Conjunctive Query Inseparability of *OWL 2 QL* TBoxes

B. Konev,[1] R. Kontchakov,[2] M. Ludwig,[1] T. Schneider,[3] F. Wolter[1] and M. Zakharyaschev[2]

[1] Department of Computer Science
University of Liverpool, UK
{konev,michel.ludwig,wolter}@liverpool.ac.uk

[2] Department of CS & IS
Birkbeck College London, UK
{roman,michael}@dcs.bbk.ac.uk

[3] Department of Computer Science
Universität Bremen, Germany
tschneider@informatik.uni-bremen.de

Abstract

The *OWL 2* profile *OWL 2 QL*, based on the *DL-Lite* family of description logics, is emerging as a major language for developing new ontologies and approximating the existing ones. Its main application is ontology based data access, where ontologies are used to provide background knowledge for answering queries over data. We investigate the corresponding notion of query inseparability (or equivalence) for *OWL 2 QL* ontologies and show that deciding query inseparability is PSPACE-hard and in EXPTIME. We give polynomial time (incomplete) algorithms and demonstrate by experiments that they can be used for practical module extraction.

Introduction

In recent years, ontology-based data access (OBDA) has emerged as one of the most interesting and challenging applications of description logic (Dolby et al. 2008; Heymans et al. 2008; Poggi et al. 2008). The key idea is to use ontologies for enriching data with additional background knowledge, and thereby enable query answering over incomplete and semistructured data from heterogeneous sources via a high-level conceptual interface. The W3C recognised the importance of OBDA by including in the *OWL 2* Web Ontology Language the profile *OWL 2 QL*, which was designed for OBDA with standard relational database systems. *OWL 2 QL* is based on a description logic (DL) that was originally introduced under the name $\textit{DL-Lite}_{\mathcal{R}}$ (Calvanese et al. 2006; 2007) and called $\textit{DL-Lite}^{\mathcal{H}}_{core}$ in the more general classification of (Artale et al. 2009). It can be described as an optimal sub-language of the DL $\mathcal{SROIQ}$, underlying *OWL 2*, which includes most of the features of conceptual models, and for which conjunctive query answering can be done in AC^0 for data complexity.

Thus, $\textit{DL-Lite}^{\mathcal{H}}_{core}$ is becoming a major language for developing ontologies, and a target language for translation and approximation of existing ontologies formulated in more expressive DLs (Pan and Thomas 2007; Botoeva, Calvanese, and Rodriguez-Muro 2010). One of the consequences of this development is that $\textit{DL-Lite}^{\mathcal{H}}_{core}$ ontologies turn out to be larger and more complex than originally envisaged. As a result, reasoning support for ontology engineering tasks such as composing, re-using, comparing, and extracting ontologies—which so far has been only analysed for expressive DLs (Cuenca Grau et al. 2008; Stuckenschmidt, Parent, and Spaccapietra 2009), $\mathcal{EL}$ (Lutz and Wolter 2010) and *DL-Lite* dialects (Kontchakov, Wolter, and Zakharyaschev 2010) without role inclusions—is becoming increasingly important for $\textit{DL-Lite}^{\mathcal{H}}_{core}$ as well.

In the context of OBDA, the basic notion underlying many ontology engineering tasks is Σ-*query inseparability*: for a signature (a set of concept and role names) Σ, two ontologies are deemed to be inseparable if they give the same answers to any conjunctive query over any data formulated in Σ. Thus, in applications using Σ-queries and data, one can safely replace any ontology by a Σ-query inseparable one. Note that the relativisation to Σ is very important here. For example, one cannot expect modules of an ontology to be query inseparable from the whole ontology for *arbitrary* queries and data sets, whereas this should be the case if we restrict the query and data language to the module's signature or a specified subset thereof. Similarly, when comparing two versions of one ontology, the subtle and potentially problematic differences are those that concern queries over their common symbols, rather than all symbols occurring in these versions. In applications where ontologies are built using imported parts, a stronger notion of inseparability is required: two ontologies are *strongly* Σ-*query inseparable* if they give the same answers to Σ-queries and data when imported to an arbitrary context ontology formulated in Σ.

The aim of this paper is to (*i*) investigate the computational complexity of deciding (strong) Σ-query inseparability for $\textit{DL-Lite}^{\mathcal{H}}_{core}$ ontologies, (*ii*) develop efficient (though incomplete) algorithms for practical inseparability checking, and (*iii*) analyse the performance of the algorithms for the challenging task of minimal module extraction.

One of our surprising discoveries is that the analysis of Σ-query inseparability for (seemingly 'harmless' and computationally well-behaved) $\textit{DL-Lite}^{\mathcal{H}}_{core}$ ontologies requires drastically different logical tools compared with the previously considered DLs. It turns out that the new syntactic ingredient—the interaction of role inclusions and inverse roles—makes deciding (strong) query inseparability PSPACE-hard, as opposed to the known CONP and Π^p_2-completeness results for *DL-Lite* dialects without role inclusions (Kontchakov, Wolter, and Zakharyaschev 2010). On the other hand, the obtained EXPTIME upper bound is actually the first known decidability result for strong inseparability, which goes beyond the 'essentially' Boolean logic and might additionally indicate a way of solving the open problem of strong Σ-query inseparability for $\mathcal{EL}$ (Lutz and Wolter 2010). For *DL-Lite*$^{\mathcal{H}}_{core}$ ontologies *without* role inclusions, strong Σ-query inseparability is shown to be only NLOGSPACE-complete. We give (incomplete) polynomial time algorithms checking (strong) Σ-inseparability and demonstrate, by a set of minimal module extraction experiments, that they are (*i*) complete for many existing *DL-Lite*$^{\mathcal{H}}_{core}$ ontologies and signatures, and (*ii*) sufficiently fast to be used in module extraction algorithms that require thousands of Σ-query inseparability checks. All omitted proofs can be found in the appendix.

Σ-Query Entailment and Inseparability

We begin by formally defining the description logic *DL-Lite*$^{\mathcal{H}}_{core}$, underlying *OWL 2 QL*, and the notions of Σ-query inseparability and Σ-query entailment. The language of *DL-Lite*$^{\mathcal{H}}_{core}$ contains countably infinite sets of *individual names* a_i, *concept names* A_i, and *role names* P_i. *Roles* R and *concepts* B of this language are defined by:

$$
\begin{array}{lcl}
R & ::= & P_i \quad | \quad P_i^-, \\
B & ::= & \bot \quad | \quad \top \quad | \quad A_i \quad | \quad \exists R.
\end{array}
$$

A *DL-Lite*$^{\mathcal{H}}_{core}$ *TBox*, $\mathcal{T}$, is a finite set of *inclusions*

$$B_1 \sqsubseteq B_2, \;\; R_1 \sqsubseteq R_2, \;\; B_1 \sqcap B_2 \sqsubseteq \bot, \;\; R_1 \sqcap R_2 \sqsubseteq \bot,$$

where B_1, B_2 are concepts and R_1, R_2 roles. An *ABox*, $\mathcal{A}$, is a finite set of *assertions* of the form $B(a_i)$, $R(a_i, a_j)$ and $a_i \neq a_j$, where a_i and a_j are individual names, B a concept and R a role. $\mathsf{Ind}(\mathcal{A})$ will stand for the set of individual names occurring in $\mathcal{A}$. Taken together, $\mathcal{T}$ and $\mathcal{A}$ constitute the *DL-Lite*$^{\mathcal{H}}_{core}$ *knowledge base* (KB, for short) $\mathcal{K} = (\mathcal{T}, \mathcal{A})$. The sub-language of *DL-Lite*$^{\mathcal{H}}_{core}$ without inclusions for roles is denoted by *DL-Lite*$_{core}$ (Calvanese et al. 2007).

The semantics of *DL-Lite*$^{\mathcal{H}}_{core}$ is defined as usual in DL (Baader et al. 2003). We only note that, in interpretations $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$, we do not have to comply with the unique name assumption, that is, we can have $a_i^{\mathcal{I}} = a_j^{\mathcal{I}}$ for $i \neq j$. We write $\mathcal{I} \models \alpha$ to say that an inclusion or assertion α is true in $\mathcal{I}$. The interpretation $\mathcal{I}$ is a *model* of a KB $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ if $\mathcal{I} \models \alpha$ for all $\alpha \in \mathcal{T} \cup \mathcal{A}$. $\mathcal{K}$ is *consistent* if it has a model. A concept B is said to be $\mathcal{T}$-*consistent* if $(\mathcal{T}, \{B(a)\})$ has a model. $\mathcal{K} \models \alpha$ means that $\mathcal{I} \models \alpha$ for all models $\mathcal{I}$ of $\mathcal{K}$.

A *conjunctive query* (CQ) $\boldsymbol{q}(x_1, \dots, x_n)$ is a first-order formula $\exists y_1 \dots \exists y_m \, \varphi(x_1, \dots, x_n, y_1, \dots, y_m)$, where φ is constructed, using only $\wedge$, from atoms of the form $B(t)$ and $R(t_1, t_2)$, with B being a concept, R a role, and t_i being an individual name or a variable from the list $x_1, \dots, x_n, y_1, \dots, y_m$. The variables in $\vec{x} = x_1, \dots, x_n$ are called *answer variables* of $\boldsymbol{q}$. We say that an n-tuple $\vec{a} \subseteq \mathsf{Ind}(\mathcal{A})$ is an *answer* to $\boldsymbol{q}$ in an interpretation $\mathcal{I}$ if $\mathcal{I} \models \boldsymbol{q}[\vec{a}]$ (here we regard $\mathcal{I}$ to be a first-order structure); $\vec{a}$ is a *certain answer* to $\boldsymbol{q}$ over a KB $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ if $\mathcal{I} \models \boldsymbol{q}[\vec{a}]$ for all models $\mathcal{I}$ of $\mathcal{K}$; in this case we write $\mathcal{K} \models \boldsymbol{q}[\vec{a}]$.

To define the main notions of this paper, consider two KBs $\mathcal{K}_1 = (\mathcal{T}_1, \mathcal{A})$ and $\mathcal{K}_2 = (\mathcal{T}_2, \mathcal{A})$. For example, the $\mathcal{T}_i$ are different versions of some ontology, or one of them is a refinement of the other by means of new axioms. The question we are interested in is whether they give the same answers to queries formulated in a certain signature, say, in the common vocabulary of the $\mathcal{T}_i$ or in a vocabulary relevant to an application. To be precise, by a *signature*, Σ, we understand any finite set of concept and role names. A concept (inclusion, TBox, etc.) all concept and role names of which are in Σ is called a Σ-*concept* (*inclusion*, etc.). We say that $\mathcal{K}_1$ Σ-*query entails* $\mathcal{K}_2$ if, for *all* Σ-queries $\boldsymbol{q}(\vec{x})$ and all $\vec{a} \subseteq \mathsf{Ind}(\mathcal{A})$, $\mathcal{K}_2 \models \boldsymbol{q}[\vec{a}]$ implies $\mathcal{K}_1 \models \boldsymbol{q}[\vec{a}]$. In other words: any certain answer to a Σ-query given by $\mathcal{K}_2$ is also given by $\mathcal{K}_1$.

As the ABox is typically not fixed or known at the ontology design stage, we may have to compare the TBoxes over *arbitrary* Σ-ABoxes rather than a fixed one, which gives the following central definition of this paper.

Definition 1. Let $\mathcal{T}_1$ and $\mathcal{T}_2$ be TBoxes and Σ a signature. $\mathcal{T}_1$ Σ-*query entails* $\mathcal{T}_2$ if $(\mathcal{T}_1, \mathcal{A})$ Σ-query entails $(\mathcal{T}_2, \mathcal{A})$ for any Σ-ABox $\mathcal{A}$. $\mathcal{T}_1$ and $\mathcal{T}_2$ are Σ-*query inseparable* if they Σ-query entail each other, in which case we write $\mathcal{T}_1 \equiv_\Sigma \mathcal{T}_2$.

In many applications, Σ-query inseparability is enough to ensure that $\mathcal{T}_1$ can be safely replaced by $\mathcal{T}_2$. However, if they are developed as part of a larger ontology or are meant to be imported in other ontologies, a stronger notion is required:

Definition 2. $\mathcal{T}_1$ *strongly* Σ-*query entails* $\mathcal{T}_2$ if $\mathcal{T} \cup \mathcal{T}_1$ Σ-query entails $\mathcal{T} \cup \mathcal{T}_2$, for all Σ-TBoxes $\mathcal{T}$. $\mathcal{T}_1$ and $\mathcal{T}_2$ are *strongly* Σ-*query inseparable* if they strongly Σ-query entail each other, in which case we write $\mathcal{T}_1 \equiv^s_\Sigma \mathcal{T}_2$.

The following example illustrates the difference between Σ-query and strong Σ-query inseparability. For further discussion and examples, we refer the reader to (Cuenca Grau et al. 2008; Kontchakov, Wolter, and Zakharyaschev 2010).

Example 3. Let $\mathcal{T}_2 = \{\top \sqsubseteq \exists R, \exists R^- \sqsubseteq B, B \sqcap A \sqsubseteq \bot\}$, $\mathcal{T}_1 = \emptyset$ and $\Sigma = \{A\}$. $\mathcal{T}_1$ and $\mathcal{T}_2$ are Σ-query inseparable. However, they are not strongly Σ-query inseparable. Indeed, for the Σ-TBox $\mathcal{T} = \{\top \sqsubseteq A\}$, $\mathcal{T}_1 \cup \mathcal{T}$ is consistent, while $\mathcal{T}_2 \cup \mathcal{T}$ is inconsistent, and so $\mathcal{T}_1 \cup \mathcal{T}$ does not Σ-query entail $\mathcal{T}_2 \cup \mathcal{T}$, as witnessed by the query $\boldsymbol{q} = \bot$.

From now on, we shall focus our attention mainly on the more basic notion of Σ-query entailment.

Σ-Query Entailment and Σ-Homomorphisms

In this section, we characterise Σ-query entailment between *DL-Lite*$^{\mathcal{H}}_{core}$ TBoxes semantically in terms of (partial) Σ-homomorphisms between certain canonical models. Then, in the next section, we use this characterisation to investigate the complexity of deciding Σ-query entailment.

The *canonical model* $\mathcal{M}_{\mathcal{K}}$ of a consistent KB $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ gives correct answers to all CQs. In general, $\mathcal{M}_{\mathcal{K}}$ is infinite; however, it can be folded up into a small *generating model* $\mathcal{G}_{\mathcal{K}} = (\mathcal{I}_{\mathcal{K}}, \leadsto_{\mathcal{K}})$ consisting of a finite interpretation $\mathcal{I}_{\mathcal{K}}$ and a *generating relation* $\leadsto_{\mathcal{K}}$ that defines the unfolding.

Let $\sqsubseteq^*_{\mathcal{T}}$ be the reflexive and transitive closure of the role inclusion relation given by $\mathcal{T}$, and let $[R] = \{S \mid R \equiv^*_{\mathcal{T}} S\}$, where $R \equiv^*_{\mathcal{T}} S$ stands for '$R \sqsubseteq^*_{\mathcal{T}} S$ and $S \sqsubseteq^*_{\mathcal{T}} R$.' We write $[R] \leq_{\mathcal{T}} [S]$ if $R \sqsubseteq^*_{\mathcal{T}} S$; thus, $\leq_{\mathcal{T}}$ is a partial order on the set $\{[R] \mid R \text{ a role in } \mathcal{T}\}$. For each $[R]$, we introduce a *witness* $w_{[R]}$ and define a *generating relation* $\leadsto_{\mathcal{K}}$ on the set of these witnesses together with $\mathsf{Ind}(\mathcal{A})$ by taking:

$a \rightsquigarrow_{\mathcal{K}} w_{[R]}$ if $a \in \mathsf{Ind}(\mathcal{A})$ and $[R]$ is $\leq_{\mathcal{T}}$-minimal such that $\mathcal{K} \models \exists R(a)$ and $\mathcal{K} \not\models R(a,b)$ for any $b \in \mathsf{Ind}(\mathcal{A})$;

$w_{[S]} \rightsquigarrow_{\mathcal{K}} w_{[R]}$ if $[R]$ is $\leq_{\mathcal{T}}$-minimal with $\mathcal{T} \models \exists S^- \sqsubseteq \exists R$ and $[S^-] \neq [R]$.

A role R is *generating in* $\mathcal{K}$ if there are $a \in \mathsf{Ind}(\mathcal{A})$ and $R_1, \dots, R_n = R$ such that $a \rightsquigarrow_{\mathcal{K}} w_{[R_1]} \rightsquigarrow_{\mathcal{K}} \cdots \rightsquigarrow_{\mathcal{K}} w_{[R_n]}$.

