Özgür Lütfü Özçep and Ralf Möller

January 28, 2012

Abstract

Complete, sound, and fast query answering for geo-thematic ontologies over geographical data is difficult to achieve due the size of the geographical database. A solution proposed in the literature for pure thematic ontologies over huge databases is to use lightweight logics such as DL-Lite or Datalog+-0. For these, query answering can be reduced to evaluating a rewritten first order logic (FOL) query on the database. In order to preserve FOL rewritability for geo-thematic domains, the interaction of the spatial calculus—such as the region connection calculus—with the lightweight logic has to be restricted. In this paper, we discuss combined geo-thematic logics for which FOL rewritability does not hold though the interaction of the thematic and spatial part seems harmless as well as combinations that allow for FOL rewritability and that are expressive enough to be used in geo-thematic scenarios involving query answering.

STS Institute for Software, Technology & Systems

TUHH Hamburg University of Technology

1 Introduction

Query answering over a database becomes far more difficult if the extensional knowledge in the database is extended by intensional knowledge in the form of, e.g., an ontology. The reason is that a database represents (in the ideal case) one specific model in the predicate logical sense, so that model checking suffices for query answering. In contrast, 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 (see the technical notion of a *certain answer* below.) But in some cases—when the expressivity of the query and ontology language is kept low—query answering w.r.t. an ontology can be reduced to model checking. In technical terms: query answering w.r.t. an ontology is *FOL (first order logic) rewritable*. The term indicates the idea of the reduction; the given query is rewritten into a FOL query in which the intensional knowledge of the ontology is captured. FOL rewritability is a desired property because in most cases the size of the database surmounts the size of the query and the ontology; and this is definitely the case for geographical databases. Hence, scalability matters. Though the rewritten queries may become exponentially bigger than the original queries, FOL rewritability really means a benefit for fast query answering as there exist optimizations [15].

Using DL-Lite [3], a lightweight logic directed towards FOL rewritability, is not sufficient for GIS scenarios, as these demand the representation and deduction over spatial concepts. Though constraint-based spatial reasoning [14] offers a well developed and well proven theory for spatial domains, it does not fill in the need for a system that combines reasoning over a spatial and a non-spatial (thematic) domain. And as far as we can see, there is still no mixed constraint based reasoning system that offers reasoning over a mixed domain of geo-thematic objects. But even in case of related work which equally considers spatial and thematic reasoning ([5], [16], [8]), these are not aimed at providing a complete, correct and fast query answering mechanism. Therefore, the construction of logics fulfilling the following conditions is highly relevant for the GIS community: The logic should provide means for representing geo-thematic concepts by, e.g., combining a lightweight logic with a spatial logic; and the combined logic should allow for efficient, complete, and correct query answering, e.g., by allowing for FOL rewritability.

Continuing previous work [10], we investigate combinations of logics in the DL-Lite family with different members of the region connection calculus (RCC) family [12], a well-known family of calculi for qualitative spatial reasoning. By using DL-Lite for the thematic part, one gets a logic with development support not only for query answering but also for all services concerning intensional knowledge processing such as subsumption or consistency testing. Extending previous work, in this paper, we use DL-Lite$^{\sqcap}_{\mathcal{F},\mathcal{R}}$for the representation of the thematic part. This logic is different from the prominent members of the DL-Lite family in the point that it allows for concept conjunctions on the left-hand side of TBox axioms. Concerning the spatial part, our approach is not restricted to RCC, but may generally allow spatial logics with an ω-admissible domain [8].

A controlled combination of DL-Lite$^{\sqcap}_{\mathcal{F},\mathcal{R}}$with the very expressive RCC8 is indeed possible in such a way that on the one hand the resulting logic is still expressive enough to be used in geographical scenarios of query answering and on the other hand FOL rewritability w.r.t. query answering of (tree shaped) queries is guaranteed. If the "flow of information" from the spatial part into the thematic part is not restricted, then the combination with even

RCC5 (see below) leads to non-FOL rewritability. A strong coupling of DL-Lite$^{\sqcap}_{\mathcal{F},\mathcal{R}}$with the spatial component is at most possible if the member of RCC chosen for the spatial part is very weakly expressible, which is the case for RCC2 and RCC3.

The paper is structured as follows. In Sect. 2 we provide details on the region connection calculus and on the necessary concepts associated with the DL-Lite logic DL-Lite$^{\sqcap}_{\mathcal{F},\mathcal{R}}$. In Sect. 3 we prove how uncontrolled combinations of DL-Lite$^{\sqcap}_{\mathcal{F},\mathcal{R}}$with RCC8 (and RCC5) can hinder FOL rewritability. In Sect. 4 we look at positive cases for FOL rewritability w.r.t. to satisfiability checks of ontologies. And in the last section before the conclusion we show that DL-Lite$^{\sqcap}_{\mathcal{F},\mathcal{R}}$(RCC8) allows for FOL rewritability w.r.t. query answering.

