Combining DL-Lite with Spatial Calculi for Feasible Geo-thematic Query Answering

Özgür Lütfü Özçep and Ralf Möller

Institute for Software Systems (STS)
Hamburg University of Technology
Hamburg, Germany
{oezguer.oezcep,moeller}@tu-harburg.de

Abstract. First order logic (FOL) rewritability is a desirable feature for query answering over geo-thematic ontologies because in most geo-processing scenarios one has to cope with large data volume. Hence, there is a need for combined geo-thematic logics that provide a sufficiently expressive query language allowing for FOL rewritability. The DL-Lite family of description logics is tailored towards FOL rewritability of query answering for unions of conjunctive queries, hence it is a suitable candidate for the thematic component of a combined geo-thematic logic. We show that a weak coupling of DL-Lite with the expressive region connection calculus RCC8 allows for FOL rewritability under a spatial completeness condition for the ABox. Stronger couplings allowing for FOL rewritability are possible only for spatial calculi as weak as the low-resolution calculus RCC2. Already a strong combination of DL-Lite with the low-resolution calculus RCC3 does not allow for FOL rewritability.

Keywords: FOL rewritability, region connection calculus, qualitative spatial reasoning, GIS, combinations

1 Introduction

Query answering over a database becomes far more difficult if the extensional knowledge in the database is extended by constraints in an ontology. The reason is that a database plus an ontology may have many different models, hence ontology based query answering has to compute the answers w.r.t. to all models and build their intersection (*certain answer* semantics). But in some cases—when using a lightweight logic like DL-Lite for the representation of the ontology and a restricted query language like unions of conjunctive queries—query answering w.r.t. an ontology can be reduced to model checking. This is formalized by the notion of *FOL (first order logic) rewritability*: a given query can be rewritten into a FOL query in which the intensional knowledge of the ontology is captured. Though the rewritten queries may become exponentially bigger than the original queries, there exist optimizations [13]. So, FOL rewritability means a benefit.

DL-Lite per se [2] is not sufficient for use in scenarios of geographic information processing, as these demand among others the representation and deduction over spatial concepts. Though there exists related work combining spatial and thematic reasoning [4], [15], [6], it is not aimed at FOL rewritability. Hence, there is still a need for investigating combinations of logics that, on the one hand, are sufficiently expressive to fit the representation's needs in geographical information processing and that, on the other hand, allow for computationally feasible (in particular FOL rewritable) satisfiability checking and query answering.

Continuing previous work [7, 8], we investigate combinations of logics in the DL-Lite family with different members of the region connection calculus (RCC) family [9], a well-known family of calculi for qualitative spatial reasoning. The DL-Lite logic we are using as thematic part differs from the known members of the DL-Lite family as it also allows for concept conjunctions on the left-hand side of general inclusion axioms. In previous work [8], we focussed on the FOL rewritability aspects for weak combinations of DL-Lite with RCC8; these combinations are weak in so far as they do not allow for the construction of arbitrary RCC8 constraint networks in the intensional part (TBox) of the ontology. In this paper we focus on strong combinations of DL-Lite with the weaker fragments RCC3 and RCC2, and prove that DL-Lite$_{\mathcal{F},\mathcal{R}}^{\square}$(RCC3) does not allow for FOL rewritability of satisfiability checking while the weaker DL-Lite$_{\mathcal{F},\mathcal{R}}^{\square}$(RCC2) does.

The paper is structured as follows. Section 2 collects technical details on the region connection calculus and the DL-Lite family of description logics. Weak combinations of DL-Lite with the region connection calculus are described in Sect. 3. In Sect. 4, the last section before the conclusion, we consider strong combinations of DL-Lite with weaker fragments of the region connection calculus.

2 Logical Preliminaries

We recapitulate the main logical notation and concepts used in this paper; the region connection calculus and DL-Lite.

2.1 The Region Connection Calculus

We will consider different fragments of the region connection calculus [9] as potential candidates for the spatial logic to be combined with DL-Lite. Randell and colleagues' axiom system [9] is based on a primitive binary relation C intended to denote a connectedness relation which is specified to be reflexive and symmetric. On the basis of C other binary relations between regions which are called *base relations* are explained. One set of base relations is the set $\mathcal{B}_{RCC8}$, which is the main component of the most expressive region connection calculus RCC8. The base relations of $\mathcal{B}_{RCC8}$ and their intended meanings are given as follows: $\mathcal{B}_{RCC8} = \{\mathsf{DC}$ (disconnected), EC (externally connected), EQ (equal), PO (partially overlapping), NTPP (non-tangential proper part), TPP (tangential proper part), NTPPi (inverse of NTPP), TPPi (inverse of TPP)$\}$. We skip

the concrete definitions of the base relations by the connectedness relation C (see, e.g., [11, p. 45]), as we—in contrast to the axiom system of Randell and colleagues—consider the following axiom system schema Ax_{RCCi}, which directly specifies the properties of the base relations in $\mathcal{B}_{RCCi}$.