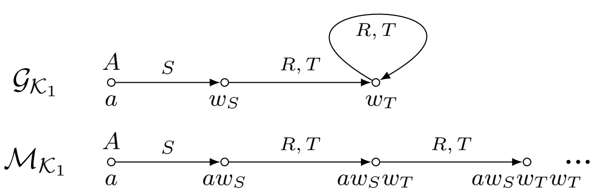
The interpretation $\mathcal{I}_{\mathcal{K}}$ is now defined as follows:

$$\begin{aligned}
\Delta^{\mathcal{I}_{\mathcal{K}}} &= \mathsf{Ind}(\mathcal{A}) \cup \{w_{[R]} \mid R \text{ is generating in } \mathcal{K}\},\\
a^{\mathcal{I}_{\mathcal{K}}} &= a, \text{ for all } a \in \mathsf{Ind}(\mathcal{A}),\\
A^{\mathcal{I}_{\mathcal{K}}} &= \{a \mid \mathcal{K} \models A(a)\} \cup \{w_{[R]} \mid \mathcal{T} \models \exists R^- \sqsubseteq A\},\\
P^{\mathcal{I}_{\mathcal{K}}} &= \{(a,b) \mid \text{there is } R(a,b) \in \mathcal{A} \text{ s.t. } R \sqsubseteq^*_{\mathcal{T}} P\} \cup\\
&\quad\ \{(x, w_{[R]}) \mid x \rightsquigarrow_{\mathcal{K}} w_{[R]} \text{ and } [R] \leq_{\mathcal{T}} [P]\} \cup\\
&\quad\ \{(w_{[R]}, x) \mid x \rightsquigarrow_{\mathcal{K}} w_{[R]} \text{ and } [R] \leq_{\mathcal{T}} [P^-]\}.
\end{aligned}$$

$\mathcal{G}_{\mathcal{K}}$ can be constructed in polynomial time in $|\mathcal{K}|$, and it is not hard to see that $\mathcal{I}_{\mathcal{K}} \models \mathcal{K}$. To construct the *canonical model* $\mathcal{M}_{\mathcal{K}}$ giving the correct answers to all CQs, we unfold the generating model $\mathcal{G}_{\mathcal{K}} = (\mathcal{I}_{\mathcal{K}}, \rightsquigarrow_{\mathcal{K}})$ along $\rightsquigarrow_{\mathcal{K}}$. A *path* in $\mathcal{G}_{\mathcal{K}}$ is a finite sequence $a w_{[R_1]} \cdots w_{[R_n]}$, $n \geq 0$, such that $a \in \mathsf{Ind}(\mathcal{A})$, $a \rightsquigarrow_{\mathcal{K}} w_{[R_1]}$ and $w_{[R_i]} \rightsquigarrow_{\mathcal{K}} w_{[R_{i+1}]}$, for $i < n$. Denote by $\mathsf{path}(\mathcal{G}_{\mathcal{K}})$ the set of all paths in $\mathcal{G}_{\mathcal{K}}$ and by $\mathsf{tail}(\sigma)$ the last element in $\sigma \in \mathsf{path}(\mathcal{G}_{\mathcal{K}})$. $\mathcal{M}_{\mathcal{K}}$ is defined by taking:

$$\begin{aligned}
\Delta^{\mathcal{M}_{\mathcal{K}}} &= \mathsf{path}(\mathcal{G}_{\mathcal{K}}),\\
a^{\mathcal{M}_{\mathcal{K}}} &= a, \text{ for all } a \in \mathsf{Ind}(\mathcal{A}),\\
A^{\mathcal{M}_{\mathcal{K}}} &= \{\sigma \mid \mathsf{tail}(\sigma) \in A^{\mathcal{I}_{\mathcal{K}}}\},\\
P^{\mathcal{M}_{\mathcal{K}}} &= \{(a,b) \in \mathsf{Ind}(\mathcal{A}) \times \mathsf{Ind}(\mathcal{A}) \mid (a,b) \in P^{\mathcal{I}_{\mathcal{K}}}\} \cup\\
&\quad\ \{(\sigma, \sigma \cdot w_{[R]}) \mid \mathsf{tail}(\sigma) \rightsquigarrow_{\mathcal{K}} w_{[R]},\ [R] \leq_{\mathcal{T}} [P]\} \cup\\
&\quad\ \{(\sigma \cdot w_{[R]}, \sigma) \mid \mathsf{tail}(\sigma) \rightsquigarrow_{\mathcal{K}} w_{[R]},\ [R] \leq_{\mathcal{T}} [P^-]\}.
\end{aligned}$$

Example 4. The models $\mathcal{G}_{\mathcal{K}_1}$ for $\mathcal{K}_1 = (\mathcal{T}_1, \{A(a)\})$ with $\mathcal{T}_1 = \{A \sqsubseteq \exists S, \exists S^- \sqsubseteq \exists T, \exists T^- \sqsubseteq \exists T, T \sqsubseteq R\}$, and $\mathcal{M}_{\mathcal{K}_1}$ look as follows ($\rightsquigarrow_{\mathcal{K}_1}$ in $\mathcal{G}_{\mathcal{K}_1}$ is depicted as $\to$):



Our first result states that $\mathcal{M}_{\mathcal{K}}$ gives correct answers to all conjunctive queries:

Theorem 5. *For all consistent DL-Lite$^{\mathcal{H}}_{core}$ KBs $\mathcal{K}$, CQs $\boldsymbol{q}(\vec{x})$ and tuples $\vec{a} \subseteq \mathsf{Ind}(\mathcal{A})$, where $\mathcal{K} = (\mathcal{T}, \mathcal{A})$, we have $\mathcal{K} \models \boldsymbol{q}[\vec{a}]$ iff $\mathcal{M}_{\mathcal{K}} \models \boldsymbol{q}[\vec{a}]$.*

Thus, to decide Σ-query entailment between KBs $\mathcal{K}_1$ and $\mathcal{K}_2$, it suffices to check whether $\mathcal{M}_{\mathcal{K}_2} \models \boldsymbol{q}[\vec{a}]$ implies $\mathcal{M}_{\mathcal{K}_1} \models \boldsymbol{q}[\vec{a}]$ for all Σ-queries $\boldsymbol{q}(\vec{x})$ and tuples $\vec{a}$. This relationship between $\mathcal{M}_{\mathcal{K}_2}$ and $\mathcal{M}_{\mathcal{K}_1}$ can be characterised semantically in terms of finite Σ-homomorphisms.

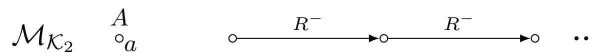
For an interpretation $\mathcal{I}$ and a signature Σ, the Σ-*types* $\boldsymbol{t}^{\mathcal{I}}_{\Sigma}(x)$ and $\boldsymbol{r}^{\mathcal{I}}_{\Sigma}(x,y)$, for $x, y \in \Delta^{\mathcal{I}}$, are given by:

$$\boldsymbol{t}^{\mathcal{I}}_{\Sigma}(x) = \{\Sigma\text{-concept } B \mid x \in B^{\mathcal{I}}\},$$
$$\boldsymbol{r}^{\mathcal{I}}_{\Sigma}(x,y) = \{\Sigma\text{-role } R \mid (x,y) \in R^{\mathcal{I}}\}.$$

A Σ-*homomorphism* from an interpretation $\mathcal{I}$ to $\mathcal{I}'$ is a function $h\colon \Delta^{\mathcal{I}} \to \Delta^{\mathcal{I}'}$ such that $h(a^{\mathcal{I}}) = a^{\mathcal{I}'}$, for all individual names a interpreted in $\mathcal{I}$, $\boldsymbol{t}^{\mathcal{I}}_{\Sigma}(x) \subseteq \boldsymbol{t}^{\mathcal{I}'}_{\Sigma}(h(x))$ and $\boldsymbol{r}^{\mathcal{I}}_{\Sigma}(x,y) \subseteq \boldsymbol{r}^{\mathcal{I}'}_{\Sigma}(h(x),h(y))$, for all $x, y \in \Delta^{\mathcal{I}}$.

It is well-known that answers to Σ-CQs are preserved under Σ-homomorphisms. Thus, if there is a Σ-homomorphism from $\mathcal{M}_{\mathcal{K}_2}$ to $\mathcal{M}_{\mathcal{K}_1}$, then $\mathcal{K}_1$ Σ-query entails $\mathcal{K}_2$. However, the converse does not hold in general.

Example 6. Take $\mathcal{T}_1$ from Example 4, and let $\mathcal{T}_2$ be the result of replacing R in $\mathcal{T}_1$ with R^-. Let $\Sigma = \{A, R\}$ and $\mathcal{K}_i = (\mathcal{T}_i, \{A(a)\})$. Then the Σ-reduct of $\mathcal{M}_{\mathcal{K}_1}$ does not contain a Σ-homomorphic image of the Σ-*reduct* of $\mathcal{M}_{\mathcal{K}_2}$, depicted below. On the other hand, it is easily seen that



$\mathcal{T}_1$ and $\mathcal{T}_2$ are Σ-query inseparable. Note that the Σ-reduct of $\mathcal{M}_{\mathcal{K}_2}$ contains points that are not reachable from the ABox by Σ-roles. In fact, using König's Lemma, one can show that if every point in $\mathcal{M}_{\mathcal{K}_2}$ is reachable from the ABox by a path of Σ-roles, then $\mathcal{K}_1$ Σ-query entails $\mathcal{K}_2$ iff there exists a Σ-homomorphism from $\mathcal{M}_{\mathcal{K}_2}$ to $\mathcal{M}_{\mathcal{K}_1}$.

Because of this, we say that $\mathcal{I}$ is *finitely Σ-homomorphically embeddable into $\mathcal{I}'$* if, for every finite sub-interpretation $\mathcal{I}_1$ of $\mathcal{I}$, there exists a Σ-homomorphism from $\mathcal{I}_1$ to $\mathcal{I}'$. Now one can show:

Theorem 7. *Let $\mathcal{K}_1$ and $\mathcal{K}_2$ be consistent DL-Lite$^{\mathcal{H}}_{core}$ KBs. Then $\mathcal{K}_1$ Σ-query entails $\mathcal{K}_2$ iff $\mathcal{M}_{\mathcal{K}_2}$ is finitely Σ-homomorphically embeddable into $\mathcal{M}_{\mathcal{K}_1}$.*

Theorem 7 does not yet give a satisfactory semantic characterisation of Σ-query entailment between TBoxes, as one still has to consider infinitely many Σ-ABoxes. However, using the fact that inclusions in *DL-Lite*$^{\mathcal{H}}_{core}$, different from disjointness axioms, involve only *one* concept or role in the left-hand side and making sure that the TBoxes entail the same Σ-inclusions, one can show that it is enough to consider *singleton* Σ-ABoxes of the form $\{B(a)\}$. Denote the models $\mathcal{G}_{(\mathcal{T},\{B(a)\})}$ and $\mathcal{M}_{(\mathcal{T},\{B(a)\})}$ by $\mathcal{G}^B_{\mathcal{T}}$ and $\mathcal{M}^B_{\mathcal{T}}$, respectively. We thus obtain the following characterisation of Σ-entailment between *DL-Lite*$^{\mathcal{H}}_{core}$ TBoxes $\mathcal{T}_1$, $\mathcal{T}_2$:

Theorem 8. *$\mathcal{T}_1$ Σ-query entails $\mathcal{T}_2$ iff*

(p) *$\mathcal{T}_2 \models \alpha$ implies $\mathcal{T}_1 \models \alpha$, for all Σ-inclusions α;*

(h) *$\mathcal{M}^B_{\mathcal{T}_2}$ is finitely Σ-homomorphically embeddable into $\mathcal{M}^B_{\mathcal{T}_1}$, for all $\mathcal{T}_1$-consistent Σ-concepts B.*

By applying condition **(p)** to $B \sqsubseteq \bot$, we obtain that every $\mathcal{T}_1$-consistent Σ-concept B is also $\mathcal{T}_2$-consistent.

Complexity of Σ-Query Entailment

We use Theorem 8 to show that deciding Σ-query entailment for *DL-Lite*$^{\mathcal{H}}_{core}$ TBoxes is PSPACE-hard and in EXPTIME.

Recall that subsumption in *DL-Lite*$^{\mathcal{H}}_{core}$ is NLOGSPACE-complete (Calvanese et al. 2007; Artale et al. 2009); so condition **(p)** of Theorem 8 can be checked in polynomial time. And, since there are at most $2 \cdot |\Sigma|$ singleton Σ-ABoxes, we can concentrate on the complexity of checking finite Σ-homomorphic embeddability of canonical models for singleton ABoxes.

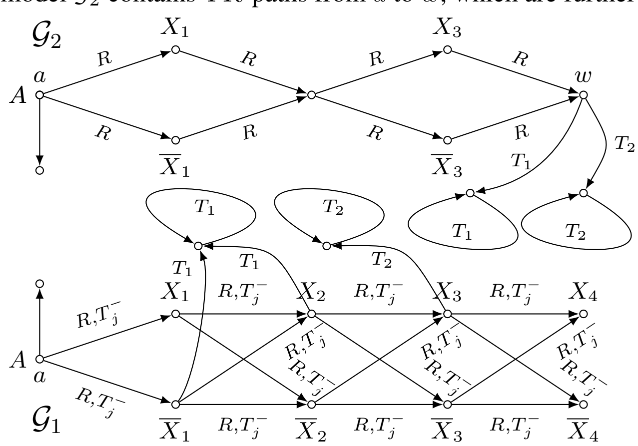
We begin by considering *DL-Lite$_{core}$*, which does not contain role inclusions. In this case, the existence of Σ-homomorphisms between canonical models can be expressed solely in terms of the types of the points in these models; cf. (Kontchakov, Wolter, and Zakharyaschev 2010). Let $\mathcal{T}_1$ and $\mathcal{T}_2$ be *DL-Lite$_{core}$* TBoxes and Σ a signature.

Theorem 9. *$\mathcal{T}_1$ Σ-query entails $\mathcal{T}_2$ iff* **(p)** *holds and, for every $\mathcal{T}_1$-consistent Σ-concept B and every $x \in \Delta^{\mathcal{I}^B_{\mathcal{T}_2}}$, there is $x' \in \Delta^{\mathcal{I}^B_{\mathcal{T}_1}}$ with $\boldsymbol{t}_\Sigma^{\mathcal{I}^B_{\mathcal{T}_2}}(x) \subseteq \boldsymbol{t}_\Sigma^{\mathcal{I}^B_{\mathcal{T}_1}}(x')$.*

The criterion of Theorem 9 can be checked in polynomial time, in NLOGSPACE, to be more precise. Thus:

Theorem 10. *Checking Σ-query entailment for TBoxes in DL-Lite$_{core}$ is* NLOGSPACE*-complete.*

However, if role inclusions become available, the picture changes dramatically: not only do we have to compare the Σ-types of points in the canonical models, but also the Σ-*paths* to these points. To illustrate, consider the generating models $\mathcal{G}_1, \mathcal{G}_2$ below, where the arrows represent the generating relations, and the concept names A, X_i, $\overline{X}_i$ and the role names R and T_j are all symbols in Σ. The model $\mathcal{G}_2$ contains 4 R-paths from a to w, which are further



extended by the infinite T_j-paths. The paths π from a to w can be homomorphically mapped to distinct R-paths $h(\pi)$ in $\mathcal{G}_1$ starting from a. But the extension of such a π with the infinite T_j-chain can only be mapped first to a *suffix of* $h(\pi)$ (backward, along T_j^-)—because we have to map paths in the unfolding $\mathcal{M}_2$ of $\mathcal{G}_2$ to paths in $\mathcal{M}_1$—and then to a T_j-loop in $\mathcal{G}_1$. But to check whether this can be done, we may have to 'remember' the whole path π.

To see that $\mathcal{G}_1$ and $\mathcal{G}_2$ can be given by *DL-Lite$^{\mathcal{H}}_{core}$* TBoxes, fix a quantified Boolean formula $\mathsf{Q}_1 X_1 \dots \mathsf{Q}_n X_n \bigwedge_{j=1}^m C_j$, where $\mathsf{Q}_i \in \{\forall, \exists\}$ and the C_j are clauses over the variables X_i. Let $\Sigma = \{A, X_i, \overline{X}_i, R, T_j \mid i \le n,\ j \le m\}$ and let $\mathcal{T}_1$ contain the inclusions

$$A \sqsubseteq \exists S_0^-, \quad \exists S_{i-1}^- \sqsubseteq \exists Q_i^k,$$
$$\exists (Q_i^k)^- \sqsubseteq X_i^k, \qquad Q_i^k \sqsubseteq S_i, \qquad S_i \sqsubseteq R,$$
$$X_i^k \sqsubseteq \exists R_j \quad \text{if } k = 0, \neg X_i \in C_j \text{ or } k = 1, X_i \in C_j,$$
$$\exists R_j^- \sqsubseteq \exists R_j, \qquad R_j \sqsubseteq T_j, \qquad S_i \sqsubseteq T_j^-,$$

and $\mathcal{T}_2$ the inclusions

$$A \sqsubseteq \exists S_0^-, \qquad \exists S_{i-1}^- \sqsubseteq \begin{cases} \exists Q_i^k, & \text{if } \mathsf{Q}_i = \forall, \\ \exists S_i, & \text{if } \mathsf{Q}_i = \exists, \end{cases}$$
$$\exists (Q_i^k)^- \sqsubseteq X_i^k, \qquad Q_i^k \sqsubseteq S_i, \qquad S_i \sqsubseteq R,$$
$$\exists S_n^- \sqsubseteq \exists P_j, \qquad \exists P_j^- \sqsubseteq \exists P_j, \qquad P_j \sqsubseteq T_j,$$

for all $i \le n$, $j \le m$ and $k = 1, 2$. The generating models $\mathcal{G}^A_{\mathcal{T}_1}$ and $\mathcal{G}^A_{\mathcal{T}_2}$, restricted to Σ, look like $\mathcal{G}_1$ and $\mathcal{G}_2$ in the picture above, respectively. Moreover, one can show that $\mathcal{M}^A_{\mathcal{T}_2}$ is (finitely) Σ-homomorphically embeddable into $\mathcal{M}^A_{\mathcal{T}_1}$ iff the QBF above is satisfiable. As satisfiability of QBFs is known to be PSPACE-complete, we obtain:

Theorem 11. *Σ-query entailment for DL-Lite$^{\mathcal{H}}_{core}$ TBoxes is* PSPACE*-hard.*

On the other hand, the problem whether $\mathcal{M}_{\mathcal{K}_2}$ is finitely Σ-homomorphically embeddable into $\mathcal{M}_{\mathcal{K}_1}$ can be reduced to the emptiness problem for alternating two-way automata, which belongs to EXPTIME (Vardi 1998). In a way similar to (Vardi 1998; Grädel and Walukiewicz 1999), where these automata were employed to prove EXPTIME-decidability of the modal μ-calculus with converse and the guarded fixed point logic of finite width, one can use their ability to 'remember' paths (in the sense illustrated in the example above) to obtain the EXPTIME upper bound:

Theorem 12. *Checking Σ-query entailment for DL-Lite$^{\mathcal{H}}_{core}$ TBoxes is in* EXPTIME.

The precise complexity of Σ-query entailment for *DL-Lite$^{\mathcal{H}}_{core}$* TBoxes is still unknown. To put the obtained results into perspective, let us recall that deciding Σ-query entailment for ontologies in the DL *DL-Lite$^{\mathcal{N}}_{horn}$* is coNP-complete (Kontchakov, Wolter, and Zakharyaschev 2010). Compared to *DL-Lite$^{\mathcal{H}}_{core}$*, *DL-Lite$^{\mathcal{N}}_{horn}$* allows (unqualified) number restrictions and conjunctions in the left-hand side of concept inclusions, but does not have role inclusions, that is: *DL-Lite$^{\mathcal{N}}_{horn}$* $\cap$ *DL-Lite$^{\mathcal{H}}_{core}$* $=$ *DL-Lite$_{core}$*. The data complexity of answering CQs is the same for all three languages under the UNA: AC^0. However, the computational properties of these logics become different as far as Σ-query entailment is concerned: NLOGSPACE-complete for *DL-Lite$_{core}$*, coNP-complete for *DL-Lite$^{\mathcal{N}}_{horn}$*, and between PSPACE and EXPTIME for *DL-Lite$^{\mathcal{H}}_{core}$*. It may be of interest to note that Σ-query entailment for *DL-Lite$^{\mathcal{N}}_{bool}$*, allowing full Booleans as concept constructs, is Π^p_2-complete.

