I}}$, whence there is $w \in \Delta^{\mathcal{I}}$ with $(h_m(u), w) \in S^{\mathcal{I}}$. Set $h_{m+1}(v) = w$. As $(\exists inv(S))^{\mathcal{I}_0} \neq \emptyset$, $cp(v) = inv(ds)$ and $w \in (\exists inv(S))^{\mathcal{I}}$, by (2), if $v \in B^{\mathcal{I}_0}$ then $w \in B^{\mathcal{I}}$, for all $B \in Bcon(\mathcal{T})$. It remains to show that $(w, v) \in R^{\mathcal{I}_0}$ implies $(h_{m+1}(w), h_{m+1}(v)) \in R^{\mathcal{I}}$. By **(role)**, we have $(w, v) \in R^{\mathcal{I}_0}$, for $w \in W_{m+1}$ and $v \in W_{m+1} \setminus W_m$, just in two cases: either $w \in W_{m+1} \setminus W_m$, and then $w = v$ with $Id \sqsubseteq^*_{\mathcal{T}} R$, or $w \in W_m$, and then $w = u$ with $S \sqsubseteq^*_{\mathcal{T}} R$. In the former case, $(h_{m+1}(v), h_{m+1}(v)) \in Id^{\mathcal{I}} \subseteq R^{\mathcal{I}}$. In the latter case, $(u, v) \in S^{\mathcal{I}_0}$; hence $(h_{m+1}(u), h_{m+1}(v)) \in S^{\mathcal{I}} \subseteq R^{\mathcal{I}}$. □

Our next lemma shows that to check whether $\mathcal{I}_0 \models \mathsf{q}(\boldsymbol{a})$ it suffices to consider only the set of points W_{m_0} of depth $\leq m_0$ in $\Delta^{\mathcal{I}}$, for some m_0 that does not depend on $|\mathcal{A}|$ (see [4, Lemma 7.4] for the proof):

Lemma 5. *Let $m_0 = k + |role^{\pm}(\mathcal{T})|$. If $\mathcal{I}_0 \models \exists \boldsymbol{y}\, \varphi(\boldsymbol{a}, \boldsymbol{y})$ then there is an assignment $\mathfrak{a}_0$ in W_{m_0} (i.e., $\mathfrak{a}_0(y_i) \in W_{m_0}$ for all i) such that $\mathcal{I}_0 \models^{\mathfrak{a}_0} \varphi(\boldsymbol{a}, \boldsymbol{y})$.*

To complete the proof of Theorem 3, we encode the problem '$\mathcal{K} \models \mathsf{q}(\boldsymbol{a})$?' as a model checking problem for first-order formulas. We fix a signature that contains a unary predicate $A_k(x)$ for each concept name A_k and a binary predicate $P_k(x, y)$ for each role name P_k, and then represent the ABox $\mathcal{A}$ of $\mathcal{K}$ as a first-order model $\mathfrak{A}_{\mathcal{A}}$ with domain $ob(\mathcal{A})$: for each $a_i, a_j \in ob(\mathcal{A})$,

$$\mathfrak{A}_{\mathcal{A}} \models A_k[a_i] \ \text{ iff } \ A_k(a_i) \in \mathcal{A} \qquad \text{and} \qquad \mathfrak{A}_{\mathcal{A}} \models P_k[a_i, a_j] \ \text{ iff } \ P_k(a_i, a_j) \in \mathcal{A}.$$

Now we define a first-order formula $\varphi_{\mathcal{T},\mathsf{q}}(\boldsymbol{x})$ in the above signature such that (i) $\varphi_{\mathcal{T},\mathsf{q}}(\boldsymbol{x})$ depends on $\mathcal{T}$ and q but not on $\mathcal{A}$, and (ii) $\mathfrak{A}_{\mathcal{A}} \models \varphi_{\mathcal{T},\mathsf{q}}(\boldsymbol{a})$ iff $\mathcal{I}_0 \models \mathsf{q}(\boldsymbol{a})$.