Definition 1 (Axiom system schema Ax_{RCCi}). *For all $i \in \{2, 3, 5, 8\}$ the axiom set Ax_{RCCi} contains the following axioms:*

$$\{\forall x, y.\ \textstyle\bigvee_{r \in \mathcal{B}_{RCCi}} r(x, y)\} \cup \qquad\qquad\qquad\text{(joint exhaustivity)}$$
$$\{\forall x, y.\ \textstyle\bigwedge_{r_1, r_2 \in \mathcal{B}_{RCCi}, r_1 \neq r_2} r_1(x, y) \to \neg r_2(x, y)\} \cup \qquad\text{(pairwise disjointness)}$$
$$\{\forall x, y, z. r_1(x, y) \wedge r_2(y, z) \to r_3^1(x, z) \vee \cdots \vee r_3^k(x, z) \mid r_1; r_2 = \{r_3^1, \ldots, r_3^k\}\}$$
$$\text{(weak composition axioms)}$$

For $i \in \{3, 5, 8\}$ additionally the axiom $\forall x \mathsf{EQ}(x, x)$ (reflexivity of EQ) is contained. For $i = 2$ one has instead the axiom $\forall x \mathsf{O}(x, x)$ (reflexivity of O).

In particular, the axioms state the JEPD-property of the base relations (each pair of regions x, y is related over exactly one base relation) and describe the (weak) composition of two base relations (denoted by ;) according to the composition table for RCCi. With the composition of two base relations, in most cases, only indefinite knowledge of spatial configurations follows. The spatial configuration $r_1(x, z) \vee \cdots \vee r_n(x, z)$ for base relations r_j in $\mathcal{B}_{RCCi}$ is also written as $\{r_1, \ldots, r_n\}(x, z)$, and the set $\{r_1, \ldots, r_n\}$ is called a general RCCi relation. Let Rel_{RCCi} be the set of all 2^i general RCCi relations. An RCCi *(constraint) network* consists of assertions of the form $\{r_1, \ldots, r_n\}(x, y)$.

We mention here the composition table for the low resolution logics RCC2 and RCC3. Their base relations are given by the sets $\mathcal{B}_{RCC3} = \{\mathsf{DR}, \mathsf{EQ}, \mathsf{ONE}\}$ and $\mathcal{B}_{RCC2} = \{\mathsf{DR}, \mathsf{O}\}$, and their weak compositions are defined as shown in Fig. 1. The discreteness relation DR is the same as $\{\mathsf{DC}, \mathsf{EC}\}$, the overlapping-but-not-equal relation ONE is equal to $\{\mathsf{PO}, \mathsf{NTPP}, \mathsf{TPP}, \mathsf{NTPPi}, \mathsf{TPPi}\}$ and the overlapping relation O is given by $\{\mathsf{ONE}, \mathsf{EQ}\}$.

;	DR	O
DR	$\mathcal{B}_{RCC2}$	$\mathcal{B}_{RCC2}$
O	$\mathcal{B}_{RCC2}$	$\mathcal{B}_{RCC2}$

;	DR	ONE	EQ
DR	$\mathcal{B}_{RCC3}$	$\{\mathsf{DR}, \mathsf{ONE}\}$	DR
ONE	$\{\mathsf{DR}, \mathsf{ONE}\}$	$\mathcal{B}_{RCC3}$	ONE
EQ	DR	ONE	EQ

Fig. 1. Composition tables for RCC2 and RCC3

Note that in the definitions of the base relations (of RCC3 and RCC2) we followed the author of [14]. The base relations of the low resolution calculus of [3] are different from those of RCC2; the authors of [3] consider the two base relations DC and $\{\mathsf{EC}, \mathsf{O}\}$. But as we do not deal with the exact definitions of the base relations referring to the connectedness relation C but with the axiom systems referring to the composition table this difference has no effect—the low resolution calculus of [3] has the same trivial composition table as RCC2. For the composition tables of RCC5 and RCC8 have a look at [12, p. 45].

2.2 DL-Lite

The family of DL-Lite description logics [2] is an appropriate candidate for the thematic component of the envisioned geo-thematic logic as it offers computationally feasible satisfiability checking and query answering over ontologies and data stored in a relational database. More concretely, satisfiability checking and query answering (for unions of conjunctive queries) are first order logic (FOL) rewritable. In this paper, we will mainly deal with a member of the extended DL-Lite family which we termed DL-Lite$_{\mathcal{F},\mathcal{R}}^{\sqcap}$, and which allows functional roles, role hierarchies and role inverses as well as the conjunction of basic concepts on the left-hand side of GCIs (general concept inclusions). The syntax is given in Def. 2. The semantics of this logic is defined in the usual way—but imposing the unique name assumption (UNA).

Definition 2 (DL-Lite$_{\mathcal{F},\mathcal{R}}^{\sqcap}$). *Let RN be the set of role symbols and $P \in RN$, CN be a set of concept symbols and $A \in CN$, Const be a set of individual constants and $a, b \in Const$.*

$$R \longrightarrow P \mid P^- \quad B \longrightarrow A \mid \exists R \quad C_l \longrightarrow B \mid C_l \sqcap B \quad C_r \longrightarrow B \mid \neg B$$

TBox*): $C_l \sqsubseteq C_r, (\text{funct } R), R_1 \sqsubseteq R_2$

ABox: $A(a), R(a, b)$

*) Restriction: If R occurs in a functionality axiom, then R and its inverse do not occur on the right-hand side of a role inclusion axiom $R_1 \sqsubseteq R_2$.