Strong Σ-Query Entailment

It is pretty straightforward to construct an exponential time algorithm checking strong Σ-query entailment between *DL-Lite$^{\mathcal{H}}_{core}$* TBoxes $\mathcal{T}_1$ and $\mathcal{T}_2$: enumerate all Σ-TBoxes $\mathcal{T}$ and check whether $\mathcal{T}_1 \cup \mathcal{T}$ Σ-query entails $\mathcal{T}_2 \cup \mathcal{T}$. As there are quadratically many Σ-inclusions, this algorithm calls the Σ-query entailment checker $2^{|\Sigma|^2}$ times, in the worst case. We now show that one can do much better than that.

First, it turns out that instead of expensive Σ-query entailment checks for the TBoxes $\mathcal{T}_i \cup \mathcal{T}$, it is enough to check consistency (in polynomial time). More precisely, suppose

$\mathcal{T}_1$ Σ-query entails $\mathcal{T}_2$. One can show then that $\mathcal{T}_1$ does not strongly Σ-query entail $\mathcal{T}_2$ iff there exist a Σ-TBox $\mathcal{T}$ and a Σ-concept B such that $(\mathcal{T}_1 \cup \mathcal{T}, \{B(a)\})$ is consistent but $(\mathcal{T}_2 \cup \mathcal{T}, \{B(a)\})$ is not (see Example 3 above).

Moreover, checking consistency for all Σ-TBoxes $\mathcal{T}$ can further be reduced—using the primitive form of *DL-Lite*$^{\mathcal{H}}_{core}$ axioms—to checking consistency for all *singleton* Σ-TBoxes $\mathcal{T}$. Thus, we obtain the following:

Theorem 13. *Suppose that $\mathcal{T}_1$ Σ-query entails $\mathcal{T}_2$. Then $\mathcal{T}_1$ does not strongly Σ-query entail $\mathcal{T}_2$ iff there is a Σ-concept B and a Σ-TBox $\mathcal{T}$ with a single inclusion of the form $B_1 \sqsubseteq B_2$ or $R_1 \sqsubseteq R_2$ such that $(\mathcal{T}_1 \cup \mathcal{T}, \{B(a)\})$ is consistent but $(\mathcal{T}_2 \cup \mathcal{T}, \{B(a)\})$ is inconsistent.*

So, if we already know that $\mathcal{T}_1$ Σ-query entails $\mathcal{T}_2$, then checking whether this entailment is actually *strong* can be done in polynomial time (and NLogSpace). The proof, based on both semantical and proof-theoretic constructions, is given in the full version of the paper.

Theorem 13 is crucial for the implementation of an efficient strong Σ-query entailment checker, as discussed in the section on our experiments below.

Incomplete algorithm for Σ-query entailment

The complex interplay between role inclusions and inverse roles, required in the proof of PSpace-hardness, appears to be too artificial compared to how roles are used in 'real-world' ontologies. For example, in conceptual modelling, the number of roles is comparable with the number of concepts, but the number of role inclusions is normally very small (see the table in the next section). For this reason, instead of a complete exponential time Σ-query entailment checker, we have implemented a polynomial time correct but incomplete algorithm, which is based on testing simulations between transition systems.

Let $\mathcal{T}_1$ and $\mathcal{T}_2$ be *DL-Lite*$^{\mathcal{H}}_{core}$ TBoxes, Σ a signature, B a Σ-concept. Denote $\mathcal{K}_i = (\mathcal{T}_i, \{B(a)\})$ and $\mathcal{I}_i = \mathcal{I}_{\mathcal{K}_i}$, $i = 1, 2$. A relation $\rho \subseteq \Delta^{\mathcal{I}_2} \times \Delta^{\mathcal{I}_1}$ is called a *Σ-simulation of $\mathcal{G}_{\mathcal{K}_2}$ in $\mathcal{G}_{\mathcal{K}_1}$* if the following conditions hold:

(s1) the domain of ρ is $\Delta^{\mathcal{I}_2}$ and $(a^{\mathcal{I}_2}, a^{\mathcal{I}_1}) \in \rho$;

(s2) $\boldsymbol{t}^{\mathcal{I}_2}_{\Sigma}(x) \subseteq \boldsymbol{t}^{\mathcal{I}_1}_{\Sigma}(x')$, for all $(x, x') \in \rho$;

(s3) if $x \rightsquigarrow_{\mathcal{K}_2} w_{[R]}$ and $(x, x') \in \rho$, then there is $y' \in \Delta^{\mathcal{I}_1}$ such that $(w_{[R]}, y') \in \rho$ and $S \in \boldsymbol{r}^{\mathcal{I}_1}_{\Sigma}(x', y')$ for every Σ-role S with $[R] \leq_{\mathcal{T}_2} [S]$.

We call ρ a *forward Σ-simulation* if it satisfies **(s1)**, **(s2)** and the condition **(s3′)**, which strengthens **(s3)** with the extra requirement: $y' = w_{[T]}$, for some role T, with $x' \rightsquigarrow_{\mathcal{K}_1} w_{[T]}$ and $[T] \leq_{\mathcal{T}_1} [S]$ for every Σ-role S with $[R] \leq_{\mathcal{T}_2} [S]$.

Example 14. In Example 6, there is a Σ-simulation of $\mathcal{G}_{\mathcal{K}_2}$ in $\mathcal{G}_{\mathcal{K}_1}$, but no forward Σ-simulation exists. The same applies to $\mathcal{G}_2$ and $\mathcal{G}_1$ in the proof of the PSpace lower bound.

In contrast to finite Σ-homomorphic embeddability of $\mathcal{M}_{\mathcal{K}_2}$ in $\mathcal{M}_{\mathcal{K}_1}$, the problem of checking the existence of (forward) Σ-simulations of $\mathcal{G}_{\mathcal{K}_2}$ in $\mathcal{G}_{\mathcal{K}_1}$ is tractable and well understood from the literature on program verification (Baier and Katoen 2007). Consider now the following conditions, which can be checked in polynomial time:

(y) condition **(p)** holds and there is a *forward* Σ-simulation of $\mathcal{G}^B_{\mathcal{T}_2}$ in $\mathcal{G}^B_{\mathcal{T}_1}$, for every $\mathcal{T}_1$-consistent Σ-concept B;

(n) condition **(p)** does not hold or there is no Σ-simulation of $\mathcal{G}^B_{\mathcal{T}_2}$ in $\mathcal{G}^B_{\mathcal{T}_1}$, for any $\mathcal{T}_1$-consistent Σ-concept B.

Theorem 15. *Let $\mathcal{T}_1, \mathcal{T}_2$ be DL-Lite$^{\mathcal{H}}_{core}$ TBoxes and Σ a signature. If* **(y)** *holds, then $\mathcal{T}_1$ Σ-query entails $\mathcal{T}_2$. If* **(n)** *holds, then $\mathcal{T}_1$ does not Σ-query entail $\mathcal{T}_2$.*

Thus, an algorithm checking conditions (**y**) and (**n**) can be used as a correct but incomplete Σ-query entailment checker. It cannot be complete since neither (**y**) nor (**n**) holds in Example 14. On the other hand, condition **(n)** proves to be a criterion of Σ-query entailment in two important cases:

Theorem 16. *Suppose that* (*a*) *$\mathcal{T}_1$ and $\mathcal{T}_2$ are DL-Lite$_{core}$ TBoxes, or* (*b*) *$\mathcal{T}_1 = \emptyset$ and $\mathcal{T}_2$ is a DL-Lite$^{\mathcal{H}}_{core}$ TBox. Then condition* **(n)** *holds iff $\mathcal{T}_1$ does not Σ-query entail $\mathcal{T}_2$.*

The case $\mathcal{T}_1 = \emptyset$ is of interest for module extraction and safe module import, which will be discussed in the next section.

Experiments

Checking (strong) Σ-query entailment has multiple applications in ontology versioning, re-use, and extraction. We have used the algorithms, suggested by Theorems 15 and 13, for *minimal module extraction* to see how efficient they are in practice and whether the incompleteness of the **(y)**–**(n)** conditions is problematic. Extracting minimal modules from medium-sized real-world ontologies requires thousands of calls of the (strong) Σ-query entailment checker, and thus provides a tough test for our approach.

For a TBox $\mathcal{T}$ and a signature Σ, a subset $\mathcal{M} \subseteq \mathcal{T}$ is

- a *Σ-query module* of $\mathcal{T}$ if $\mathcal{M} \equiv_{\Sigma} \mathcal{T}$;
- a *strong Σ-query module* of $\mathcal{T}$ if $\mathcal{M} \equiv^{s}_{\Sigma} \mathcal{T}$;
- a *depleting Σ-query module* of $\mathcal{T}$ if $\emptyset \equiv^{s}_{\Sigma \cup \mathsf{sig}(\mathcal{M})} \mathcal{T} \setminus \mathcal{M}$, where $\mathsf{sig}(\mathcal{M})$ is the signature of $\mathcal{M}$.

We are concerned with computing a *minimal* (w.r.t. $\subseteq$) Σ-query (MQM), a *minimal* strong Σ-query (MSQM), and the (uniquely determined) *minimal* depleting Σ-query (MDQM) module of $\mathcal{T}$. The general extraction algorithms, which call Σ-query entailment checkers, are taken from (Kontchakov, Wolter, and Zakharyaschev 2010). For MQMs and MSQMs, the number of calls to the checker coincides with number of inclusions in $\mathcal{T}$. For MDQMs (where one of the TBoxes given to the checker is empty, and so the checker is complete, by Theorem 16), the number of checker calls is quadratic in the number of inclusions in $\mathcal{T}$.

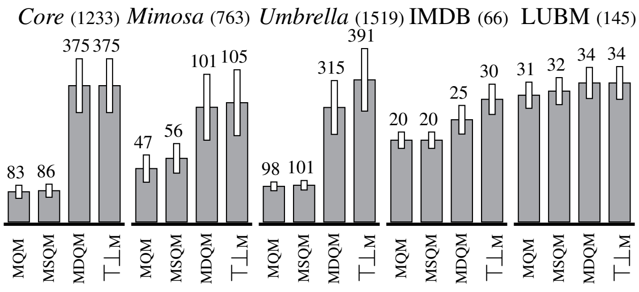
We extracted modules from *OWL 2 QL* approximations of 3 commercial software applications called *Core*, *Umbrella* and *Mimosa* (the original ontologies use a few axioms that are not expressible *OWL 2 QL*). *Mimosa* is a specialisation of the MIMOSA OSA-EAI specification[1] for container shipping. *Core* is based on a supply-chain management system used by the bookstore chain Ottakar's (now merged with Waterstone's), and *Umbrella* on a research data validation and processing system used by the Intensive Care National

[1] htpp://www.mimosa.org/?q=resources/specs/osa-eai-v321

Audit and Research Centre[2] The original *Core* and *Umbrella* were used for the experiments in (Kontchakov, Wolter, and Zakharyaschev 2010). For comparison, we extracted modules from *OWL 2 QL* approximations of the well-known IMDB and LUBM ontologies. For each of these ontologies,

ontology	*Mimosa*	*Core*	*Umbrella*	IMDB	LUBM
concept incl.	710	1214	1506	45	136
role incl.	53	19	13	21	9
concept nm.	106	82	79	14	43
role names	145	76	64	30	31

we randomly generated 20 signatures Σ of 5 concept and 5 roles names. We extracted Σ-MQMs, MSQMs, MDQMs as well as the $\top\bot$-module (Cuenca Grau et al. 2008) from the whole *Mimosa*, IMBD and LUBM ontologies. For the larger *Umbrella* and *Core* ontologies, we first computed the $\top\bot$-modules, and then employed them to further extract MQMs, MSQMs, MDQMs, which are all contained in the $\top\bot$-modules. The average size of the resulting modules and its standard deviation is shown below:



Details of the experiments and ontologies are available at `http://aaai-11.tripod.com/`. Here we briefly comment on efficiency and incompleteness. Checking Σ-query inseparability turned out to be very fast: a single call of the checker never took more than 1s for our ontologies. For strong Σ-query inseparability, the maximal time was less than 1 min. For comparisons with the empty TBox, the maximal time for strong Σ-query inseparability tests was less than 10s. For the hardest case, *Mimosa*, the average total extraction times were 2.5mins for MQMs, 140mins for MSQMs, and 317mins for MDQMs. Finally, only in 9 out of about 75,000 calls, the Σ-query entailment checker was not able to give a certain answer due to the incompleteness of the **(y)**–**(n)** condition, in which case the inclusions in question were added to the module.

Outlook

We have demonstrated that, despite its PSPACE-hardness, (strong) Σ-query inseparability can be decided efficiently for real-world *OWL 2 QL* ontologies. It would be of interest to explore (*i*) whether (some of) our techniques can be extended to more expressive DLs such as $DL\text{-}Lite^{\mathcal{N}}_{horn}$ or even $\mathcal{ELI}$, and (*ii*) how the algorithms deciding inseparability can be utilised for analysing and visualising the difference between ontology versions if two ontologies are not Σ-query inseparable, as required by ontology versioning systems (Noy and Musen 2002).

[2] http://www.icnarc.org

Acknowledgments. This work was partially supported by the U.K. EPSRC grants EP/H05099X/1 and EP/H043594/1.

References

Artale, A.; Calvanese, D.; Kontchakov, R.; and Zakharyaschev, M. 2009. The *DL-Lite* family and relations. *Journal of Artificial Intelligence Research* 36:1–69.

Baader, F.; Calvanese, D.; McGuinness, D.; Nardi, D.; and Patel-Schneider, P., eds. 2003. *The Description Logic Handbook: Theory, Implementation, and Applications*. Cambridge University Press.

Baier, C., and Katoen, J.-P. 2007. *Principles of Model Checking*. MIT Press.

Botoeva, E.; Calvanese, D.; and Rodriguez-Muro, M. 2010. Expressive approximations in *DL-Lite* ontologies. In *Proc. of AIMSA*, 21–31. Springer.

Calvanese, D.; De Giacomo, G.; Lembo, D.; Lenzerini, M.; and Rosati, R. 2006. Data complexity of query answering in description logics. In *Proc. of KR*, 260–270.

Calvanese, D.; De Giacomo, G.; Lembo, D.; Lenzerini, M.; and Rosati, R. 2007. Tractable reasoning and efficient query answering in description logics: The *DL-Lite* family. *J. of Automated Reasoning* 39(3):385–429.

Cuenca Grau, B.; Horrocks, I.; Kazakov, Y.; and Sattler, U. 2008. Modular reuse of ontologies: Theory and practice. *JAIR* 31:273–318.

Dolby, J.; Fokoue, A.; Kalyanpur, A.; Ma, L.; Schonberg, E.; Srinivas, K.; and Sun, X. 2008. Scalable grounded conjunctive query evaluation over large and expressive knowledge bases. In *Proc. of ISWC*, v. 5318 of *LNCS*, 403–418.

Grädel, E., and Walukiewicz, I. 1999. Guarded fixed point logic. In *Proc. of LICS*, 45–54.

Heymans, S.; Ma, L.; Anicic, D.; Ma, Z.; Steinmetz, *et al.* 2008. Ontology reasoning with large data repositories. In *Ontology Management, Semantic Web, Semantic Web Services, and Business Applications*, Springer. 89–128.

Kontchakov, R.; Wolter, F.; and Zakharyaschev, M. 2010. Logic-based ontology comparison and module extraction, with an application to *DL-Lite. Artif. Intell.* 174:1093–1141.

Lutz, C., and Wolter, F. 2010. Deciding inseparability and conservative extensions in the description logic $\mathcal{EL}$. *J. Symb. Comput.* 45(2):194–228.

Noy, N. F., and Musen, M. A. 2002. Promptdiff: A fixed-point algorithm for comparing ontology versions. In *Proc. of AAAI/IAAI*, 744–750.

Pan, J. Z., and Thomas, E. 2007. Approximating OWL-DL Ontologies. In *Proc. of AAAI*, 1434–1439.

Poggi, A.; Lembo, D.; Calvanese, D.; De Giacomo, G.; Lenzerini, M.; and Rosati, R. 2008. Linking data to ontologies. *J. on Data Semantics* X:133–173.

Stuckenschmidt, H.; Parent, C.; and Spaccapietra, S., eds. 2009. *Modular Ontologies: Concepts, Theories and Techniques for Knowledge Modularization*, v. 5445 of *LNCS*.

Vardi, M. Y. 1998. Reasoning about the past with two-way automata. In *Proc. of ICALP*, v. 1443 of *LNCS*, 628–641.

Proof of Theorem 5

We first prove that $\mathcal{I}_\mathcal{K} \models \mathcal{K}$ and $\mathcal{M}_\mathcal{K} \models \mathcal{K}$. We also show a lemma that we will reuse in the following sections.

Proposition 17. *For all consistent DL-Lite$^{\mathcal{H}}_{core}$ KBs $\mathcal{K}$, $\mathcal{I}_\mathcal{K} \models \mathcal{K}$ and $\mathcal{M}_\mathcal{K} \models \mathcal{K}$.*

Proof. Let $\mathcal{K} = (\mathcal{T}, \mathcal{A})$. We show $\mathcal{I}_\mathcal{K} \models \mathcal{K}$. $\mathcal{M}_\mathcal{K} \models \mathcal{K}$ can be shown similarly and is left to the reader. The part $\mathcal{I}_\mathcal{K} \models \mathcal{A}$ is obvious from the construction of $\mathcal{I}_\mathcal{K}$. For $\mathcal{I}_\mathcal{K} \models \mathcal{T}$, we take $\alpha \in \mathcal{T}$ and proceed by case distinction over the possible shapes of α. We can treat $\top$ and $\bot$ the same way as concept names because the equation

$$A^{\mathcal{I}_\mathcal{K}} = \{a \mid \mathcal{K} \models A(a)\} \cup \{w_{[R]} \mid \mathcal{T} \models \exists R^- \sqsubseteq A\},$$

which is part of the definition of $\mathcal{I}_\mathcal{K}$, holds for $\top$ and $\bot$ as well—for $\top$ trivially, and for $\bot$ using the fact that R is generating for all $w_{[R]} \in \Delta^{\mathcal{I}_\mathcal{K}}$.

$\alpha = A_1 \sqsubseteq A_2$. Let $x \in A_1^{\mathcal{I}_\mathcal{K}}$. If $x = a \in \mathsf{Ind}(\mathcal{A})$, then $\mathcal{K} \models A_1(a)$ by construction of $A_1^{\mathcal{I}_\mathcal{K}}$. Since $\alpha \in \mathcal{T}$, we have $\mathcal{K} \models A_2(a)$ and, by construction of $A_2^{\mathcal{I}_\mathcal{K}}$, $a \in A_2^{\mathcal{I}_\mathcal{K}}$. For $x = w_{[R]}$ we can argue analogously.

