Taming Complex Role Inclusions for *DL-Lite*★

Medina Andreşel, Yazmín Ibáñez-García, Magdalena Ortiz, and Mantas Šimkus

Institute of Logic and Computation, TU Wien, Austria

Abstract. Hierarchical data, where facts may refer to different categories in an ordered hierarchy, in the style of the multi-dimensional data model, arises in many applications. When these hierarchies are not captured by subclasses, but require navigation along roles, this data cannot be satisfactorily queried in the standard OBDA setting based on first-order rewritable languages like *DL-Lite*. For this reason, we study how to extend *DL-Lite* with complex role inclusions (CRIs) in a way that overcomes this limitation. Complex role inclusions (CRIs) cause the loss of first-order rewritability in general, but we study meaningful restrictions which guarantee that rewritability is preserved.

1 Introduction

In *Ontology-based Data Access* (OBDA) [15] the knowledge represented in an ontology is leveraged to retrieve more complete answers from incomplete data. For example, consider the dataset $\mathcal{A}_e$ in Figure 1 about cultural events and their locations, and the ontology in Figure 2 which includes the knowledge that both concerts and exhibitions are cultural events. Using this knowledge, all cultural events ex_1, ev_1, and c_1 can be retrieved with a the query: $q_1(x) \leftarrow \text{CulturEvent}(x)$.

In OBDA, ontologies are often written in the Description Logics (DLs) of the *DL-Lite* family [6]. These DLs are tailored in such a way that CQs mediated by *DL-Lite* ontologies are first-order (FO)-rewritable. That means that evaluating a query q over $(\mathcal{T}, \mathcal{A})$ 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_1 is

$$q_\mathcal{T}(x) \leftarrow \text{CulturEvent}(x) \vee \text{Exhibition}(x) \vee \text{Concert}(x).$$

FO rewritability is important as it allows to implement OBDA by using standard database technologies. A missing functionality in OBDA is leveraging hierarchical knowledge not captured by subclass relations. Event locations may range from a specific venue, to more general locations such as the city or the country where the event occurs. But unlike our previous example, venue is not a subclass of city, and there is no natural way to express in *DL-Lite* that *if an event occurs in a venue located in a city, then it occurs in that city.* Therefore, when evaluated over $(\mathcal{T}_e, \mathcal{A}_e)$, the query q_2 retrieves ex_1 but not the expected c_1:

$$q_2(x) \leftarrow \text{CulturEvent}(x), \text{occursIn}(x, y), y = \mathsf{Vienna}.$$

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

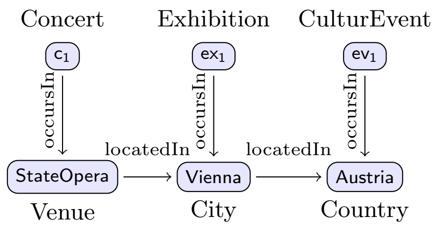


Fig. 1: Event dataset $\mathcal{A}_e$.

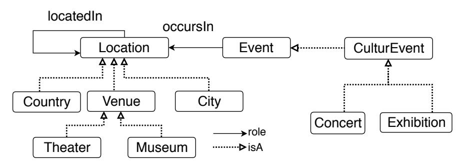
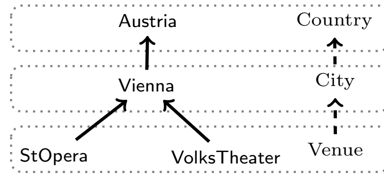


Fig. 2: Event ontology $\mathcal{T}_e$.

The previous is a prototypical example of *dimensional knowledge*: venues, cities, and countries can be seen as different levels in the Location dimension. Dimensional knowledge arises in many settings, and it is useful for storing and accessing data at different granularity levels. To store and query time-stamped data, a Time dimension including day, month, and year could be used; the physical parts of complex objects may be ordered along a hierarchy of components. These hierarchies are often called *dimensions*, and are formalized as a finite set of *categories* with a partial order between them. Figure 3, shows a formalization for a Location dimension. Modeling and leveraging dimensional knowledge in query answering has been a major research problem in the database community, we discuss some works in Section 5. In this paper, we make a step towards extending the OBDA setting to leverage dimensional knowledge. 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 the CRI occursIn $\cdot$ locatedIn $\sqsubseteq$ occursIn to our example captures the missing knowledge, and makes c_1 an answer to q_2.

While CRIs enables *DL-Lite* to capture hierarchical knowledge, their addition is in general computationally costly. They easily lead to undecidability if unrestricted [10], and critically for *DL-Lite*, even one fixed CRI destroys the FO-rewritability of CQs [1]. In this paper, we are interested in extensions of *DL-Lite* with CRIs able to capture dimensional knowledge, while still preserving FO-rewritability of CQs. Achieving both simultaneously is not easy though. While an FO-rewritable extension results from imposing suitable acyclicity conditions on the roles in CRIs, the combined complexity of standard reasoning becomes intractable, and more critically, the language cannot capture the desired scenarios: the CRIs we need to navigate along dimensions are in general recursive, as in the example above. We therefore allow recursion under some safety restrictions, in a way that we can ensure FO-rewritability and *under the assumption that datasets satisfy certain guarantees*, which effectively impose a bound on the length of paths over which CRIs provide relevant inferences. The interesting observation is that dimensions naturally provide such guarantees, bounding the propagation of dimensional knowledge. We introduce *order constraints* that naturally express dimensional information, and at the same time guarantee the

Fig. 3: Location dimension; the dashed arrows show the order between the categories Venue, City, and Country. Some members of each category are illustrated, and the solid arrows represent the role locatedIn.



boundedness required for FO-rewritability. Due to space restrictions, full proofs can be found in the long version[1].