FOL rewritability also holds for the logic DL-Lite$_{\mathcal{F},\mathcal{R}}^{\sqcap}$ which follows from the corresponding FOL rewritability results for the Datalog extension Datalog$^\pm$ [1]. We assume that the reader is familiar with the notion of FOL queries, conjunctive queries (CQ) and unions of conjunctive queries (UCQ), their semantics and the notion of certain answers $cert(Q, \mathcal{A} \cup \mathcal{T})$ of a query w.r.t. to the union of a TBox and an ABox $\mathcal{T} \cup \mathcal{A}$ [2]. Let $DB(\mathcal{A})$ be the minimal Herbrand model of $\mathcal{A}$. *Checking the satisfiability of ontologies is FOL rewritable* iff for all TBoxes $\mathcal{T}$ there is a boolean FOL query $Q_\mathcal{T}$ s.t. for all ABoxes $\mathcal{A}$: the ontology $\mathcal{T} \cup \mathcal{A}$ is satisfiable iff $DB(\mathcal{A}) \not\models Q_\mathcal{T}$. *Answering queries from a subclass $\mathcal{C}$ of FOL queries w.r.t. to ontologies is FOL rewritable* iff for all TBoxes $\mathcal{T}$ and queries $Q = \psi(\boldsymbol{x})$ in $\mathcal{C}$ there is a FOL query $Q_\mathcal{T}$ such that for all ABoxes $\mathcal{A}$ it is the case that $cert(Q, \mathcal{T} \cup \mathcal{A}) = Q_\mathcal{T}^{DB(\mathcal{A})}$. For DL-Lite, FOL-rewritability can be proved w.r.t. to satisfiability as well as w.r.t. answering UCQs [2, Thm 4.14, Thm 5.15]. A crucial part of the proof for FOL rewritability w.r.t. query answering is the perfect rewriting algorithm [2, Fig. 2]. This algorithm works by using positive inclusion axioms of the form $A_1 \sqsubseteq A_2$ as rewriting rules from right to left so that for example an atom $A_2(x)$ occurring in a CQ would produce a new CQ containing $A_1(x)$ instead of $A_2(x)$.

3 Weak Combinations of DL-Lite with RCC

In this section, we recapitulate the results concerning a weak coupling of DL-Lite with the most expressive region connection calculus fragment RCC8, which we

introduced in [7, 8], and provide an example for its use(fulness). This will give us the opportunity to introduce further concepts that are necessary to understand the following discussions on stronger couplings of DL-Lite with the weaker region connection calculi RCC2 and RCC3.

The combination paradigm follows that of Lutz and Miličič [6] who combine $\mathcal{ALC}$ with the RCC8 and, more generally, with ω-admissible concrete domains [6, Def. 5, p. 7]. The combined logic $\mathcal{ALC}(RCC8)$ of [6] is well behaved in so far as testing concept subsumption is decidable. As we aim at FOL rewritability we have to be even more careful in choosing the right combination method.

We use an axiom set T_ω with corresponding properties of an ω-admissible domain for coupling with DL-Lite because axioms are more appropriate for rewriting investigations. The axiom sets Ax_{RCCi} will instantiate T_ω.

We recapitulate the syntax and the semantics of the constructors of [6] that are used for the coupling of the thematic and the spatial domain. A path U (of length at most 2) is defined as l for a fixed attribute l ("has location") or as $R \circ l$, the composition of the role symbol R with l. We abbreviate $R \circ l$ with $\tilde{R}$ in this paper. The usual notion of an interpretation $\mathcal{I}$ in our combined logic is slightly modified by using two separate domains $\Delta^\mathcal{I}$ and $(\Delta^*)^\mathcal{I}$. All symbols of the theory T_ω are interpreted relative to $(\Delta^*)^\mathcal{I}$. Let r be an RCC-relation of some RCC-fragment. That is, let be given a set of base relations $\mathcal{B}_{RCCi}$ and $r = \{r_1, \dots r_n\} \equiv r_1 \vee \cdots \vee r_n$ for $r_i \in \mathcal{B}_{RCCi}$. Then $l^\mathcal{I} \subseteq \Delta^\mathcal{I} \times (\Delta^*)^\mathcal{I}$; $r^\mathcal{I} = r_1^\mathcal{I} \cup \cdots \cup r_n^\mathcal{I}$; $(R \circ l)^\mathcal{I} = \{(d, e^*) \in \Delta^\mathcal{I} \times (\Delta^*)^\mathcal{I} \mid$ there is an e s.t. $(d, e) \in R^\mathcal{I}$ and $(e, e^*) \in l^\mathcal{I}\}$; $(\exists U_1, U_2.r)^\mathcal{I} = \{d \in \Delta^\mathcal{I} \mid$ there exist e_1^*, e_2^* s.t. $(d, e_1^*) \in U_1^\mathcal{I}, (d, e_2^*) \in U_2^\mathcal{I}$ and $(e_1^*, e_2^*) \in r^\mathcal{I}\}$; $(\forall U_1, U_2.r)^\mathcal{I} = \{d \in \Delta^\mathcal{I} \mid$ for all e_1^*, e_2^* s.t. $(d, e_1^*) \in U_1^\mathcal{I}, (d, e_2^*) \in U_2^\mathcal{I}$ it holds that $(e_1^*, e_2^*) \in r^\mathcal{I}\}$.