$\alpha = A \sqsubseteq \exists R$. Let $x \in A^{\mathcal{I}_\mathcal{K}}$. If $x = a \in \mathsf{Ind}(\mathcal{A})$, then $\mathcal{K} \models A(a)$ and, since $\alpha \in \mathcal{T}$, we have $\mathcal{K} \models \exists R(a)$. Take some $\leq_\mathcal{T}$-minimal $[S]$ with $[S] \leq_\mathcal{T} [R]$ and $\mathcal{K} \models \exists S(a)$. Then it holds that $x \rightsquigarrow_\mathcal{K} w_{[S]}$. From the construction of $R^{\mathcal{I}_\mathcal{K}}$, we can derive $(x, w_{[S]}) \in R^{\mathcal{I}_\mathcal{K}}$, hence $x \in (\exists R)^{\mathcal{I}_\mathcal{K}}$. If $x = w_{[S]}$ for some S, then $\mathcal{T} \models \exists S^- \sqsubseteq A$. Since $\alpha \in \mathcal{T}$, we have $\mathcal{T} \models \exists S^- \sqsubseteq \exists R$ and therefore $w_{[S]} \rightsquigarrow_\mathcal{K} w_{[T]}$ for some $\leq_\mathcal{T}$-minimal $[T]$ with $[T] \leq_\mathcal{T} [R]$ and $\mathcal{T} \models \exists T^- \sqsubseteq \exists R$. The rest of the argument is the same.

$\alpha = \exists R \sqsubseteq A$. Let $x \in (\exists R)^{\mathcal{I}_\mathcal{K}}$. If $x = a \in \mathsf{Ind}(\mathcal{A})$, then the construction of $R^{\mathcal{I}_\mathcal{K}}$ implies that either (i) there is some $b \in \mathsf{Ind}(\mathcal{A})$ with $S(a, b) \in \mathcal{A}$ for $S \sqsubseteq^*_\mathcal{T} R$, or (ii) $a \rightsquigarrow_\mathcal{K} w_{[S]}$ for some $[S]$ with $[S] \leq_\mathcal{T} [R]$. In case (i), since $\alpha \in \mathcal{T}$, we obtain $\mathcal{K} \models A(a)$ and therefore $a \in A^{\mathcal{I}_\mathcal{K}}$. In case (ii), we conclude that $\mathcal{K} \models \exists R(a)$ and, since $\alpha \in \mathcal{T}$, $\mathcal{K} \models A(a)$. The rest is by construction of $A^{\mathcal{I}_\mathcal{K}}$. For $x = w_{[S]}$ we can argue analogously.

$\alpha = \exists R \sqsubseteq \exists S$. The argument is a combination of those in the previous two cases.

$\alpha = A_1 \sqcap A_2 \sqsubseteq \bot$. Let $x \in A_1^{\mathcal{I}_\mathcal{K}} \cup A_2^{\mathcal{I}_\mathcal{K}}$. If $x = a \in \mathsf{Ind}(\mathcal{A})$, then $\mathcal{K} \models A_i(a)$ and, since $\alpha \in \mathcal{T}$, we have a contradiction to $\mathcal{K}$ being consistent. If $x = w_{[R]}$ for some R, then $\mathcal{T} \models \exists R^- \sqsubseteq A_i$. Since $\alpha \in \mathcal{T}$, we obtain $\mathcal{T} \models \exists R^- \sqsubseteq \bot$, a contradiction to R being generating.

$\alpha = A \sqcap \exists R \sqsubseteq \bot$. We derive a contradiction analogously to the previous case.

$\alpha = \exists R \sqcap \exists S \sqsubseteq \bot$. Analogous.

$\alpha = R_1 \sqsubseteq R_2$. Let $(x, y) \in R_1^{\mathcal{I}_\mathcal{K}}$. By construction of $R_1^{\mathcal{I}_\mathcal{K}}$, we have that $y = w_{[S]}$ for some $[S]$ with $[S] \leq_\mathcal{T} [R_1]$. Since $\alpha \in \mathcal{T}$, we have $[S] \leq_\mathcal{T} [R_2]$ and, therefore, $(x, y) \in R_2^{\mathcal{I}_\mathcal{K}}$.

$\alpha = R_1 \sqcap R_2 \sqsubseteq \bot$. We proceed as in the previous case until we obtain a contradiction to S being generating. $\square$

For the following lemma, we say that a *homomorphism* from an interpretation $\mathcal{I}$ to $\mathcal{I}'$ is a $\Sigma_\mathcal{I}$-homomorphism from $\mathcal{I}$ to $\mathcal{I}'$, where $\Sigma_\mathcal{I}$ is the signature of $\mathcal{I}$. Analogously, $\boldsymbol{t}^\mathcal{I}(\cdot)$ and $\boldsymbol{r}^\mathcal{I}(\cdot,\cdot)$ denote $\boldsymbol{t}^\mathcal{I}_{\Sigma_\mathcal{I}}(\cdot)$ and $\boldsymbol{r}^\mathcal{I}_{\Sigma_\mathcal{I}}(\cdot,\cdot)$.

Lemma 18. *For every consistent DL-Lite$^{\mathcal{H}}_{core}$ KB $\mathcal{K}$ and every model $\mathcal{I} \models \mathcal{K}$, there exists a homomorphism from $\mathcal{M}_\mathcal{K}$ to $\mathcal{I}$.*

Proof. We define a function $h : \Delta^{\mathcal{M}_\mathcal{K}} \to \Delta^\mathcal{I}$ that guarantees $h(\sigma \cdot w_{[R]}) \in (\exists R^-)^\mathcal{I}$ for all paths σ and roles R. We simultaneously define $h(\sigma)$ and prove the guarantee by induction on $|\sigma|$. For $\sigma = a$, we set $h(a) = a^\mathcal{I}$. For longer paths, we set $h(\sigma \cdot w_{[R]})$ to some $z \in \Delta^\mathcal{I}$ with $(h(\sigma), z) \in R^\mathcal{I}$. Such a z exists: the construction of $\mathcal{M}_\mathcal{K}$ requires that $\mathsf{tail}(\sigma) \rightsquigarrow_\mathcal{K} w_{[S]}$ for some $[S]$ with $[S] \leq_\mathcal{T} [R]$, and the definition of $\rightsquigarrow_\mathcal{K}$ now admits two cases. If $\mathsf{tail}(\sigma) = a \in \mathsf{Ind}(\mathcal{A})$, then $\mathcal{K} \models \exists S(a)$, i.e., $\mathcal{K} \models \exists R(a)$. If $\mathsf{tail}(\sigma) = w_{[T]}$, then $\mathcal{T} \models \exists T^- \sqsubseteq \exists S$, i.e., $\mathcal{T} \models \exists T^- \sqsubseteq \exists R$. Due to the induction hypothesis for the above guarantee, we obtain that z with $(h(\sigma), z) \in R^\mathcal{I}$ exists and that $h(\sigma \cdot w_{[R]}) = z$ satisfies the guarantee again.

It remains to show that h is a homomorphism, i.e.,

1. $h(a^{\mathcal{M}_\mathcal{K}}) = a^\mathcal{I}$ for all $a \in \mathsf{Ind}(\mathcal{A})$,
2. For all paths $\sigma \in \Delta^{\mathcal{M}_\mathcal{K}}$: $\boldsymbol{t}^{\mathcal{M}_\mathcal{K}}(\sigma) \subseteq \boldsymbol{t}^\mathcal{I}(h(\sigma))$.
3. For all paths $\sigma_1, \sigma_2 \in \Delta^{\mathcal{M}_\mathcal{K}}$: $\boldsymbol{r}^{\mathcal{M}_\mathcal{K}}(\sigma_1, \sigma_2) \subseteq \boldsymbol{t}^\mathcal{I}(h(\sigma_1), h(\sigma_2))$.

(1) is ensured by $a^{\mathcal{M}_\mathcal{K}} = a$ and $h(a) = a^\mathcal{I}$. For (2), we proceed by induction on $|\sigma|$.

$\sigma = a$. Let $B \in \boldsymbol{t}^{\mathcal{M}_\mathcal{K}}(a)$, i.e., $a \in B^{\mathcal{M}_\mathcal{K}}$. If $B = A$, then the construction of $\mathcal{M}_\mathcal{K}$ implies that $\mathcal{K} \models A(a)$, hence $a \in A^\mathcal{I}$, i.e., $B \in \boldsymbol{t}^\mathcal{I}(a)$. If $B = \exists R$, then the construction of $\mathcal{M}_\mathcal{K}$ implies that $(a, x) \in R^{\mathcal{I}_\mathcal{K}}$ for some x. Either $x = b \in \mathsf{Ind}(\mathcal{A})$—in which case $S(a, b) \in \mathcal{A}$ for some $S \sqsubseteq^* R$, and hence $a \in (\exists R)^\mathcal{I}$—or $x = w_{[S]}$ for some $[S]$ with $a \rightsquigarrow_\mathcal{K} w_{[S]}$ and $[S] \leq_\mathcal{T} [R]$. In the latter case, we obtain $\mathcal{K} \models \exists S(a)$, i.e., $\mathcal{K} \models \exists R(a)$ and $a \in (\exists R)^\mathcal{I}$, i.e., $B \in \boldsymbol{t}^\mathcal{I}(a)$.

$\sigma = \sigma' \cdot w_{[R]}$. Let $B \in \boldsymbol{t}^{\mathcal{M}_\mathcal{K}}(\sigma)$, i.e., $\sigma \in B^{\mathcal{M}_\mathcal{K}}$. If $B = A$, then the constructions of $\mathcal{M}_\mathcal{K}$ and $\mathcal{I}_\mathcal{K}$ imply that $w_{[R]} \in A^{\mathcal{I}_\mathcal{K}}$ and $\mathcal{T} \models \exists R^- \sqsubseteq A$. With the guarantee $h(\sigma) \in (\exists R^-)^\mathcal{I}$, we obtain $h(\sigma) \in A^\mathcal{I}$, i.e., $B \in \boldsymbol{t}^\mathcal{I}(h(\sigma))$. If $B = \exists S$, then the construction of $\mathcal{M}_\mathcal{K}$ implies that $w_{[R]} \rightsquigarrow_\mathcal{K} w_{[T]}$ for some $[T]$ with $[T] \leq_\mathcal{T} [S]$. Hence, $\mathcal{T} \models \exists R^- \sqsubseteq \exists T$, i.e., $\mathcal{T} \models \exists R^- \sqsubseteq \exists S$. With the guarantee $h(\sigma) \in (\exists R^-)^\mathcal{I}$, we obtain $h(\sigma) \in (\exists S)^\mathcal{I}$, i.e., $B \in \boldsymbol{t}^\mathcal{I}(h(\sigma))$.

For (3), let $R \in \boldsymbol{r}^{\mathcal{M}_\mathcal{K}}(\sigma_1, \sigma_2)$, i.e., $(\sigma_1, \sigma_2) \in R^{\mathcal{M}_\mathcal{K}}$. Assume w.l.o.g. that $R = P$ is a role name. From the construction of $\mathcal{M}_\mathcal{K}$, we conclude that one of the following two cases occurs.

$\sigma_2 = \sigma_1 \cdot w_{[S]}$ for some $[S]$ with $\mathsf{tail}(\sigma_1) \leadsto_{\mathcal{K}} w_{[S]}$ and $[S] \leq_{\mathcal{T}} [P]$. From the construction of h, we obtain $(h(\sigma_1), h(\sigma_2)) \in S^{\mathcal{I}}$, hence $(h(\sigma_1), h(\sigma_2)) \in P^{\mathcal{I}}$, i.e. $R = P \in \boldsymbol{r}^{\mathcal{I}}(h(\sigma_1), h(\sigma_2))$.

$\sigma_1 = \sigma_2 \cdot w_{[S]}$ for some $[S]$ with $\mathsf{tail}(\sigma_2) \leadsto_{\mathcal{K}} w_{[S]}$ and $[S] \leq_{\mathcal{T}} [P^-]$. From the construction of h, we obtain $(h(\sigma_2), h(\sigma_1)) \in S^{\mathcal{I}}$, hence $(h(\sigma_1), h(\sigma_2)) \in P^{\mathcal{I}}$, i.e. $R = P \in \boldsymbol{r}^{\mathcal{I}}(h(\sigma_1), h(\sigma_2))$.

□

We are now ready to prove Theorem 5 which we formulate again.

Theorem 5. For all consistent *DL-Lite*$^{\mathcal{H}}_{core}$ KBs $\mathcal{K}$, CQs $\boldsymbol{q}(\vec{x})$ and tuples $\vec{a} \subseteq \mathsf{Ind}(\mathcal{A})$, where $\mathcal{K} = (\mathcal{T}, \mathcal{A})$, we have $\mathcal{K} \models \boldsymbol{q}[\vec{a}]$ iff $\mathcal{M}_{\mathcal{K}} \models \boldsymbol{q}[\vec{a}]$.

Proof. By Proposition 17, $\mathcal{M}_{\mathcal{K}} \models \mathcal{K}$. This implies direction "⇒".

For the "⇐" direction, let $\mathcal{M}_{\mathcal{K}} \models \boldsymbol{q}[\vec{a}]$ with $\vec{a} = a_1, \ldots, a_n \subseteq \mathsf{Ind}(\mathcal{A})$, and let $\boldsymbol{q}(\vec{x}) = \exists y_1 \ldots \exists y_m.\varphi(x_1, \ldots, x_n, y_1, \ldots, y_m)$. It is sufficient to show that $\mathcal{I} \models \boldsymbol{q}[\vec{a}]$ for every model $\mathcal{I}$ of $\mathcal{K}$. There are $\sigma_1, \ldots, \sigma_m \in \Delta^{\mathcal{M}_{\mathcal{K}}}$ such that $\mathcal{M}_{\mathcal{K}} \models \varphi[\vec{a}, \sigma_1, \ldots, \sigma_m]$. Now let $\mathcal{I} \models \mathcal{K}$. Take a homomorphism h from $\mathcal{M}_{\mathcal{K}}$ to $\mathcal{I}$, which exists due to Lemma 18. The following two properties are immediate consequences of h being a homomorphism, and they imply that $\mathcal{I} \models \boldsymbol{q}[\vec{a}, h(\sigma_1), \ldots, h(\sigma_m)]$.

1. For all concepts B and paths $\sigma \in \Delta^{\mathcal{M}_{\mathcal{K}}}$: if $\sigma \in B^{\mathcal{M}_{\mathcal{K}}}$, then $h(\sigma) \in B^{\mathcal{I}}$.
2. For all roles R and paths $\sigma_1, \sigma_2 \in \Delta^{\mathcal{M}_{\mathcal{K}}}$: if $(\sigma_1, \sigma_2) \in R^{\mathcal{M}_{\mathcal{K}}}$, then $(h(\sigma_1), h(\sigma_2)) \in R^{\mathcal{I}}$.

□

Proof of Theorem 7

Proof. Using Theorem 5, it suffices to show that the following two conditions are equivalent.

1. For all Σ-queries $\boldsymbol{q}$ and $\vec{a} \subseteq \mathsf{Ind}(\mathcal{A})$, if $\mathcal{M}_{\mathcal{K}_2} \models \boldsymbol{q}[\vec{a}]$, then $\mathcal{M}_{\mathcal{K}_1} \models \boldsymbol{q}[\vec{a}]$.
2. $\mathcal{M}_{\mathcal{K}_2}$ is finitely Σ-homomorphically embeddable into $\mathcal{M}_{\mathcal{K}_1}$.

For "(1) ⇒ (2)", assume (1). Take $\Delta' \subseteq \Delta^{\mathcal{M}_{\mathcal{K}_2}}$ and let $\Delta' = \{a_1, \ldots, a_k, \sigma_1, \ldots, \sigma_\ell\}$, where $a_i \in \mathsf{Ind}(\mathcal{A})$ and the σ_j are $\mathcal{G}_{\mathcal{K}_2}$-paths. For ease of notation, we abbreviate $\boldsymbol{t}_\Sigma^{\mathcal{M}_{\mathcal{K}_2}}(\cdot)$ and $\boldsymbol{r}_\Sigma^{\mathcal{M}_{\mathcal{K}_2}}(\cdot,\cdot)$ by $\boldsymbol{t}^2_\Sigma(\cdot)$ and $\boldsymbol{r}^2_\Sigma(\cdot,\cdot)$. Take variables $x_1, \ldots, x_k, y_1, \ldots, y_\ell$ and let $\boldsymbol{q} = \exists y_1 \ldots \exists y_\ell \varphi$, where

$$\varphi = \bigwedge_{\substack{i=1,\ldots,k \\ B \in \boldsymbol{t}^2_\Sigma(a_i)}} B(x_i) \wedge \bigwedge_{\substack{i,j=1,\ldots,k \\ B \in \boldsymbol{r}^2_\Sigma(a_i,a_j)}} R(x_i, x_j) \wedge \bigwedge_{\substack{i=1,\ldots,k \\ j=1,\ldots,\ell \\ B \in \boldsymbol{r}^2_\Sigma(a_i,\sigma_j)}} R(x_i, y_j)$$
$$\wedge \bigwedge_{\substack{i=1,\ldots,\ell \\ B \in \boldsymbol{t}^2_\Sigma(\sigma_i)}} B(y_i) \wedge \bigwedge_{\substack{i,j=1,\ldots,\ell \\ B \in \boldsymbol{r}^2_\Sigma(\sigma_i,\sigma_j)}} R(y_i, y_j).$$

Since the Σ-concepts B in the definition of $\boldsymbol{q}$ include $\top$, the query $\boldsymbol{q}$ uses all variables x_i and y_j. Clearly, $\mathcal{M}_{\mathcal{K}_2} \models \boldsymbol{q}[a_1, \ldots, a_k, \sigma_1, \ldots, \sigma_\ell]$. Due to (1), we have $\mathcal{M}_{\mathcal{K}_1} \models \boldsymbol{q}[a_1, \ldots, a_k, \sigma'_1, \ldots, \sigma'_\ell]$ for some $\sigma'_1, \ldots, \sigma'_\ell$ in $\Delta^{\mathcal{M}_{\mathcal{K}_1}}$. We define a function $h : \Delta' \to \Delta^{\mathcal{M}_{\mathcal{K}_1}}$ via $h(a_i) = a_i$ and $h(\sigma_i) = \sigma'_i$. This function is a Σ-homomorphism because it maps every a_i to a_i and the conjuncts in $\boldsymbol{q}$ explicitly require that all Σ-(role-)types are preserved.