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

The FOL theory that we use in order to model spatial configurations is given by the set Ax_{RCC8} (Def. 1) derived from the axioms of [12] for their family of region connection calculi (RCC).

Definition 1 (Ax_{RCC8}). $\{\forall x, y. \bigvee_{r \in \mathcal{B}_{RCC8}} r(x, y)\} \cup$ (joint exhaustivity)

$\{\forall x, y. \bigwedge_{r_1, r_2 \in \mathcal{B}_{RCC8}, r_1 \neq r_2} r_1(x, y) \rightarrow \neg r_2(x, y)\} \cup$

(pairwise disjointness)

$\forall x, y, z. r_1(x, y) \wedge r_2(y, z) \rightarrow r_3^1(x, z) \vee \cdots \vee r_r^k(x, z) \mid r; s = \{r_3^1, \ldots, r_3^k\}\} \cup$
(weak composition axioms)

$\{\forall x. \mathsf{eq}(x, x)\}$ (reflexivity of eq)

Randell and colleagues use a connectedness relation C to describe spatial relations, among them the set of eight *base relations* $\mathcal{B}_{RCC8} = \{\mathsf{dc}$ (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)$\}$. For an overview of the base relations' definitions see, e.g., [13, p. 42]. In contrast to the axiom set of [12], Ax_{RCC8} directly specifies the properties of the base relations in $\mathcal{B}_{RCC8}$. In particular, the axioms state the JEPD-property of the base relations (all base relations are jointly exhaustive and pairwise disjoint) and describe the (weak) composition of two base relations according to the composition table for RCC8 [13, p. 45]. Take for example the table entry for the pair $(\mathsf{tpp}, \mathsf{tppi})$ which is $\mathsf{tpp}; \mathsf{tppi} = \{\mathsf{dc}, \mathsf{ec}, \mathsf{po}, \mathsf{tpp}, \mathsf{tppi}, \mathsf{eq}\}$. It is described in Ax_{RCC8} by $\forall x, y, z. \mathsf{tpp}(x, y) \wedge \mathsf{tppi}(y, z) \rightarrow \{\mathsf{dc}, \mathsf{ec}, \mathsf{po}, \mathsf{tpp}, \mathsf{tppi}, \mathsf{eq}\}(x, z)$. Ax_{RCC8} does not contain the axiom of non-atomicity $\forall x \exists y. \mathsf{ntpp}(y, x)$, which asserts that all regions have a non-tangential proper part; in combination with the transitivity and irreflexivity of ntpp, the non-atomicity axiom restricts the class of models to infinite ones. Excluding this axiom has the consequence that Ax_{RCC8} allows for discrete spatial models. Thus we work with a generalized version of RCC—similar to the approach of [7]. The variables in Ax_{RCC8} are intended to denote bounded plane regions. In case of practical applications of the combined logic DL-Lite$^{\sqcap}_{\mathcal{F},\mathcal{R}}$(RCC8) (see below) we will further presume that the regions, which are extracted from a geographical database, even do not contain holes. However, this restriction will not be reflected in the axiom set which is used as the spatial component for the combined logic DL-Lite$^{\sqcap}_{\mathcal{F},\mathcal{R}}$(RCC8).

Indefinite knowledge on spatial configurations can be modelled by disjunctions $r_1 \vee \cdots \vee r_n$ of base relations r_i, also written as the set $\{r_1, \ldots r_n\}$ and called general RCC8 relation. Let Rel_{RCC8} be the set of all 2^8 RCC8 relations.

Besides RCC8, we also consider the coarser calculi RCC5, RCC3 and RCC2 which are defined on the sets of base relations: $\mathcal{B}_{RCC5}$ $=\{\mathsf{dr}, \mathsf{eq}, \mathsf{pp}, \mathsf{ppi}, \mathsf{po}\}$, where $\mathsf{dr} = \{\mathsf{dc}, \mathsf{ec}\}$, $\mathsf{pp} = \{\mathsf{tpp}, \mathsf{ntpp}\}$ and $\mathsf{ppi} = \{\mathsf{tppi}, \mathsf{ntppi}\}$; $\mathcal{B}_{RCC3} = \{\mathsf{dr}, \mathsf{eq}, \mathsf{one}\}$ where $\mathsf{one} = \{\mathsf{pp}, \mathsf{ppi}, \mathsf{po}\}$ (overlapping but not equal); and $\mathcal{B}_{RCC2} = \{\mathsf{dr}, \mathsf{o}\}$ where $\mathsf{o} = \{\mathsf{one}, \mathsf{eq}\}$. The axiom sets Ax_{RCCi} for $i \in \{2, 3, 5\}$ are similar to Ax_{RCC8} but use composition tables for $\mathcal{B}_{RCCi}$. In case of RCC2 the reflexivity for eq (see Def. 1) is replaced by the reflexivity axiom for o: $\forall x \mathsf{o}(x, x)$.