2 Preliminaries

We consider the core fragment of *DL-Lite* with role inclusions $DL\text{-}Lite^{\mathcal{H}}$ [5,1]. As usual, $\mathsf{N_C}$,$\mathsf{N_R}$, and $\mathsf{N_I}$ are countable infinite alphabets of *concept*, *role*, and *individual* names, respectively. $DL\text{-}Lite^{\mathcal{H}}$ expressions are constructed according to the following grammar:

$$B := \bot \mid A \mid \exists r \qquad\qquad r := p \mid p^{-},$$

where $A \in \mathsf{N_C}$, $p \in \mathsf{N_R}$, B is called a *concept*, and p^{-} 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 $DL\text{-}Lite^{\mathcal{H}}$ TBox $\mathcal{T}$ is a finite set of concept inclusion axioms taking any of the following *normal forms*:

$$A \sqsubseteq A', \qquad A \sqsubseteq \exists p, \qquad \exists p \sqsubseteq A, \qquad p \sqsubseteq s, \qquad p \sqsubseteq s^{-},$$

together with a set of disjointness axioms of the form $\mathbf{disj}(A, A')$, and $\mathbf{disj}(p, p')$. For example, the ontology in Figure 2 is expressed with the $DL\text{-}Lite^{\mathcal{H}}$ TBox:

$$\begin{array}{rrr}
\exists\text{occursIn} \sqsubseteq \text{Event} & \exists\text{locationOf} \sqsubseteq \text{Location} & \text{Theater} \sqsubseteq \text{Venue} \\
\text{CulturEvent} \sqsubseteq \text{Event} & \text{City} \sqsubseteq \text{Location} & \text{Museum} \sqsubseteq \text{Venue} \\
\text{Exhibition} \sqsubseteq \text{CulturEvent} & \text{Country} \sqsubseteq \text{Location} & \text{locationOf} \sqsubseteq \text{occursIn}^{-} \\
\text{Concert} \sqsubseteq \text{CulturEvent} & \text{Venue} \sqsubseteq \text{Location} &
\end{array}$$

A $DL\text{-}Lite^{\mathcal{H}}$ *ABox* (or *dataset*) is a finite set of assertions $A(a)$, and $p(a, b)$, with $a, b \in \mathsf{N_I}$, $A \in \mathsf{N_C}$, and $p \in \mathsf{N_R}$, and we denote $ind(\mathcal{A})$ as the set of individuals occurring in $\mathcal{A}$. A knowledge base (KB) is a pair $\mathcal{K} = (\mathcal{T}, \mathcal{A})$. The semantics is defined as usual 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* in the sense that $a^{\mathcal{I}} = a$ for every $a \in \mathsf{N_I}$.

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 $\boldsymbol{x}$ and existential variables $\boldsymbol{y}$ that takes the form $q(\boldsymbol{x}) \leftarrow \varphi(\boldsymbol{x}, \boldsymbol{y})$, with φ a conjunction of *atoms* of the form $A(x), r(x, y)$, and $t = t'$, where $A \in \mathsf{N_C}$, $r \in \mathsf{N_R}$, and t, t' range over terms. *Instance queries* are CQs with exactly one atom and no existential variables. The terms occurring

[1] https://arxiv.org/abs/1808.02850

in q are denoted $terms(q)$, and the variables $vars(q)$. The free variables $\boldsymbol{x}$ of a query are called *answer variables*.

Let $\mathcal{I}$ be an interpretation, $q(\boldsymbol{x})$ a CQ. An *answer to* q *in* $\mathcal{I}$ is a tuple $\boldsymbol{a}$ from $\Delta^{\mathcal{I}}$ of length $|\boldsymbol{x}|$ such that there is a map $\pi : terms(q) \mapsto \Delta^{\mathcal{I}}$ satisfying (i) $\pi(\boldsymbol{x}) = \boldsymbol{a}$, (ii) $\pi(b) = b$ for each individual b, (iii) $\mathcal{I} \models P(\pi(\boldsymbol{z}))$ for each atom $P(\boldsymbol{z})$ in q, and (iv) $\pi(t) = \pi(t')$ for each atom $t = t'$ in q, and in that case we write $\mathcal{I} \models q(\boldsymbol{a})$. The map π is called a *match* for q in $\mathcal{I}$. We denote $ans(q(\boldsymbol{x}), \mathcal{I})$ as the set of all answers to q in $\mathcal{I}$. The *certain answers* of $q(\boldsymbol{x})$ over $\mathcal{A}$ w.r.t. $\mathcal{T}$, denoted $cert(q, \mathcal{T}, \mathcal{A})$, is defined as the tuples of individuals that are an answer to q in $\mathcal{I}$, for every model $\mathcal{I}$ of $(\mathcal{T}, \mathcal{A})$.

3 DL-Lite with Complex Role Inclusions

A *complex role inclusion* (CRI) is an expression of the form $r \cdot s \sqsubseteq t$, with $r, s, t \in \mathsf{N_R}$. An interpretation $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ *satisfies* a CRI $r \cdot s \sqsubseteq t$ if for all $d_1, d_2, d_3 \in \Delta^{\mathcal{I}}$, $(d_1, d_2) \in r^{\mathcal{I}}$, $(d_2, d_3) \in s^{\mathcal{I}}$ imply $(d_1, d_3) \in t^{\mathcal{I}}$. 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 $s^- \in \mathsf{N_{R_s}}$); for each $r \in \mathsf{N_R}^{\pm} \setminus \mathsf{N_{R_s}}$, r is a *non-simple* role.

