Relaxing and Restraining Queries for OBDA

Medina Andreşel and **Yazmín Ibáñez-García** and **Magdalena Ortiz** and **Mantas Šimkus**
{andresel,ibanez,ortiz}@kr.tuwien.ac.at | simkus@dbai.tuwien.ac.at
Institute of Logic and Computation, TU Wien, Austria

Abstract

We advocate the use of ontologies for relaxing and restraining queries, so that they retrieve either more or less answers, enabling the exploration of a given dataset. We propose a set of rewriting rules to relax and restrain conjunctive queries (CQs) over datasets mediated by an ontology written in a dialect of DL-Lite with complex role inclusions (CRIs). The addition of CRI enables the representation of knowledge about data involving ordered hierarchies of categories, in the style of multi-dimensional data models. Although CRIs in general destroy the first-order rewritability of CQs, we identify settings in which CQs remain rewritable.

Introduction

In *Ontology-based data access (OBDA)* an ontology provides a conceptual view of a collection of data sources, and can describe knowledge about the domain of interest at a high level of abstraction, using a familiar vocabulary (Poggi et al. 2008). An advantage of having an ontology is that the knowledge can be leveraged to retrieve more complete answers from incomplete data. For example, we consider the following dataset about cultural events and their locations

$$\begin{aligned}\mathcal{A}_e = \{&\mathrm{Concert}(\mathsf{c_1}), \mathrm{CulturEvt}(\mathsf{ev_1}), \mathrm{Exhibition}(\mathsf{ex_1}),\\ &\mathrm{Venue}(\mathsf{StOpera}), \mathrm{Venue}(\mathsf{VkTheater}), \mathrm{City}(\mathsf{Vienna}),\\ &\mathrm{Country}(\mathsf{Austria}), \mathrm{occIn}(\mathsf{c_1}, \mathsf{StOpera}), \mathrm{occIn}(\mathsf{ex_1}, \mathsf{Vienna}),\\ &\mathrm{occIn}(\mathsf{ev_1}, \mathsf{Austria}), \mathrm{locIn}(\mathsf{StOpera}, \mathsf{Vienna}),\\ &\mathrm{locIn}(\mathsf{VkTheater}, \mathsf{Vienna}), \mathrm{locIn}(\mathsf{Vienna}, \mathsf{Austria})\}\end{aligned}$$

and the following ontology that, among other knowledge, says that concerts and exhibitions are cultural events

$$\begin{aligned}\mathcal{T}_e = \{ \quad \mathrm{Concert} \sqsubseteq \mathrm{CulturEvt}, \quad &\mathrm{Country} \sqsubseteq \mathrm{Location},\\ \mathrm{Exhibition} \sqsubseteq \mathrm{CulturEvt}, \quad &\mathrm{City} \sqsubseteq \mathrm{Location},\\ \mathrm{CulturEvt} \sqsubseteq \mathrm{Event}, \quad &\mathrm{Venue} \sqsubseteq \mathrm{Location},\\ \exists \mathrm{occIn} \sqsubseteq \mathrm{Event}, \quad &\exists \mathrm{occIn}^- \sqsubseteq \mathrm{Location}\}.\end{aligned}$$

Using this knowledge, all cultural events ($\mathsf{ex_1}$, $\mathsf{ev_1}$ and $\mathsf{c_1}$) can be retrieved with a simple *conjunctive query (CQ)*:

$$q(x) \leftarrow \mathrm{CulturEvt}(x).$$

In OBDA, ontologies are often written in the Description Logics (DLs) of the *DL-Lite* family (Calvanese et al. 2007). These DLs are tailored towards FO-rewritability of CQs. This means that evaluating a query q over a dataset $\mathcal{A}$ mediated by an ontology $\mathcal{T}$ can be reduced to evaluating a query $q_{\mathcal{T}}$, incorporating knowledge from $\mathcal{T}$, over $\mathcal{A}$ alone, which amounts to standard query evaluation in relational databases. In our example, a rewriting of q is the following query

$$q_{\mathcal{T}_e}(x) \leftarrow \mathrm{CulturEvt}(x) \vee \mathrm{Exhibition}(x) \vee \mathrm{Concert}(x).$$

In this paper, we advocate a novel use of ontologies complementary to OBDA query answering. Namely, we use ontologies to modify queries, relaxing or restraining them, so that they can retrieve either more or less answers over a given dataset. This can help users reflect their information needs and flexibly explore datasets. We build on the observation that query *restrictions* can be obtained using the standard *DL-Lite* rewriting rules (Calvanese et al. 2007). For example, $q_c(x) \leftarrow \mathrm{Concert}(x)$ is a restriction of q, and it occurs as a disjunct in its rewriting $q_{\mathcal{T}_e}$. Moreover, these rewriting rules have natural 'counterparts' that produce *relaxations*. In our running example, if answers to a query for *concerts* are too scarce, one might get more answers by relaxing it to one asking for *all cultural events*. Conversely, if a query for cultural events produces too many answers, it is possible to restrict it to events of a specific type, for instance concerts.

Notably, there are intuitive answers and reformulations that cannot be produced with the standard *DL-Lite* rewriting rules and their counterparts. For example, consider a query retrieving concerts occurring in Vienna:

$$q_2(x) \leftarrow \mathrm{Concert}(x), \mathrm{occIn}(x, y), y = \mathsf{Vienna}.$$

In the presence of standard *DL-Lite* ontologies, there are no answers to q_2 when evaluated over $(\mathcal{T}_e, \mathcal{A}_e)$, although $\mathsf{c_1}$ may be considered an answer to q_2 according to the intuition that *if an event occurs in a venue located in a city, then it occurs in that city*. In order to capture this kind of knowledge, we propose to extend the expressive power of *DL-Lite* with *complex role inclusions* (CRIs). For instance, adding

$$\mathrm{occIn} \cdot \mathrm{locIn} \sqsubseteq \mathrm{occIn} \tag{1}$$

to our example captures the intuition above, and makes $\mathsf{c_1}$ an answer of q_2. The addition of CRIs enables *DL-Lite* to leverage hierarchical knowledge not captured by subclass relations. Indeed, venues, cities, and countries can be seen as different levels of a *dimension* we can call Location. Similarly, a Time dimension could include days, months, and years, while the physical parts of complex objects may be

hierarchically ordered along a Component dimension. *Dimensions* lie at the core of the so-called *multi-dimensional data model* (Hurtado and Mendelzon 2002) used for storing and accessing data at different granularity levels. We show that the addition of CRIs enables *DL-Lite* to leverage dimensional knowledge. Unfortunately, CRIs in DLs are computationally costly: unrestricted they easily lead to undecidability (Horrocks and Sattler 2004), and critically for *DL-Lite*, even one fixed CRI destroys FO-rewritability of CQs. For this reason we devote a section to defining an expressive setting that supports CRIs and enjoys FO-rewritability.

Along with the addition of CRIs, we propose a set of reformulation rules operating not only along the subclass and subrole relations, but also along CRIs. For example we can use (1) to reformulate the query

$$q_3(x) \leftarrow \mathrm{Concert}(x), \mathrm{occIn}(x, y), \mathrm{City}(y),$$

from *all concerts occurring in a city*, to those occurring in some more specific location:

$$q_3'(x) \leftarrow \mathrm{Concert}(x), \mathrm{occIn}(x, z), \mathrm{locIn}(z, y), \mathrm{City}(y).$$

We propose another set of rules that not only use the knowledge from the ontology, but also use instances of concepts and relations, as well as inclusions between concepts guaranteed to hold in the current dataset. For example, we can restrict q_2 to ask for concerts in the State Opera, or relax it to concerts in Austria. Such reformulations are similar to the *drilling down* and *rolling up* operations used for navigating along a dimension. Note that these reformulations are not data independent, but instead rely on the current dataset $\mathcal{A}_e$.

Preliminaries

As usual, $\mathsf{N_C}$, $\mathsf{N_R}$, and $\mathsf{N_I}$ are countable infinite alphabets of *concept*, *role*, and *individual* names, respectively. In what follows, we will use A, A' to denote elements in $\mathsf{N_C}$, s, p, p', elements in $\mathsf{N_R}$ and a, b, elements in $\mathsf{N_I}$. In *DL-Lite*$^{\mathcal{H}}$(Artale et al. 2009), concepts B are built according to the grammar

$$B := A \mid \exists r; \qquad r := p \mid p^-,$$

where p^- is called an *inverse role*. The set of *roles* is defined as $\mathsf{N_R}^{\pm} = \mathsf{N_R} \cup \{p^- \mid p \in \mathsf{N_R}\}$.