2.2 DL-Lite

In this section we describe the lightweight description logic DL-Lite$^{\sqcap}_{\mathcal{F},\mathcal{R}}$, a member of the DL-Lite family that 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), one form of TBox axioms. 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). (See, e.g., [1].)

Definition 2 (DL-Lite$^{\sqcap}_{\mathcal{F},\mathcal{R}}$)**.** 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*): $\quad C_l \sqsubseteq C_r, (\text{funct } R), R_1 \sqsubseteq R_2$

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

*) 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 $R_1 \sqsubseteq R_2$.

An outstanding feature of logics in the DL-Lite family is that they allow for *first order logic (FOL) rewritability* of satisfiability checks and of query answering. This means that these problems can be reduced to the evaluation of FOL queries over a database. This fact holds also for the logic DL-Lite$^{\sqcap}_{\mathcal{F},\mathcal{R}}$. Though [3] discuss only logics without concept conjunction $\sqcap$ on the left-hand side of GCIs, adding $\sqcap$ does not lead to the loss of FOL rewritability—as [2] show in the context of their comparison of Datalog$^{\pm}$ and DL-Lite [2].

We recapitulate the technical notions needed for defining FOL rewritability. An ontology $\mathcal{O}$ is a tuple $(Sig, \mathcal{T}, \mathcal{A})$, with a signature Sig (i.e., a set of concept symbols, role symbols and constants also denoted by $Sig(\mathcal{O})$), with a TBox $\mathcal{T}$, and with an ABox $\mathcal{A}$. An *FOL query* $Q = \psi(\vec{x})$ is an FOL formula $\psi(\vec{x})$ with free variables $\vec{x}$ called *distinguished variables.* If $\vec{x}$ is empty, the query is called *boolean.* Let $\vec{a}$ be a vector of constants from $Sig(\mathcal{O})$. The set of answers w.r.t. $\mathcal{I}$ is defined by $Q^{\mathcal{I}} = \{\vec{d} \in (\Delta^{\mathcal{I}})^n \mid \mathcal{I}_{[\vec{x} \mapsto \vec{d}]} \models \psi(\vec{x})\}$. (We use $Q^{\mathcal{I}}$ later on for a specific model $\mathcal{I}$, namely a Herbrand model.) The set of certain answers w.r.t. to $\mathcal{O}$ is defined by $cert(Q, \mathcal{T} \cup \mathcal{A}) = \{\vec{a} \mid \mathcal{T} \cup \mathcal{A} \models \psi[\vec{x}/\vec{a}]\}$. A *conjunctive query (CQ)* is a FOL query in which $\psi(\vec{x})$ is an existentially quantified conjunction of atomic formulas $at(\cdot)$,

$\psi(\vec{x}) = \exists \vec{y} \bigwedge_i at_i(\vec{x}, \vec{y})$. A *union of conjunctive queries (UCQ)* is a disjunction of CQs, i.e., a formula of the form $\exists \vec{y_1} \bigwedge_{i_1} at_{i_1}(\vec{x}, \vec{y_1}) \vee \cdots \vee \exists \vec{y_n} \bigwedge_{i_n} at_{i_n}(\vec{x}, \vec{y_n})$. We conceive a UCQ as a set of CQs. The existential quantifiers in UCQs are interpreted in the same way as for FOL formulas (natural domain semantics) and not w.r.t. a given set of constants mentioned in the signature (active domain semantics).

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 $Q = \psi(\vec{x})$ 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 [3, Thm 4.14, Thm 5.15].

The rewritability results are proved with the so called chase construction known from database theory. The idea of the chase construction is to repair the ABox with respect to the constraints formulated in the TBox. If, e.g., the TBox contains the axiom $A_1 \sqsubseteq A_2$ and the ABox contains $A_1(a)$ but not $A_2(a)$, then it is enriched by the atom $A_2(a)$. This procedure is applied stepwise to yield a sequence of ABoxes S_i starting with the original ABox as S_0. The resulting set of ABox axioms $\bigcup S_i$ may be infinite but induces a canonical model $can(\mathcal{O})$ for the ABox and the TBox axioms being used in the chasing process (see below). We will summarize the chase construction for DL-Lite.