Now we can define the following combined geo-thematic logic (where a^*, b^* stand for constants intended to be interpreted by regions):

Definition 3 (DL-Lite$_{\mathcal{F},\mathcal{R}}^\sqcap$(RCC8)). *Let $r \in Rel_{RCC8}$ and $T_\omega = Ax_{RCC8}$.*

$$R \longrightarrow P \mid P^- \qquad U \longrightarrow R \mid \tilde{R} \quad B \longrightarrow A \mid \exists R \mid \exists l$$
$$C_l \longrightarrow B \mid C_l \sqcap B \quad C_r \longrightarrow B \mid \neg B \mid \exists U_1, U_2.r$$
$$TBox^{*)}: \qquad C_l \sqsubseteq C_r, (\text{funct } l), (\text{funct } R), R_1 \sqsubseteq R_2$$
$$ABox: \qquad A(a), R(a, b), l(a, a^*), r(a^*, b^*)$$

*) *Restriction: If* $(\text{funct } R) \in \mathcal{T}$, *then* R *and* R^- *do not occur on the right-hand side of a role inclusion axiom or in a concept of the form* $\exists U_1, U_2.r$.

As satisfiability checking of RCC8 constraint networks is NP-complete, there is only a chance to reach FOL rewritability if we assume within the ABox a constraint network which is consistent and complete, i.e., only base relations are allowed as labels; in this case the ABox is called *spatially complete*. It seems to us that for cadastral maps or maps containing areas of administration one can assume pretty safely (almost) spatial completeness. The coupling with RCC8 is so weak that FOL rewritability of satisfiability follows.

Proposition 1. *Checking the satisfiability of* $DL-Lite_{\mathcal{F},\mathcal{R}}^\sqcap(RCC8)$ *ontologies that have a spatially complete ABox is FOL rewritable.*

The FOL rewritability holds also for query rewriting w.r.t. to a restricted class of treelike queries (termed GCQ^+). The result is achieved by an adapted perfect rewriting algorithm PerfectRef [2, Fig. 13].

Theorem 1. *Answering GCQ^+-queries w.r.t. $DL-Lite^{\square}_{\mathcal{F},\mathcal{R}}$(RCC8) ontologies that have a spatially complete ABox is FOL rewritable.*

DL-Lite$^{\square}_{\mathcal{F},\mathcal{R}}$(RCC8) as well as the logics introduced below are suited for use in scenarios such as that of an engineering bureau planning additional parks in a city [7]. Assume, the bureau has stored geographical data in some database and declares relevant concepts in the TBox: $Park+Lake \sqsubseteq Park$; $Park4Playing \sqsubseteq Park$; $Park+Lake \sqsubseteq \exists hasLake \circ l, l.\mathsf{TPP}$; $Park4Playing \sqsubseteq \exists hasPlAr \circ l, l.\mathsf{TPP}$. The ABox $\mathcal{A}$ is derived virtually by mappings from geographical data in a database. In particular, assume that $\{Park+Lake(a), Park4Playing(a)\} \subseteq \mathcal{A}$. According to $\mathcal{A}$, the object a is a park with a lake and a park with a playing area but its spatial extension is not known. Now, the engineering bureau asks for all parks with lakes and playing areas such that the playing area is not contained as island in the lake. This can be formalized by the GCQ^+ query: $Q = Park(x) \wedge \exists hasLake \circ l, hasPlAr \circ l.(\mathcal{B}_{RCC8} \setminus \{\mathsf{NTPP}\})(x)$. Using the composition entry $\mathsf{TPP};\mathsf{TPPi} = \{\mathsf{DC}, \mathsf{EC}, \mathsf{PO}, \mathsf{TPP}, \mathsf{TPPi}, \mathsf{EQ}\} \subseteq \mathcal{B}_{RCC8} \setminus \{\mathsf{NTPP}\}$, the reformulation algorithm produces a UCQ that contains the following CQ: $Q' = (\exists hasLake \circ l, l.\mathsf{TPP})(x) \wedge (\exists l, hasPlAr \circ l.\mathsf{TPPi})(x)$. Rewriting $\exists l, hasPlAr \circ l.\mathsf{TPPi}$ to $\exists hasPlAr \circ l, l.\mathsf{TPP}$ in combination with the rewriting rule for $A_1 \sqsubseteq A_2$ we get another CQ $Q'' = Park+Lake(x) \wedge Park4Playing(x)$. Now, Q'' captures (as desired) the object a.