To simplify the presentation, we denote by $\mathsf{e}(\mathcal{T})$ the extension of $\mathcal{T}$ with:

- $\geq q'\, R \sqsubseteq\, \geq q\, R$, for all $R \in role^{\pm}(\mathcal{T})$ and $q, q' \in Q^R_{\mathcal{T}}$ with $q' > q$, and
- $\geq q\, R \sqsubseteq\, \geq q\, R'$, for all $q \in Q^R_{\mathcal{T}}$ and $R \sqsubseteq R' \in \mathcal{T}$ or $inv(R) \sqsubseteq inv(R') \in \mathcal{T}$.

It follows from the definition of $\cdot^{\ddagger_{\mathsf{e}}}$ and Lemma 1 that, for a Horn concept inclusion $C \sqsubseteq B$, we have $\mathcal{T} \models C \sqsubseteq B$ iff $(C^*(x) \to B^*(x))$ is a logical consequence of $\{(C_i^*(x) \to B_i^*(x)) \mid C_i \sqsubseteq B_i \in \mathsf{e}(\mathcal{T})\}$.

We begin by defining formulas $\psi_B(x)$, $B \in Bcon(\mathcal{T})$, that describe the types of the elements of $ob(\mathcal{A})$ in the model $\mathcal{I}_0$ in the following sense (cf. (1)):

$$\mathfrak{A}_{\mathcal{A}} \models \psi_B[a_i] \quad \text{iff} \quad a_i^{\mathcal{I}_0} \in B^{\mathcal{I}_0}, \qquad \text{for } B \in Bcon(\mathcal{T}) \text{ and } a_i \in ob(\mathcal{A}). \tag{3}$$

These formulas are defined as the 'fixed-points' of sequences $\psi_B^0(x), \psi_B^1(x), \dots$ of formulas with one free variable, where

$$\psi_B^0(x) = \begin{cases} A(x), & \text{if } B = A,\\ \exists y_1 \dots \exists y_q \Big[\bigwedge\limits_{1 \le i < j \le q} (y_i \neq y_j) \wedge \bigwedge\limits_{1 \le i \le q} R^{\mathcal{T}}(x, y_i) \Big], & \text{if } B = \geq q\, R, \end{cases}$$

$$\psi_B^i(x) = \psi_B^0(x) \quad \vee \bigvee_{B_1 \sqcap \dots \sqcap B_k \sqsubseteq B \in \mathsf{e}(\mathcal{T})} \big(\psi_{B_1}^{i-1}(x) \wedge \dots \wedge \psi_{B_k}^{i-1}(x)\big), \qquad \text{for } i \geq 1,$$

and $R^{\mathcal{T}}(x,y) = \bigvee_{P_k \sqsubseteq_{\mathcal{T}}^* R} P_k(x,y) \ \vee \bigvee_{P_k^- \sqsubseteq_{\mathcal{T}}^* R} P_k(y,x)$. Clearly, if there is an i such that, for all $B \in Bcon(\mathcal{T})$, $\psi_B^i(x) \equiv \psi_B^{i+1}(x)$, i.e., every $\psi_B^i(x)$ is equivalent to $\psi_B^{i+1}(x)$ in first-order logic, then $\psi_B^i(x) \equiv \psi_B^j(x)$ for all $B \in Bcon(\mathcal{T})$, $j \geq i$. The minimum such i does not exceed $N = |Bcon(\mathcal{T})|$, so we set $\psi_B(x) = \psi_B^N(x)$.

Next we introduce sentences $\theta_{B,dr}$, for $B \in Bcon(\mathcal{T})$ and $dr \in dr(\mathcal{T})$, that describe the types of the elements of $dr(\mathcal{T})$ in $\mathcal{I}_0$ in the following sense (cf. (2)):

$$\mathfrak{A}_{\mathcal{A}} \models \theta_{B,dr} \quad \text{iff} \quad w \in B^{\mathcal{I}_0}, \text{ for each (some) } w \in \Delta^{\mathcal{I}_0} \text{ with } cp(w) = dr. \tag{4}$$