For "(2) ⇒ (1)", assume that $\mathcal{M}_{\mathcal{K}_2}$ is finitely Σ-homomorphically embeddable into $\mathcal{M}_{\mathcal{K}_1}$. Let $\boldsymbol{q}$ be a Σ-query with $\mathcal{M}_{\mathcal{K}_2} \models \boldsymbol{q}[a_1, \ldots, a_k, \sigma_1, \ldots, \sigma_\ell]$ for some $\sigma_1, \ldots, \sigma_\ell$ in $\Delta^{\mathcal{M}_{\mathcal{K}_2}}$. Let $\Delta' = \{a_1, \ldots, a_k, \sigma_1, \ldots, \sigma_\ell\}$. Take a Σ-homomorphism $h : \Delta' \to \Delta^{\mathcal{M}_{\mathcal{K}_1}}$ with $h(a_i) = a_i$ for all i. The homomorphism laws imply

- For all Σ-concepts B and $\sigma_i \in \Delta'$: if $\sigma_i \in B^{\mathcal{M}_{\mathcal{K}_2}}$, then $h(\sigma_i) \in B^{\mathcal{M}_{\mathcal{K}_1}}$.
- For all Σ-roles R and $\sigma_i, \sigma_j \in \Delta'$: if $(\sigma_i, \sigma_j) \in R^{\mathcal{M}_{\mathcal{K}_2}}$, then $(h(\sigma_i), h(\sigma_j)) \in R^{\mathcal{M}_{\mathcal{K}_1}}$.

Therefore, $\mathcal{M}_{\mathcal{K}_1} \models \boldsymbol{q}[a_1, \ldots, a_k, h(\sigma_1), \ldots, h(\sigma_\ell)]$. □

Proof of Theorem 8

We divide the proof into two parts. An ABox $\mathcal{A}$ is called a *single individuum ABox* if $\mathcal{A}$ is of the form $\{B_1(a), \ldots, B_n(a)\}$, where a is an individual name and $B_1, \ldots, B_n$ are concepts.

Lemma 19. *$\mathcal{T}_1$ Σ-query entails $\mathcal{T}_2$ if*

(p) *$\mathcal{T}_2 \models \alpha$ implies $\mathcal{T}_1 \models \alpha$, for all Σ-inclusions α;*

(i) *$\mathcal{M}_{(\mathcal{T}_2,\mathcal{A})}$ is finitely Σ-homomorphically embeddable into $\mathcal{M}_{(\mathcal{T}_1,\mathcal{A})}$, for all consistent $(\mathcal{T}_i, \mathcal{A})$, $i = 1, 2$, such that $\mathcal{A}$ is a single individuum Σ-ABox.*

Proof. Assume that **(p)** and **(i)** hold. To prove that $\mathcal{T}_1$ Σ-query entails $\mathcal{T}_2$ let $\mathcal{A}$ be a Σ-ABox. Assume first $(\mathcal{T}_2, \mathcal{A})$ is inconsistent. We have to show that $(\mathcal{T}_1, \mathcal{A})$ is inconsistent. We may assume that $\exists R(a) \in \mathcal{A}$ whenever there exists b with $R(a, b) \in \mathcal{A}$. Similarly, we assume $R(a, b) \in \mathcal{A}$ iff $R^-(b, a) \in \mathcal{A}$. Then one can readily show that from $(\mathcal{T}_2, \mathcal{A})$ inconsistent it follows that at least one of the following applies:

- there exist $a \in \mathsf{Ind}(\mathcal{A})$ and $B_1(a), B_2(a) \in \mathcal{A}$ such that $\mathcal{T}_2 \models B_1 \sqcap B_2 \sqsubseteq \bot$. But then, by **(p)**, $\mathcal{T}_1 \models B_1 \sqcap B_2 \sqsubseteq \bot$ and so $(\mathcal{T}_1, \mathcal{A})$ is inconsistent.
- there exist $a, b \in \mathsf{Ind}(\mathcal{A})$ and $R(a, b), S(a, b) \in \mathcal{A}$ such that $\mathcal{T}_2 \models R \sqcap S \sqsubseteq \bot$. But then, by **(p)**, $\mathcal{T}_1 \models S \sqcap R \sqsubseteq \bot$ and so $(\mathcal{T}_1, \mathcal{A})$ is inconsistent.

Assume now that $(\mathcal{T}_i, \mathcal{A})$, $i = 1, 2$, are consistent. We show that $\mathcal{M}_{(\mathcal{T}_2,\mathcal{A})}$ is finitely Σ-homomorphically embeddable into $\mathcal{M}_{(\mathcal{T}_1,\mathcal{A})}$. First observe that we can relax the condition defining $a \leadsto_{\mathcal{K}} w_{[R]}$ to requiring $a \in \mathsf{Ind}(\mathcal{A})$ and $[R]$ being $\leq_{\mathcal{T}}$-minimal such that $\mathcal{K} \models \exists R.(a)$ (but not necessarily $\mathcal{K} \not\models R(a, b)$ for any $b \in \mathsf{Ind}(\mathcal{A})$). Denote the modified canonical model by $\mathcal{M}'_{\mathcal{K}}$. Then there exists a Σ-homomorphism from $\mathcal{M}'_{\mathcal{K}}$ to $\mathcal{M}_{\mathcal{K}}$ and vice versa.

Thus, it is sufficient to show that $\mathcal{M}'_{(\mathcal{T}_2,\mathcal{A})}$ is finitely Σ-homomorphically embeddable into $\mathcal{M}'_{(\mathcal{T}_1,\mathcal{A})}$. Let D be a finite subset of $\Delta^{\mathcal{M}'_{(\mathcal{T}_2,\mathcal{A})}}$. We may assume that $\mathsf{Ind}(\mathcal{A}) \subseteq D$. By **(i)**, for $a \in \mathsf{Ind}(\mathcal{A})$, there are Σ-homomorphisms h_a from the interpretations induced by $D \cap \Delta^{\mathcal{M}_{(\mathcal{T}_2,\Gamma(a))}}$ in $\mathcal{M}_{(\mathcal{T}_2,\Gamma(a))}$ to $\mathcal{M}_{(\mathcal{T}_1,\Gamma(a))}$, where $\Gamma(a) = \{B(a) \mid B(a) \in \mathcal{A}\}$. It remains to show that $h = \bigcup_{a\in\mathsf{Ind}(\mathcal{A})} h_a$ is a partial Σ homomorphism with domain D from $\mathcal{M}'_{(\mathcal{T}_2,\mathcal{A})}$ into $\mathcal{M}'_{(\mathcal{T}_1,\mathcal{A})}$. But this follows immediately if

- $a \in B^{\mathcal{M}'_{(\mathcal{T}_2,\mathcal{A})}}$ implies $a \in B^{\mathcal{M}'_{(\mathcal{T}_1,\mathcal{A})}}$ for all Σ-concepts B and $a \in \mathsf{Ind}(\mathcal{A})$;
- $(a,b) \in R^{\mathcal{M}'_{(\mathcal{T}_2,\mathcal{A})}}$ implies $(a,b) \in R^{\mathcal{M}'_{(\mathcal{T}_1,\mathcal{A})}}$ for all Σ-roles B and $a,b \in \mathsf{Ind}(\mathcal{A})$.

The first condition follows from: $\mathcal{T}_2 \models B_1 \sqsubseteq B_2$ implies $\mathcal{T}_1 \models B_1 \sqsubseteq B_2$, for all Σ-inclusions $B_1 \sqsubseteq B_2$, by **(p)**. The second condition follows from: $\mathcal{T}_2 \models R_1 \sqsubseteq R_2$ implies $\mathcal{T}_1 \models R_1 \sqsubseteq R_2$, for all Σ-inclusions $R_1 \sqsubseteq R_2$, by **(p)**. □

We are now ready to prove Theorem 8 which we formulate again.

Theorem 8 $\mathcal{T}_1$ Σ-query entails $\mathcal{T}_2$ iff

(p) $\mathcal{T}_2 \models \alpha$ implies $\mathcal{T}_1 \models \alpha$, for all Σ-inclusions α;

(h) $\mathcal{M}^B_{\mathcal{T}_2}$ is finitely Σ-homomorphically embeddable into $\mathcal{M}^B_{\mathcal{T}_1}$, for all $\mathcal{T}_1$-consistent Σ-concepts B.

Proof. The proof that **(p)** and **(h)** follow if $\mathcal{T}_1$ Σ-query entails $\mathcal{T}_2$ is straightforward and omitted. Conversely, assume that **(p)** and **(h)** hold. It is sufficient to prove that **(i)** from Lemma 19 holds. So, assume that $\mathcal{A} = \{B_1(a),\dots,B_n(a)\}$ is a single individuum Σ-ABoxes such that $(\mathcal{T}_i,\mathcal{A})$, $i = 1,2$, are consistent. We have to show that $\mathcal{M}_{(\mathcal{T}_2,\mathcal{A})}$ is finitely Σ-homomorphically embeddable into $\mathcal{M}_{(\mathcal{T}_1,\mathcal{A})}$. But this follows directly from the fact that every $\mathcal{M}_{(\mathcal{T}_2,\{B_i(a)\})}$ is finitely Σ-homomorphically embeddable into $\mathcal{M}_{(\mathcal{T}_1,\{B_i(a)\})}$, for $1 \leq i \leq n$. □

Proof of Theorem 9

Theorem 9 $\mathcal{T}_1$ Σ-query entails $\mathcal{T}_2$ iff **(p)** holds and, for every $\mathcal{T}_1$-consistent Σ-concept B and every $x \in \Delta^{\mathcal{I}^B_{\mathcal{T}_2}}$, there is $x' \in \Delta^{\mathcal{I}^B_{\mathcal{T}_1}}$ with $\boldsymbol{t}_\Sigma^{\mathcal{I}^B_{\mathcal{T}_2}}(x) \subseteq \boldsymbol{t}_\Sigma^{\mathcal{I}^B_{\mathcal{T}_1}}(x')$.

Proof. The implication $(\Rightarrow)$ follows immediately from Theorem 8 and the definition of Σ-homomorphism.

To prove the converse, we show that, for every $\mathcal{T}_1$-consistent Σ-concept B, there exists a Σ-homomorphism h from $\mathcal{M}^B_{\mathcal{T}_2}$ to $\mathcal{M}^B_{\mathcal{T}_1}$. First, we observe that $\boldsymbol{t}_\Sigma^{\mathcal{M}^B_{\mathcal{T}_2}}(a) \subseteq \boldsymbol{t}_\Sigma^{\mathcal{M}^B_{\mathcal{T}_1}}(a)$. Indeed, otherwise we would have a Σ-concept C such that $\mathcal{T}_2 \models B \sqsubseteq C$ but $\mathcal{T}_1 \not\models B \sqsubseteq C$, contrary to condition **(p)**. Thus, we can set $h(a^{\mathcal{M}^B_{\mathcal{T}_2}}) = a^{\mathcal{M}^B_{\mathcal{T}_1}}$.

Suppose now that we already have $h(x) = x'$ and $\boldsymbol{r}_\Sigma^{\mathcal{M}^B_{\mathcal{T}_2}}(x,y) = \{R\}$ (remember, we are dealing with the language *DL-Lite$_{core}$* which does not contain role inclusions). Then $\exists R \in \boldsymbol{t}_\Sigma^{\mathcal{M}^B_{\mathcal{T}_2}}(x)$, whence $\exists R \in \boldsymbol{t}_\Sigma^{\mathcal{M}^B_{\mathcal{T}_1}}(x')$, and so there is y' such that $\boldsymbol{r}_\Sigma^{\mathcal{M}^B_{\mathcal{T}_1}}(x',y') = \{R\}$. Set $h(y) = y'$. By the observation above, we then have $\boldsymbol{t}_\Sigma^{\mathcal{M}^B_{\mathcal{T}_2}}(y) \subseteq \boldsymbol{t}_\Sigma^{\mathcal{M}^B_{\mathcal{T}_1}}(y')$ (it suffices to consider the case $B = \exists R^-$).

Finally, if a point x is not reachable from the root of $\mathcal{M}^B_{\mathcal{T}_2}$ by a Σ-path. In this case, we take any x' in $\mathcal{M}^B_{\mathcal{T}_2}$ with $\boldsymbol{t}_\Sigma^{\mathcal{M}^B_{\mathcal{T}_2}}(x) \subseteq \boldsymbol{t}_\Sigma^{\mathcal{M}^B_{\mathcal{T}_1}}(x')$ and set $h(x) = x'$.

The resulting map h is clearly a Σ-homomorphism h from $\mathcal{M}^B_{\mathcal{T}_2}$ to $\mathcal{M}^B_{\mathcal{T}_1}$.

□

Proof of Theorem 11

In this section, we show that Σ-query entailment for *DL-Lite$^{\mathcal{H}}_{core}$* TBoxes is PSPACE-hard.

Proof. The proof is by reduction of the satisfiability problem for quantified Boolean formulas (QBFs), which is known to be PSPACE-complete. Suppose we are given a QBF

$$\varphi = \mathsf{Q}_1 X_1 \dots \mathsf{Q}_n X_n \bigwedge_{j=1}^{m} C_j$$

where $\mathsf{Q}_i \in \{\forall,\exists\}$ and the C_j, $1 \leq j \leq m$, are clauses over the variables X_i, $1 \leq i \leq n$. Let

$$\Sigma = \{A, X_i^0, X_i^1, R, T_j \mid 1 \leq i \leq n,\ 1 \leq j \leq m\},$$

where the $X_i, \overline{X}_i$ and A are concept names and the T_j and R are role names. Let $\mathcal{T}_1$ contain the following axioms, for all $1 \leq i \leq n$, $1 \leq j \leq m$ and $k = 0,1$:

$$\begin{gathered} A \sqsubseteq \exists S_0^-, \quad \exists S_{i-1}^- \sqsubseteq \exists Q_i^k, \\ \exists (Q_i^k)^- \sqsubseteq X_i^k, \qquad Q_i^k \sqsubseteq S_i, \qquad S_i \sqsubseteq R, \\ X_i^k \sqsubseteq \exists R_j \quad \text{if } k = 0, \neg X_i \in C_j \text{ or } k = 1, X_i \in C_j, \\ \exists R_j^- \sqsubseteq \exists R_j, \qquad R_j \sqsubseteq T_j, \qquad S_i \sqsubseteq T_j^-, \end{gathered}$$

Consider the TBox $\mathcal{T}_2$ with the following axioms, for all $1 \leq i \leq n$, $1 \leq j \leq m$ and $k = 0,1$:

$$A \sqsubseteq \exists S_0^-, \qquad \exists S_{i-1}^- \sqsubseteq \begin{cases} \exists Q_i^k, & \text{if } \mathsf{Q}_i = \forall, \\ \exists S_i, & \text{if } \mathsf{Q}_i = \exists, \end{cases}$$

$$\begin{gathered} \exists (Q_i^k)^- \sqsubseteq X_i^k, \qquad Q_i^k \sqsubseteq S_i, \qquad S_i \sqsubseteq R, \\ \exists S_n^- \sqsubseteq \exists P_j, \qquad \exists P_j^- \sqsubseteq \exists P_j, \qquad P_j \sqsubseteq T_j, \end{gathered}$$

We show that $\models \varphi$ if, and only if, $\mathcal{T}_1$ Σ-query entails $\mathcal{T}_2$.

$(\Rightarrow)$ Suppose $\models \varphi$. Observe first that the canonical model of $(\mathcal{T}_2, \{B(a)\})$, where B is one of X_i^0, X_i^1, $\exists R$, $\exists R^-$, $\exists T_j$, $\exists T_j^-$ ($1 \leq i \leq n$, $1 \leq j \leq m$), coincides with the canonical model of $(\emptyset, \{B(a)\})$. Therefore, it is trivially finite Σ-homomorphically embeddable in the canonical model of $(\mathcal{T}_1, \{B(a)\})$, which is consistent as there are no occurrences of $\bot$ in $\mathcal{T}_1$. Thus, it remains to show that the canonical model $\mathcal{M}_2$ of $\mathcal{K}_2 = (\mathcal{T}_2, \{A(a)\})$ can be finitely Σ-homomorphically embedded in the canonical model $\mathcal{M}_1$ of $\mathcal{K}_1 = (\mathcal{T}_1, \{A(a)\})$. (The corresponding Σ-reducts of the generating models $\mathcal{G}_2 = (\mathcal{I}_2, \rightsquigarrow_2)$ and

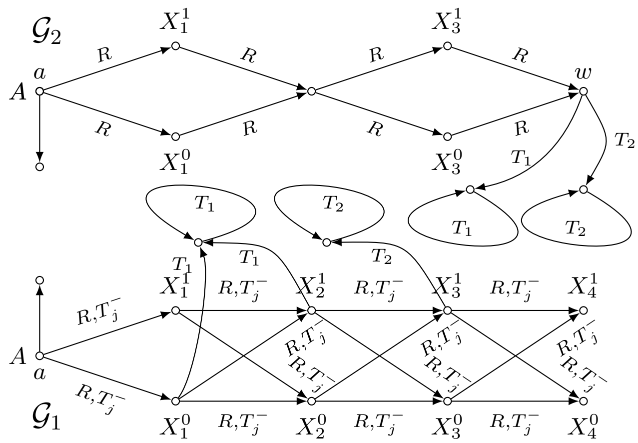


Figure 1: Σ-reducts of generating models $\mathcal{G}_2$ and $\mathcal{G}_1$.

$\mathcal{G}_1 = (\mathcal{I}_1, \rightsquigarrow_1)$ are shown in Fig. 1 for $n = 4$ and $m = 2$ with $\varphi = \forall X_1 \exists X_2 \forall X_3 \exists X_4((\neg X_1 \vee X_2) \wedge X_3)$.) In fact, we construct a Σ-homomorphism $h\colon \Delta^{\mathcal{M}_2} \to \Delta^{\mathcal{M}_1}$. We denote by $w^i_{[R]}$ the R-witness in the generating model $\mathcal{G}_i$, for $i = 1, 2$.

We begin by setting

$$h(a^{\mathcal{I}_2}) = a^{\mathcal{I}_1} \quad \text{and} \quad h(a^{\mathcal{I}_2} w^2_{[S_0^-]}) = a^{\mathcal{I}_1} \cdot w^1_{[S_0^-]}.$$

In what follows we exclude the path $a^{\mathcal{I}_2} w^2_{[S_0^-]}$ from our considerations. We define h in such a way that, for each path π in $\mathcal{G}_2$ of length $i + 1 \leq n$, $h(\pi)$ is a path $a^{\mathcal{I}_1} w_1 \dots w_i$ of length $i + 1$ in $\mathcal{G}_1$ and it defines an assignment $\mathfrak{a}_{h(\pi)}$ to the variables $X_1, \dots, X_i$ by taking, for all $1 \leq i' \leq i$,

$$\mathfrak{a}_{h(\pi)}(X_{i'}) = \top \quad \Leftrightarrow \quad w_{i'} \in (X^1_{i'})^{\mathcal{I}_1},$$
$$\mathfrak{a}_{h(\pi)}(X_{i'}) = \bot \quad \Leftrightarrow \quad w_{i'} \in (X^0_{i'})^{\mathcal{I}_1}.$$

Such assignments $\mathfrak{a}_{h(\pi)}$ will satisfy the following:

($\mathfrak{a}$) the QBF obtained from φ by removing $\mathsf{Q}_1 X_1 \dots \mathsf{Q}_i X_i$ from its prefix is true under $\mathfrak{a}_{h(\pi)}$.