Let $\mathcal{T}$ be a DL-Lite TBox, let $\mathcal{T}_p$ be the subset of positive inclusion (PI) axioms in $\mathcal{T}$ (no negation symbol allowed) and let $\mathcal{A}$ be an ABox and $\mathcal{O} = \mathcal{T} \cup \mathcal{A}$. Chasing will be carried out with respect to PIs only. Let $S_0 = \mathcal{A}$. Let S_i be the set of ABox axioms constructed so far and α be a PI axiom in $\mathcal{T}_p$. Let α be of the form $A_1 \sqsubseteq A_2$ and let $\beta \in S_i$ (resp. $\beta \subseteq S_i$) be an ABox axiom (resp. set of ABox axioms). The PI axiom α is called applicable to β if β is of the form $A_1(a)$ and $A_2(a)$ is not in S_i. The applicability of other PI axioms of the form $B \sqsubseteq C$ is defined similarly [3, Def. 4.1, p. 287]. If the left-hand side of the PI is a conjunction of base concepts, e.g., if the PI is of the form $A_1 \sqcap \cdots \sqcap A_n \sqsubseteq A_0$, and if β is $\{A_1(a), \ldots, A_n(a)\}$ and $A_0(a)$ is not in S_i, then PI is applicable to β.

As there may be many possible applications of PI axioms to atoms and sets of atoms, one has to impose an order on the TBox axioms and the (finite) subsets of the ABox. So we assume that all strings over the signature $Sig(\mathcal{O})$ of the ontology and some countably infinite set of new constants C_{ch} are well ordered. Such a well ordering exists and has the order type of the natural numbers $\mathbb{N}$. This ordering is different from the one of [3]; but it can also be used also for infinite ABoxes and it can handle concept conjunction. If there is a PI axiom α applicable to an atom β in S_i, one takes the minimal pair (α, β) with respect to the ordering and produces the next level $S_{i+1} = S_i \cup \{\beta_{new}\}$; here β_{new} is the atom that results from applying the chase rule for (α, β) as listed in Def. 3. The primed constants are the chasing constants from C_{ch}.

Definition 3 (Chasing rules for DL-Lite$^{\sqcap}_{\mathcal{F},\mathcal{R}}$)**.** If $\alpha = A_1 \sqsubseteq A_2$ and $\beta = A_1(a)$ then $\beta_{new} = A_2(a)$

If $\alpha = A_1 \sqsubseteq \exists R$ and $\beta = A_1(a)$ then $\beta_{new} = R(a, a')$

If $\alpha = \exists R \sqsubseteq A$ and $\beta = R(a, b)$ then $\beta_{new} = A(a)$

If $\alpha = \exists R_1 \sqsubseteq \exists R_2$ and $\beta = R_1(a, b)$ then $\beta_{new} = R_2(a, a')$

If $\alpha = R_1 \sqsubseteq R_2$ and $\beta = R_1(a, b)$ then $\beta_{new} = R_2(a, b)$

If $\alpha = A_1 \sqcap \cdots \sqcap A_n \sqsubseteq A_0$ and $\beta = \{A_1(a), \ldots, A_n(a)\}$ then $\beta_{new} = A_2(a)$ (and similarly for other PIs of the form $B_1 \sqcap \cdots \sqcap B_n \sqsubseteq C$)

The chase is defined by $chase(\mathcal{O}) = chase(\mathcal{T}_p \cup \mathcal{A}) = \bigcup_{i \in \mathbb{N}} S_i$. The canonical model $can(\mathcal{O})$ is the minimal Herbrand model of $chase(\mathcal{O})$.

The canonical model $can(\mathcal{O})$ is a universal model of $\mathcal{T}_p \cup \mathcal{A}$ with respect to homomorphisms. In particular this implies that answering a UCQ Q w.r.t. to $\mathcal{T}_p \cup \mathcal{A}$ can be reduced to answering $Q^{can(\mathcal{O})}$ w.r.t. to $DB(\mathcal{A})$. More concretely, (some finite closure $cln(\mathcal{T})$ of) the negative inclusions axioms and the functionality axioms are only relevant for checking the satisfiability of the ontology which can be tested by a simple FOL query. We leave out the details of the construction (see the extended version of this paper [11]).

3 Uncontrolled Combinations

We intend to construct geo-thematic description logics in which the combinations between the thematic domain and the spatial domain are maintained by constructors available in the logic $\mathcal{ALC}$(RCC8) of [8]. The reason for focusing on these constructors is that they are well behaved in so far as they control the "information flow" from the spatial domain (more generally: from the ω-admissible domain) to the thematic domain. In case of the logic $\mathcal{ALC}$(RCC8) this controlled information flow has the effect that subsumption w.r.t. $\mathcal{ALC}$(RCC8) TBoxes is decidable, i.e., there is a complete and correct algorithm to test whether $\mathcal{T} \models C_1 \sqsubseteq C_2$ holds for arbitrary $\mathcal{ALC}$(RCC8) concepts C_1, C_2. But $\mathcal{ALC}$(RCC8) does not allow for FOL rewritability w.r.t. to satisfiability checks as already $\mathcal{ALC}$ and RCC8 do not allow for FOL rewritability. Weakening the expressivity of the thematic component ($\mathcal{ALC}$) or of the spatial component (RCC8) does not change this fact. This can be seen as follows. If the way in which both components are combined is not (further) restricted either, it is possible to construct complex networks of the spatial component in the combined logic—and testing the consistency of these (in case of RCC5 and RCC8) is NP-hard. We will give two examples of logics for which FOL rewritability does not hold even though the thematic part is weakened to a lightweight logic (Propositions 4, 5).