These sentences are defined similarly to the $\psi_B(x)$: namely, for each $B \in Bcon(\mathcal{T})$ and $dr \in dr(\mathcal{T})$, we consider a sequence $\theta_{B,dr}^0, \theta_{B,dr}^1, \dots$ by taking

$$\theta_{B,dr}^0 = \rho_{B,dr}^0, \qquad \theta_{B,dr}^i = \rho_{B,dr}^i \ \vee \bigvee_{B_1 \sqcap \dots \sqcap B_k \sqsubseteq B \in \mathsf{e}(\mathcal{T})} \big(\theta_{B_1,dr}^{i-1} \wedge \dots \wedge \theta_{B_k,dr}^{i-1}\big), \qquad \text{for } i \geq 1,$$

where $\rho_{B,dr}^i = \bot$, for all $B \neq \exists R$ and $i \geq 0$, and

$$\rho_{\exists R,dr}^0 = \exists x\, \psi_{\exists inv(R)}(x) \qquad \text{and} \qquad \rho_{\exists R,dr}^i = \bigvee_{ds \in dr(\mathcal{T})} \theta_{\exists inv(R),ds}^{i-1}, \qquad \text{for } i \geq 1.$$

We have $\theta_{B,dr}^i \equiv \theta_{B,dr}^{i+1}$ for some $i \leq M = N \cdot |role^{\pm}(\mathcal{T})|$. So, let $\theta_{B,dr} = \theta_{B,dr}^M$.

Now we consider the directed graph $G_{\mathcal{T}} = (V_{\mathcal{T}}, E_{\mathcal{T}})$, where $V_{\mathcal{T}}$ is the set of all equivalence classes $[R]$, $[R] = \{R' \mid R \sqsubseteq_{\mathcal{T}}^* R', R' \sqsubseteq_{\mathcal{T}}^* R\}$, such that $\exists R$ is not empty in *some* model of $\mathcal{T}$, and $E_{\mathcal{T}}$ is the set of all pairs $([R_i],[R_j])$ such that

(path) $\mathcal{T} \models \exists inv(R_i) \sqsubseteq \geq q\, R_j$ and either $inv(R_i) \not\sqsubseteq_{\mathcal{T}}^* R_j$ or $q \geq 2$,

and R_j has no proper sub-role satisfying **(path)**. We have $([R_i],[R_j]) \in E_{\mathcal{T}}$ iff, for any ABox $\mathcal{A}'$, whenever the minimal untangled model $\mathcal{I}_0$ of $(\mathcal{T},\mathcal{A}')$ contains a copy w of $inv(dr_i')$, for $R_i' \in [R_i]$, then w is connected to a copy of $inv(dr_j')$, for $R_j' \in [R_j]$, by all relations S with $R_j \sqsubseteq_{\mathcal{T}}^* S$. Let $\Sigma_{\mathcal{T},m_0}$ be the set of all paths in $G_{\mathcal{T}}$ of length $\leq m_0$ (as in Lemma 5):

$$\Sigma_{\mathcal{T},m_0} = \{\varepsilon\} \cup \{([R_1],\dots,[R_n]) \mid 1 \leq n \leq m_0 \ \&\ ([R_j],[R_{j+1}]) \in E_{\mathcal{T}}, \text{ for } j < n\}.$$

For $\sigma, \sigma' \in \Sigma_{\mathcal{T},m_0}$ and $R \in \mathit{role}^{\pm}(\mathcal{T})$, we write $\sigma \xrightarrow{R} \sigma'$ if (i) $\sigma = \sigma'$ and $\mathit{Id} \sqsubseteq^*_{\mathcal{T}} R$ or (ii) $\sigma.[S] = \sigma'$ or (iii) $\sigma = \sigma'.[\mathit{inv}(S)]$, for some S with $S \sqsubseteq^*_{\mathcal{T}} R$.