For the paths of length 0 the Σ-homomorphism h has been defined and ($\mathfrak{a}$) trivially holds. Suppose that we have defined h for all paths in $\mathcal{G}_2$ of length $i + 1 \leq n$. We extend h to all paths of length $i + 2$ in $\mathcal{G}_2$ such that ($\mathfrak{a}$) holds. Let π be a path of length $i + 1$. In $\mathcal{G}_1$ we have

$$\mathsf{tail}(h(\pi)) \rightsquigarrow_1 w^1_{[Q_i^k]} \in (X_i^k)^{\mathcal{I}_1}, \qquad \text{for } k = 0, 1.$$

If $\mathsf{Q}_i = \forall$ then in $\mathcal{G}_2$ we have

$$\mathsf{tail}(\pi) \rightsquigarrow_2 w^2_{[Q_i^k]} \in (X_i^k)^{\mathcal{I}_2}, \qquad \text{for } k = 0, 1.$$

So, we set $h(\pi \cdot w^2_{[Q_i^k]}) = h(\pi) \cdot w^1_{[Q_i^k]}$, for both $k = 0, 1$. Clearly, ($\mathfrak{a}$) holds. Otherwise, $\mathsf{Q}_i = \exists$ and in $\mathcal{G}_2$ we have

$$\mathsf{tail}(\pi) \rightsquigarrow_2 w^2_{[S_i]}.$$

We know that $\models \varphi$ and so, by, ($\mathfrak{a}$), the QBF obtained from φ by removing $\mathsf{Q}_1 X_1 \dots \mathsf{Q}_i X_i$ is true under either $\mathfrak{a}_{h(\pi)} \cup \{X_i = \top\}$ or $\mathfrak{a}_{h(\pi)} \cup \{X_i = \bot\}$. We set $h(\pi \cdot w^2_{[S_i]}) = h(\pi) \cdot w^1_{[Q_i^k]}$ with $k = 1$ in the former case and $k = 0$ in the latter case. Either way, ($\mathfrak{a}$) holds.

Consider now a path π from $a^{\mathcal{I}_2}$ to w^2_n in $\mathcal{G}_2$. By construction, we have

$$h(\pi) = a^{\mathcal{I}_1} w^1_{[Q_1^{k_1}]} \dots w^1_{[Q_n^{k_n}]}.$$

On the one hand, the path π in $\mathcal{G}_2$ has m infinite extensions of the form $\pi w^2_{[P_j]} w^2_{[P_j]} \dots$, for $1 \leq j \leq m$. On the other hand, as $\models \varphi$, by ($\mathfrak{a}$), for each clause C_j, there is some $1 \leq i_j \leq n$ such that $h(\pi)$ contains either $w^1_{[Q^1_{i_j}]}$ with $X_{i_j} \in C_j$ or $w^1_{[Q^0_{i_j}]}$ with $\neg X_{i_j} \in C_j$. We set, for each $1 \leq k \leq n - i_j$,

$$h(\pi \underbrace{w^2_{[P_j]} \cdots w^2_{[P_j]}}_{k \text{ times}}) = a^{\mathcal{I}_1} w_1 \dots w_{n-k},$$

and, for each $k > n - i_j$,

$$h(\pi \underbrace{w^2_{[P_j]} \cdots w^2_{[P_j]}}_{k \text{ times}}) = a^{\mathcal{I}_1} w_1 \dots w_{i_j} \underbrace{w^1_{[R_j]} \dots w^1_{[R_j]}}_{k-(n-i_j) \text{ times}},$$

where $w_i = w^1_{[Q_i^{k_i}]}$, i.e., the $(i + 1)$ element of $h(\pi)$. It is straightforward to check that h is a Σ-homomorphism from $\mathcal{M}_2$ to $\mathcal{M}_1$.

($\Rightarrow$) Suppose now that h is a Σ-homomorphism from a part of $\mathcal{M}_2$ that contains all points of distance $\leq 2n+1$ from the root $a^{\mathcal{I}_2}$ to $\mathcal{M}_1$, where as before $\mathcal{M}_i$ is the canonical model and $\mathcal{G}_i$ is the generating model of $(\mathcal{T}_i, \{A(a)\})$, $i = 1, 2$. We have to show that $\models \varphi$.

Let π be an R-path of length $n + 1$ from $a^{\mathcal{I}_2}$ in $\mathcal{G}_2$ (i.e., a path from $a^{\mathcal{I}_2}$ to w) and let $X_1^{k_1}, X_2^{k_2}, \dots, X_n^{k_n}$ be the concepts containing some nodes on $h(\pi)$. We show that, for every $1 \leq j \leq m$, the clause C_j contains at least one of the literals

$$\{X_i \mid k_i = 1, 1 \leq i \leq n\} \cup \{\neg X_i \mid k_i = 0, 1 \leq i \leq n\}.$$

The path π ends in w. So, for each $1 \leq j \leq m$, consider now the path

$$\pi \underbrace{w^2_{[P_j]} w^2_{[P_j]}}_{n+1 \text{ times}}$$

in $\mathcal{G}_2$. It should be clear that its h-image in $\mathcal{M}_1$ can only be of the form $\sigma w^1_{[Q_i^k]} w^1_{[R_j]} w^1_{[R_j]} \dots w^1_{[R_j]}$, for $k = 0$ or $k = 1$. If $k = 0$ then C_j must contain $\neg X_i$, otherwise X_i. $\square$

Proof of Theorem 12

In this section, we show that deciding Σ-query-entailment between *DL-Lite*$^{\mathcal{H}}_{core}$ TBoxes is in ExpTime. Fix Σ and $\mathcal{K}_i = (\mathcal{T}_i, \{B(a)\})$, $i = 1, 2$, where B is a Σ-concept. It is sufficient to show that it is decidable in exponential time whether $\mathcal{M}_{\mathcal{K}_2}$ is finitely Σ-homomorphically embeddable into $\mathcal{M}_{\mathcal{K}_1}$. This can be proved by a reduction to the emptiness problem for alternating two-way automata, which is in ExpTime (Vardi 1998). The simplest way of sketching the proof is indirect, via a reduction to the validity of guarded fixpoint logic of finite width which has been shown to be decidable is ExpTime using alternating two-way automata in (Grädel and Walukiewicz 1999). In our encoding in the guarded fixpoint logic of finite width, we do not use first-order syntax but the syntax of the extension $\mathcal{ALCI}_\nu^\sqcap$ of *DL-Lite*$^{\mathcal{H}}_{core}$, i.e., the standard description logic $\mathcal{ALC}$ extended with inverse roles, role inclusions, intersections of roles, and greatest simultaneous fixpoints. $\mathcal{ALCI}_\nu^\sqcap$ TBoxes are easily translated to guarded fixpoint logic of finite width.

Thus, we use concept variables $X_1, \ldots$ that can be used like concept names, we have simultaneous fixpoints $\nu_i X_1 \cdots X_m.C_1 \cdots C_m$, where $1 \leq i \leq m$, and we can construct, for a finite set Γ of roles, the concept $\exists\Gamma.C$ which is interpreted by setting $d \in (\exists\Gamma.C)^\mathcal{I}$ iff there exists $d' \in C^\mathcal{I}$ with $(d, d') \in R^\mathcal{I}$ for all $R \in \Gamma$. The semantics of the greatest fixpoint constructor is as follows, where $\mathcal{V}$ is an *assignment* that maps concept variables to subsets of $\Delta^\mathcal{I}$ and $\mathcal{V}[X \mapsto W]$ denotes $\mathcal{V}$ modified by setting $\mathcal{V}(X) = W$. Then $(\nu_i X_1 \cdots X_m.C_1 \cdots C_m)^{\mathcal{I},\mathcal{V}}$ is interpreted as

$$\bigcup\{W_i \mid \quad \exists W_1, \ldots, W_{i-1}, W_{i+1}, \ldots, W_m \text{ s.t. for } 1 \leq j \leq m: \; W_j \subseteq C_j^{\mathcal{I},\mathcal{V}[X_1 \mapsto W_1, \ldots, X_m \mapsto W_m]}\}$$

For $\mathcal{I}_{\mathcal{K}_2}$ and $\mathcal{M}_{\mathcal{K}_2}$ we use the following notation. We assume that $\Delta^{\mathcal{I}_{\mathcal{K}_2}} = \{1, \ldots, n\}$ with $a = 1$. We set $i \rightsquigarrow^\Sigma j$ iff $i \rightsquigarrow_{\mathcal{K}_2} j$ and there exists $R \in \Sigma$ with $(i, j) \in R^{\mathcal{I}_2}$. Let $D \subseteq \Delta^{\mathcal{I}_{\mathcal{K}_2}}$ be the set of all $m \in \Delta^{\mathcal{I}_{\mathcal{K}_2}}$ such that there does not exist k with $k \rightsquigarrow^\Sigma m$.

For $m \in D$, denote by $\mathcal{I}_m$ the Σ-reduct of the interpretation induced by the set of points reachable from m along $\rightsquigarrow^\Sigma$-paths in $\mathcal{I}_{\mathcal{K}_2}$.

Similarly, call π' a Σ-son of π in $\mathcal{M}_{\mathcal{K}_2}$, in symbols $\pi \rightsquigarrow^\Sigma \pi'$, if $\pi' = \pi \cdot w_{[R]}$ for some R and $(\pi, \pi') \in S^{\mathcal{M}_{\mathcal{K}_2}}$ for some $S \in \Sigma$. Denote by $\mathcal{N}_m$ the Σ-reduct of the interpretation induced by the set of points reachable from some (fixed) $\pi \in \Delta^{\mathcal{M}_{\mathcal{K}_2}}$ with $\mathsf{tail}(\pi) = m$ along $\rightsquigarrow^\Sigma$. Note that, up to isomorphims, $\mathcal{N}_m$ does not depend on π, but on m only. We denote the root of $\mathcal{N}_m$ by π_m.

Clearly, if is sufficient to decide whether, for every $m \in D$, $\mathcal{N}_m$ is finitely Σ-homomorphically embeddable into $\mathcal{M}_{\mathcal{K}_1}$. Recall that, by Example 6 this is not equivalent to having a Σ-homomorphism from each $\mathcal{N}_m$ to $\mathcal{M}_{\mathcal{K}_1}$. However, as the image of $a \in \Delta^{\mathcal{N}_1}$ is fixed as a, for $\mathcal{N}_1$ finite Σ-homomorphic embeddability is equivalent to the existence of a Σ-homomorphism from $\mathcal{N}_1$, by König's Lemma. For the remaining $\mathcal{N}_m$, we require a slightly more general condition.

A partial Σ-homomorphism $h : \mathcal{N}_m \rightarrow \mathcal{M}_{\mathcal{K}_1}$ is e-complete, for some $e \in \Delta^{\mathcal{M}_{\mathcal{K}_1}}$, if

- $h(\pi_m)$ is defined;
- the range of h is contained in the set of points that are $\rightsquigarrow^\Sigma$-reachable from e;
- if $h(\pi)$ is defined and $h(\pi)$ is not a son of e, then $h(\pi')$ is defined, for all sons π' of π.

In $\mathcal{M}_{\mathcal{K}_1}$, we call π' a son of π, in symbols $\pi \rightsquigarrow \pi'$ if $\pi' = \pi \cdot w_{[R]}$ for some R. Now the following can be proved using König's Lemma.

Lemma 20. *(i) $\mathcal{N}_1$ is finitely Σ-homomorphically embeddable into $\mathcal{M}_{\mathcal{K}_1}$ iff there exists a Σ-homomorphism $h : \mathcal{N}_1 \rightarrow \mathcal{M}_{\mathcal{K}_1}$ (note that $h(a) = a$).*

(ii) Let $m \in D \setminus \{a\}$. $\mathcal{N}_m$ is finitely Σ-homomorphically embeddable into $\mathcal{M}_{\mathcal{K}_1}$ iff at least one of the following conditions holds:

- *there exists a $\rightsquigarrow$-path p in $\mathcal{M}_{\mathcal{K}_1}$ through some e such that for infinitely many $\sigma \in p$ there there exists a partial Σ-homomorphism $h_\sigma : \mathcal{N}_m \rightarrow \mathcal{M}_{\mathcal{K}_1}$ with $h_\sigma(\pi_m)) = \sigma$ that is e-complete.*
- *there exists a Σ-homomorphism $h : \mathcal{N}_m \rightarrow \mathcal{M}_{\mathcal{K}_1}$.*

We now encode the existence of such Σ-homomorphisms into $\mathcal{ALCIH}_\nu^\sqcap$. Let, for $1 \leq i, j \leq n$ with $i \rightsquigarrow^\Sigma j = w_S$, $\Gamma_{i,j}$ denote the set of Σ-roles R with $\mathcal{T}_2 \models S \sqsubseteq R$. Let X_i be concept variables and set

$$C_i = (\bigsqcap_{A \in \Sigma, i \in A^{\mathcal{I}_{\mathcal{K}_2}}} A) \sqcap (\bigsqcap_{i \rightsquigarrow^\Sigma j} \exists\Gamma_{i,j}.X_j)$$

For every $i \in D$ let $\vec{X}_i$ and $\vec{C}_i$ denote the subsequences $X_{i_1}, \ldots, X_{i_m}$ and $C_{i_1}, \ldots, C_{i_m}$ of $X_1, \ldots, X_n$ and $C_1, \ldots, C_n$, respectively, with $i_j \in \Delta^{\mathcal{I}_i}$ for $1 \leq j \leq m$. Now let

$$C(\mathcal{I}_i) = \nu_i \vec{X}_i.\vec{C}_i.$$

We first encode part (i) of Lemma 20:

Lemma 21. *There exists a Σ-homomorphism $h : \mathcal{N}_1 \rightarrow \mathcal{M}_{\mathcal{K}_1}$ (with $h(a) = a$) iff $\mathcal{T}_1 \models B \sqsubseteq C(\mathcal{I}_i)$.*

Proof. Assume first that $\mathcal{T}_1 \models B \sqsubseteq C(\mathcal{I}_1)$. $\mathcal{M}_{\mathcal{K}_1}$ is a model of $\mathcal{T}_1$ satisfying B in a. Thus, $1 = a \in C(\mathcal{I}_1)^{\mathcal{M}_{\mathcal{K}_1}}$. Take $\mathcal{V}_1$ with $a \in \mathcal{V}_1(X_1)$ and $\mathcal{V}_1(X_j) \subseteq C_j^{\mathcal{M}_{\mathcal{K}_1},\mathcal{V}_1}$ for all $j \in \Delta^{\mathcal{I}_1}$.

Now we define a Σ-homomorphism $f : \mathcal{N}_1 \rightarrow \mathcal{M}_{\mathcal{K}_1}$ in such a way that for all $\pi \in \Delta^{\mathcal{N}_1}$ with $\mathsf{tail}(\pi) = j$ we have $f(\pi) \in \mathcal{V}_1(X_j)$. First set $f(a) = a \in \mathcal{V}_1(X_1)$.

Assume now that $\pi \in \Delta^{\mathcal{N}_1}$, $f(\pi)$ has not yet been defined, $\mathsf{tail}(\pi) = k$, $\pi_{pre} \rightsquigarrow^\Sigma \pi$, and that $f(\pi_{pre})$ has been defined such that $f(\pi_{pre}) \in \mathcal{V}_1(X_j)$ for $\mathsf{tail}(\pi_{pre}) = j$ and $j \in \mathcal{I}_1$. Then $f(\pi_{pre}) \in C_j^{\mathcal{M}_{\mathcal{K}_1},\mathcal{V}_1}$. Thus, $f(\pi_{pre}) \in (\exists\Gamma_{j,k}.X_k)^{\mathcal{M}_{\mathcal{K}_1},\mathcal{V}_1}$ and so we can define $f(\pi)$ as a member of $\mathcal{V}_1(X_k)$ with $(\pi_{pre}, \pi) \in R^{\mathcal{M}_{\mathcal{K}_1}}$ for all $R \in \Gamma_{j,k}$.

It is readily checked that f is a Σ-homomorphism.

Conversely, let $f : \mathcal{N}_1 \rightarrow \mathcal{M}_{\mathcal{K}_1}$ we a Σ-homomorphism with $f(a) = a$. Let $\mathcal{I} \models \mathcal{T}_1$ and assume $d \in B^\mathcal{I}$. We have a Σ-homomorphism $g : \mathcal{M}_1 \rightarrow \mathcal{I}$ with $g(a) = d$.

Let, for $j \in \Delta^{\mathcal{I}_1}$,

$$\mathcal{V}_1(X_j) = \{g(f(\pi)) \mid \mathsf{tail}(\pi) = j\}.$$

We show that $\mathcal{V}_1(X_j) \subseteq C_j^{\mathcal{I},\mathcal{V}_1}$ for all $j \in \Delta^{\mathcal{I}_1}$. From $gf(1) = gf(a) = d$ we then obtain $d \in C(\mathcal{I}_1)^\mathcal{I}$, as required. Let $e \in \mathcal{V}_1(X_j)$. Then $e = g(f(\pi))$ for some π with $\mathsf{tail}(\pi) = j$. As gf is a Σ-homomorphism, $e \in A^\mathcal{I}$ for all $A \in \Sigma$ with $\pi \in A^{\mathcal{N}_1}$. From $\mathsf{tail}(\pi) \in A^{\mathcal{I}_{\mathcal{K}_2}}$ iff $\pi \in A^{\mathcal{N}_1}$ we obtain $e \in A^\mathcal{I}$ for all $A \in \Sigma$ with $j \in A^{\mathcal{I}_{\mathcal{K}_2}}$. Now assume $j \rightsquigarrow^\Sigma j'$. Then $(\pi, \pi \cdot j') \in R^{\mathcal{N}_1}$ for all $R \in \Gamma_{j,j'}$. Thus $(gf(\pi), gf(\pi \cdot j')) \in R^\mathcal{I}$. By definition, $gf(\pi \cdot j') \in \mathcal{V}_1(X_{j'})$. Thus $e \in (\exists\Gamma_{j,j'}.X_{j'})^{\mathcal{I},\mathcal{V}_1}$, as required. $\square$

To encode Condition (ii) of Lemma 20, we have to recognize $\rightsquigarrow$-paths. To this end, we consider a fresh role ρ and add to $\mathcal{T}_1$ the inclusions

$$B \sqsubseteq \exists(\Gamma_S \cup \{\rho\}).\exists S^-,$$

for every $w_{[S]} \in \Delta^{\mathcal{I}_1}$ with $a \rightsquigarrow w_{[S]}$ and $R \in \Gamma_S$ iff $(a, w_{[S]}) \in R^{\mathcal{I}_{\mathcal{K}_1}}$, and

$$\exists R^- \sqsubseteq \exists(\Gamma_{R,S} \cup \{\rho\}).\exists S^-$$

for every $w_{[S]}, w_{[R]} \in \Delta^{\mathcal{I}_1}$ with $w_{[R]} \rightsquigarrow w_{[S]}$ and $R \in \Gamma_{R,S}$ iff $(w_{[R]}, w_{[S]}) \in R^{\mathcal{I}_{\mathcal{K}_1}}$. Call the resulting TBox $\mathcal{T}_1'$.