4 Strong Combinations of DL-Lite with RCC

Another way of reaching FOL rewritability for combinations of DL-Lite with RCC is weakening the expressivity of the spatial component. Hence, one may ask whether a combination with the calculus RCC3 or RCC2 [15], both fragments with weak expressibility, allows for weak FOL rewritability w.r.t. satisfiability checks (and query answering). Their potential use as logics for approximating [5] ontologies in more expressible combined logics like $\mathcal{ALC}$(RCC8) makes the investigation valuable. The logics DL-Lite$^{\square,+}_{\mathcal{F},\mathcal{R}}$(RCC2) and DL-Lite$^{\square,+}_{\mathcal{F},\mathcal{R}}$(RCC3) are defined as follows ('+' indicates the strong combination):

Definition 4 (DL-Lite$^{\square,+}_{\mathcal{F},\mathcal{R}}$(RCC2) and DL-Lite$^{\square,+}_{\mathcal{F},\mathcal{R}}$(RCC3)). *Let $T_\omega = Ax_{RCC2}$ resp. $T_\omega = Ax_{RCC3}$ and $r \in \mathcal{B}_{RCC2}$ resp. $r \in \mathcal{B}_{RCC3}$*

$$R \longrightarrow P \mid P^- \qquad U \longrightarrow l \mid \tilde{R} \qquad B \longrightarrow A \mid \exists R$$
$$C_l \longrightarrow B \mid C_l \sqcap B \qquad C_r \longrightarrow B \mid \neg B \mid \exists U_1, U_2.r$$
$$TBox^{*)}: \qquad C_l \sqsubseteq C_r, (\mathrm{funct}\ l, R), R_1 \sqsubseteq R_2$$
$$ABox: \qquad A(a), R(a,b), l(a,a^*), r(a^*,b^*)$$

*) Restriction: If $(\mathrm{funct}\ R) \in \mathcal{T}$, then R and R^- do not occur on the right-hand side of a role inclusion axiom.

For RCC3 the strong combination with DL-Lite$^{\square}_{\mathcal{F},\mathcal{R}}$ leads to non-FOL rewritability. The reason lies in the fact that testing the satisfiability of RCC3 is not in the complexity class AC^0 as shown by the following lemma.

Lemma 1. *Checking satisfiability of RCC3 networks is* LOGSPACE *hard.*

Proof. As is known, the reachability problem in symmetric (undirected) graphs is logspace complete [10]—where graph reachability asks whether for nodes s, t in G there is a path between s and t. By reducing this problem to the satisfiability test for RCC3 networks we will have shown that the latter problem is LOGSPACE hard itself. So let be given a (symmetric) graph $G = (V, E)$ and nodes $s, t \in V$. We define the network N in the following way (see Figure 2): Let $V = \{v_1, \ldots, v_n\}$ be an enumeration of the nodes of G; w.l.o.g. let $s = v_1$ and $t = v_n$ and let $\mathcal{B} = \mathcal{B}_{RCC3}$. Nodes of N are given by $V \cup \{a\}$ where $a \notin V$. Labelled edges of N are given by: $s\{\mathsf{DR}\}a$; $t\{\mathsf{ONE}\}a$; $v_i\{\mathcal{B}\}a$ for all $i \neq 1, n$; $v_i\{\mathsf{EQ}\}v_j$ if $E(v_i, v_j)$; $v_i\{\mathcal{B}\}v_j$ if $\neg E(v_i, v_j)$. Now we show that the network N is satisfiable iff s and t are
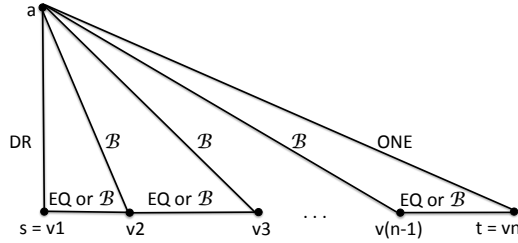

Fig. 2. Network N used in proof of Lemma 1

connected in G. Assume that s and t are connected; then there is an EQ-path in N between them, hence $s\{\mathsf{EQ}\}t$ follows. But this contradicts $s\{\mathsf{DR}\}a$ and $t\{\mathsf{ONE}\}a$. Now assume that s and t are not connected; then there is no path consisting only of EQ-labels between s and t. The graph G consists of at least 2 components, and s, t are in different components. We define a consistent configuration as follows: For all nodes v, v' in the component in which s is contained, let $v\{\mathsf{DR}\}a$ and $v\{\mathsf{EQ}\}v'$. For all nodes v, v' in the component of t let $v\{\mathsf{ONE}\}a$ and $v\{\mathsf{EQ}\}v'$. For all nodes v, v' in the other components let $v\{\mathsf{DR}\}a$ and $v\{\mathsf{EQ}\}v'$. For all nodes v, v' which have not a label yet, let $v\{\mathsf{DR}\}v'$. (Two remarks : 1) EQ-edges for edges $E(v_i, v_j)$ in G with $j > i + 1$ are not shown in Fig. 2. 2) We inserted edges labelled $\mathcal{B}$ for better illustrations. But these are not needed.)

This lemma immediately entails the fact that satisfiability checking for ontologies over the logic DL-Lite$_{\mathcal{F},\mathcal{R}}^{\sqcap,+}$(RCC3) is not FOL rewritable. This problem does not vanish if we presuppose that the ABox $\mathcal{A}$ is spatially complete—as shown by the following proposition.