Let $\Sigma^k_{\mathcal{T},m_0}$ be the set of all k-tuples of the form $\boldsymbol{\sigma} = (\sigma_1, \dots, \sigma_k)$, $\sigma_i \in \Sigma_{\mathcal{T},m_0}$. Intuitively, when evaluating the query $\exists \boldsymbol{y}\, \varphi(\boldsymbol{x}, \boldsymbol{y})$ over $\mathcal{I}_0$, each bound, or non-distinguished, variable y_i is mapped to a point w in W_{m_0}. However, the first-order model $\mathfrak{A}_{\mathcal{A}}$ does not contain the points from $W_{m_0} \setminus W_0$, and to represent them, we use the following 'trick.' By **(uniq)**, every point w in W_{m_0} is uniquely determined by the pair (a, σ), where $a^{\mathcal{I}_0}$ is the root of the tree $\mathfrak{T}_a$ containing w, and σ is the sequence of labels $\ell_a(u, v)$ on the path from $a^{\mathcal{I}_0}$ to w. It follows from the unraveling procedure and **(path)** that $\sigma \in \Sigma_{\mathcal{T},m_0}$. So, in the formula $\varphi_{\mathcal{T},\mathsf{q}}$ we are about to define we assume that the y_i range over W_0 and represent the first component of the pairs (a, σ), whereas the second component is encoded in the ith member of $\boldsymbol{\sigma}$ (these y_i should not be confused with the y_i in the original query q, which range over all of W_{m_0}). In order to treat arbitrary terms t occurring in $\varphi(\boldsymbol{x}, \boldsymbol{y})$ in a uniform way, we set $t^{\boldsymbol{\sigma}} = \varepsilon$, if $t = a \in \mathit{ob}(\mathcal{A})$ or $t = x_i$, and $t^{\boldsymbol{\sigma}} = \sigma_i$, if $t = y_i$ (the distinguished variables x_i and the object names a are mapped to W_0 and do not require the second component of the pairs).

Given an assignment $\mathfrak{a}_0$ in W_{m_0} we denote by $\mathit{split}(\mathfrak{a}_0)$ the pair $(\mathfrak{a}, \boldsymbol{\sigma})$, where $\mathfrak{a}$ is an assignment in $\mathfrak{A}_{\mathcal{A}}$ and $\boldsymbol{\sigma} = (\sigma_1, \dots, \sigma_k) \in \Sigma^k_{\mathcal{T},m_0}$ are such that

- for each distinguished variable x_i, $\mathfrak{a}(x_i) = a$ with $a^{\mathcal{I}_0} = \mathfrak{a}_0(x_i)$;
- for each bound variable y_i, $\mathfrak{a}(y_i) = a$ and $\sigma_i = ([R_1], \dots, [R_n])$, $n \le m_0$, with $a^{\mathcal{I}_0}$ being the root of the tree containing $\mathfrak{a}_0(y_i)$ and $R_1, \dots, R_n$ being the sequence of labels $\ell_a(u, v)$ on the path from $a^{\mathcal{I}_0}$ to $\mathfrak{a}_0(y_i)$.

Not every pair $(\mathfrak{a}, \boldsymbol{\sigma})$, however, corresponds to an assignment in W_{m_0} because some paths in $\boldsymbol{\sigma}$ may not exist in our $\mathcal{I}_0$: $G_{\mathcal{T}}$ represents possible paths in *all* models for the fixed TBox $\mathcal{T}$ and varying ABox. As follows from the unraveling procedure, a point in $W_{m_0} \setminus W_0$ corresponds to $a \in \mathit{ob}(\mathcal{A})$ and $\sigma \in \Sigma_{\mathcal{T},m_0}$, $\sigma = ([R], \dots)$, iff a has not enough R-witnesses in $\mathcal{A}$: $\mathfrak{A}_{\mathcal{A}} \models \neg\psi^0_{\ge q\, R}[a] \wedge \psi_{\ge q\, R}[a]$, for some $q \in Q^R_{\mathcal{T}}$. Thus, for every $(\mathfrak{a}, \boldsymbol{\sigma})$ with $\boldsymbol{\sigma} = (\sigma_1, \dots, \sigma_k)$, there is an assignment $\mathfrak{a}_0$ in W_{m_0} with $\mathit{split}(\mathfrak{a}_0) = (\mathfrak{a}, \boldsymbol{\sigma})$ iff $\mathfrak{A}_{\mathcal{A}} \models^{\mathfrak{a}} \eta^{\boldsymbol{\sigma}}(\boldsymbol{y})$, where

$$\eta^{\boldsymbol{\sigma}}(\boldsymbol{y}) = \bigwedge_{\substack{1 \le i \le k \\ \sigma_i \ne \varepsilon}} \ \bigvee_{q \in Q^{R_i}_{\mathcal{T}}} \big(\neg\psi^0_{\ge q\, R_i}(y_i) \ \wedge \ \psi_{\ge q\, R_i}(y_i)\big)$$

and each R_i, for $1 \le i \le k$ with $\sigma_i \ne \varepsilon$, is such that $\sigma_i = ([R_i], \dots)$.