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 such that:*

- *For every CRI* $r \cdot s \sqsubseteq t \in \mathcal{T}$, s *is simple and* t *is non-simple.*
- *If* $s \sqsubseteq t \in \mathcal{T}$ *and* $t \in \mathsf{N_{R_s}}$, *then* $s \in \mathsf{N_{R_s}}$.

The restriction to a simple role s in Definition 1 guarantees that recursion is *linear*, avoiding a possible explosion in the size of rewritings.
Properties such as FO-rewritability are affected by CRIs. In the case of *DL-Lite*, even one single fixed CRI $r \cdot s \sqsubseteq r$ destroys first-order rewritability, since it can easily enforce r to capture reachability along the s-edges of a given graph.

Lemma 1. *[1] Instance checking in DL-Lite*$^{\mathcal{HR}}$ *is* NLogSpace*-hard in data complexity, already for TBoxes consisting of the CRI* $r \cdot s \sqsubseteq r$ *only.*

3.1 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.

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

Restricting CRIs to be non-recursive indeed guarantees FO-rewritability. For a CQ q, we denote by z^q an arbitrary but fixed variable not occurring in q. 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 applying $\rightsquigarrow_{\mathcal{T}}$ exhaustively, we obtain a FO-rewriting of a given query q.

Definition 4. *The* rewriting of q w.r.t. $\mathcal{T}$ *is the $rew(q, \mathcal{T}) = \{q' \mid q \rightsquigarrow_{\mathcal{T}}^* q'\}$ such that for each $q' \in rew(q, \mathcal{T})$ there is no $q'' \in rew(q, \mathcal{T})$ isomorphic to q', where $q \rightsquigarrow_{\mathcal{T}}^* q'$ is the reflexive, transitive closure of $q \rightsquigarrow_{\mathcal{T}} q'$.*

For any CQ q, $rew(q, \mathcal{T})$ is a finite query that can be effectively computed.

Lemma 2. *Let $\mathcal{T}$ be a DL-Lite$^{\mathcal{HR}}_{\mathsf{non\text{-}rec}}$ TBox and let q a CQ. Each $q' \in rew(q, \mathcal{T})$ is polynomially bounded in the size of $\mathcal{T}$ and q, and can be obtained in a polynomial number of steps.*

Proof (sketch). Due to the non-recursiveness of the dependency graph and the restriction on simple roles, we show that we can assign to queries a (suitably bounded) degree that roughly corresponds to the number of rewriting steps that can be further applied. We prove that for each q' such that $q \rightsquigarrow_{\mathcal{T}}^* q'$, the degree does not increase, and after polynomially many steps we will reach $q \rightsquigarrow_{\mathcal{T}}^* q''$ such that the degree strictly decreases. □

The next result is shown analogously as in [6], extended to the new rule **S6**.

Theorem 1. *Let $\mathcal{T}$ be a DL-Lite$^{\mathcal{HR}}_{\mathsf{non\text{-}rec}}$ TBox, q a CQ. For every ABox $\mathcal{A}$ consistent with $\mathcal{T}$ we have that $cert(q, \mathcal{T}, \mathcal{A}) = \bigcup_{q' \in rew(q, \mathcal{T})} cert(q', \emptyset, \mathcal{A})$.*

Non-recursive CRIs preserve FO-rewritability of *DL-Lite*, but their addition is far from harmless. Indeed, unlike the extension with transitive roles, even non-recursive CRIs increase the complexity of testing KB consistency.

Theorem 2. *Consistency checking in DL-Lite$^{\mathcal{HR}}_{\mathsf{non\text{-}rec}}$ is coNP-complete.*

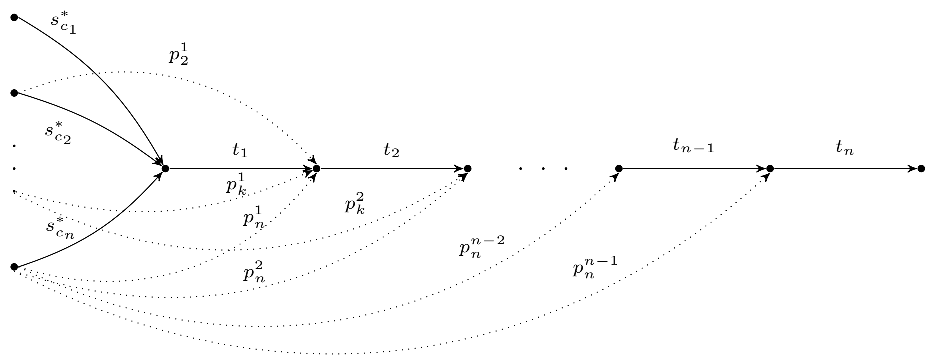


Fig. 4: Propagating the clauses that are satisfied under a variable assignment.

Proof. Upper-bound: Similarly as for standard *DL-Lite*, inconsistency checking can be reduced to UCQ answering, using a CQ q_α for testing whether each disjointness axiom α is violated. By Lemmas 2 and 1, an NP procedure can guess one such q_α, guess a q'_α in its rewriting, and evaluate q'_α over $\mathcal{A}$.