Our approach diverges from the one of [8] in the point that we do not presuppose an ω-admissible domain [8, Def. 5, p. 7] but a finite set T_ω of FOL formulas (Ax_{RCCi}) that has the corresponding properties of ω-admissible domains. So we explicitly represent the axioms of the domain rather than make calls to an oracle. The main reason for this shift from a concrete domain to a theory is the intended use of known techniques for query answering such as the chase construction.

Let Rel be a finite set of binary relation symbols, $Const$ be a set of constants and T_ω be a finite set of sentences with respect to a signature containing Rel and $Const$. A network $\mathcal{N}$ is a set of sentences over $Rel \cup Const$ of the form $r_1(a^*, b^*) \vee \cdots \vee r_k(a^*, b^*)$ for $r_1, \ldots, r_k \in Rel$ and $a^*, b^* \in Const$. $\mathcal{N}$ is called complete if it contains only atomic sentences $r(a^*, b^*)$ and for all constants a^*, b^* in $\mathcal{N}$ there is a $r \in Rel$ such that $r(a^*, b^*) \in \mathcal{N}$. For two complete finite networks $\mathcal{N}, \mathcal{M}$ let $I_{\mathcal{N},\mathcal{M}}$ denote the atoms $r(a^*, b^*) \in \mathcal{N}$ such that a^*, b^* occur in $\mathcal{M}$. The restriction $\mathcal{N}_{Const'}$ of a network to the set of constants $Const'$ is the subset of $\mathcal{N}$ restricted to those sentences containing only constants from $Const'$.

T_ω is an ω-admissible theory iff it fulfills the following conditions: T_ω is satisfiable; T_ω implies the JEPD-property for the relations in *Rel*; testing whether a finite complete syntactic network $\mathcal{N}$ is satisfiable with respect to T_ω, i.e., testing whether $T_\omega \cup \mathcal{N}$ is satisfiable, is decidable;if $\mathcal{N}, \mathcal{M}$ are finite complete networks that are satisfiable relative to T_ω, respectively, and if $I_{\mathcal{N},\mathcal{M}} = I_{\mathcal{M},\mathcal{N}}$, then $\mathcal{N} \cup \mathcal{M}$ is satisfiable relative to T_ω, too (*patchwork property*).

We recapitulate the syntax and the semantics of the constructors of [8] 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 B_{RCCi} and $r = \{r_1, \ldots r_n\} \equiv r_1 \vee \cdots \vee r_n$ for $r_i \in B_{RCCi}$. 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 $(e_1^*, e_2^*) \in r^{\mathcal{I}}\}$.

In the following we will restrict our attention to the most used fragments of the region connection calculus, RCC5 and RCC8. Testing the satisfiability of arbitrary RCC8 (RCC5) constraint networks is not FOL rewritable; therefore we have to weaken the notion of FOL rewritability for the combined logics using RCC8 (RCC5) as the spatial part. We call an ABox $\mathcal{A}$ *spatially complete* iff the constraint network $\mathcal{N}_{\mathcal{A}}$ contained in $\mathcal{A}$ is complete and satisfiable.

Real-world (geographical) databases may be incomplete. Take, e.g., a database in which a hospital in a park is not modelled as a polygon but as a point. In this case, more than one of the base relations may hold between the hospital and the park. For instance, one does not know, whether the hospital touches the park from within or whether it resides as an island in the park. Hence, we have to assume that this kind of incompleteness is removed in a preprocessing step.

Though spatial completeness of ABoxes can be used for the definition of a weaker notion of FOL rewritability, i.e, FOL rewritability w.r.t. to ABoxes fulfilling spatial completeness, uncontrolled combinations of a lightweight description logic with RCC still may not allow for weak FOL rewritability. In fact, the naively combined logics in the following propositions are examples of such logics. In the proofs we use the fact that FOL rewritability of solving a problem implies (time) data complexity AC^0 for solving it [3]. *Data complexity* is the complexity w.r.t. to the size of the ABox, i.e., the TBox and the query are fixed.