We define now, for every $\boldsymbol{\sigma} \in \Sigma^k_{\mathcal{T},m_0}$, concept name A and role name R,

$$A^{\boldsymbol{\sigma}}(t) \ = \ \begin{cases} \psi_A(t), & \text{if } t^{\boldsymbol{\sigma}} = \varepsilon, \\ \theta_{A,\mathit{inv}(ds)}, & \text{if } t^{\boldsymbol{\sigma}} = \sigma'.[S], \text{ for some } \sigma' \in \Sigma_{\mathcal{T},m_0}, \end{cases}$$

$$R^{\boldsymbol{\sigma}}(t_1, t_2) \ = \ \begin{cases} R^{\mathcal{T}}(t_1, t_2), & \text{if } t_1^{\boldsymbol{\sigma}} = t_2^{\boldsymbol{\sigma}} = \varepsilon, \\ (t_1 = t_2), & \text{if } t_1^{\boldsymbol{\sigma}} \xrightarrow{R} t_2^{\boldsymbol{\sigma}} \text{ and either } t_1^{\boldsymbol{\sigma}} \ne \varepsilon \text{ or } t_2^{\boldsymbol{\sigma}} \ne \varepsilon, \\ \bot, & \text{otherwise.} \end{cases}$$

We claim that, for each assignment $\mathfrak{a}_0$ in W_{m_0}, $(\mathfrak{a}, \sigma) = \mathit{split}(\mathfrak{a}_0)$ and term t,

$$\mathcal{I}_0 \models^{\mathfrak{a}_0} A(t) \quad \text{iff} \quad \mathfrak{A}_{\mathcal{A}} \models^{\mathfrak{a}} A^{\boldsymbol{\sigma}}(t), \qquad \text{for all concept names } A, \tag{5}$$

$$\mathcal{I}_0 \models^{\mathfrak{a}_0} R(t_1, t_2) \quad \text{iff} \quad \mathfrak{A}_{\mathcal{A}} \models^{\mathfrak{a}} R^{\boldsymbol{\sigma}}(t_1, t_2), \qquad \text{for all roles } R. \tag{6}$$

For $A(a)$, $A(x_i)$ or $A(y_i)$ with $\sigma_i = \varepsilon$ the claim follows from (3). For $A(y_i)$ with $\sigma_i = \sigma'.[S]$, by **(cp)**, we have $\mathit{cp}(\mathfrak{a}(y_i)) = \mathit{inv}(dr)$, for some $R \in [S]$; the claim then follows from (4). For $R(y_{i_1}, y_{i_2})$ with $\sigma_{i_1} = \sigma_{i_2} = \varepsilon$, the claim follows from **(abox)**. Let us consider the case of $R(y_{i_1}, y_{i_2})$ with $\sigma_{i_2} \neq \varepsilon$: we have $\mathfrak{a}_0(y_{i_2}) \notin W_0$ and thus, by **(role)**, $\mathcal{I}_0 \models^{\mathfrak{a}_0} R(y_{i_1}, y_{i_2})$ iff

- $\mathfrak{a}_0(y_{i_1}), \mathfrak{a}_0(y_{i_2})$ are in the same tree $\mathfrak{T}_a$, for $a \in \mathit{ob}(\mathcal{A})$, i.e., $\mathfrak{A}_{\mathcal{A}} \models^{\mathfrak{a}} (y_{i_1} = y_{i_2})$,
- and either $(\mathfrak{a}_0(y_{i_1}), \mathfrak{a}_0(y_{i_2})) \in E_a$ and then $\ell_a(\mathfrak{a}_0(y_{i_1}), \mathfrak{a}_0(y_{i_2})) = S$ for some $S \sqsubseteq^*_{\mathcal{T}} R$, or $(\mathfrak{a}_0(y_{i_2}), \mathfrak{a}_0(y_{i_1})) \in E_a$ and then $\ell_a(\mathfrak{a}_0(y_{i_2}), \mathfrak{a}_0(y_{i_1})) = S$ for some $\mathit{inv}(S) \sqsubseteq^*_{\mathcal{T}} R$, or $\mathfrak{a}_0(y_{i_1}) = \mathfrak{a}_0(y_{i_2})$ and then $\mathit{Id} \sqsubseteq^*_{\mathcal{T}} R$, i.e., $\sigma_{i_1} \xrightarrow{R} \sigma_{i_2}$.