Lower-bound: We reduce the complement of 3SAT to KB satisfiability. Suppose we are given a conjunction $\varphi = c_1 \wedge \cdots \wedge c_n$ of clauses of the form $\ell_{i_1} \vee \ell_{i_2} \vee \ell_{i_3}$, where the ℓ_k are literals, i.e., propositional variables or their negation. Let $x_0, \ldots, x_m$ be all the propositional variables occurring in φ. In order to encode the possible truth assignments of each variable x_i, we take two fresh roles r_{x_i} and $\bar{r}_{x_i}$, intended to be disjoint. We construct a $\textit{DL-Lite}^{\mathcal{HR}}_{\textsf{non-rec}}$ TBox $\mathcal{T}_\varphi$ containing, for every $0 \leq i \leq m$, the following axioms:

$$\begin{array}{llll} \mathbf{disj}(r_{x_i}, \bar{r}_{x_i}), & A_i \sqsubseteq \exists r_{x_i} \sqcap \exists \bar{r}_{x_i}, & \exists r^-_{x_i} \sqsubseteq A_{i+1}, & \exists(\bar{r}_{x_i})^- \sqsubseteq A_{i+1}, \\ & r_{x_i} \sqsubseteq t, & \bar{r}_{x_i} \sqsubseteq t & \end{array}$$

These axioms have a model that is a full binary tree, rooted at A_0 and whose edges are labeled with the role t, and with different combinations of the roles r_i and $\bar{r}_{x_i}$. Intuitively, each path represents a possible variable truth assignment. Further, $\mathcal{T}_\varphi$ contains axioms relating each variable assignment with the clauses it satisfies, using roles $s_{c_1}, \ldots, s_{c_n}$. More precisely, we have the following role inclusions for $0 \leq i \leq m$, and $1 \leq j \leq n$:

$$r_{x_i} \sqsubseteq s_{c_j}, \quad \text{if } x_i \in c_j \qquad\qquad \bar{r}_{x_i} \sqsubseteq s_{c_j}, \quad \text{if } \neg x_i \in c_j \tag{1}$$

To encode the evaluation of all clauses, we have axioms propagating down the tree all clauses satisfied by some assignment. Note that we could do this easily using a CRI such as $s_{c_j} \cdot t \sqsubseteq s_{c_j}$. However, this would need a recursive role s_{c_j}. Since the depth of the assignment tree is bounded by m, we can encode this (bounded) propagation using at most m roles $s^i_{c_j}$ $(1 \leq i \leq n)$ for each clause c_j, which will be declared as subroles of another role $s^*_{c_j}$. For $1 \leq j \leq n$ and $1 \leq i < m$, we have the CRIs: $s_{c_j} \cdot t \sqsubseteq s^1_{c_j}$, $\quad s^i_{c_j} \cdot t \sqsubseteq s^{i+1}_{c_j}$, $\quad s^i_{c_j} \sqsubseteq s^*_{c_j}$.

Thus, if c_j is satisfied in a t-branch of the assignment tree, its leaf will have an incoming $s^*_{c_j}$ edge. Now, in order to encode that there is at least one clause that is not satisfied, we need to forbid the existence of a leaf satisfying the concept $\exists(s^*_{c_1})^- \sqcap \cdots \sqcap \exists(s^*_{c_n})^-$. This cannot be straightforwardly written in $\mathit{DL\text{-}Lite}^{\mathcal{HR}}_{\mathsf{non\text{-}rec}}$, but we resort again to CRIs to propagate information:

$$\exists(s^*_{c_1})^- \sqsubseteq \exists t_1 \qquad s^*_{c_k} \cdot t_1 \sqsubseteq p^1_k, \text{ for } 2 \leq k \leq n \tag{2}$$

$$\exists(p^{i-1}_i)^- \sqsubseteq \exists t_i \qquad p^{i-1}_k \cdot t_i \sqsubseteq p^i_k, \text{ for } 2 \leq i \leq n,\ i < k \leq n \tag{3}$$

Figure 4 shows how the satisfaction of clauses is propagated using these axioms. Lastly, by adding the axiom $\exists t_n \sqsubseteq \bot$, we obtain the required restriction. We use this in the full proof that φ is unsatisfiable iff $(\mathcal{T}_\varphi, \{A_0(a)\})$ is satisfiable. □

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

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

Definition 5 (recursion safe *DL-Lite*$^{\mathcal{HR}}$). *A* $\mathit{DL\text{-}Lite}^{\mathcal{HR}}_{\mathsf{rec\text{-}safe}}$ *TBox* $\mathcal{T}$ *is a* $\mathit{DL\text{-}Lite}^{\mathcal{HR}}$ *TBox where every CRI* $r_1 \cdot s \sqsubseteq r_2 \in \mathcal{T}$ *satisfies the following conditions:*

- *If* r_2 *is recursive, 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* $s_2 \in \mathsf{N_{R_s}}$*.*

The key idea behind recursion safety is that every recursive CRI is 'guarded' by a simple role that is not existentially implied. For query answering, we can then assume that only ABox individuals are connected by these guarding roles, and thus CRIs only 'fire' close to the ABox (that is, each pair in the extension of a recursive role has at least one individual). In fact, we show below that every consistent recursion-safe KB has a model where both conditions hold.

Example 1. $\mathcal{K}_e$ is recursion safe, since occursIn $\cdot$ locatedIn $\sqsubseteq$ occursIn is the only CRI, and locatedIn is not implied by any existential axiom in $\mathcal{T}_e$.

In $\mathit{DL\text{-}Lite}^{\mathcal{HR}}_{\mathsf{rec\text{-}safe}}$, consistency checking and instance query answering are tractable. In fact, for a given KB, we can build a polynomial-sized interpretation that is a model whenever the KB is consistent, and that can be used for testing entailment of assertions and of disjointness axioms.