We assume w.l.o.g. that a *DL-Lite*$^{\mathcal{H}}$*TBox* (or *ontology*) $\mathcal{T}$ is a finite set of concept inclusion axioms taking the following *normal form* $A \sqsubseteq A'$, $A \sqsubseteq \exists p$, $\exists p \sqsubseteq A$, $p \sqsubseteq s$, $p \sqsubseteq s^-$, together with a set of disjointness axioms of the form $\mathbf{disj}(A, A')$, and $\mathbf{disj}(p, p')$. An *ABox* (or *dataset*) $\mathcal{A}$ is a finite set of assertions $A(a)$, and $p(a, b)$. We denote the set of individuals occurring in $\mathcal{A}$ as $ind(\mathcal{A})$. A knowledge base (KB) is a pair $\mathcal{K} = (\mathcal{T}, \mathcal{A})$. The semantics is defined in terms of interpretations $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ consisting of a non-empty *domain* $\Delta^{\mathcal{I}}$ and an *interpretation function* $\cdot^{\mathcal{I}}$, that complies with the *standard name assumption* i.e., $a^{\mathcal{I}} = a$ for every individual. Satisfaction is defined as usual. For $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ we write $\mathcal{I} \models \mathcal{K}$ if $\mathcal{I}$ satisfies every axiom in $\mathcal{T}$ and every assertion in $\mathcal{A}$, and in that case we say that $\mathcal{I}$ is a *model of* $\mathcal{K}$. We say $\mathcal{K}$ is *satisfiable* if it has a model.

We consider the class of conjunctive queries and unions thereof. A *term* is either an individual name or a variable. A *conjunctive query* (CQ) is a first order formula with free variables $\vec{x}$ and existential variables $\vec{y}$ that takes the form $q(\vec{x}) \leftarrow \varphi(\vec{x}, \vec{y})$, with φ a conjunction of *atoms* of the form $A(x), r(x, y)$, and $t = t'$, where t, t' range over terms. *Instance queries* are CQs with exactly one atom and no existential variables. The terms occurring in q are denoted $terms(q)$, and the variables $vars(q)$. The free variables $\vec{x}$ of a query are called *answer variables*.

Let $\mathcal{I}$ be an interpretation, $q(\vec{x})$ a CQ. An *answer to* q *in* $\mathcal{I}$ is a tuple $\vec{a}$ of elements from $\Delta^{\mathcal{I}}$ of length $|\vec{x}|$ such that there is a map $\pi : terms(q) \mapsto \Delta^{\mathcal{I}}$ satisfying (i) $\pi(\vec{x}) = \vec{a}$, (ii) $\pi(b) = b$ for each individual b, (iii) $\mathcal{I} \models P(\pi(\vec{z}))$ for each atom $P(\vec{z})$ in q, and (iv) $\pi(t) = \pi(t')$ for each atom $t = t'$ in q, and in that case we write $\mathcal{I} \models q(\vec{a})$. The map π is called a *match* for q in $\mathcal{I}$. The *certain answers* of $q(\vec{x})$ over $\mathcal{A}$ w.r.t. $\mathcal{T}$ are denoted $cert(q, \mathcal{T}, \mathcal{A})$ and defined as set of the tuples $\vec{a}$ such that $\mathcal{I} \models q(\vec{a})$ for every model $\mathcal{I}$ of $(\mathcal{T}, \mathcal{A})$. When we talk about the computational complexity of *query answering*, we mean the following decision problem: given a KB $\mathcal{K}$, a query q, and a tuple of individuals $\vec{a}$, check whether $\vec{a} \in cert(q, \mathcal{T}, \mathcal{A})$.

DL-Lite with Complex Role Inclusions

In this section, we study an extension of *DL-Lite*$^{\mathcal{H}}$ with complex role inclusions, which are critically important in our approach to ontology-based relaxing and restraining of queries. A *complex role inclusion* (CRI) is an expression of the form $r \cdot s \sqsubseteq t$, with $r, s, t \in \mathsf{N_R}^{\pm}$. An interpretation $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ *satisfies* $r \cdot s \sqsubseteq t$ if $(d_1, d_2) \in r^{\mathcal{I}}$, $(d_2, d_3) \in s^{\mathcal{I}}$ imply $(d_1, d_3) \in t^{\mathcal{I}}$, for all $d_1, d_2, d_3 \in \Delta^{\mathcal{I}}$.

The addition of CRIs to DLs may lead to undecidability of reasoning (Horrocks and Sattler 2004), hence syntactic conditions such as *regularity* (Kazakov 2010) are often needed.

In what follows, we assume a set $\mathsf{N_{R_s}} \subseteq \mathsf{N_R}^{\pm}$ of *simple roles* closed w.r.t. inverses (i.e. $s \in \mathsf{N_{R_s}}$ implies $s^- \in \mathsf{N_{R_s}}$); each $r \in \mathsf{N_R}^{\pm} \setminus \mathsf{N_{R_s}}$ is a *non-simple* role. We extend *DL-Lite*$^{\mathcal{H}}$ with CRIs as follows:

Definition 1 (*DL-Lite*$^{\mathcal{HR}}$)**.** *A DL-Lite*$^{\mathcal{HR}}$ *TBox* $\mathcal{T}$ *is a DL-Lite*$^{\mathcal{H}}$ *TBox that may also contain CRIs, and satisfies*

- $s \in \mathsf{N_{R_s}}$ *and* $t \in \mathsf{N_R}^{\pm} \setminus \mathsf{N_{R_s}}$*, for every* $r \cdot s \sqsubseteq t \in \mathcal{T}$*;*
- *if* $s \sqsubseteq t \in \mathcal{T}$ *and* $t \in \mathsf{N_{R_s}}$*, then* $s \in \mathsf{N_{R_s}}$.

Properties such as FO-rewritability are affected by CRIs. Indeed, using a single CRI $r \cdot s \sqsubseteq r$, it is possible to express that r corresponds to the transitive closure of role s on a given graph. Since KB satisfiability in *DL-Lite* with transitive roles is NLOGSPACE-hard, the following follows easily.

Lemma 1. *(Artale et al. 2009) Answering instance queries in DL-Lite*$^{\mathcal{HR}}$ *is* NLOGSPACE*-hard in data complexity, already for TBoxes consisting of the CRI* $r \cdot s \sqsubseteq r$ *only.*

Therefore CQs are not FO-rewritable w.r.t. *DL-Lite*$^{\mathcal{HR}}$ TBoxes. Our goal next is to present a restricted *DL-Lite*$^{\mathcal{HR}}$-based setting that is expressive enough to capture the desired scenarios and that still supports FO-rewritability. To this end, we first define a fragment of *DL-Lite*$^{\mathcal{HR}}$ that disallows cyclic dependencies among roles in CRIs. This *non-recursive* fragment of *DL-Lite*$^{\mathcal{HR}}$ is quite expressive

(e.g., KB satisfiability is not tractable) and supports FO-rewritability of CQs, yet it is not sufficient to cover our motivating example that involves recursion. To overcome this, we carefully relax the non-recursiveness requirement so that desired cyclic dependencies in CRIs are allowed, obtaining *recursion-safe DL-Lite*$^{\mathcal{HR}}$. We then identify a sufficient condition (over ABoxes) for eliminating recursive CRIs, return to the non-recursive setting, and regain FO-rewritability. Moreover, we show that satisfiability of such knowledge bases is tractable.

Non-recursive *DL-Lite*$^{\mathcal{HR}}$

We start by defining a suitable notion of recursive CRIs. For a *DL-Lite*$^{\mathcal{HR}}$ TBox $\mathcal{T}$, the *recursion graph* $\mathcal{G}_\mathcal{T}$ *of* $\mathcal{T}$ is the directed graph that contains (i) a node v_A for each concept name A in $\mathcal{T}$, (ii) a node v_r for each role name r in $\mathcal{T}$, and (iii) there exists an edge from a node $v_{P'}$ to a node v_P whenever P occurs on the left-hand-side and P' on the right-hand-side of an axiom in $\mathcal{T}$. A CRI $t \cdot s \sqsubseteq r$ is *recursive w.r.t.* a TBox $\mathcal{T}$ if $\mathcal{G}_\mathcal{T}$ has a path from v_t or v_s to v_r. In this case we also say that r is *a recursive role in* $\mathcal{T}$.

Definition 2. *A DL-Lite*$^{\mathcal{HR}}_{\mathsf{non\text{-}rec}}$ *TBox is a DL-Lite*$^{\mathcal{HR}}$ *TBox* $\mathcal{T}$ *without recursive CRIs.*

We now define query rewriting rules for non-recursive *DL-Lite*$^{\mathcal{HR}}$. For a CQ q, we denote by z^q an arbitrary variable not occurring in q; we will use z^q in the query rewriting rules through the rest of the paper. An *atom substitution* $\theta = [\Gamma_1/\Gamma_2]$ can be applied to q if $\Gamma_1 \subseteq q$ and the effect is to replace atoms Γ_1 with atoms Γ_2 in q.