Other cases are similar and left to the reader.

Finally, let $\varphi^{\boldsymbol{\sigma}}(\boldsymbol{x}, \boldsymbol{y})$ be the result of attaching the superscript $^{\boldsymbol{\sigma}}$ to each atom of φ and

$$\varphi_{\mathcal{T},\mathsf{q}}(\boldsymbol{x}) \;=\; \exists \boldsymbol{y} \bigvee_{\boldsymbol{\sigma} \in \Sigma^k_{\mathcal{T},m_0}} \Big(\varphi^{\boldsymbol{\sigma}}(\boldsymbol{x}, \boldsymbol{y}) \quad \wedge \eta^{\boldsymbol{\sigma}}(\boldsymbol{y}) \Big).$$

As follows from (5)–(6), for every assignment $\mathfrak{a}_0$ in W_{m_0}, we have $\mathcal{I}_0 \models^{\mathfrak{a}_0} \varphi(\boldsymbol{x}, \boldsymbol{y})$ iff $\mathfrak{A}_{\mathcal{A}} \models^{\mathfrak{a}} \varphi^{\boldsymbol{\sigma}}(\boldsymbol{x}, \boldsymbol{y})$ for $(\mathfrak{a}, \sigma) = \mathit{split}(\mathfrak{a}_0)$. For the converse direction notice that, if $\mathfrak{A}_{\mathcal{A}} \models^{\mathfrak{a}} \eta^{\boldsymbol{\sigma}}(\boldsymbol{y})$ then there is an assignment $\mathfrak{a}_0$ in W_{m_0} with $\mathit{split}(\mathfrak{a}_0) = (\mathfrak{a}, \boldsymbol{\sigma})$.

Clearly, $\mathfrak{A}_{\mathcal{A}} \models \varphi_{\mathcal{T},\mathsf{q}}(\boldsymbol{a})$ iff $\mathcal{I}_0 \models \mathsf{q}(\boldsymbol{a})$, for every tuple $\boldsymbol{a}$. We also note that, for every pair of tuples $\boldsymbol{a}$ and $\boldsymbol{b}$ of object names in $\mathit{ob}(\mathcal{A})$, $\varphi^{\boldsymbol{\sigma}}(\boldsymbol{a}, \boldsymbol{b})$ is a positive existential sentence with inequalities, and so domain-independent.[4] It is also easily seen that, for each $\boldsymbol{b}$, $\eta^{\boldsymbol{\sigma}}(\boldsymbol{b})$ is domain-independent. It follows from the minimality of $\mathcal{I}_0$ that $\varphi_{\mathcal{T},\mathsf{q}}(\boldsymbol{a})$ is domain-independent, for each tuple $\boldsymbol{a}$ of object names in $\mathit{ob}(\mathcal{A})$.

Finally, we note that the resulting query contains $\leq m^{k\cdot(k+m)}$ disjuncts, where $m = |\mathit{role}^{\pm}(\mathcal{T})|$ and k is the number of bound variables in q. □

We also remark that although extending the $\mathit{DL\text{-}Lite}^{(\mathcal{RN})}_{\alpha}$ languages with transitive roles does not change the combined complexity of reasoning, it does change the data complexity: instance checking and satisfiability in $\mathit{DL\text{-}Lite}^{(\mathcal{RN})}_{\alpha}$, for $\alpha \in \{\mathit{core}, \mathit{krom}, \mathit{horn}, \mathit{bool}\}$, are NLogSpace-complete (rather than in AC^0) and query answering over $\mathit{DL\text{-}Lite}^{(\mathcal{RN})}_{\mathit{horn}}$ and $\mathit{DL\text{-}Lite}^{(\mathcal{RN})}_{\mathit{core}}$ KBs is NLogSpace-complete for data complexity (see Section 5.4 of [4]).