Proposition 2. *Satisfiability checking of ontologies in DL-Lite$_{\mathcal{F},\mathcal{R}}^{\sqcap,+}$ (RCC3) with spatially complete ABoxes is not FOL rewritable.*

Proof. We construct a generic TBox $\mathcal{T}_g$ that allows one to encode any RCC3 constraint network so that checking the consistency of RCC3 constraint networks

is reducible to a satisfiability check of this TBox and a spatially complete ABox. Let for every $r \in Rel_{RCC3}$ be given role symbols R_r^1, R_r^2. The generic TBox $\mathcal{T}_g$ has for every $r \in Rel_{RCC3}$ a concept symbol A_r and a corresponding axiom with the content that all instances of A_r have paths over the abstract features R_1 resp. R_2 to regions that are r-related.

$$\mathcal{T}_g = \{A_r \sqsubseteq \exists \tilde{R}_r^1, \tilde{R}_r^2.r, (\text{funct } l, R_r^1, R_r^2) \mid r \in Rel_{RCC3}\} \qquad (1)$$

Now, let N be an arbitrary RCC3 constraint network which has to be tested for relational consistency. Let $\mathcal{A}_N$ be an ABox such that for every $r(a,b)$ in N three new constants are introduced: x_{ab}, x_a, x_b.

$$\mathcal{A}_N = \{A_r(x_{ab}), R_r^1(x_{ab}, x_a), R_r^2(x_{ab}, x_b) \mid r(a,b) \in N\} \qquad (2)$$

The construction entails: $\mathcal{T}_g \cup \mathcal{A}_N \cup Ax_{RCC3}$ is satisfiable iff $N \cup Ax_{RCC3}$ is satisfiable. If the data complexity of the satisfiability check for DL-Lite$_{\mathcal{F},\mathcal{R}}^{\sqcap,+}$(RCC3)-ontologies were in AC^0, then the consistency of constraint networks could be tested in AC^0, too. (Note that $\mathcal{T}_g$ is a fixed TBox.) But checking the consistency of RCC3 constraint networks is LOGSPACE-hard and $AC^0 \subsetneq$ LOGSPACE.

As a corollary to this proposition we get the assertion that strong combinations of RCC5 and RCC8 into DL-Lite$_{\mathcal{F},\mathcal{R}}^{\sqcap,+}$(RCC5) and DL-Lite$_{\mathcal{F},\mathcal{R}}^{\sqcap,+}$(RCC8), respectively—which are defined in the same manner as in Definition 4—do not allow for FOL rewritability of satisfiability checking.

The low resolution calculus RCC2 is quite more inexpressive than RCC3 due to the fact that the composition table does not allow for the propagation of information: All compositions of DR, O result in the maximally unspecified relation $\{\mathsf{DR}, \mathsf{O}\}$. Hence, FOL rewritability of satisfiability testing follows easily considering the query $Q = \exists x, y[\mathsf{O}(x,y) \wedge \mathsf{DR}(x,y)] \vee \exists x[\mathsf{DR}(x,x)]$.

Proposition 3. *Testing the satisfiability of RCC2 networks is FOL rewritable.*

But in combination with functionality axioms of the TBox one could have the problem that the ABox may lead to identifications of regions. The identified regions are not allowed to have edges labelled O, DR resp. to the same region. Though this can be tested, the problem arises when a chain of regions is identified by the TBox and the ABox, because we do not know the length of the chain in advance. More formally: In addition to RCC2 constraint-network assertions we allow identity assertions $v = w$ for regions v, w. As we can assume that all nodes in a RCC2 network are connected by an edge labelled O, DR or $\mathcal{B}_{RCC2}$ we use a more intuitive formalism where, for every assertion $v = w$, the label of the edge between v and w is marked with an $=$; e.g., an edge between v, w with label $\mathsf{DR}^=$ stands for $\mathsf{DR}(v,w) \wedge v = w$. We call such a network an $=$-marked RCC2 network (a RCC$^=$2 network for short). Let $\mathcal{B} = \mathcal{B}_{RCC2}$ in the following.

Proposition 4. *An RCC$^=$2 constraint network N is unsatisfiable iff*

1. *N contains $\mathsf{DR}(v,v)$ or $\mathsf{DR}^=(v,v)$ for some node v; or*

Proof. Sufficiency of the condition for unsatisfiability is clear. The proof for necessity is done by induction on the number of =-marked labels. So let be given a RCC$^=$ network N. We may assume, that the network has for every pair of nodes exactly one labelled edge between; we assume the edges to be undirected as all relations are symmetric.

Base case: Assume that none of the four conditions hold. As there are no marked labels, then N unsatisfiability can occur only, if N contains $\mathsf{DR}(v, v)$ for some node v (first condition) or $\mathsf{DR}(v, w)$ and $\mathsf{O}(v, w)$ (fourth condition). But these cases are excluded by assumption.