Proposition 4. *Consider the following simple description logic, which we denote by $\mathcal{L}_{\mathcal{F}}^0(RCC8)$. Let $r \in Rel_{RCC8}$ and let $T_\omega = Ax_{RCC8}$.*

$$U \longrightarrow l \mid \tilde{R} \qquad C_l \longrightarrow A \qquad C_r \longrightarrow \exists U_1, U_2.r$$

$$\textit{TBox:} \qquad C_l \sqsubseteq C_r, (\text{funct } R), (\text{funct } l)$$

$$\textit{ABox:} \qquad A(a), R(a,b), l(a,a^*), r(a^*,b^*)$$

The data complexity for checking satisfiability of $\mathcal{L}_{\mathcal{F}}^0(RCC8)$ ontologies w.r.t. spatially complete ABoxes is NP*-hard, hence satisfiability checking is not FOL rewritable.*

The reason for this negative result is the fact that functional roles make it possible to identify the regions of the abstract objects, and therefore arbitrary constraint networks (for which consistency checking is NP-complete) can be constructed. (For full proofs see [11].) A similar effect is caused by using all-quantifier constructs of the form $\forall U_1, U_2.r$ on the right-hand side of TBox axioms as well as inverse roles.

Proposition 5. *Consider the following simple combined description logic, called $\mathcal{L}^1_{\mathcal{F}}(RCC8)$. Let $r \in Rel_{RCC8}$ and $T_\omega = Ax_{RCC8}$.*

$$\begin{array}{lll} U \longrightarrow l \mid \tilde{R} & C_l \longrightarrow A | \exists R \mid \exists R^{-1} \mid \exists l & C_r \longrightarrow \forall U_1, U_2.r \\ \textit{TBox:} & C_l \sqsubseteq C_r, (\text{funct } l) & \\ \textit{ABox:} & A(a), R(a,b), l(a,a^*), r(a^*,b^*) & \end{array}$$

The data complexity for checking satisfiability of $\mathcal{L}^1_{\mathcal{F}}(RCC8)$ ontologies w.r.t. spatially complete ABoxes is NP*-hard, hence satisfiability checking is not FOL rewritable.*

The same ideas in the proof of the propositions above can be applied to show that logics $\mathcal{L}^0_{\mathcal{F}}$(RCC5) and $\mathcal{L}^1_{\mathcal{F}}$(RCC5) are not weakly FOL-rewritable because checking consistency of RCC5 constraint networks is NP-hard, too [4].

4 Combinations Allowing for FOL Rewritability w.r.t. Satisfiability

There are at least four ways in which to further restrict the combined logic so that FOL rewritability with respect to satisfiability checks can be guaranteed. First, one may presuppose further conditions on the ABox, which may go beyond the spatial completeness condition used in this paper. In imposing these conditions, information on ABoxes is conveyed before the actual ABox is presented for the satisfiability test. In this paper we will stick to the classical (though slightly weakened) notion of FOL rewritability mentioned above and try to weaken the combined logic. This leads to three new ways to reach FOL rewritable combined logics, i.e., either weakening the spatial calculus or weakening the interaction of the combined components or weakening the thematic part.

We first look at the options of weakening the thematic part together with restricting the interaction with the spatial component. We consider the very weak extension DL-Lite$^{\sqcap}_{\mathcal{F},\mathcal{R}}$(RCC8) of DL-Lite in which concepts of the form $\exists U_1, U_2.r$ may appear on the right-hand side of TBox axioms; but the role symbols in U_i are not allowed to be functional.

Definition 6 (DL-Lite$^{\sqcap}_{\mathcal{F},\mathcal{R}}$(RCC8))**.** Let $r \in Rel_{RCC8}$ and $T_\omega = Ax_{RCC8}$.

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

*) Restriction: If (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 $\exists U_1, U_2.r$.

Concepts of the form $\exists U_1, U_2.r$ may lead to inconsistencies that do not already follow from the thematic part. But these can be captured by FOL queries.

Proposition 7. *Checking the satisfiability of DL-Lite$^{\sqcap}_{\mathcal{F},\mathcal{R}}$(RCC8)-ontologies that have a spatially complete ABox is FOL-rewritable.*

We turn now to the other way of reaching FOL rewritability mentioned above: weakening the expressivity of the spatial component. One may ask whether a combination with the calculus RCC3 or RCC2 [16], both very weakly expressible fragments, allows for weak FOL rewritability w.r.t. satisfiability checks. Though the interest for such combinations cannot be justified by their use as fine-grained means for representing a geo-thematic domain, their potential use as logics for approximating ontologies in more expressible combined logics like $\mathcal{ALC}$(RCC8) makes the investigation valuable [6].

Definition 8 (DL-Lite$^{\sqcap,+}_{\mathcal{F},\mathcal{R}}$(RCC2) and DL-Lite$^{\sqcap,+}_{\mathcal{F},\mathcal{R}}$(RCC3))**.** Let $T_\omega = Ax_{RCC2}$ resp. $T_\omega = Ax_{RCC3}$ and $r \in \mathcal{B}_{RCC2}$ resp. $r \in \mathcal{B}_{RCC3}$

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

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

Proposition 9 shows that the combination of RCC2 (resp. RCC3) with DL-Lite$^{\sqcap}_{\mathcal{F},\mathcal{R}}$is FOL rewritable w.r.t. to satisfiability. The reason is that RCC2 and RCC3 have very simple composition tables: In case of RCC2 for all combinations of the two base relations dr, o the maximally unspecified relation $\{\mathsf{dr}, \mathsf{o}\}$ results, hence any RCC2 constraint network is satisfiable. In case of of RCC3 only eq may cause inconsistencies, but these can be identified by a FOL query.