By ρ^* we denote the transitive reflexive closure of the role ρ. Then $\exists\rho^*.C$ has the standard PDL semantics: $d \in (\exists\rho^*.C)^{\mathcal{I}}$ iff there exists a ρ-path from d to some d' with $d' \in C^{\mathcal{I}}$. Clearly, $\exists\rho^*.C$ can be expressed in the guarded fixpoint logic of finite width. The following lemma can be proved in the same way as Lemma 21.

Lemma 22. *Let $i \in D \setminus \{a\}$. The following conditions are equivalent:*

- *There exists a Σ-homomorphism h from $\mathcal{N}_i$ to $\mathcal{M}_1$*
- $\mathcal{T}_1' \models B \sqsubseteq \exists\rho^*.C(\mathcal{I}_i)$

So far, we have encoded the existence of Σ-homomorphisms. Now we encode e-complete partial Σ-homomorphisms. It remains to encode the first condition of (ii) in Lemma 20. Let P be a fresh concept name. We use P to denote the point e at which an e-complete partial Σ-homomorphism does not have to be "expanded" further. Define

$$C_i(P) = (\bigsqcap_{A\in\Sigma, i\in A^{\mathcal{I}_{\mathcal{K}_2}}} A) \sqcap \bigsqcap_{i \rightsquigarrow^{\Sigma} j} \exists\Gamma_{i,j}.(X_j \sqcup P).$$

For every $i \in D$ let $\vec{C}_i(P)$ denote the subsequence $C_{i_1}(P), \dots, C_{i_m}(P)$ of $C_1, \dots, C_n$ with $i_j \in \Delta^{\mathcal{I}_i}$ for $1 \leq j \leq m$. Now let

$$C_P(\mathcal{I}_i) = \nu_i \vec{X}_i.\vec{C}_i(P).$$

Now the following can be proved in a straightforward way.

Lemma 23. *Let $m \in D \setminus \{a\}$. The following conditions are equivalent:*

- *there exists a $\rightsquigarrow$-path p in $\mathcal{M}_{\mathcal{K}_1}$ starting at some σ_0 with* $\mathsf{tail}(\sigma_0) = w_{[R]}$ *such that for infinitely many $\sigma \in p$ there there exists a partial Σ-homomorphism $h_\sigma : \mathcal{N}_m \to \mathcal{M}_{\mathcal{K}_1}$ with $h_\sigma(\pi_m)) = \sigma$ that is σ_0-complete.*
- *We have*
$$\mathcal{T}_1' \models (P \sqcap \exists R^-) \sqsubseteq \gamma(C_i, P)$$
where $\gamma(C_i, P)$ is the guarded fixpoint formula stating "there is a ρ-path p such that $C_P(\mathcal{I}_i)$ holds infinitely often on p.

Lemmas 21, 22, and 23 together give the reduction to TBox reasoning in $\mathcal{ALCIH}_\nu^\sqcap$, as required.

Proof of Theorem 13

Let us turn now to strong Σ-query entailment. Suppose that we are given two *DL-Lite*$_{core}^{\mathcal{H}}$ TBoxes $\mathcal{T}_1$ and $\mathcal{T}_2$ and a signature Σ such that $\mathcal{T}_1$ Σ-query entails $\mathcal{T}_2$. Our task is to find out whether there exists a Σ-TBox $\mathcal{T}$ such that $\mathcal{T}_1 \cup \mathcal{T}$ does not Σ-query entail $\mathcal{T}_2 \cup \mathcal{T}$. In our constructions, we will be using the following operation of gluing models.

Suppose we have two models $\mathcal{I}_1$ and $\mathcal{I}_2$ of a Σ-TBox $\mathcal{T}$ with $\Delta^{\mathcal{I}_1} \cap \Delta^{\mathcal{I}_2} = \emptyset$. A bijection g between subsets of $\Lambda_1 \subseteq \Delta^{\mathcal{I}_1}$ and $\Lambda_2 \subseteq \Delta^{\mathcal{I}_2}$ induces an equivalence relation $\sim$ on $\Delta^{\mathcal{I}_1} \cup \Delta^{\mathcal{I}_2}$ by taking $x \sim y$ in case $x = y$ or $y = g(x)$ or $x = g(y)$. We call these interpretations *$\mathcal{T}$-compatible w.r.t.* such an equivalence relation $\sim$ if

$$\mathcal{T} \not\models B_1 \sqcap B_2 \sqsubseteq \bot, \text{ for all } B_i \in \boldsymbol{t}_\Sigma^{\mathcal{I}_i}(x_i), \;\; i = 1, 2,$$
$$\mathcal{T} \not\models R_1 \sqcap R_2 \sqsubseteq \bot, \text{ for all } R_i \in \boldsymbol{r}_\Sigma^{\mathcal{I}_i}(x_i, y_i), \;\; i = 1, 2,$$

for all $x_1, y_1 \in \Lambda_1$ and $x_2, y_2 \in \Lambda_2$ with $x_1 \sim x_2$ and $y_1 \sim y_2$. Denote by $[x]$ the $\sim$-equivalence class of x. Then $\mathcal{I}$ is said to be the *glueing of $\mathcal{I}_1$ and $\mathcal{I}_2$ along $\sim$* if

$$\begin{aligned}
\Delta^{\mathcal{I}} &= \{[x] \mid x \in \Delta^{\mathcal{I}_1} \cup \Delta^{\mathcal{I}_2}\}, \\
a^{\mathcal{I}} &= [a^{\mathcal{I}_1}], \\
A^{\mathcal{I}} &= \{[x] \mid x \in A^{\mathcal{I}_1} \cup A^{\mathcal{I}_2}\}, \\
P^{\mathcal{I}} &= \{([x],[y]) \mid (x,y) \in P^{\mathcal{I}_1} \cup P^{\mathcal{I}_2}\}.
\end{aligned}$$

In particular, we have $x \in A^{\mathcal{I}}$ iff $x \in A^{\mathcal{I}_i} \setminus \Lambda_i$ for $i = 1$ or $i = 2$ or $x \in \Lambda_1$ with either $x \in A^{\mathcal{I}_1}$ or $g(x) \in A^{\mathcal{I}_2}$.

Lemma 24. *Let $\mathcal{I}$ be the glueing along $\sim$ of $\mathcal{T}$-compatible models $\mathcal{I}_1$ and $\mathcal{I}_2$ of $\mathcal{T}$. Then $\mathcal{I} \models \mathcal{T}$.*

Proof. Here we use the form of the axioms in *DL-Lite*$_{core}^{\mathcal{H}}$. If we glue two points from different models, then all axioms of the form $B_1 \sqsubseteq B_2$ or $R_1 \sqsubseteq R_2$ will be satisfied. Only axioms of the form $B_1 \sqcap B_2 \sqsubseteq \bot$ or $R_1 \sqcap R_2$ may become false in the glueing, but that would mean that $\mathcal{I}_1$ and $\mathcal{I}_2$ are not $\mathcal{T}$-compatible w.r.t. $\sim$. $\square$

Theorem 25. *Suppose that $\mathcal{T}_1$ Σ-query entails $\mathcal{T}_2$, for DL-Lite$_{core}^{\mathcal{H}}$ TBoxes $\mathcal{T}_1$, $\mathcal{T}_2$ and a signature Σ. Then $\mathcal{T}_1$ does not strongly Σ-query entail $\mathcal{T}_2$ iff there exist a Σ-TBox $\mathcal{T}$ and a Σ-concept C such that $(\mathcal{T}_1 \cup \mathcal{T}, \{C(a)\})$ is consistent but $(\mathcal{T}_2 \cup \mathcal{T}, \{C(a)\})$ is inconsistent.*

Proof. The implication $(\Leftarrow)$ is trivial. To prove the converse, suppose that, for any Σ-TBox $\mathcal{T}$ and Σ-concept C, if $(\mathcal{T}_1 \cup \mathcal{T}, \{C(a)\})$ is consistent then $(\mathcal{T}_2 \cup \mathcal{T}, \{C(a)\})$ is consistent as well.

It suffices to show that, for all Σ-inclusions α and Σ-concepts C, if $\mathcal{K}_{1\alpha} = (\mathcal{T}_1 \cup \{\alpha\}, \{C(a)\})$ is consistent and the canonical model $\mathcal{M}_2$ of $\mathcal{K}_2 = (\mathcal{T}_2, \{C(a)\})$ is finitely Σ-homomorphically embeddable in the canonical model $\mathcal{M}_1$ of $\mathcal{K}_1 = (\mathcal{T}_1, \{C(a)\})$, then the canonical model $\mathcal{M}_{2\alpha}$ of $\mathcal{K}_{2\alpha} = (\mathcal{T}_2 \cup \{\alpha\}, \{C(a)\})$ is finitely Σ-homomorphically embeddable in the canonical model $\mathcal{M}_{1\alpha}$ of $\mathcal{K}_{1\alpha}$ as well.

For any of our canonical models $\mathcal{M}$ in this proof (with a single named individual a) and $n > 0$, we denote by $t_n(\mathcal{M})$ the sub-interpretation consisting of all points of distance $\leq$

n from $a^{\mathcal{M}}$ (we note that such $\mathcal{M}$ are connected in the sense that every point has a uniquely defined distance from the root). Our aim is to show that, for each $n > 0$, there is a Σ-homomorphism from $t_n(\mathcal{M}_{2\alpha})$ into $\mathcal{M}_{1\alpha}$. We do this by 'lifting' Σ-homomorphisms from $t_n(\mathcal{M}_2)$ into $\mathcal{M}_1$. The construction depends on the form of α.

Case 1: $\alpha = A \sqcap B \sqsubseteq \bot$ or $\alpha = R \sqcap S \sqsubseteq \bot$. In this case, $\mathcal{M}_i = \mathcal{M}_{i\alpha}$ and the Σ-homomorphisms coincide.

Case 2: $\alpha = A \sqsubseteq B$. If $A^{\mathcal{M}_2} = \emptyset$ then $\mathcal{M}_{2\alpha} = \mathcal{M}_2$ and again the Σ-homomorphisms coincide.

So, suppose $A^{\mathcal{M}_2} \neq \emptyset$, and so, as $\mathcal{T}_1$ Σ-query entails $\mathcal{T}_2$, we have $A^{\mathcal{M}_1} \neq \emptyset$. As $\mathcal{K}_{i\alpha}$ is consistent, $(\mathcal{T}_i, \{B(b)\})$ is also consistent, for $i = 1, 2$. Denote their canonical models by $\mathcal{M}_i^B$. Since $\mathcal{T}_1$ Σ-query entails $\mathcal{T}_2$, $\mathcal{M}_2^B$ is finitely Σ-embeddable into $\mathcal{M}_1^B$.

Let $h\colon t_n(\mathcal{M}_2) \to \mathcal{M}_1$ be a Σ-homomorphism. Set $\mathcal{M}_1^0 = \mathcal{M}_1$, $\mathcal{M}_2^0 = \mathcal{M}_2$ and $h^0 = h$. We repeat the following procedure until there is $x_2 \in A^{\mathcal{M}_2^k} \setminus B^{\mathcal{M}_2^k}$ (a 'defect') with a distance $\leq n$ from $a^{\mathcal{M}_2^k}$. We select such an x_2 so that no other such point is located at a smaller distance from the root $a^{\mathcal{M}_2^k}$. Define $\mathcal{M}_2^{k+1}$ to be the glueing of $\mathcal{M}_2^k$ and a fresh copy $\mathcal{M}_2^B$ along $x_2 \sim b^{\mathcal{M}_2^B}$, which are clearly $\mathcal{T}_2$-compatible w.r.t. this equivalence relation. So, $\mathcal{M}_2^{k+1} \models \mathcal{T}_2$. (Note, however, that $\mathcal{M}_2^{k+1}$ may contain 'redundant' successors of x witnessing some concepts of the form $\exists R$ despite that there are such R-witnesses in $\mathcal{M}_2^k$.) Also, we define $\mathcal{M}_1^{k+1}$ to be the glueing of $\mathcal{M}_1^k$ and a fresh copy $\mathcal{M}_1^B$ along $h^k(x_2) \sim b^{\mathcal{M}_1^B}$, which are clearly $\mathcal{T}_1$-compatible w.r.t. this equivalence relation. So, $\mathcal{M}_1^{k+1} \models \mathcal{T}_1$. We also define h^{k+1} by extending h^k to the set of new points in the copy of $\mathcal{M}_2^B$ with the help of the Σ-homomorphism from $t_n(\mathcal{M}_2^B)$ into $\mathcal{M}_1^B$.

Since we always select a 'defect' closest to the root, there is a k such that a subtree of $t_n(\mathcal{M}_2^k)$ is Σ-isomorphic to $t_n(\mathcal{M}_{2\alpha})$ (i.e., $\mathcal{M}_2^k$ has no 'defects' up to depth n). On the other hand, h^k is a Σ-homomorphism from $t_n(\mathcal{M}_2^k)$ to $\mathcal{M}_1^k$. So, there is a Σ-homomorphism from $t_n(\mathcal{M}_{2\alpha})$ into $\mathcal{M}_1^k$. Repeating the above procedure for $\mathcal{M}_1^k$ alone, we can extend the Σ-homomorphism from $t_n(\mathcal{M}_{2\alpha})$ to $\mathcal{M}_{1\alpha}$.

Case 3: $\alpha = R \sqsubseteq S$. If $R^{\mathcal{M}_2} = \emptyset$ then $\mathcal{M}_{2\alpha} = \mathcal{M}_2$ and the Σ-homomorphisms coincide.

So, suppose $R^{\mathcal{M}_2} \neq \emptyset$, and so, as $\mathcal{T}_1$ Σ-query entails $\mathcal{T}_2$, $R^{\mathcal{M}_1} \neq \emptyset$. As the $\mathcal{K}_{i\alpha}$ are consistent, $(\mathcal{T}_i, \{S(b, c)\})$ is also consistent. Denote their canonical models by $\mathcal{M}_i^S$. Since $\mathcal{T}_1$ Σ-query entails $\mathcal{T}_2$, $\mathcal{M}_2^S$ is finitely Σ-homomorphically embeddable into $\mathcal{M}_1^S$.

As in the previous case, for each n, we can construct $\mathcal{M}_1^k$, $\mathcal{M}_2^k$ and h^k such that a subtree of $t_n(\mathcal{M}_2^k)$ is Σ-isomorphic to $t_n(\mathcal{M}_{2\alpha})$ and h^k is a Σ-homomorphism from $t_n(\mathcal{M}_2^k)$ to $\mathcal{M}_1^k$ (this time though, defects are pairs $(x, y) \in R^{\mathcal{M}_2^k} \setminus S^{\mathcal{M}_2^k}$, which are 'cured' by glueing $\mathcal{M}_i^k$ and a fresh copy of $\mathcal{M}_i^S$ along $x \sim b^{\mathcal{M}_i^S}$ and $y \sim c^{\mathcal{M}_i^S}$). The rest of the proof is the same as in Case 2. □

We show now that, if $\mathcal{T}_1$ does not strongly Σ-query entail $\mathcal{T}_2$, then this can be detected using a Σ-TBox $\mathcal{T}$ with a single axiom.

Theorem 13. *Suppose that $\mathcal{T}_1$ Σ-query entails $\mathcal{T}_2$. Then $\mathcal{T}_1$ does not strongly Σ-query entail $\mathcal{T}_2$ iff there is a Σ-concept C and Σ-TBox $\mathcal{T}$ with a single inclusion of the form $B_1 \sqsubseteq B_2$ or $R_1 \sqsubseteq R_2$ such that $(\mathcal{T}_1 \cup \mathcal{T}, \{C(a)\})$ is consistent but $(\mathcal{T}_2 \cup \mathcal{T}, \{C(a)\})$ is inconsistent.*

Proof. For any TBox $\mathcal{T}$, denote by $\sqsubseteq_{\mathcal{T}}^*$ the reflexive and transitive closure of $\{(R_1, R_2), (R_1^-, R_2^-) \mid R_1 \sqsubseteq R_2 \in \mathcal{T}\}$, where $R^- = P$ if $R = P^-$ and $R^- = P^-$ if $R = P$, for a role name P. We define an analogous relation on concepts: $\sqsubseteq_{\mathcal{T}}^*$ is the reflexive and transitive closure of

$$\{(B, \top) \mid B \text{ a concept in } \mathcal{T}\} \cup \{(B_1, B_2) \mid B_1 \sqsubseteq B_2 \in \mathcal{T}\} \cup \{(\exists R_1, \exists R_2) \mid R_1 \sqsubseteq_{\mathcal{T}}^* R_2\}.$$

We say that a concept C is locally $\mathcal{T}$-consistent if the following three conditions hold:

- $C \not\sqsubseteq_{\mathcal{T}}^* \bot$,
- there is no B_1, B_2 with $C \sqsubseteq_{\mathcal{T}}^* B_i$ for $i = 1, 2$ and $B_1 \sqcap B_2 \sqsubseteq \bot$,
- there is no R, R_1, R_2 with $C \sqsubseteq_{\mathcal{T}}^* \exists R$, $R \sqsubseteq_{\mathcal{T}}^* R_i$ for $i = 1, 2$ and $R_1 \sqcap R_2 \sqsubseteq \bot$.

Otherwise, C is said to be locally $\mathcal{T}$-inconsistent.