Definition 3. *Let* $\mathcal{T}$ *be a DL-Lite*$^{\mathcal{HR}}_{\mathsf{non\text{-}rec}}$ *TBox. For CQs* q, q'*, we write* $q \rightsquigarrow_\mathcal{T} q'$ *whenever* q' *is obtained by*

B1 *replacing* x *by* y *in* q*, for* $x, y \in vars(q)$

or by applying an atom substitution θ *to* q*, as follows:*

S1 $\theta = [A_2(x)/A_1(x)]$*, if* $A_1 \sqsubseteq A_2 \in \mathcal{T}$ *and* $A_2(x) \in q$*;*

S2 $\theta = [r(x, y)/A(x)]$*, if* $A \sqsubseteq \exists r \in \mathcal{T}$*,* $r(x, y) \in q$ *and* y *is a non-answer variable occurring only once in* q*;*

S3 $\theta = [A(x)/r(x, z^q)]$*, if* $\exists r \sqsubseteq A \in \mathcal{T}$ *and* $A(x) \in q$*;*

S4 $\theta = [s(x, y)/r(x, y)]$*, if* $r \sqsubseteq s \in \mathcal{T}$ *and* $s(x, y) \in q$*;*

S5 $\theta = [s(x, y)/r(y, x)]$*, if* $r \sqsubseteq s^- \in \mathcal{T}$ *and* $s(x, y) \in q$*;*

S6 $\theta = [r(x, y)/\{t(x, z^q), s(z^q, y)\}]$*, if* $t \cdot s \sqsubseteq r \in \mathcal{T}$ *and* $r(x, y) \in q$*;*

By $q \rightsquigarrow_\mathcal{T}{}^* q'$ we denote the reflexive, transitive closure of $q \rightsquigarrow_\mathcal{T} q'$. The *rewriting of* q *w.r.t.* $\mathcal{T}$ is the set $rew(q, \mathcal{T})$ of all queries (modulo isomorphisms) q' such that $q \rightsquigarrow_\mathcal{T}{}^* q'$. Moreover, the absence of recursive CRIs in $\mathcal{T}$ ensures that $rew(q, \mathcal{T})$ is finite and can be effectively computed.

Lemma 2. *For a DL-Lite*$^{\mathcal{HR}}_{\mathsf{non\text{-}rec}}$ *TBox* $\mathcal{T}$ *and CQ* q*, the size of each* $q' \in rew(q, \mathcal{T})$ *is bounded by a polynomial, and can be computed in polynomial time, in the size of* $\mathcal{T}$ *and* q*.*

We can now show FO-rewritability of CQs in *DL-Lite*$^{\mathcal{HR}}_{\mathsf{non\text{-}rec}}$.

Theorem 1. *Let* $\mathcal{T}$ *be a DL-Lite*$^{\mathcal{HR}}_{\mathsf{non\text{-}rec}}$ *TBox,* q *a CQ. For every ABox* $\mathcal{A}$ *such that* $(\mathcal{T}, \mathcal{A})$ *is satisfiable, we have:*

$$cert(q, \mathcal{T}, \mathcal{A}) = \bigcup_{q' \in rew(q, \mathcal{T})} cert(q', \emptyset, \mathcal{A}).$$

Surprisingly, unlike the extension with transitive roles, KB satisfiability for non-recursive CRIs is intractable.

Theorem 2. *(i) Satisfiability of DL-Lite*$^{\mathcal{HR}}_{\mathsf{non\text{-}rec}}$ *KBs is* coNP*-complete for combined complexity, and (ii) CQ answering over consistent DL-Lite*$^{\mathcal{HR}}_{\mathsf{non\text{-}rec}}$ *KBs is in* AC^0 *for data, and* NP*-complete for combined complexity.*

For *(i)*, the upper bound is obtained similarly as for standard *DL-Lite*: KB satisfiability can be reduced to UCQ answering, using a CQ q_α for testing whether each disjointness axiom α is violated. By Lemma 2 and Theorem 1, an NP procedure can guess one such q_α, guess a q'_α in its rewriting, and evaluate q'_α over $\mathcal{A}$. The lower bound can be shown by a reduction of the complement of 3SAT to KB satisfiability.

For *(ii)*, data complexity follows from Theorem 1, while for combined complexity, NP-hardness is inherited from CQ evaluation in relational databases. An NP procedure for answering q over $\mathcal{A}$ w.r.t. $\mathcal{T}$ can guess some $q' \in rew(q, \mathcal{T})$ (which is polynomial in q and $\mathcal{T}$) and a map $\pi : vars(q') \to ind(\mathcal{A})$, and then verify whether π is a match of q' in $\mathcal{A}$, which can be done in polynomial time.

Recursion-safe *DL-Lite*$^{\mathcal{HR}}$

In *DL-Lite*$^{\mathcal{HR}}_{\mathsf{non\text{-}rec}}$ we cannot express CRIs like the one in our motivating example. To overcome this, we introduce an FO-rewritable fragment of *DL-Lite*$^{\mathcal{HR}}$ able to express certain kind of recursive CRIs.

Definition 4 (*DL-Lite*$^{\mathcal{HR}}_{\mathsf{rec\text{-}safe}}$)**.** *A* recursion safe *DL-Lite*$^{\mathcal{HR}}$ *TBox is a TBox* $\mathcal{T}$ *where every CRI* $r_1 \cdot s \sqsubseteq r_2 \in \mathcal{T}$ *satisfies the following conditions.*

- *If* r_2 *is a recursive role, then every cycle in* $\mathcal{G}_\mathcal{T}$ *containing* r_2 *has length at most one, and* $r_1 = r_2$*.*
- *There is no axiom of the form* $B \sqsubseteq \exists t \in \mathcal{T}$ *with* $t \sqsubseteq^{\mathsf{s}}_\mathcal{T} s$ *or* $t \sqsubseteq^{\mathsf{s}}_\mathcal{T} s^-$*, where* $\sqsubseteq^{\mathsf{s}}_\mathcal{T}$ *denotes the reflexive and transitive closure of* $s_1 \sqsubseteq s_2 \in \mathcal{T}$ *with* $s_2 \in \mathsf{N}_{\mathsf{R_s}}$*.*

The first condition restricts recursion to a simple form; we show later that this form of recursion can be eliminated when the ABox satisfies certain conditions. The second, on the other hand ensures that every CRI is 'guarded' by a simple role that is not existentially implied. Thus, for query answering, we can assume that only ABox individuals are connected by these guarding simple roles, and thus edges in the extension of recursive roles 'produced' by CRIs will always contain at least one individual. Note, for example, that $\mathcal{T}_e$ extended with (1) is recursion safe, since the simple role locIn in the CRI is not existentially implied by other axioms in $\mathcal{T}_e$. In contrast to non-recursive *DL-Lite*$^{\mathcal{HR}}$, in combined complexity KB satisfiability and instance query answering are tractable for *DL-Lite*$^{\mathcal{HR}}_{\mathsf{rec\text{-}safe}}$ KBs.

Theorem 3. *KB satisfiability and answering instance queries in DL-Lite*$^{\mathcal{HR}}_{\mathsf{rec\text{-}safe}}$ *are in* PTIME *for combined complexity.*

The proof of the theorem above relies on the construction of a particular *canonical* model. Let $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ be a *DL-Lite*$^{\mathcal{HR}}_{\mathsf{rec\text{-}safe}}$ KB. We define an interpretation $\mathcal{E}_{\mathcal{T},\mathcal{A}}$ with domain $\Delta^{\mathcal{E}_{\mathcal{T},\mathcal{A}}} = D_0 \cup D_1 \cup D_2$, where

$$D_0 = ind(\mathcal{A}), \quad D_1 = \{c_{ar} \mid a \in D_0, B \sqsubseteq \exists r \in \mathcal{T}\},$$
$$D_2 = \{c_r \mid B \sqsubseteq \exists r \in \mathcal{T}\},$$

and such that each concept, respectively each role name in $\mathcal{K}$ is interpreted as the minimal subset of $\Delta^{\mathcal{E}_{\mathcal{T},\mathcal{A}}}$, respectively of $\Delta^{\mathcal{E}_{\mathcal{T},\mathcal{A}}} \times \Delta^{\mathcal{E}_{\mathcal{T},\mathcal{A}}}$, such that for all concepts A, B and all roles r, r_1, r_2, s, t in $\mathcal{K}$, the following conditions hold:

- If $A(d) \in \mathcal{A}$ then $d \in A^{\mathcal{E}_{\mathcal{T},\mathcal{A}}}$, and if $r(d,d') \in \mathcal{A}$ then $(d,d') \in r^{\mathcal{E}_{\mathcal{T},\mathcal{A}}}$.
- If $B \sqsubseteq \exists r \in \mathcal{T}$, then $(d, c_{ar}) \in r^{\mathcal{E}_{\mathcal{T},\mathcal{A}}}$ if $d \in B^{\mathcal{E}_{\mathcal{T},\mathcal{A}}} \cap D_0$, and $(d, c_r) \in r^{\mathcal{E}_{\mathcal{T},\mathcal{A}}}$ if $d \in B^{\mathcal{E}_{\mathcal{T},\mathcal{A}}} \cap (D_1 \cup D_2)$.
- If $B \sqsubseteq A \in \mathcal{T}$ and $d \in B^{\mathcal{E}_{\mathcal{T},\mathcal{A}}}$ then $d \in A^{\mathcal{E}_{\mathcal{T},\mathcal{A}}}$.
- If $r_1 \sqsubseteq r_2 \in \mathcal{T}$ and $(d,d') \in r_1^{\mathcal{E}_{\mathcal{T},\mathcal{A}}}$ then $(d,d') \in r_2^{\mathcal{E}_{\mathcal{T},\mathcal{A}}}$.
- If $r_1 \sqsubseteq r_2^- \in \mathcal{T}$ and $(d,d') \in r_1^{\mathcal{E}_{\mathcal{T},\mathcal{A}}}$ then $(d',d) \in r_2^{\mathcal{E}_{\mathcal{T},\mathcal{A}}}$.
- If $r \cdot s \sqsubseteq t \in \mathcal{T}$ and $(d,d') \in r^{\mathcal{E}_{\mathcal{T},\mathcal{A}}}$ and $(d',d'') \in s^{\mathcal{E}_{\mathcal{T},\mathcal{A}}}$ then $(d,d'') \in t^{\mathcal{E}_{\mathcal{T},\mathcal{A}}}$.

It can be readily verified that $\mathcal{E}_{\mathcal{T},\mathcal{A}}$ is of polynomial size in $\mathcal{K}$, and that whenever $\mathcal{K}$ is satisfiable then $\mathcal{E}_{\mathcal{T},\mathcal{A}} \models \mathcal{K}$. Moreover, canonicity of $\mathcal{E}_{\mathcal{T},\mathcal{A}}$ is given by the following.

Claim 1. *For any (satisfiable) DL-Lite$^{\mathcal{HR}}_{\mathsf{rec\text{-}safe}}$ KB $(\mathcal{T},\mathcal{A})$ and any instance query q,*

$$cert(q,\mathcal{T},\mathcal{A}) = ans(q,\mathcal{E}_{\mathcal{T},\mathcal{A}}).$$

Recall that by Lemma 1 *DL-Lite$^{\mathcal{HR}}_{\mathsf{rec\text{-}safe}}$* TBoxes are not FO-rewritable in general. However, we will show that recursive CRIs can be eliminated provided that they are only relevant on paths of bounded length in models of $\mathcal{K}$. We formalize this intuition next.

Definition 5 (k-bounded ABox)**.** *Let $\mathcal{T}$ be a DL-Lite$^{\mathcal{HR}}$ TBox. For S a set of simple roles, an S-*path of length n between a and b in $ind(\mathcal{A})$ w.r.t. $\mathcal{T}$ *is a sequence of different pairs of individuals $(d_0,d_1),(d_1,d_2),\ldots,(d_{n-1},d_n)$ such that $d_0 = a$, $d_n = b$ and for each (d_i,d_{i+1}), $0 \le i < n$, there exists $s'(d_i,d_{i+1}) \in \mathcal{A}$ such that $s' \sqsubseteq^{\mathsf{s}}_{\mathcal{T}} s$ for some $s \in S$.*

*Given an ABox $\mathcal{A}$ and some $k \ge 0$, we say that $\mathcal{A}$ is k-*bounded for $\mathcal{T}$ *if for each $S_r = \{s \mid r \cdot s \sqsubseteq r \in \mathcal{T}\}$ there is no S_r-path of length larger than k in $\mathcal{A}$.*

If the given ABox is k-bounded for the given k, recursive CRIs can be unfolded into non-recursive ones. Therefore, queries can be rewritten using a TBox in which all recursive CRIs have been unfolded.

Definition 6 (k-unfolding, k-rewriting)**.** *For an arbitrary DL-Lite$^{\mathcal{HR}}_{\mathsf{rec\text{-}safe}}$ TBox $\mathcal{T}$, and fixed $k \ge 0$, a k-*unfolding of $\mathcal{T}$ *is a DL-Lite$^{\mathcal{HR}}_{\mathsf{non\text{-}rec}}$ TBox $\mathcal{T}_k$ obtained by replacing each $r \cdot s \sqsubseteq r \in \mathcal{T}$ with the axioms*

$$r_{j-1} \cdot s \sqsubseteq r_j \quad r \sqsubseteq r_0 \quad r_j \sqsubseteq \hat{r} \quad (1 \le j \le k),$$

where $\hat{r}$ and r_j are fresh role names. For a CQ q, let $\hat{q}$ be the query obtained from q by replacing, for every $r \cdot s \sqsubseteq r \in \mathcal{T}$, each $r(x,y) \in q$ by $\hat{r}(x,y)$.

For k-bounded ABoxes, $rew(\hat{q},\mathcal{T}_k)$ is an FO-rewriting of q.

Theorem 4. *Let $\mathcal{T}$ be a DL-Lite$^{\mathcal{HR}}_{\mathsf{rec\text{-}safe}}$ TBox, $\mathcal{T}_k$ a k-unfolding of $\mathcal{T}$, for some $k \ge 0$, and q a CQ over the signature of $\mathcal{T}$. Then, for every k-bounded ABox $\mathcal{A}$:*

$$cert(q,\mathcal{T},\mathcal{A}) = \bigcup_{q' \in rew(\hat{q},\mathcal{T}_k)} cert(q',\emptyset,\mathcal{A}).$$

Query Reformulations

In this section we propose two sets of rules for relaxing and restraining queries. The first one uses axioms in the ontology to guide the reformulation. Essentially, these are based on the usual rules for query rewriting and on suitable counterparts, resulting in restrictions and relaxations, respectively. The second set of rules are analogous but use dependencies that hold for a given dataset instead of axioms in the ontology. The goal of these rules is to provide a simple approach for reformulating queries, which is intuitive, computationally inexpensive, and that can leverage multidimensional knowledge.

Ontology-based Reformulations

The query rewriting rules **B1** and **S1**-**S6** for *DL-Lite$^{\mathcal{HR}}_{\mathsf{non\text{-}rec}}$* produce *restrainings* of a given query in the sense that the answers of the resulting query are necessarily contained in the answers of the original one.

Definition 7. *Let $\mathcal{T}$ be a DL-Lite$^{\mathcal{HR}}$ TBox. Given a pair of CQs q, q', we write $q \rightsquigarrow^{s}_{\mathcal{T}} q'$ if $q \rightsquigarrow_{\mathcal{T}} q'$, and call q' a* restraining of *q w.r.t. $\mathcal{T}$.*

These reformulations are *ontology-based* because they depend on the axioms of $\mathcal{T}$ only, and they are restrainings for every dataset mediated by $\mathcal{T}$.

Proposition 1. *Let $\mathcal{T}$ be a DL-Lite$^{\mathcal{HR}}$ TBox. For any two CQs such that $q_1 \rightsquigarrow^{s}_{\mathcal{T}} q_2$ and every ABox $\mathcal{A}$, we have that $cert(q_2,\mathcal{T},\mathcal{A}) \subseteq cert(q_1,\mathcal{T},\mathcal{A})$.*

Example 1. *Let $\mathcal{T}_e$ be as above. For the following queries*

$$q(x) \leftarrow \mathrm{CulturEvt}(x), \mathrm{occIn}(x,y), \mathrm{City}(y)$$
$$q_1(x) \leftarrow \mathrm{Concert}(x), \mathrm{occIn}(x,y), \mathrm{City}(y)$$
$$q_2(x) \leftarrow \mathrm{Concert}(x), \mathrm{occIn}(x,z), \mathrm{locIn}(z,y), \mathrm{City}(y).$$

it holds that $q \rightsquigarrow^{s}_{\mathcal{T}_e} q_1 \rightsquigarrow^{s}_{\mathcal{T}_e} q_2$ since by applying **S1** *using* $\mathrm{Concert} \sqsubseteq \mathrm{CulturEvt} \in \mathcal{T}_e$ *we obtain q_1, and further by applying* **S6** *using* (1) *we obtain q_2.* △

We have seen that the query rewriting rules that 'apply' the axioms in a right-to-left fashion provide natural means to restrain queries. The natural next step is to define analogous rules that use the axioms in a left-to-right fashion to relax queries. Note that in the next definition, rules **G1**–**G6** are, essentially, the dual of rules **S1**–**S6**, while rule **R1** simply allows us to relax a query by dropping an atom.