Proposition 9. *Satisfiability checking of ontologies in the logics DL-Lite$^{\sqcap,+}_{\mathcal{F},\mathcal{R}}$(RCC2) and DL-Lite$^{\sqcap,+}_{\mathcal{F},\mathcal{R}}$(RCC3) is FOL rewritable.*

5 Combinations Allowing for FOL Rewritability w.r.t. Query Answering

We mentioned above four factors that are relevant for the question whether FOL rewritability w.r.t. to satisfiability checks is given. These have to be supplemented by another factor when rewritability is considered w.r.t. to query answering, namely the expressivity of the query language. The query language which we consider is derived from grounded conjunctive queries and is denoted by GCQ^+.

Definition 10. A *GCQ^+ atom w.r.t. DL-Lite$^{\sqcap}_{\mathcal{F},\mathcal{R}}$(RCC8)* is a formula of one of the following forms:

- $C(x)$, where C is a DL-Lite$^{\sqcap}_{\mathcal{F},\mathcal{R}}$(RCC8) concept without the negation symbol and x is a variable or a constant.
- $(\exists R_1 \dots R_n.C)(x)$ for role symbols or inverses of role symbols R_i, a DL-Lite$^{\sqcap}_{\mathcal{F},\mathcal{R}}$(RCC8) concept C without the negation symbol, and a variable or a constant x
- $R(x, y)$ for a role symbol R or an inverse thereof

- $l(x, y^*)$, where x is a variable or constant and y^* is a variable or constant intended to denote elements of Ax_{RCC8}
- $r(x^*, y^*)$, where $r \in Rel_{RCC8}$ and x^*, y^* are variables or constants intended to denote elements of Ax_{RCC8}

A *GCQ^+ query w.r.t. DL-Lite$^{\sqcap}_{\mathcal{F},\mathcal{R}}$(RCC8)* is a query of the form $\tilde{\exists}\vec{y}\vec{z}^* \bigwedge C_i(\vec{x}, \vec{w}^*, \vec{y}, \vec{z}^*)$ where all $C_i(\vec{x}, \vec{w}^*, \vec{y}, \vec{z}^*)$ are GCQ^+ atoms and $\tilde{\exists}\vec{y}\vec{z}^* = \tilde{\exists}y_1 \dots \tilde{\exists}y_n \tilde{\exists}z_1^* \dots \tilde{\exists}z_m^*$ is a sequence of $\exists$-quantifiers interpreted w.r.t. the active domain semantics.

We want to adapt the algorithm PerfectRef [3, Fig. 13] for reformulating UCQs w.r.t. DL-Lite ontologies to our setting in which GCQ^+-queries are issued to DL-Lite$^{\sqcap}_{\mathcal{F},\mathcal{R}}$(RCC8) ontologies. The first step for the adaptation is to translate GCQ^+-queries into CQs. This can be handled by introducing a new concept symbol intended to denote all elements in the database. Let $\tau_1(Q)$ be the result of the transformation to a UCQ and define the answers of a GCQ^+-query w.r.t. $\mathcal{O}$ by: $cert(\mathcal{O}, Q) = cert(\mathcal{O}, \tau_1(Q))$.

For the second step of the adaptation of the PerfectRef algorithm we consider a partial type of transformation $\tau_2(\cdot, \cdot)$ with two arguments which transforms only the atom occurrences given as second argument to classical predicate logical atoms. For example, let $Q = \exists R.A(x) \wedge \exists \tilde{R}_1, l.\mathsf{tpp}(y)$ and $X = \tau_2(Q, \{\exists \tilde{R}_1, l.\mathsf{tpp}(y)\})$ for short. Then $X = \exists z, z^*, x^*.\exists R.A(x) \wedge R(y, z) \wedge l(z, z^*) \wedge l(x, x^*) \wedge \mathsf{tpp}(z^*, x^*)$. The resulting query $\tau_2(Q)$ is a hybrid union of conjunctive queries whose conjuncts are either classical predicate logical atoms or GCQ^+-atoms. The unary $\tau_2(\cdot)$ can be defined by the binary transformation $\tau_2(\cdot, \cdot)$ by $\tau_2(Q, \{at \mid at \text{ is a } GCQ^+\text{-atom in } Q \text{ not of the form } \exists U_1, U_2.r(x)\})$.