Definition 6. *Let* $(\mathcal{T}, \mathcal{A})$ *be a* $\mathit{DL\text{-}Lite}^{\mathcal{HR}}_{\mathsf{rec\text{-}safe}}$ *KB. We define an interpretation* $\mathcal{E}_{\mathcal{T},\mathcal{A}}$ *as follows. As domain* $\Delta^{\mathcal{E}_{\mathcal{T},\mathcal{A}}} = D_0 \cup D_1 \cup D_2$ *we use the individuals in* $\mathcal{A}$*, fresh individuals* c_{ar} *that serve as* r*-fillers for individual* a*, and fresh individuals* c_r *that serve as shared* r*-fillers for the objects that are not individuals in* $\mathcal{A}$*:*

$$D_0 = ind(\mathcal{A}), \qquad D_1 = \{c_{ar} \mid a \in D_0, r \text{ occurs on the rhs of a CI in } \mathcal{T}\},$$
$$D_2 = \{c_r \mid r \text{ occurs on the rhs of a CI in } \mathcal{T}\}.$$

The interpretation function has $a^{\mathcal{E}_{\mathcal{T},\mathcal{A}}} = a$ for each $a \in \Delta^{\mathcal{E}_{\mathcal{T},\mathcal{A}}}$, and assigns to each concept name A and each role name r in $\Sigma_{\mathcal{T}}$ the minimal set of the form $A^{\mathcal{E}_{\mathcal{T},\mathcal{A}}} \subseteq \Delta^{\mathcal{E}_{\mathcal{T},\mathcal{A}}}$, $r^{\mathcal{E}_{\mathcal{T},\mathcal{A}}} \subseteq \Delta^{\mathcal{E}_{\mathcal{T},\mathcal{A}}} \times \Delta^{\mathcal{E}_{\mathcal{T},\mathcal{A}}}$ such that the following conditions hold, for all $A \in \mathsf{N_C}$, B a concept, and $r, r_1, r_2, s, t \in \mathsf{N_R}$:

1. *$A(a) \in \mathcal{A}$ implies $a \in A^{\mathcal{E}_{\mathcal{T},\mathcal{A}}}$, and $r(a,b) \in \mathcal{A}$ implies $(a,b) \in r^{\mathcal{E}_{\mathcal{T},\mathcal{A}}}$.*
2. *For each $B \sqsubseteq \exists r \in \mathcal{T}$, $a \in B^{\mathcal{E}_{\mathcal{T},\mathcal{A}}} \cap D_0$ implies $(a, c_{ar}) \in r^{\mathcal{E}_{\mathcal{T},\mathcal{A}}}$.*
3. *For each $B \sqsubseteq \exists r \in \mathcal{T}$, $d \in B^{\mathcal{E}_{\mathcal{T},\mathcal{A}}} \cap (D_1 \cup D_2)$ implies $(d, c_r) \in r^{\mathcal{E}_{\mathcal{T},\mathcal{A}}}$.*
4. *For each $B \sqsubseteq A \in \mathcal{T}$, $d \in B^{\mathcal{E}_{\mathcal{T},\mathcal{A}}}$ implies $d \in A^{\mathcal{E}_{\mathcal{T},\mathcal{A}}}$.*
5. *For each $r_1 \sqsubseteq r_2 \in \mathcal{T}$, $(a,b) \in r_1^{\mathcal{E}_{\mathcal{T},\mathcal{A}}}$ implies $(a,b) \in r_1^{\mathcal{E}_{\mathcal{T},\mathcal{A}}}$.*
6. *For each $r \cdot s \sqsubseteq t \in \mathcal{T}$, $(a,b) \in r^{\mathcal{E}_{\mathcal{T},\mathcal{A}}}$ and $(b,c) \in s^{\mathcal{E}_{\mathcal{T},\mathcal{A}}}$ imply $(a,c) \in t^{\mathcal{E}_{\mathcal{T},\mathcal{A}}}$.*

For $\mathcal{E}_{\mathcal{T},\mathcal{A}}$, we can show the following useful properties:

Proposition 1. *Let $\mathcal{T} = \mathcal{T}_p \cup \mathcal{T}_n$ be a DL-Lite$^{\mathcal{HR}}_{\mathsf{rec\text{-}safe}}$ TBox, where $\mathcal{T}_p$ contains only positive inclusions, and $\mathcal{T}_n$ contains only disjointness axioms. Then, for every ABox $\mathcal{A}$:*

P1 *If $(\mathcal{T},\mathcal{A})$ is consistent, then $\mathcal{E}_{\mathcal{T},\mathcal{A}} \models (\mathcal{T},\mathcal{A})$.*
P2 *$(\mathcal{T},\mathcal{A})$ is inconsistent iff $\mathcal{E}_{\mathcal{T},\mathcal{A}} \not\models \alpha$ for some $\alpha \in \mathcal{T}_n$.*
P3 *If $(\mathcal{T},\mathcal{A})$ is consistent and q is an instance query, $cert(q,\mathcal{T},\mathcal{A}) = ans(q,\mathcal{E}_{\mathcal{T},\mathcal{A}})$.*

Proof (sketch). To prove **P1**, we assume that $(\mathcal{T},\mathcal{A})$ is consistent. Verifying that $\mathcal{E}_{\mathcal{T},\mathcal{A}}$ satisfies all but the disjointness axioms is straightforward. Let $\mathcal{I}$ be an arbitrary model of $(\mathcal{T},\mathcal{A})$. For $d, d' \in \Delta^{\mathcal{I}}$, let $\mathsf{tp}_{\mathcal{I}}(d) = \{B \mid d \in B^{\mathcal{I}}\}$ the set of concepts satisfied at d in $\mathcal{I}$, and $\mathsf{tp}_{\mathcal{I}}(d,d') = \{r \mid (d,d') \in r^{\mathcal{I}}\}$, the set of roles connecting d and d' in $\mathcal{I}$. We show the following claim:

Claim. For any given $d \in \Delta^{\mathcal{E}_{\mathcal{T},\mathcal{A}}}$ (i) there exists $e \in \Delta^{\mathcal{I}}$ such that $\mathsf{tp}_{\mathcal{E}_{\mathcal{T},\mathcal{A}}}(d) \subseteq \mathsf{tp}_{\mathcal{I}}(e)$ and (ii) for each $d' \in \Delta^{\mathcal{E}_{\mathcal{T},\mathcal{A}}}$ such that $\mathsf{tp}_{\mathcal{E}_{\mathcal{T},\mathcal{A}}}(d,d') \neq \emptyset$ we have that there exists $e' \in \Delta^{\mathcal{I}}$ such that $\mathsf{tp}_{\mathcal{E}_{\mathcal{T},\mathcal{A}}}(d,d') \subseteq \mathsf{tp}_{\mathcal{I}}(e,e')$.

Towards a contradiction, assume there is $\alpha = \mathbf{disj}(B_1,B_2) \in \mathcal{T}$ such that $\mathcal{E}_{\mathcal{T},\mathcal{A}} \not\models \alpha$; the case of role disjointness axioms is analogous. Then there is $d \in \Delta^{\mathcal{E}_{\mathcal{T},\mathcal{A}}}$ with $B_1, B_2 \in \mathsf{tp}_{\mathcal{E}_{\mathcal{T},\mathcal{A}}}$, and by the claim above, $B_1, B_2 \in \mathsf{tp}_{\mathcal{I}}(d)$ for each model $\mathcal{I}$. Hence $\mathcal{E}_{\mathcal{T},\mathcal{A}} \models \alpha$, and this concludes proof of **P1**; **P2** and **P3** can also be shown using the above claim and the fact that $\mathcal{E}_{\mathcal{T},\mathcal{A}}$ is a model of the KB. □

This proposition allows us to establish the following results:

Theorem 3. *KB consistency and instance query answering in recursion safe DL-Lite$^{\mathcal{HR}}$ are in* PTIME *for combined complexity.*

The recursion safe fragment of *DL-Lite*$^{\mathcal{HR}}$ is not FO-rewritable: indeed, the TBox in the proof of Lemma 1 is recursion safe. However, we can get rid of recursive CRIs and regain rewritability if we have guarantees that they will only be relevant on paths of bounded length. We formalize this rough intuition next.

Input: $(\mathcal{T}, \mathcal{A})$ satisfiable recursion safe $DL\text{-}Lite^{\mathcal{HR}}$ KB, $\mathcal{C}$ - order constraints;
Output: **true** if $(\mathcal{A}, \mathcal{T})$ is $\mathcal{C}$-admissible, **false** otherwise;
foreach $ord(s, \mathbf{A}, \prec) \in \mathcal{C}$ **do**
    $q_1(x, y) \leftarrow s(x, y), \quad q_2(x, y) \leftarrow \bigvee_{A_1 \prec A_2} A_1(x), s(x, y), A_2(y)$;
    if $ans(q_1, \mathcal{E}_{\mathcal{T},\mathcal{A}}) \not\subseteq ans(q_2, \mathcal{E}_{\mathcal{T},\mathcal{A}})$ **then** **return false** ;
    $q_3(x, y) \leftarrow \bigvee_{A_1 \not\prec A_2} A_1(x), s(x, y), A_2(y)$;
    if $ans(q_3, \mathcal{E}_{\mathcal{T},\mathcal{A}}) \neq \emptyset$ **then** **return false** ;
return true.

Algorithm 1: CheckAdmissibility

Definition 7 (k-bounded ABox). *Let $\mathcal{T}$ be a $DL\text{-}Lite^{\mathcal{HR}}$ TBox and $\mathcal{A}$ an ABox. Let S be a set of simple roles. Given $a, b \in ind(\mathcal{A})$, we say that there exists an S-*path of length n between a and b (in $\mathcal{A}$ w.r.t. $\mathcal{T}$) *if there exist pairwise distinct $d_1, \ldots, d_{n-1} \in ind(\mathcal{A})$ with $d_i \notin \{a, b\}$, and $s_1(a, d_1), \ldots, s_i(d_{i-1}, d_i), \ldots, s_n(d_{n-1}, b) \in \mathcal{A}$ such that $s_i \sqsubseteq^{\mathsf{s}}_{\mathcal{T}} s$ and $s \in S$, $1 \leq i < n$. Let $S_r = \{s \mid r \cdot s \sqsubseteq r \in \mathcal{T}\}$. We say that $\mathcal{A}$ is* k-bounded for $\mathcal{T}$ *if for each recursive $r \in \mathcal{T}$ there is no S_r-path of size larger than k.*

We simulate recursive CRIs by unfolding them into k non-recursive ones.

Definition 8 (k-unfolding). *For an arbitrary $DL\text{-}Lite^{\mathcal{HR}}_{\mathsf{rec\text{-}safe}}$ TBox $\mathcal{T}$, and fixed $k \geq 0$, a* k-unfolding of $\mathcal{T}$ *is a $DL\text{-}Lite^{\mathcal{HR}}_{\mathsf{non\text{-}rec}}$ TBox $\mathcal{T}_k$ obtained by replacing each $r \cdot s \sqsubseteq r \in \mathcal{T}$ with the axioms:*

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

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.