Definition 8. *Let $\mathcal{T}$ be a DL-Lite$^{\mathcal{HR}}$ TBox. For CQs q, q', we write $q \rightsquigarrow^{g}_{\mathcal{T}} q'$ whenever q' is obtained from q by*

R1 *removing an atom $x = a$ or an atom $A(x)$ with x a non-answer variable,*

or by applying an atom substitution θ as follows:

G1 $\theta = [A_1(x)/A_2(x)]$, *if $A_1 \sqsubseteq A_2 \in \mathcal{T}$ and $A_1(x) \in q$;*
G2 $\theta = [A(x)/r(x,z^q)]$, *if $A \sqsubseteq \exists r \in \mathcal{T}$ and $A(x) \in q$;*
G3 $\theta = [r(x,y)/A(x)]$, *if $\exists r \sqsubseteq A \in \mathcal{T}$, $r(x,y) \in q$ and y is a non-answer variable occurring only once in q;*
G4 $\theta = [r(x,y)/s(x,y)]$, *if $r \sqsubseteq s \in \mathcal{T}$ and $r(x,y) \in q$;*
G5 $\theta = [r(x,y)/s(y,x)]$, *if $r \sqsubseteq s^- \in \mathcal{T}$ and $r(x,y) \in q$;*

G6 $\theta = [\{t(x,y), s(y,z)\}/r(x,z)]$, *if* $t \cdot s \sqsubseteq r \in \mathcal{T}$, $t(x,y), s(y,z) \in q$ *and* y *is a non-answer variable that does not occur elsewhere in* q*;*

We call q' *a* query relaxation of q *w.r.t* $\mathcal{T}$ *whenever* $q \rightsquigarrow^{g}_{\mathcal{T}} q'$.

Example 2. *Let* $\mathcal{T}_e$ *be as above, and take the queries*

$$q(x) \leftarrow \mathrm{Concert}(x), \mathrm{occIn}(x,y), \mathrm{locIn}(y,z), z = \mathsf{Vienna},$$
$$q_1(x) \leftarrow \mathrm{CulturEvt}(x), \mathrm{occIn}(x,y), \mathrm{locIn}(y,z), z = \mathsf{Vienna};$$
$$q_2(x) \leftarrow \mathrm{CulturEvt}(x), \mathrm{occIn}(x,z), z = \mathsf{Vienna}.$$

it holds that $q \rightsquigarrow^{g}_{\mathcal{T}_e} q_1 \rightsquigarrow^{g}_{\mathcal{T}_e} q_2$*, since by applying* **G1** *using* $\mathrm{Concert} \sqsubseteq \mathrm{CulturEvt} \in \mathcal{T}_e$ *we obtain* q_1 *and further by applying* **G6** *using* (1) *we get* q_2. △

The following result is the analogous of Proposition 1.

Proposition 2. *Let* $\mathcal{T}$ *be a DL-Lite*$^{\mathcal{HR}}$ *TBox. For any two CQs, such that* $q_1 \rightsquigarrow^{g}_{\mathcal{T}} q_2$*, and every ABox* $\mathcal{A}$*, we have that* $cert(q_1, \mathcal{T}, \mathcal{A}) \subseteq cert(q_2, \mathcal{T}, \mathcal{A})$.

Data-dependent Query Reformulations

The query reformulation rules above might miss interesting reformulations. For instance, in our running example, a query relaxation from *concerts in Vienna*, to *concerts at some location in Austria*, or a restraining to *concerts at the State Opera* would be meaningful. For this reason, we define *data-dependent* query refomulations that assume a fixed dataset, and use *assertions and dependencies* that hold for it as if they were axioms in the ontology. In our running example, since a quick inspection at $\mathcal{A}_e$ tells us that *every existing venue is located in a city*, we could use this to relax the query

$$q(x) \leftarrow \mathrm{Event}(x), \mathrm{occIn}(x,y), \mathrm{Venue}(y) \qquad \text{into}$$
$$q'(x) \leftarrow \mathrm{Event}(x), \mathrm{occIn}(x,y), \mathrm{locIn}(y,z), \mathrm{City}(z).$$

We note that such a reformulation could be done using rule **S2**, if we had an inclusion $\mathrm{Venue} \sqsubseteq \exists \mathrm{locIn}.\mathrm{City}$ in the TBox. But if we do not have such an axiom and it is not possible or desirable to add it, the rules below rely on

$$cert(\mathrm{Venue}(x), \mathcal{T}_e, \mathcal{A}_e) \subseteq cert(\exists \mathrm{locIn}.\mathrm{City}(x), \mathcal{T}_e, \mathcal{A}_e)$$

to enable such a reformulation.

We define first data-dependent restraining rules of two kinds. First we have rules that use entailed assertions. For example, if $\mathcal{K} \models A(a)$ and q contains an atom $A(x)$, then we can restrain q by equating x and a, as done by rule (**SA1**). Similarly with role assertions: if $\mathcal{K} \models r(a,b)$ and $r(x,y)$ is in q, with (**SA2**) we can equate x to a, or y to b. We can also use $\mathcal{K} \models r(a,b)$ to replace $r(x,y), y = b$ by $x = a$ if y is a non-answer variable that does not occur elsewhere in q; see (**SA3**).

The second kind of rules (**SD1**–**SD5**) use dependencies of the form $cert(q_1, \mathcal{T}, \mathcal{A}) \subseteq cert(q_2, \mathcal{T}, \mathcal{A})$, written $q_1 \subseteq_{(\mathcal{T},\mathcal{A})} q_2$ for short. They are similar to the rules in the previous section, but we may now replace $B(x)$ by $A(x)$ not only when $A \sqsubseteq B$ is in $\mathcal{T}$, but also when the weaker condition $A(x) \subseteq_{\mathcal{K}} B(x)$ holds. Such replacements are also possible for some more complex pairs of atoms. For example, if $\exists r.B(x) \subseteq_{\mathcal{K}} A(x)$, then $A(x)$ can be replaced by $r(x,y), B(y)$ to restrain q. This would be similar to having in Definition 3 a rule for (non-*DL-Lite*) axioms $\exists r.B \sqsubseteq A$.

The relaxation rules are dual to the restraining ones. For example, (**GA1**) is dual to (**SA1**): if the query is equating some variable x to a and a is an instance of A, then we can replace $x = a$ by $A(x)$, thus allowing x to be any instance of A, rather than just a. Note that there is no dual to (**SA2**) since it would simply need to drop $x = a$ or $y = b$ from q, but this does not depend on the data and is captured by (**R1**).

Definition 9. *Let* $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ *be a DL-Lite*$^{\mathcal{HR}}$ *KB.*

For CQs q, q'*, we write* $q \rightsquigarrow^{s}_{\mathcal{K}} q'$ *if* $q \rightsquigarrow^{s}_{\mathcal{T}} q'$ *or* q' *is obtained from* q *using atom substitution* θ *as follows:*

SA1 $\theta = [\emptyset/x = a]$, *if* $A(x) \in q$ *and* $\mathcal{K} \models A(a)$*;*

SA2 $\theta = [\emptyset/x = a]$ *or* $\theta = [\emptyset/y = b]$, *if* $r(x,y) \in q$ *and* $\mathcal{K} \models r(a,b)$*;*

SA3 $\theta = [\{r(x,y), y = b\}/x = a]$, *if* $r(x,y), y = b \in q$*,* y *does not occur elsewhere in* q *and* $\mathcal{K} \models r(a,b)$*;*

SD1 $\theta = [A_2(x)/A_1(x)]$, *if* $A_1(x) \subseteq_{\mathcal{K}} A_2(x)$ *and* $A_2(x) \in q$*;*

SD2 $\theta = [\{r(x,y), A'(y)\}/A(x)]$, *if* $A(x) \subseteq_{\mathcal{K}} \exists r.A'(x)$*, and* $r(x,y), A'(y) \in q$ *such that* y *does not occur elsewhere in* q*;*

SD3 $\theta = [A(x)/\{r(x,z^q), A'(z^q)\}]$, *if* $\exists r.A'(x) \subseteq_{\mathcal{K}} A(x)$ *and* $A(x) \in q$*;*

SD4 $\theta = [s(x,y)/r(x,y)]$, *if* $r(x,y) \subseteq_{\mathcal{K}} s(x,y)$ *and* $s(x,y) \in q$*;*

SD5 $\theta = [\{r(x,y), A(y)\}/\{p(x,y), A'(y)\}]$, *if* $r(x,y)$*,* $A(y) \in q$*,* y *does not occur elsewhere in* q *and* $\exists p.A'(x) \subseteq_{\mathcal{K}} \exists r.A(x)$*.*