Induction step: Let N contain n marked labelled edges and assume that for all networks N' with $n-1$ marked labelled edges unsatisfiability of N' implies one of the conditions. Now, assume that for N no one of the four conditions holds. We have to show that N is satisfiable. Take an arbitrary =-marked labelled edge between nodes v, w. The label λ of this edge is either $\mathsf{O}^=$ or $\mathcal{B}^=$. We define a new network N' which results as a contraction from N by identifying v and w to the node z. For N' we may assume again that it contains for every pair of nodes exactly one labelled edge: If in N we have $r_1(v, k)$ and $r_2(w, k)$, then the edge between z and k in N' results as $r_1 \cap r_2$ and is =-marked iff r_1 or r_2 is marked. Clearly $r_1 \cap r_2$ is not empty, as otherwise we would have a contradiction to the fact that N does not fulfill conditions 3 and 4. Clearly N is satisfiable iff N' is satisfiable. Assume to the contrary that N' is unsatisfiable. Hence, one of the four conditions holds for N': 1) Assume N' contains $\mathsf{DR}(v', v')$ or $\mathsf{DR}^=(v', v')$ for some node v'. If v' is not z, then also $\mathsf{DR}(v', v') \in N$ resp. $\mathsf{DR}^=(v', v') \in N$—contradicting the fact that N does not fulfill condition 1. Otherwise $v' = z$, but this cannot be the case either, as there are no self-loops $\mathsf{DR}(v, v)$, $\mathsf{DR}^=(v, v)$, $\mathsf{DR}(w, w)$, $\mathsf{DR}^=(w, w)$ in N nor $\mathsf{DR}(v, w)$ or $\mathsf{DR}(v, w)$. 2) Assume N' contains $\mathsf{DR}^=(v', w')$. If neither v' nor w' is z, this contradicts the fact that N does not fulfill the 2. condition. Otherwise $v' = w' = z$, leading to a contradiction with the fact that N does not fulfill condition 1. 3) Assume N' contains a cycle in which there is $\mathsf{DR}(v', w')$ and there is a path from v' to w' such that every label on the path is $\mathcal{B}^=$ or $\mathsf{O}^=$. The case that the path does not contain z immediately leads to a contradiction. Otherwise, the path extends to a path in N fulfilling the 3. condition—contradiction. Similarly, if N' fulfills condition 4, the verifying path can be extended to a path in N fulfilling condition 4—contradiction.

Proposition 4 shows that adding identity assertions to an RCC2 network may require checking the existence of identity chains of arbitrary length. Hence, in principle it is possible that the functional roles used in DL-Lite$^{\sqcap,+}_{\mathcal{F},\mathcal{R}}$(RCC2) may lead to identity chains. But as we want to show in the following proposition, this cannot be the case: The identity paths induced by functionalities in DL-Lite$^{\sqcap,+}_{\mathcal{F},\mathcal{R}}$(RCC2) can have only a maximal length of one.

Proposition 5. *Satisfiability checking of ontologies in DL-Lite$_{\mathcal{F},\mathcal{R}}^{\sqcap,+}$ (RCC2) is FOL rewritable.*

We give a proof sketch. The idea is to find concepts that are unsatisfiable according to the TBox (this amounts to constructing the negative closure as in the proofs for pure DL-Lite [2, Def. 4.7, p. 292]); these are formulated as boolean queries, and the (finite) disjunctions of these queries is answered over the ABox; e.g., if $A_1 \sqsubseteq \neg A_2$ is in the TBox, then the query would contain $\exists x(A_1(x) \wedge A_2(x))$ as disjunct. The introduction of concepts of the form $\exists U_1, U_2.r$ enlarges the potential conflicts of a TBox $\mathcal{T}$ with an ABox $\mathcal{A}$. So in addition to the FOL queries that result from the negative closure of the TBox, we have to find queries for the potential conflicts in the four conditions of Proposition 4. The resulting conjunctive queries have further to be fed into a perfect rewriting algorithm like PerfectRef for pure DL-Lite [2] in order to capture the implications of the TBox.

Concerning the first two conditions we have to deal with axioms of the form $B \sqsubseteq \exists l, l.\mathsf{DR}$. If this axiom occurs in the TBox $\mathcal{T}$, then one has to produce the CQ $\exists x.B(x)$. Similarly, if $\{B \sqsubseteq \exists \tilde{R}, \tilde{R}.\mathsf{DR}, (\mathrm{funct}\ R)\} \subseteq \mathcal{T}$, then the CQ $\exists x.B(x)$ has to be added. Also if $\{B \sqsubseteq \exists \tilde{R}_1, \tilde{R}_2.\mathsf{DR}, (\mathrm{funct}\ R_1, \mathrm{funct}\ R_2)\} \subseteq \mathcal{T}$, then we may get a conflict of the first kind and hence we have to add a CQ $\exists x, y[B(x) \wedge R_1(x,y) \wedge R_2(x,y)]$. Concerning the third and fourth condition we note that this can be only the case if the $=$-marked paths have maximal length one. Otherwise we already have a contradiction of the ABox with the functionality assertions in the TBox. Hence, one considers only the case of pairs of TBox axioms of the general form $A \sqsubseteq \exists \tilde{R}_1, \tilde{R}_2.\mathsf{DR}$ and $B \sqsubseteq \exists \tilde{R}_3, \tilde{R}_4.\mathsf{O}$ with $(\mathrm{funct}\ R_1, R_2, R_3, R_4)$. In this case we feed also into the PerfectRef algorithm the CQ $\exists x, y, z, w[A(x) \wedge B(y) \wedge R_1(x,z) \wedge R_3(y,z) \wedge R_2(x,w) \wedge R_4(y,w)]$.