Lemma 3. *Let $\mathcal{T}$ be a $DL\text{-}Lite^{\mathcal{HR}}_{\mathsf{rec\text{-}safe}}$ TBox, $\mathcal{T}_k$ a k-unfolding of $\mathcal{T}$ for some $k \geq 0$, and let q be a CQ over $\Sigma_{\mathcal{T}}$. Then, for every k-bounded ABox $\mathcal{A}$:*

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

Proof (sketch). In a nutshell, recursion-safety ensures that recursive CRIs in $\mathcal{T}$ can only 'fire' in the chase along S_r-paths in the ABox. If $\mathcal{A}$ is k-bounded for $\mathcal{T}$, then such paths have length $\leq k$, so we get that every pair (d, d') that should be added to a recursive role r is added to some r_j, and hence to $\hat{r}$. □

4 Taming CRIs for Dimensional Data

We introduce *order constraints* to express the order between categories of a dimension.

Definition 9. *An* order constraint *takes the form* $ord(s, \mathbf{A}, \prec)$, *with* $s \in \mathsf{N}_{\mathsf{R_s}}$, $\mathbf{A} \subseteq \mathsf{N_C}$ *finite, and* $\prec$ *a strict partial order over* $\mathbf{A}$. $\mathcal{I}$ *satisfies* $ord(s, \mathbf{A}, \prec)$ *if*

$$s^{\mathcal{I}} \subseteq \bigcup_{A_1, A_2 \in \mathbf{A}} (A_1^{\mathcal{I}} \times A_2^{\mathcal{I}}), \qquad (4) \qquad\qquad s^{\mathcal{I}} \cap \bigcup_{A_1 \not\prec A_2} (A_1^{\mathcal{I}} \times A_2^{\mathcal{I}}) = \emptyset. \qquad (5)$$

Intuitively, if $ord(s, \mathbf{A}, \prec)$ is satisfied in $\mathcal{I}$, then all objects connected via role s are instances of $\mathbf{A}$-concepts, in away that is compliant with the order $\prec$. Equation 5 disallows also s-paths that are order compliant but connect instances of $\mathbf{A}$-concepts which are incomparable w.r.t. $\prec$. If such paths are allowed, one cannot guarantee k-boundednes. An alternative solution would have been to restrict concepts in $\mathbf{A}$ to be disjoint.

Example 2. The Location dimension from Figure 3 is captured by $\mathcal{K}_e = (\mathcal{T}_e, \mathcal{A}_e)$ and the constraint

$$\mathrm{c} = ord(\mathrm{locatedIn}, \{\mathrm{Venue}, \mathrm{City}, \mathrm{Country}\}, \prec) \qquad (6)$$

with Venue $\prec$ City $\prec$ Country. In each model of $\mathcal{K}_e$ satisfying c, the role locatedIn 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. △

An useful insight is that order constraints can provide k-bounded guarantees.

Definition 10. $\mathcal{C}$ 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 ordering constraint c from (6), covers $\mathcal{T}_e$, and $\mathcal{K}_e$ is $\{\mathrm{c}\}$-admissible since $\mathcal{E}_{\mathcal{T}_e,\mathcal{A}_e}$ satisfies c.

Lemma 4. *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* $\ell(\mathcal{C}) := max\{|\mathbf{A}| \mid ord(s, \mathbf{A}, \prec) \in \mathcal{C}\}$.

Proof (sketch). For any $\mathcal{I}$, if $\mathcal{I} \models ord(s, \mathbf{A}, \prec)$, for each chain of individuals $a_1, \ldots, a_n$ with $(a_i, a_{i+1}) \in s^{\mathcal{I}}$ for all $1 \leq j < n$, we have $n \leq |\mathbf{A}|$. This applies to $\mathcal{E}_{\mathcal{T},\mathcal{A}}$, as $(\mathcal{T}, \mathcal{A})$ is $\mathcal{C}$-admissible. Further, $\mathcal{C}$ covers $\mathcal{T}$, so for each $S_r = \{s \mid r \cdot s \sqsubseteq r \in \mathcal{T}\}$, all S_r-paths in $\mathcal{E}_{\mathcal{T},\mathcal{A}}$ have size $\leq \ell(\mathcal{C})$. Finally, all S_r-paths in $\mathcal{A}$ w.r.t. $\mathcal{T}$ are also in $\mathcal{E}_{\mathcal{T},\mathcal{A}}$, so their length is $\leq \ell(\mathcal{C})$. □

Lemmas 3 and 4 give us the desired result: we obtain FO-rewritability in the presence of CRIs, whenever order constraints allow us to guarantee boundedness.

Theorem 4. *Let $\mathcal{T}$ be a DL-Lite$^{\mathcal{HR}}_{\mathsf{rec\text{-}safe}}$ TBox, $\mathcal{C}$ a set of order constraints that covers $\mathcal{T}$, and q a CQ. Let $q_{\mathcal{C}}$ 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 consistent and $\mathcal{C}$-admissible, $cert(q, \mathcal{T}, \mathcal{A}) = cert(q_{\mathcal{C}}, \emptyset, \mathcal{A})$.*

Finally, we note that $\mathcal{C}$-admissibility amounts to evaluate simple queries on $\mathcal{E}_{\mathcal{T},\mathcal{A}}$. This can be done in time that is polynomial in $\mathcal{C}$, $\mathcal{T}$, and $\mathcal{A}$, using the procedure in Algorithm 1. Moreover, although testing $\mathcal{C}$-admissibility is data dependent, once it is established, FO-rewritability is guaranteed for any CQ.