Further, we write $q \rightsquigarrow^{g}_{\mathcal{K}} q'$ *if* $q \rightsquigarrow^{g}_{\mathcal{T}} q'$ *or* q' *is obtained from* q *using atom substitution* θ *as follows:*

GA1 $\theta = [x = a/A(x)]$, *if* $x = a \in q$ *and* $\mathcal{K} \models A(a)$*;*

GA3 $\theta = [x = a/\{r(x,y), y = b\}]$, *if* $x = a \in q$ *and* $\mathcal{K} \models r(a,b)$*;*

GD1 $\theta = [A_1(x)/A_2(x)]$, *if* $A_1(x) \subseteq_{\mathcal{K}} A_2(x)$ *and* $A_1(x) \in q$*;*

GD2 $\theta = [A(x)/\{r(x,z^q), A'(z^q)\}]$, *if* $A(x) \subseteq_{\mathcal{K}} \exists r.A'(x)$ *and* $A(x) \in q$*;*

GD3 $\theta = [\{r(x,y), A'(y)\}/A(x)]$, *if* $\exists r.A'(x) \subseteq_{\mathcal{K}} A(x)$ *and* $r(x,y), A'(y) \in q$ *such that* y *does not occur elsewhere in* q*;*

GD4 $\theta = [r(x,y)/s(x,y)]$, *if* $r(x,y) \subseteq_{\mathcal{K}} s(x,y)$ *and* $r(x,y) \in q$*;*

GD5 $\theta = [\{p(x,y), A'(y)\}/\{r(x,y), A(y)\}]$, *if* $p(x,y)$*,* $A'(y) \in q$*,* y *does not occur elsewhere in* q*, and* $\exists p.A'(x) \subseteq_{\mathcal{K}} \exists r.A(x)$*.*

Example 3. *Let* $\mathcal{K}_e = (\mathcal{T}_e, \mathcal{A}_e)$ *be as above. For CQs*

$$q(x) \leftarrow \mathrm{Concert}(x), \mathrm{occIn}(x,y), y = \mathsf{Vienna},$$
$$q_1(x) \leftarrow \mathrm{Concert}(x), \mathrm{occIn}(x,y), \mathrm{locIn}(y,z), z = \mathsf{Austria},$$
$$q_2(x) \leftarrow \mathrm{Concert}(x), \mathrm{occIn}(x,y), \mathrm{locIn}(y,z), \mathrm{Country}(z),$$
$$q_3(x) \leftarrow \mathrm{Concert}(x), \mathrm{occIn}(x,y), \mathrm{City}(y),$$

we have that $q \rightsquigarrow^{g}_{\mathcal{K}_e} q_1 \rightsquigarrow^{g}_{\mathcal{K}_e} q_2$ *since* q_1 *is obtained by applying* **GA3** *to* q *using* $\mathcal{K}_e \models \mathrm{locIn}(\mathsf{Vienna}, \mathsf{Austria})$ *and* q_2 *is obtained from* q_1 *by applying* **GA1** *using* $\mathcal{K}_e \models \mathrm{Country}(\mathsf{Austria})$.

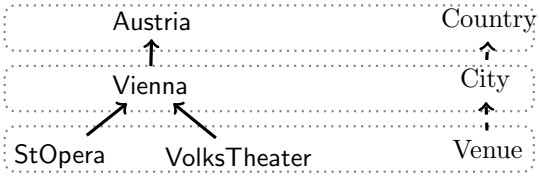

Figure 1: Dimension Location.

We can now choose to restrain q_2 and obtain q_3 by applying **SD2** *on q_2 using* $\mathrm{City}(x) \subseteq_{\mathcal{K}_e} \exists \mathrm{locIn}.\mathrm{Country}(x)$*, hence $q_2 \rightsquigarrow^s_{\mathcal{K}_e} q_3$. Note that q_3 can also be obtained as a relaxation of q with* **GA1** *and $\mathcal{K}_e \models \mathrm{City}(\mathsf{Vienna})$, thus $q \rightsquigarrow^g_{\mathcal{K}_e} q_3$ also holds.* △

Our data-driven rules indeed relax and restrain queries when evaluated over $(\mathcal{T}, \mathcal{A})$, but not for any arbitrary ABox.

Proposition 3. *For CQs q_1, q_2 and DL-Lite$^{\mathcal{HR}}$ KB $(\mathcal{T}, \mathcal{A})$:*

(g) *$q_1 \rightsquigarrow^g_{\mathcal{K}} q_2$ implies $cert(q_1, \mathcal{T}, \mathcal{A}) \subseteq cert(q_2, \mathcal{T}, \mathcal{A})$, and*

(s) *$q_1 \rightsquigarrow^s_{\mathcal{K}} q_2$ implies $cert(q_2, \mathcal{T}, \mathcal{A}) \subseteq cert(q_1, \mathcal{T}, \mathcal{A})$.*

Modeling Multi-dimensional Data

In this section we show that recursion safe *DL-Lite*$^{\mathcal{HR}}$ together with k-bounded ABoxes, are well-suited to describe this kind of *multi-dimensional knowledge* in the setting of OBDA. In the *multi-dimensional data model* (Hurtado and Mendelzon 2002), the data-schema is usually formalized as a set of *dimensions*, comprising a finite set of *categories* and a partial order between them, sometimes called *child-parent relation*. The model also considers a representation at the instance level that defines *members* for each category, and a child-parent relation between members of connected categories. In Figure 1 a dimension and instance of some Location hierarchy are illustrated, which makes use of concepts from $\mathcal{T}_e$ as categories. The dashed arrows show the order between categories, rectangles contain members for each category and solid arrows represent the role locIn.

We define order constraints to encode dimensions.

Definition 10. *For a simple role s, an* order constraint (along s) *takes the form $ord(s, \mathbf{A}, \prec)$, with $\mathbf{A} \subseteq \mathsf{N_C}$ finite, and $\prec$ a strict partial order over $\mathbf{A}$. An interpretation $\mathcal{I}$ satisfies $ord(s, \mathbf{A}, \prec)$ if*

$$s^{\mathcal{I}} \subseteq \bigcup_{A_1, A_2 \in \mathbf{A}} (A_1^{\mathcal{I}} \times A_2^{\mathcal{I}}), \quad s^{\mathcal{I}} \cap \bigcup_{A_1 \not\prec A_2} (A_1^{\mathcal{I}} \times A_2^{\mathcal{I}}) = \emptyset.$$

Intuitively, whenever $ord(s, \mathbf{A}, \prec)$ is satisfied in $\mathcal{I}$, all objects connected by s are instances of $\mathbf{A}$-concepts, in a way that is compliant with the order $\prec$. Further, s-paths connecting instances of concepts in $\mathbf{A}$ which are incomparable w.r.t. $\prec$ will be disallowed in $\mathcal{I}$. The latter is important as otherwise one cannot guarantee k-boundedness.

Example 4. *The* Location *dimension from Figure 1 is captured by $\mathcal{K}_e = (\mathcal{T}_e, \mathcal{A}_e)$ and the order constraint*

$$c = ord(\mathrm{locIn}, \{\mathrm{Venue}, \mathrm{City}, \mathrm{Country}\}, \prec) \qquad (2)$$

with $\mathrm{Venue} \prec \mathrm{City} \prec \mathrm{Country}$*. In each model of $\mathcal{K}_e$ satisfying* c*, the role* locIn *only connects instances of* Venue *with those of* City *or* Country*, and instances of* City *only with those of* Country*, thus capturing the intended semantics of the dimension.* △

Definition 11. *A* multi-dimensional KB *is a triple $(\mathcal{T}, \mathcal{A}, \mathcal{C})$ where $(\mathcal{T}, \mathcal{A})$ is a recursion safe DL-Lite$^{\mathcal{HR}}$ KB and $\mathcal{C}$ a set of order constraints.*

We will show that for multi-dimensional KBs providing certain guarantees w.r.t. $\mathcal{C}$, there is a k that ensures k-boundedness and allows construction of a k-unfolding of $\mathcal{T}$.

Definition 12. *$\mathcal{C}$* covers a role r in $\mathcal{T}$ *if there exists a partial order $(\mathbf{A}, \prec)$ such that for every role s in the set $\{s \mid r \cdot s \sqsubseteq r \in \mathcal{T}\}$, $ord(s, \mathbf{A}', \prec) \in \mathcal{C}$ for some $\mathbf{A} \subseteq \mathbf{A}'$. We say that $\mathcal{C}$ covers $\mathcal{T}$, if it covers every role r in $\mathcal{T}$.*
Further, $(\mathcal{T}, \mathcal{A})$ is $\mathcal{C}$-admissible *if $\mathcal{E}_{\mathcal{T},\mathcal{A}}$ satisfies each $c \in \mathcal{C}$.*

For example, the singleton set containing the order constraint c of Example 4 covers $\mathcal{T}_e$, and $\mathcal{K}_e$ is $\{\mathrm{c}\}$-admissible since $\mathcal{E}_{\mathcal{T}_e,\mathcal{A}_e}$ satisfies c.