Suppose $(\mathcal{T}_1 \cup \mathcal{T}, C(a))$ is consistent, and so is $(\mathcal{T}_2, C(a))$, but $(\mathcal{T}_2 \cup \mathcal{T}, C(a))$ is not. Then either C is locally $\mathcal{T}_2 \cup \mathcal{T}$-inconsistent or there is a sequence of roles $R_1, \ldots, R_n$ such that

- $C \sqsubseteq_{\mathcal{T}_2 \cup \mathcal{T}}^* \exists R_1$,
- $\exists R_i^- \sqsubseteq_{\mathcal{T}_2 \cup \mathcal{T}}^* \exists R_{i+1}$ and $\exists R_i^-$ is locally $\mathcal{T}_2 \cup \mathcal{T}$-consistent, for all $1 \leq i < n$,
- $\exists R_n^-$ is locally $\mathcal{T}_2 \cup \mathcal{T}$-inconsistent.

Denote $\exists R_n^-$ by D in the latter case and C by D in the former case. We know that D is locally $\mathcal{T}_2 \cup \mathcal{T}$-inconsistent. In the latter case without loss of generality we can even assume that

- there is no Σ-concept B with $C \sqsubseteq_{\mathcal{T}_2 \cup \mathcal{T}}^* B \sqsubseteq_{\mathcal{T}_2 \cup \mathcal{T}}^* \exists R_1$ and $B \not\sqsubseteq_{\mathcal{T}_2 \cup \mathcal{T}}^* C$,
- for all $1 \leq i < n$, there is no Σ-concept B with $\exists R_i^- \sqsubseteq_{\mathcal{T}_2 \cup \mathcal{T}}^* B \sqsubseteq_{\mathcal{T}_2 \cup \mathcal{T}}^* \exists R_{i+1}$;

for if it is not the case, we could start from $(\mathcal{T}_2 \cup \mathcal{T}, B(a))$. It follows that none of the axioms from $\mathcal{T}$ is involved in 'deriving' the $\exists R_i$ an so,

$$C \sqsubseteq_{\mathcal{T}_2}^* \exists R_1 \text{ and } \exists R_i^- \sqsubseteq_{\mathcal{T}_2}^* \exists R_{i+1}, \text{ for } 1 \leq i < n.$$

Therefore, the canonical model of $(\mathcal{T}_2, C(a))$ must contain a point in $\exists R_n^-$. So, in either case (i.e., whether D is C or $\exists R_n^-$) the canonical model of $(\mathcal{T}_2, C(a))$ contains a point in D (thus, D is $\mathcal{T}_2$-consistent). So, we need to pin-point a single axiom that is a reason for local $\mathcal{T}_2 \cup \mathcal{T}$-inconsistency of D. Consider all possible cases:

Case 1: Suppose there are B_1, B_2 such that $D \sqsubseteq^*_{\mathcal{T}_2\cup\mathcal{T}} B_i$, for $i = 1, 2$, and $B_1 \sqcap B_2 \sqsubseteq \bot \in \mathcal{T}_2 \cup \mathcal{T}$. Suppose further that, for both $i = 1, 2$, we have Σ-concepts $B_i^1, \dots, B_i^n$ with

$$D \sqsubseteq^*_{\mathcal{T}_2} B_i^1 \sqsubseteq^*_{\mathcal{T}} B_i^2 \sqsubseteq^*_{\mathcal{T}_2} B_i^3 \sqsubseteq^*_{\mathcal{T}} \cdots \sqsubseteq^*_{\mathcal{T}} B_i^n \sqsubseteq^*_{\mathcal{T}_2} B_i;$$

that is, B_i^1 is a minimal and B_i^n a maximal (w.r.t. $\sqsubseteq^*_{\mathcal{T}_2\cup\mathcal{T}}$) Σ-concepts between D and B_i and each consecutive B_i^{j+1} is derived from B_i^j by axioms in either $\mathcal{T}_2$ or $\mathcal{T}$. As $\mathcal{T}_1$ Σ-query entails $\mathcal{T}_2$, by **(h)**, the canonical model of $(\mathcal{T}_1, C(a))$ has a point in both B_1^1 and B_2^1. It also follows that the canonical model of $(\mathcal{T}_1 \cup \mathcal{T}, C(a))$ has a point in both B_1^1 and B_2^1. On the other hand, for each $1 \le j < n$, the inclusion $B_i^j \sqsubseteq B_i^{j+1}$ is a consequence of either $\mathcal{T}_2$ or $\mathcal{T}$. The latter case is trivial and in the former case, as $\mathcal{T}_1$ Σ-query entails $\mathcal{T}_2$, by **(p)**, we have $\mathcal{T}_1 \models B_i^j \sqsubseteq B_i^{j+1}$. Summing up, we have $\mathcal{T}_1 \cup \mathcal{T} \models B_i^1 \sqsubseteq B_i^n$ and so, the canonical model of $(\mathcal{T}_1 \cup \mathcal{T}, C(a))$ has a point in both B_1^n and B_2^n. Now, we have either $B_1 \sqsubseteq B_2 \in \mathcal{T}$ or $B_1 \sqsubseteq B_2 \in \mathcal{T}_2$. In the former case $B_1^n = B_1$ and $B_2^n = B_2$ and thus, $(\mathcal{T}_1 \cup \mathcal{T}, C(a))$ is inconsistent. In the latter case, provided that both both sequences B_1^j and B_2^j exist, we have $\mathcal{T}_2 \models B_1^n \sqcap B_2^n \sqsubseteq \bot$, and so, by **(p)**, $\mathcal{T}_1 \models B_1^n \sqcap B_2^n \sqsubseteq \bot$, whence $(\mathcal{T}_1 \cup \mathcal{T}, C(a))$ is inconsistent. Either way, we arrive at a contradiction provided that both both sequences of Σ-concepts exist. On the other hand, one of the sequences must exist for otherwise D is not $\mathcal{T}_2$-consistent.

Suppose that $D \sqsubseteq^*_{\mathcal{T}_2} B_2$. Let B_1^1 and B_1^n be a minimal and a maximal Σ-concepts as above. We then take $\mathcal{T}' = \{B_1^1 \sqsubseteq B_1^n\}$, which gives us $D \sqsubseteq_{\mathcal{T}_2\cup\mathcal{T}'} B_1$ and so, $(\mathcal{T}_2 \cup \mathcal{T}', C(a))$ is inconsistent. On the other hand, $(\mathcal{T}_1 \cup \mathcal{T}', C(a))$ is clearly consistent because $(\mathcal{T}_1 \cup \mathcal{T}, C(a))$ is consistent and $\mathcal{T}_1 \cup \mathcal{T} \models \mathcal{T}'$.

Case 2: Suppose there is a concept B such that $D \sqsubseteq^*_{\mathcal{T}_2\cup\mathcal{T}} B$ and $B \sqsubseteq \bot \in \mathcal{T}_2 \cup \mathcal{T}$. By the argument of Case 1 with $B_1 = B_2$, we obtain a contradiction if assume that there is a Σ-concept B' such that $D \sqsubseteq_{\mathcal{T}_2} B'$. So, this case is in fact impossible.

Case 3: Suppose there are R, R_1, R_2 such that $D \sqsubseteq_{\mathcal{T}_2\cup\mathcal{T}^*} \exists R$, $R \sqsubseteq^*_{\mathcal{T}_2\cup\mathcal{T}} R_i$, for $i = 1, 2$ and $R_1 \sqcap R_2 \sqsubseteq \bot \in \mathcal{T}_2 \cup \mathcal{T}$. The argument of Case 1 is repeated for roles R_1 and R_2. □

Proof of Theorem 15

In this section, we prove the following

Theorem 15 Let $\mathcal{T}_1, \mathcal{T}_2$ be *DL-Lite*$^{\mathcal{H}}_{core}$ TBoxes and Σ a signature. If **(y)** holds, then $\mathcal{T}_1$ Σ-query entails $\mathcal{T}_2$. If **(n)** holds, then $\mathcal{T}_1$ does not Σ-query entail $\mathcal{T}_2$.

Proof. By Theorem 8 it suffices to show that

1. If $\mathcal{M}^B_{\mathcal{T}_2}$ is finitely Σ-homomorphically embeddable into $\mathcal{M}^B_{\mathcal{T}_1}$ then there exists a Σ-simulation of $\mathcal{G}^B_{\mathcal{T}_2}$ in $\mathcal{G}^B_{\mathcal{T}_1}$;
2. If there exists a forward simulation of $\mathcal{G}^B_{\mathcal{T}_2}$ in $\mathcal{G}^B_{\mathcal{T}_1}$ then there exists a Σ-homomorphism from $\mathcal{M}^B_{\mathcal{T}_2}$ to $\mathcal{M}^B_{\mathcal{T}_1}$.

Let for $i = 1, 2$, $\mathcal{G}_i = (\mathcal{I}_i, \rightsquigarrow_i)$ denote $\mathcal{G}^B_{\mathcal{T}_i}$ and $\mathcal{M}_i$ denote $\mathcal{M}^B_{\mathcal{T}_i}$.

We prove 1 first. For any $\pi \in \Delta^{\mathcal{M}_2}$ and $n \ge 0$ by $\mathcal{M}_2(n, \pi)$ we denote the induced subinterpretation of $\mathcal{M}_2$ with domain $\Delta^{\mathcal{M}_2(n,\pi)} = \{\pi \cdot \pi' \mid |\pi'| \le n\}$. Intuitively, $\mathcal{M}_2(n, \pi)$ is a tree rooted at π unravelled up to level n. Notice that $\mathcal{M}_2(n, \pi)$ is a finite structure for any n and, hence, there exists a Σ-homomorphism from $\mathcal{M}_2(n, \pi)$ to $\mathcal{M}_1$.

For $n \ge 0$ we define now a sequence of relations $\rho_n \subseteq \Delta^{\mathcal{I}_2} \times \Delta^{\mathcal{I}_1}$ as follows: let $(x, x') \in \rho_n$ iff there exist $\pi \in \Delta^{\mathcal{M}_2}$ and a Σ-homomorphism h from $\mathcal{M}_2(n, \pi)$ to $\mathcal{M}_1$ such that

- $x = \mathsf{tail}(\pi)$ and $x' = \mathsf{tail}(h(\pi))$.

It can be readily seen that

- for every n the domain of ρ_n is $\Delta^{\mathcal{I}_2}$ and $(a^{\mathcal{I}_2}, a^{\mathcal{I}_1}) \in \rho_n$;
- $\boldsymbol{t}^{\mathcal{I}_2}_{\Sigma}(x) \subseteq \boldsymbol{t}^{\mathcal{I}_1}_{\Sigma}(x')$, for all $(x, x') \in \rho_n$;
- if $x \rightsquigarrow_{\mathcal{K}_2} w_{[R]}$ and $(x, x') \in \rho_{n+1}$, then there is $y' \in \Delta^{\mathcal{I}_1}$ such that $(w_{[R]}, y') \in \rho_n$ and $S \in \boldsymbol{r}^{\mathcal{I}_1}_{\Sigma}(x', y')$ for every Σ-role S with $[R] \le_{\mathcal{T}_2} [S]$.
- $\rho_{n+1} \subseteq \rho_n$.

Since ρ_n is a relation on finite structures, there must exist N such that $\rho_N = \rho_i$ for all $i > N$. It can be readily seen that ρ_N is a simulation of $\mathcal{G}_2$ in $\mathcal{G}_1$.

We prove now item 2. Let η be a forward simulation of $\mathcal{G}_2$ in $\mathcal{G}_1$. For every $\pi = aw_{[R_1]} \cdots w_{[R_n]} \in \Delta^{\mathcal{M}_2}$ we define $h(\pi)$ by induction on n. The constructed function h has the additional property that $\mathsf{tail}(\pi)\eta h(\mathsf{tail}(\pi))$ for any π.

If $\pi = a^{\mathcal{M}_2}$ set $h(a^{\mathcal{M}_2}) = a^{\mathcal{M}_1}$. Next, suppose that for all $k < n$ and all paths of the form $\tau = aw_{[R_1]} \cdots w_{[R_k]}$ $h(\tau)$ is defined. Consider cases.

Suppose $\boldsymbol{r}^{\mathcal{I}_2}_{\Sigma}(w_{[R_{n-1}]}, w_{[R_n]}) \neq \emptyset$. Let $aw_{[S_1]} \cdots w_{[S_k]} = h(aw_{[R_1]} \cdots w_{[R_{n-1}]})$. Since $w_{[R_{n-1}]}\eta w_{[S_k]}$, η is a forward simulation, and $\boldsymbol{r}^{\mathcal{I}_2}_{\Sigma}(w_{[R_{n-1}]}, w_{R_n}) \neq \emptyset$ there must exist $w_{[S_{k+1}]} \in \Delta^{\mathcal{I}_1}$ such that $w_{[S_k]} \rightsquigarrow_1 w_{[S_{k+1}]}$, $\boldsymbol{r}^{\mathcal{I}_2}_{\Sigma}(w_{[R_{n-1}]}, w_{[R_n]}) \subseteq \boldsymbol{r}^{\mathcal{I}_1}_{\Sigma}(w_{[S_k]}, w_{[S_{k+1}]})$ and $w_{[R_n]}\eta w_{[S_{k+1}]}$. We define $h(\pi)$ as $aw_{[S_1]} \cdots w_{[S_k]} w_{[S_{k+1}]}$, which is a path in $\mathcal{I}_1$.

Suppose now that $\boldsymbol{r}^{\mathcal{I}_2}_{\Sigma}(w_{[R_{n-1}]}, w_{[R_n]}) = \emptyset$. Let $w_{[S_k]} \in \Delta^{\mathcal{I}_1}$ be such that $w_{[R_n]}\eta w_{[S_k]}$. Let π' be an arbitrary path in $\mathcal{I}_1$ such that $\mathsf{tail}(\pi') = w_{[S_k]}$. We define $h(\pi) = \pi'$.

One can see that h is a Σ-homomorphism from $\mathcal{M}_2$ to $\mathcal{M}_1$.

□

Proof of Theorem 16

Direction "⇒" follows from Theorem 15. For "⇐", it suffices to prove the following lemma.

Lemma 26. *Suppose that* (*a*) $\mathcal{T}_1$ *and* $\mathcal{T}_2$ *are DL-Lite*$_{core}$ *TBoxes, or* (*b*) $\mathcal{T}_1 = \emptyset$ *and* $\mathcal{T}_2$ *is a DL-Lite*$^{\mathcal{H}}_{core}$ *TBox. Suppose further that* **(p)** *holds and that, for every* Σ*-concept* B*, there is a* Σ*-simulation of* $\mathcal{G}^B_{\mathcal{T}_2}$ *in* $\mathcal{G}^B_{\mathcal{T}_1}$*. Then* $\mathcal{T}_1$ Σ*-query entails* $\mathcal{T}_2$.

Proof. In case (a), the claim follows from Theorem 9, using **(s1)** and **(s2)** to obtain the required x' associated with $x \in \Delta^{\mathcal{I}^B_{\mathcal{T}_2}}$.

In case (b), let B be a Σ-concept and ρ a Σ-simulation of $\mathcal{G}^B_{\mathcal{T}_2}$ in $\mathcal{G}^B_{\mathcal{T}_1}$. Let $\mathcal{K}_i = (\mathcal{T}_i, B(a))$. We show that $\mathcal{M}_{\mathcal{K}_2}$ is Σ-homomorphically embeddable into $\mathcal{M}_{\mathcal{K}_1}$. If $B = A$, then $\Delta^{\mathcal{I}_{\mathcal{K}_1}} = \{a\}$. Therefore, $\rho(x) = a$ for all $x \in \Delta^{\mathcal{I}_{\mathcal{K}_2}}$, and, whenever $[R] \leq_{\mathcal{T}_2} [S]$ for some $S \in \Sigma$, the role R is not generating, due to **(s3)**. Set $h(\sigma) = a$, for all $\sigma \in \mathcal{M}_{\mathcal{K}_2}$. Then h is a homomorphism because

- $h(a) = a$ trivially;
- $\boldsymbol{t}_\Sigma^{\mathcal{M}_{\mathcal{K}_2}}(\sigma) = \boldsymbol{t}_\Sigma^{\mathcal{I}_{\mathcal{K}_2}}(\mathsf{tail}(\sigma)) \subseteq \boldsymbol{t}_\Sigma^{\mathcal{I}_{\mathcal{K}_1}}(a) = \boldsymbol{t}_\Sigma^{\mathcal{M}_{\mathcal{K}_1}}(h(\sigma))$, due to the construction of $\mathcal{M}_{\mathcal{K}_2}$, **(s2)** and the construction of $\mathcal{M}_{\mathcal{K}_1}$;
- if $S \in \Sigma$ and $S \in \boldsymbol{r}_\Sigma^{\mathcal{M}_{\mathcal{K}_2}}(\sigma_1, \sigma_2)$, then $\sigma_2 = \sigma_1 \cdot w_{[R]}$ for some $[R]$ with $[R] \leq_{\mathcal{T}_2} [S]$ which is a contradiction to R not being generating.

If $B = \exists S$, then $\Delta^{\mathcal{I}_{\mathcal{K}_1}} = \{a, w_{[S]}\}$. Due to **(s3)**, whenever $[R] \leq_{\mathcal{T}_2} [T]$ for some $T \in \Sigma$, then R can only be generating if $T = S$ and $[R] \not\leq_{\mathcal{T}_2} [T']$ for $T' \neq S$. Then, for every $\sigma = \sigma' x w_{[R]}$ with $[R] \leq_{\mathcal{T}_2} [S]$, we have that $\rho(x) = \{a\}$ and $\rho(w_{[R]}) = \{w_{[S]}\}$, where $\rho(z) = \{z' \mid (z, z') \in \rho\}$. We now set $h(\sigma) = a$ if $(\mathsf{tail}(\sigma), a) \in \rho$, and $h(\sigma) = a w_{[R]}$ otherwise. Then h is a homomorphism because

- $h(a) = a$ from **(s1)**;
- $\boldsymbol{t}_\Sigma^{\mathcal{M}_{\mathcal{K}_2}}(\sigma) \subseteq \boldsymbol{t}_\Sigma^{\mathcal{M}_{\mathcal{K}_1}}(h(\sigma))$ follows from **(s2)** as above;
- if $T \in \Sigma$ and $T \in \boldsymbol{r}_\Sigma^{\mathcal{M}_{\mathcal{K}_2}}(\sigma_1, \sigma_2)$, then $\sigma_2 = \sigma_1 \cdot w_{[R]}$ for some $[R]$ with $[R] \leq_{\mathcal{T}_2} [T]$. Due to the above said, $T = S$. Then the observations about ρ and the definition of h entail that $h(\sigma_1) = a$ and $h(\sigma_2) = w_{[S]}$, hence $T = S \in \boldsymbol{r}_\Sigma^{\mathcal{M}_{\mathcal{K}_1}}(h(\sigma_1), h(\sigma_2))$.

$\square$