Proposition 2. *Checking $\mathcal{C}$-admissibility for recursion-safe DL-Lite$^{\mathcal{HR}}$ KBs is in* PTIME *in combined complexity.*

5 Related Work

CRIs have been studied since the earliest DL research, when *role value maps* where considered very desirable [16]. Indeed, CRIs are part of the OWL standard, both in the OWL EL profile which is based on $\mathcal{EL}^{++}$ [2], and in full OWL 2 which is based on $\mathcal{SROIQ}$ [9]. Our work is also related to *regular path queries* (RPQs) and their extensions. In fact, ontology mediated query answering where the DL has CRIs, is also supported in settings where the query language contains conjunctive RPQs; many such settings have been considered in the literature and their complexity is well understood, see [13,14] for references. The latter are necessarily NLOGSPACE-hard in data complexity, and PSPACE-hard in combined complexity even for lightweight DLs [4]. We have focused on FO-rewritable settings with tractable combined complexity. Rule-based formalisms for OBDA that can leverage dimensional knowledge include weakly-sticky Datalog$^{\pm}$ [3], but those formalisms are not FO-rewritable.

The notion of dimension used here is basis of the *multi-dimensional data model* used for online-analytical processing (OLAP) [11]. Logic-based formalizations of dimensions and multi-dimensional data schemata have been proposed in the literature. Some works focus on modeling such data and use DLs to reason about the models, rather than for querying [8,7]. A recent work in the database area focuses on operators for taxonomy-based relaxation of queries over relational data [12]. Our work is closely related to [3], but as mentioned, they rely on an expressive fragment of Datalog$^{\pm}$ for which CQs are not FO-rewritable.

6 Conclusions

In this paper we have advocated to use CRIs for getting more complete answers in the OBDA setting, particularly in the presence of dimensional knowledge. We extended *DL-Lite*$^{\mathcal{H}}$ with CRIs. Severe restrictions are needed to preserve FO-rewritability, but we have identified a setting that is both natural and useful for the desired use case. An investigation of *DL-Lite*$^{\mathcal{HR}}$ without the simple roles restriction is left for future work. Finding effective ways to guarantee $\mathcal{C}$-admissibility, and identifying settings in which it can be tested via FO-queries, also seems to be an interesting challenge.

References

1. Artale, A., Calvanese, D., Kontchakov, R., Zakharyaschev, M.: The DL-Lite family and relations. J. Artif. Int. Res. **36**(1), 1–69 (Sep 2009)
2. Baader, F., Brandt, S., Lutz, C.: Pushing the EL envelope. In: Proceedings of the 19th International Joint Conference on Artificial Intelligence. pp. 364–369. IJCAI'05 (2005)
3. Bertossi, L., Milani, M.: Ontological multidimensional data models and contextual data quality. J. Data and Information Quality **9**(3), 14:1–14:36 (Jan 2018)
4. Bienvenu, M., Ortiz, M., Šimkus, M.: Regular path queries in lightweight description logics: Complexity and algorithms. J. Artif. Int. Res. **53**(1), 315–374 (May 2015)
5. Calvanese, D., De Giacomo, G., Lembo, D., Lenzerini, M., Rosati, R.: Data complexity of query answering in description logics. In: KR. pp. 260–270. AAAI Press (2006)
6. Calvanese, D., De Giacomo, G., Lembo, D., Lenzerini, M., Rosati, R.: Tractable reasoning and efficient query answering in description logics: The *DL-Lite* family. J. Autom. Reasoning **39**(3), 385–429 (2007)
7. Franconi, E., Kamble, A.: The $\mathcal{GMD}$ data model and algebra for multidimensional information. In: Persson, A., Stirna, J. (eds.) Advanced Information Systems Engineering. pp. 446–462. Springer Berlin Heidelberg, Berlin, Heidelberg (2004)
8. Franconi, E., Sattler, U.: A data warehouse conceptual data model for multidimensional aggregation. In: In Proceedings of the Workshop on Design and Management of Data Warehouses (DMDW99 (1999)
9. Horrocks, I., Kutz, O., Sattler, U.: The even more irresistible SROIQ. In: Proceedings of the Tenth International Conference on Principles of Knowledge Representation and Reasoning. pp. 57–67. KR'06, AAAI Press (2006)
10. Horrocks, I., Sattler, U.: Decidability of SHIQ with complex role inclusion axioms. Artif. Intell. **160**(1-2), 79–104 (Dec 2004)
11. Hurtado, C.A., Mendelzon, A.O.: OLAP dimension constraints. In: Proceedings of the Twenty-first ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems. pp. 169–179. PODS '02, ACM, New York, NY, USA (2002)
12. Martinenghi, D., Torlone, R.: Taxonomy-based relaxation of query answering in relational databases. VLDB J. **23**(5), 747–769 (2014)
13. Ortiz, M.: Ontology based query answering: The story so far. In: AMW. CEUR Workshop Proceedings, vol. 1087. CEUR-WS.org (2013)
14. Ortiz, M., Simkus, M.: Reasoning and query answering in description logics. In: Reasoning Web. Lecture Notes in Computer Science, vol. 7487, pp. 1–53. Springer (2012)
15. Poggi, A., Lembo, D., Calvanese, D., De Giacomo, G., Lenzerini, M., Rosati, R.: Linking data to ontologies. J. Data Semantics **10**, 133–173 (2008)
16. Schmidt-Schaubß, M.: Subsumption in KL-ONE is undecidable. In: Proceedings of the First International Conference on Principles of Knowledge Representation and Reasoning. pp. 421–431. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA (1989)