Lemma 3. *Let $(\mathcal{T}, \mathcal{A})$ be a recursion-safe DL-Lite$^{\mathcal{HR}}$ KB, and let $\mathcal{C}$ be a set of order constraints covering $\mathcal{T}$. If $(\mathcal{T}, \mathcal{A})$ is $\mathcal{C}$-admissible, then $\mathcal{A}$ is $\ell(\mathcal{C})$-bounded for $\mathcal{T}$, where $\ell(\mathcal{C}) := max\{|\mathbf{A}| \mid ord(s, \mathbf{A}, \prec) \in \mathcal{C}\}$.*

Lemma 3 together with Theorem 4 yield FO-rewritability of queries over multi-dimensional KBs.

Theorem 5. *Let $\mathcal{T}$ be a recursion safe DL-Lite$^{\mathcal{HR}}$ TBox, $\mathcal{C}$ a set of order constraints that covers $\mathcal{T}$, and q a CQ. Let Q be the $\ell(\mathcal{C})$-rewriting of q w.r.t. $\mathcal{T}$. Then, for each ABox $\mathcal{A}$ such that $(\mathcal{T}, \mathcal{A})$ is $\mathcal{C}$-admissible,*

$$cert(q, \mathcal{T}, \mathcal{A}) = \bigcup_{q' \in Q} cert(q', \emptyset, \mathcal{A}).$$

Finally, we note that $\mathcal{C}$-admissibility amounts to evaluating simple queries on $\mathcal{E}_{\mathcal{T},\mathcal{A}}$. This can be done in time that is polynomial in $\mathcal{C}$, $\mathcal{T}$, and $\mathcal{A}$. Moreover, although testing $\mathcal{C}$-admissibility is data dependent, once it is established, FO-rewritability is guaranteed for any CQ.

Proposition 4. *Checking $\mathcal{C}$-admissibility for recursion safe DL-Lite$^{\mathcal{HR}}$ KBs is feasible in polynomial time in combined complexity.*

Reformulations for Dimensional Navigation

The multi-dimensional data model enables data navigation along different axes given by the dimensions, similarly to points in a multi-dimensional space. This view lies at the core of OLAP and similar data analytic applications. For instance, using dimension Location, we can navigate from events occurring in some city to those occurring in some country, or in some venue. This navigation mechanism allows users to either zoom in or zoom out on the particular data, and it is usually realized by the so-called *drill-down* and and *roll-up* operators.

In our setting we can define similar navigation mechanisms for multi-dimensional KBs. We do this by means of queries containing an 'entry point' to some dimension, represented by some order constraint. A CQ q *refers to* $(\mathcal{T}, \mathcal{A}, \mathcal{C})$ if there are $ord(s, \mathbf{A}, \prec) \in \mathcal{C}$, $r \cdot s \sqsubseteq r \in \mathcal{T}$ and $A \in \mathbf{A}$ such that one of the following conditions is satisfied

1. $\{r(x,y), A(y)\} \subseteq q$ or
2. $\{r(x,y), y = a\} \subseteq q$ and $(\mathcal{T}, \mathcal{A}) \models A(a)$

where y is a non-answer variable which does not occur elsewhere in q.

We now define our version of the roll-up and drill-down operators. We restrict their application to *coherent* KBs where paths along the dimension exist. A multi-dimensional KB $(\mathcal{T}, \mathcal{A}, \mathcal{C})$ is *coherent* if $\mathcal{C}$ covers $\mathcal{T}$, $(\mathcal{T}, \mathcal{A})$ is $\mathcal{C}$-admissible and for each $ord(s, \mathbf{A}, \prec) \in \mathcal{C}$ and each $A \in \mathbf{A}$ such that there is some $A' \in \mathbf{A}$ with $A \prec A'$ we have $A(x) \subseteq_{\mathcal{K}} \exists s(x)$.

Definition 13 (Roll-up, drill-down)**.** *Let $(\mathcal{T}, \mathcal{A}, \mathcal{C})$ be a coherent multi-dimensional KB, q a CQ that refers to $(\mathcal{T}, \mathcal{A}, \mathcal{C})$, and $\Gamma \subseteq q$ the set of atoms witnessing this.*

A roll-up *of q w.r.t. Γ is a CQ obtained by applying a substitution θ_r on q, and a* drill-down *of q w.r.t. Γ is a CQ obtained by applying substitution θ_d on q, where θ_r, θ_d are as follows:*

- *$\theta_r = [A(x)/B(x)]$, if $A(x) \in \Gamma$, $A \prec B$ in some $c \in \mathcal{C}$;*
- *$\theta_r = [x = a/x = b]$, if $x = a \in \Gamma$ and $(\mathcal{T}, \mathcal{A}) \models s(a, b)$;*
- *$\theta_d = [A(x)/B(x)]$, if $A(x) \in \Gamma$, $B \prec A$ in some $c \in \mathcal{C}$;*
- *$\theta_d = [x = a/x = b]$, if $x = a \in \Gamma$ and $(\mathcal{T}, \mathcal{A}) \models s(b, a)$.*

Example 5. *Consider the KB $\mathcal{K}_e$ and order constraint c be as in Example 4. We have that $(\mathcal{K}_e, \{c\})$ is coherent. Now, consider the following set of queries:*

$$q_1(x) \leftarrow \text{Event}(x), \text{occIn}(x, y), \text{City}(y);$$
$$q_1^r(x) \leftarrow \text{Event}(x), \text{occIn}(x, y), \text{Country}(y);$$
$$q_1^d(x) \leftarrow \text{Event}(x), \text{occIn}(x, y), \text{Venue}(y).$$

The fact that q_1 refers to $(\mathcal{K}_e, \{c\})$ is witnessed by the atoms $\Gamma_1 = \{\text{occIn}(x, y), \text{City}(y)\}$. Since City $\prec$ Country *in c, q_1^r is a roll-up of q_1 w.r.t. Γ_1, and since* Venue $\prec$ City *in c, q_1^d is a drill-down of q_1. Now, consider the following queries:*

$$q_2(x) \leftarrow \text{Event}(x), \text{occIn}(x, y), y = \mathsf{Vienna};$$
$$q_2^r(x) \leftarrow \text{Event}(x), \text{occIn}(x, y), y = \mathsf{Austria};$$
$$q_2^d(x) \leftarrow \text{Event}(x), \text{occIn}(x, y), y = \mathsf{StOpera}.$$

The fact that q_2 refers to $(\mathcal{K}_e, \{c\})$ is witnessed by the atoms $\Gamma_2 = \{\text{occIn}(x, y), y = \mathsf{Vienna}\}$ and using $\mathcal{K}_e \models \text{locIn}(\mathsf{Vienna}, \mathsf{Austria})$ and $\mathcal{K}_e \models \text{locIn}(\mathsf{StOpera}, \mathsf{Vienna})$ we get that q_2^r is a roll-up q_2 w.r.t. Γ_2 and q_2^d is a drill-down of q_2 w.r.t. Γ_2. △

For coherent multi-dimensional KBs, the roll-up operation can be seen as a sequence of relaxing rules, and similarly, drill-down can be seen as a sequence of restrainings. Let q be a CQ that refers to a coherent $(\mathcal{T}, \mathcal{A}, \mathcal{C})$. If q_r is a roll-up of q w.r.t. $\{r(x,y), A(y)\}$, then q_r can be obtained as follows:

1. Coherence guaranties that **GD2** can be applied, obtaining query q' by replacing $A(y) \in q$ with $s(y,z), B(z)$, where z is a fresh variable; due to $\mathcal{C}$-admissibility it must be that $A \prec B$.
2. Since q refers to $(\mathcal{T}, \mathcal{A}, \mathcal{C})$ we have that $r \cdot s \sqsubseteq r \in \mathcal{T}$ and $r(x,y) \in q'$ with y not occuring elsewhere in q'. Thus we can apply **G6** on q' to obtain q_r.

Likewise, if q_r is a roll-up of q w.r.t. $\{r(x,y), y = a\}$, then q_r can be obtained as follows:

1. Since q refers to $(\mathcal{T}, \mathcal{A}, \mathcal{C})$ it must be that $(\mathcal{T}, \mathcal{A}) \models A(a)$, for some A in some order constraint in $\mathcal{C}$; coherence ensures that there exists $b \in ind(\mathcal{A})$ such that $(\mathcal{T}, \mathcal{A}) \models s(a, b)$, therefore we can apply **GA3** to obtain query q' by replacing $y = a$ with $s(y,z), z = b$, where z is a fresh variable.
2. Next, again we can apply **G6** as above and obtain q_r.