5 Conclusion

As recapitulated in this paper, combining DL-Lite with expressive fragments of the region calculus like RCC8 into logics that preserve the property of FOL rewritability is possible if the coupling is weak: Constraints of the RCC8 network contained in the ABox are not transported over to the implicitly constructed constraint network resulting from the constructors of the form $\exists U_1, U_2.r$. In this paper we mainly dealt with strong combinations for weaker calculi like RCC2 or RCC3. As we have shown by a reduction proof, a strong combination with RCC3 destroys the FOL rewritability of satisfiability checking. The reason is that checking the satisfiability of RCC3 networks needs to test for reachability along EQ paths, which can be reproduced by the TBox. For the low resolution calculus RCC2, FOL rewritability of satisfiability checking is provable—though checking the satisfiability of RCC2 networks with additional identity assertions is at least as hard as checking RCC3 networks. We plan to investigate whether DL-Lite$_{\mathcal{F},\mathcal{R}}^{\sqcap,+}$ (RCC2) and DL-Lite$_{\mathcal{F},\mathcal{R}}^{\sqcap}$ (RCC8) can be used for approximation—following the complete but not necessarily correct approximation method of [5]. Moreover we want to check whether DL-Lite$_{\mathcal{F},\mathcal{R}}^{\sqcap,+}$ (RCC2) allows for FOL rewritability of query answering w.r.t. unions of conjunctive queries.

References

1. Calì, A., Gottlob, G., Lukasiewicz, T.: A general datalog-based framework for tractable query answering over ontologies. Technical Report CL-RR-10-21, Oxford University Computing Laboratory (November 2010)
2. Calvanese, D., De Giacomo, G., Lembo, D., Lenzerini, M., Poggi, A., Rodríguez-Muro, M., Rosati, R.: Ontologies and databases: The DL-Lite approach. In: Tessaris, S., Franconi, E. (eds.) Semantic Technologies for Informations Systems – 5th Int. Reasoning Web Summer School (RW-2009), LNCS, vol. 5689, pp. 255–356. Springer (2009)
3. Grigni, M., Papadias, D., Papadimitriou, C.H.: Topological inference. In: IJCAI (1). pp. 901–907 (1995)
4. Haarslev, V., Lutz, C., Möller, R.: A description logic with concrete domains and a role-forming predicate operator. J. Log. Comput. 9(3), 351–384 (1999)
5. Kaplunova, A., Moeller, R., Wandelt, S., Wessel, M.: Towards scalable instance retrieval over ontologies. In: Yaxin, B., Mary-Anne, W. (eds.) Knowledge Science, Engineering and Management, Fourth International Conference, KSEM 2010, Proceedings. Lecture Notes in Computer Science, vol. 6291. Springer (2010)
6. Lutz, C., Miličić, M.: A tableau algorithm for description logics with concrete domains and general TBoxes. J. Autom. Reasoning 38(1-3), 227–259 (2007)
7. Özçep, O.L., Möller, R.: Computationally feasible query answering over spatio-thematic ontologies. In: Proceedings of GEOProcessing 2012, The Fourth International Conference on Advanced Geographic Information Systems, Applications, and Services (2012), available online at `http://www.thinkmind.org/index.php?view=article&articleid=geoprocessing_2012_7_10_30059`
8. Özçep, Ö.L., Möller, R.: Scalable geo-thematic query answering. Technical report, Institute for Softwaresystems (STS), Hamburg University of Technology (2012), available online at `http://www.sts.tu-harburg.de/tech-reports/papers.html`
9. Randell, D.A., Cui, Z., Cohn, A.G.: A spatial logic based on regions and connection. In: Proceedings of the 3rd International Conferecence on Knowledge Representation and Reasoning. pp. 165–176 (1992)
10. Reingold, O.: Undirected connectivity in log-space. J. ACM 55(4), 17:1–17:24 (Sep 2008), `http://doi.acm.org/10.1145/1391289.1391291`
11. Renz, J.: Qualitative Spatial Reasoning with Topological Information, Lecture Notes in Computer Science, vol. 2293. Springer (2002)
12. Renz, J., Nebel, B.: Qualitative spatial reasoning using constraint calculi. In: Aiello, M., Pratt-Hartmann, I., Benthem, J. (eds.) Handbook of Spatial Logics, pp. 161–215. Springer Netherlands (2007)
13. Rodriguez-Muro, M., Calvanese, D.: Semantic index: Scalable query answering without forward chaining or exponential rewritings. In: Posters of the 10th Int. Semantic Web Conf. (ISWC 2011) (2011)
14. Wessel, M.: On spatial reasoning with description logics - position paper. In: Ian Horrocks, S.T. (ed.) Proceedings of the International Workshop on Description Logics (DL-2002). CEUR Workshop Proceedings, vol. 53 (2002), http://CEUR-WS.org/Vol-53/
15. Wessel, M.: Qualitative spatial reasoning with the $\mathcal{ALCI}_{RCC}$- family — first results and unanswered questions. Technical Report FBI–HH–M–324/03, University of Hamburg, Department for Informatics (2003)