[4] A query $\mathsf{q}(\boldsymbol{x})$ is said to be domain-independent in case $\mathfrak{A}_{\mathcal{A}} \models^{\mathfrak{a}} \mathsf{q}(\boldsymbol{x})$ iff $\mathfrak{A} \models^{\mathfrak{a}} \mathsf{q}(\boldsymbol{x})$, for each $\mathfrak{A}$ such that the domain of $\mathfrak{A}$ contains $\mathit{ob}(\mathcal{A})$, the active domain of $\mathfrak{A}_{\mathcal{A}}$, and $A^{\mathfrak{A}} = A^{\mathfrak{A}_{\mathcal{A}}}$ and $P^{\mathfrak{A}} = P^{\mathfrak{A}_{\mathcal{A}}}$, for all concept and role names A and P.

References

1. S. Abiteboul, R. Hull, and V. Vianu. *Foundations of Databases*. Addison-Wesley, 1995.
2. K. Apt. Logic programming. In J. van Leeuwen, editor, *Handbook of Theoretical Computer Science, Volume B: Formal Models and Sematics*, pages 493–574. Elsevier and MIT Press, 1990.
3. A. Artale, D. Calvanese, R. Kontchakov, and M. Zakharyaschev. *DL-Lite* in the light of first-order logic. In *Proc. of the 22nd Nat. Conf. on Artificial Intelligence (AAAI 2007)*, pages 361–366, 2007.
4. A. Artale, D. Calvanese, R. Kontchakov, and M. Zakharyaschev. The *DL-Lite* family and relations. Technical Report BBKCS-09-03, SCSIS, Birkbeck College, London, 2009 (available at `http://www.dcs.bbk.ac.uk/research/techreps/2009/bbkcs-09-03.pdf`).
5. E. Börger, E. Grädel, and Y. Gurevich. *The Classical Decision Problem*. Perspectives in Mathematical Logic. Springer, 1997.
6. D. Calvanese, G. De Giacomo, D. Lembo, M. Lenzerini, and R. Rosati. *DL-Lite*: Tractable description logics for ontologies. In *Proc. of the 20th Nat. Conf. on Artificial Intelligence (AAAI 2005)*, pages 602–607, 2005.
7. D. Calvanese, G. De Giacomo, D. Lembo, M. Lenzerini, and R. Rosati. Data complexity of query answering in description logics. In *Proc. of the 10th Int. Conf. on the Principles of Knowledge Representation and Reasoning (KR 2006)*, pages 260–270, 2006.
8. D. Calvanese, G. De Giacomo, D. Lembo, M. Lenzerini, and R. Rosati. Tractable reasoning and efficient query answering in description logics: The *DL-Lite* family. *J. of Automated Reasoning*, 39(3):385–429, 2007.
9. B. Glimm, I. Horrocks, C. Lutz, and U. Sattler. Conjunctive query answering for the description logic $\mathcal{SHIQ}$. In *Proc. of the 20th Int. Joint Conf. on Artificial Intelligence (IJCAI 2007)*, pages 399–404, 2007.
10. N. Immerman. *Descriptive Complexity*. Springer, 1999.
11. R. Kontchakov and M. Zakharyaschev. *DL-Lite* and role inclusions. In J. Domingue and C. Anutariya, editors, *Proc. of the 3rd Asian Semantic Web Conf. (ASWC 2008)*, volume 5367 of *Lecture Notes in Computer Science*, pages 16–30. Springer, 2008.
12. D. Kozen. *Theory of Computation*. Springer, 2006.
13. M. Ortiz, D. Calvanese, and T. Eiter. Data complexity of query answering in expressive description logics via tableaux. *J. of Automated Reasoning*, 41(1):61–98, 2008.
14. M. M. Ortiz, D. Calvanese, and T. Eiter. Characterizing data complexity for conjunctive query answering in expressive description logics. In *Proc. of the 21st Nat. Conf. on Artificial Intelligence (AAAI 2006)*, pages 275–280, 2006.
15. C. Papadimitriou. *Computational Complexity*. Addison-Wesley, 1994.
16. A. Poggi, D. Lembo, D. Calvanese, G. De Giacomo, M. Lenzerini, and R. Rosati. Linking data to ontologies. *J. on Data Semantics*, X:133–173, 2008.
17. W. Rautenberg. *A Concise Introduction to Mathematical Logic*. Springer, 2006.
18. A. Schaerf. On the complexity of the instance checking problem in concept languages with existential quantification. *J. of Intelligent Information Systems*, 2:265–278, 1993.