For drill-down queries, we apply the dual rules in the reverse order. From this observation, we obtain:

Proposition 5. *Let $(\mathcal{T}, \mathcal{A}, \mathcal{C})$ be a coherent multi-dimensional KB, q a CQ that refers to $(\mathcal{T}, \mathcal{A}, \mathcal{C})$ and $\Gamma \subseteq q$ a subset of atoms witnessing this. We have that the following hold:*

(r) *$q \subseteq_{(\mathcal{T}, \mathcal{A})} q_r$ for each roll-up q_r of q w.r.t. Γ, and*
(d) *$q_d \subseteq_{(\mathcal{T}, \mathcal{A})} q$ for each drill-down q_d of q w.r.t. Γ.*

Related Work

Query reformulations based on similarity measures have been considered for RDF data and SPARQL queries by Reddy and Kumar (2010), Huang and Liu (2010) and Virgilio et. al (2013), whereas approaches using simple ontological knowledge for such task have been proposed by Hurtado, Poulovassilis and Wood (2008), Elbassuoni et. al (2011), Dolog et al. (2009) and Frosini et. al (2017). Relaxations of SQL queries in relational databases using concept taxonomies have been studied by Martinenghi and Torlone (2014). For cooperative KBs, Inoue and Wiese (2011) define relaxations of CQs following a principled logic-based approach. Interactive faceted search techniques implement effective drill-down for refining queries (Roy et al. 2008; Kashyap, Hristidis, and Petropoulos 2010); in the context of RDF and knowledge graphs Arenas et. al (2016) and Sherkhonov et. al (2017) addressed theoretical underpinnings of faceted search for data/ontolgy exploration. Query evaluation minimizing data access in an OBDA under generalization/specialization relations is studied by Andresel, Ortiz and Šimkus (2016).

Logic-based formalizations of the multi-dimensional data model of Hurtado and Mendelzon (2002) have been proposed in the literature. Franconi and Sattler (1999), and Franconi and Kamble (2004) use DLs for modeling and reasoning about multi-dimensional data without considering querying, while Bertossi and Milani (2018) rely on an expressive fragment of Datalog$^{\pm}$ to capture dimensional knowledge, although at the expense of higher complexity (i.e., not FO-rewritable).

Query answering using CRIs is supported in ontology mediated settings where the DL has CRIs, or the query language contains conjunctive regular path queries (see (Ortiz 2013) and (Ortiz and Šimkus 2012) for references). However, query answering in all those settings is necessarily NLOGSPACE-hard in data complexity, and usually PSPACE-hard in combined complexity even for lightweight DLs (Bienvenu, Ortiz, and Šimkus 2015), while our goal in this paper is to design FO-rewritable *DL-Lite* extensions.

Discussion and Conclusions

We presented query reformulation rules that are data independent, as well as more fine-grained rules leveraging current dataset. To capture multi-dimensional knowledge, we extended *DL-Lite*$^{\mathcal{H}}$ with a restricted use of CRIs that preserve FO-rewritability, and yet cover the desired use case.

Our data-driven rules take into account a particular kind of dependencies in the data. We chose these somewhat subjectively, aiming at enhanced dimensional navigation, while keeping the complexity in check. Indeed, testing those dependencies in *DL-Lite*$^{\mathcal{HR}}$ KBs consisting of a recursion safe TBox and a k-bounded ABox is not computationally expensive (AC^0 in data complexity). Clearly, other approaches for data-driven reformulation might be feasible.

With the approach we have described there may be many possible reformulations of a given query, and not all of them are always equally interesting. In our future research we will investigate relaxations/restrictions that minimally modify the query answers, properties of our reformulation rules, and algorithms to effectively compute preferred reformulations. We are also investigating mechanisms for compiling the data and the ontology to support efficient answering of reformulated queries, and the definition of a declarative query language considering relaxing/restraining operators as first-class citizens. Finally, we also plan to consider aggregation and the definition of operators suitable for data analysis tasks in the spirit of OLAP systems.

Acknowledgments

This work was supported by the Austrian Science Fund (FWF) projects P30360, P30873 and W1255.

References

Andresel, M.; Ortiz, M.; and Šimkus, M. 2016. A compilation technique for interactive ontology-mediated data exploration. In *In Proc. of Description Logics 2016*, volume 1577 of *CEUR*.

Arenas, M.; Cuenca Grau, B.; Kharlamov, E.; Marciuska, S.; and Zheleznyakov, D. 2016. Faceted search over RDF-based knowledge graphs. *J. Web Sem.* 37-38:55–74.

Artale, A.; Calvanese, D.; Kontchakov, R.; and Zakharyaschev, M. 2009. The DL-Lite family and relations. *J. Artif. Intell. Res.* 36:1–69.

Bertossi, L. E., and Milani, M. 2018. Ontological multidimensional data models and contextual data quality. *J. Data and Information Quality* 9(3):14:1–14:36.

Bienvenu, M.; Ortiz, M.; and Šimkus, M. 2015. Regular path queries in lightweight description logics: Complexity and algorithms. *J. Artif. Int. Res.* 53(1):315–374.

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. Autom. Reasoning* 39(3):385–429.

Dolog, P.; Stuckenschmidt, H.; Wache, H.; and Diederich, J. 2009. Relaxing RDF queries based on user and domain preferences. *J. Intell. Inf. Syst.* 33(3):239–260.

Elbassuoni, S.; Ramanath, M.; and Weikum, G. 2011. Query relaxation for entity-relationship search. In *Proc. ESWC 2011*, volume 6644 of *LNCS*, 62–76. Springer.

Franconi, E., and Kamble, A. 2004. The GMD data model and algebra for multidimensional information. In *CAiSE*, volume 3084 of *LNCS*, 446–462. Springer.

Franconi, E., and Sattler, U. 1999. A data warehouse conceptual data model for multidimensional aggregation. In *Proc. DMDW 1999*, volume 19 of *CEUR*.

Frosini, R.; Calì, A.; Poulovassilis, A.; and Wood, P. T. 2017. Flexible query processing for SPARQL. *Semantic Web* 8(4):533–563.

Horrocks, I., and Sattler, U. 2004. Decidability of SHIQ with complex role inclusion axioms. *Artif. Intell.* 160(1-2):79–104.

Huang, H., and Liu, C. 2010. Query relaxation for star queries on RDF. In *Proc. WISE 2010*, volume 6488 of *LNCS*, 376–389. Springer.

Hurtado, C. A., and Mendelzon, A. O. 2002. OLAP dimension constraints. In *Proc. PODS 2002*, 169–179. ACM.

Hurtado, C. A.; Poulovassilis, A.; and Wood, P. T. 2008. Query relaxation in RDF. *J. Data Semantics* 10:31–61.

Inoue, K., and Wiese, L. 2011. Generalizing conjunctive queries for informative answers. In *Proc. FQAS 2011*, volume 7022 of *LNCS*, 1–12. Springer.

Kashyap, A.; Hristidis, V.; and Petropoulos, M. 2010. Facetor: cost-driven exploration of faceted query results. In *Proc. CIKM 2010*, 719–728. ACM.

Kazakov, Y. 2010. An extension of complex role inclusion axioms in the description logic SROIQ. In *Proc. IJCAR 2010*.

Martinenghi, D., and Torlone, R. 2014. Taxonomy-based relaxation of query answering in relational databases. *VLDB J.* 23(5):747–769.

Ortiz, M., and Šimkus, M. 2012. Reasoning and query answering in description logics. In *Reasoning Web*, volume 7487 of *LNCS*, 1–53. Springer.

Ortiz, M. 2013. Ontology based query answering: The story so far. In *Proc. AMW 2013*, volume 1087 of *CEUR*.

Poggi, A.; Lembo, D.; Calvanese, D.; De Giacomo, G.; Lenzerini, M.; and Rosati, R. 2008. Linking data to ontologies. *J. Data Semantics* 10:133–173.

Reddy, B. R. K., and Kumar, P. S. 2010. Efficient approximate SPARQL querying of web of linked data. In *Proc. URSW 2010*, volume 654 of *CEUR*, 37–48.

Roy, S. B.; Wang, H.; Das, G.; Nambiar, U.; and Mohania, M. K. 2008. Minimum-effort driven dynamic faceted search in structured databases. In *Proc. CIKM 2008*, 13–22. ACM.

Sherkhonov, E.; Cuenca Grau, B.; Kharlamov, E.; and Kostylev, E. V. 2017. Semantic faceted search with aggregation and recursion. In *Proc. ISWC 2017*, volume 10587 of *LNCS*, 594–610. Springer.

Virgilio, R. D.; Maccioni, A.; and Torlone, R. 2013. A similarity measure for approximate querying over RDF data. In *Proc. EDBT/ICDT Workshops 2013*, 205–213. ACM.