We give a description of our adapted algorithm in the following. In Algorithm 1 we use the notation "$g = F$" for "g is of the form F". The original algorithm PerfectRef operates on the PI axioms of a DL-Lite ontology by using them as rewriting aids for the atomic formulas in the UCQ. Lines 5–12 and 28–34 of our adapted algorithm (Algorithm 1) make up the original PerfectRef. Roughly, the PerfectRef algorithm acts in the inverse direction with respect to the chasing process. For example, if the TBox contains the PI axiom $A_1 \sqcap A_2 \sqsubseteq A_3$, and the UCQ contains the atom $A_3(x)$ in a CQ, then the new rewritten UCQ query contains a CQ in which $A_3(x)$ is substituted by $A_1(x) \wedge A_2(x)$. The applicability of a PI axiom to an atom is restricted in those cases where the variables of an atom are either distinguished variables or also appear in another atom of the CQ at hand. To handle these cases, PerfectRef—as well as also our adapted version—uses anonymous variables $\_$ to denote all non-distinguished variables in an atom that do not occur in other atoms of the same CQ. The function anon (line 31 in Algorithm 1) implements the anonymization. The application conditions for PI axioms α and atoms are as follows: α is applicable to $A(x)$ if A occurs on the right-hand side; and α is applicable to $R(x_1, x_2)$, if $x_2 = \_$ and the right-hand side of α is $\exists R$; or $x_1 = \_$ and the right-hand side of α is $\exists R^-$; or α is a role inclusion assertion and its right-hand side is either R or R^-. The outcome $gr(g, \alpha)$ of applying an applicable PI α to an atom g corresponds to the outcome of resolving α with g. For example, if α is $A \sqsubseteq \exists R$ and the atom g is $R(x, \_)$, then the result of the application is $gr(g, \alpha) = A(x)$. We leave out the details here [3, Fig.12, p. 307]. In the algorithm PerfectRef, atoms in a CQ are rewritten with the PI axioms (lines 6–11) and if possible merged by the function reduce (line 31) which unifies the atoms with the most general unifier (lines 28–34).

input : a hybrid query $\tau_1(Q) \cup \tau_2(Q)$, DL-Lite(RCC8) TBox $\mathcal{T}$
output: a UCQ *pr*
$pr := \tau_1(Q) \cup \tau_2(Q)$;
repeat
$pr' := pr$;
foreach *query* $q' \in pr'$ **do**
foreach *atom* g *in* q' **do**
if g *is a FOL-atom* **then**
foreach *PI* α *in* $\mathcal{T}$ **do**
if α *is applicable to* g **then**
$pr := pr \cup \{q'[g/gr(g, \alpha)]\}$;
end
end
else
if $g = \exists \tilde{R}_1, \tilde{R}_2.r_3(x)$ *and* $r_1; r_2 \subseteq r_3$ **then**
$X := q'[g/(\exists \tilde{R}_1, l.r_1(x) \wedge \exists l, \tilde{R}_2.r_2(x))]$;
$pr := pr \cup \{X\} \cup \{\tau_2\big(X, \{\exists \tilde{R}_1, l.r_1(x), \exists l, \tilde{R}_2.r_2(x)\}\big)\}$
end
if $g = \exists U_1, U_2.r_1(x)$ *and* $B \sqsubseteq \exists U_1, U_2.r_2(x) \in \mathcal{T}$ *for* $r_2 \subseteq r_1$ **then**
$pr := pr \cup \{q'[g/B(x)]\}$;
end
if $g = \exists U_1, U_2.r_1(x)$ *and* $B \sqsubseteq \exists U_1, U_2.r_2(x) \in \mathcal{T}$ *for* $r_2^{-1} \subseteq r_1$ **then**
$pr := pr \cup \{q'[g/B(x)]\}$;
end
if $g = \exists \tilde{R}_1, U_1.r(x)$ *(resp.* $\exists U_1, \tilde{R}_1.r(x)$*) and (*$R_2 \sqsubseteq R_1 \in \mathcal{T}$ *or* $R_2^{-1} \sqsubseteq R_1^{-1} \in \mathcal{T}$*)* **then**
$X := q'[g/(g[R_1/R_2])]$;
$pr := pr \cup \{X\} \cup \{\tau_2\big(X, \{g[R_1/R_2]\}\big)\}$;
end
end
end
foreach *pair of FOL-atoms* g_1, g_2 *in* q' **do**
if g_1 *and* g_2 *unify* **then**
$pr := pr \cup \{anon(reduce(q', g_1, g_2))\}$;
end
end
end
until $pr' = pr$;
return $drop(pr)$

Algorithm 1: Adapted PerfectRef

The modification of PerfectRef concerns the handling of GCQ^+-atoms that have the the form $\exists U_1, U_2.r(x)$. These atoms may have additional implications that are accounted for with four cases (lines 12–26 of the algorithm). At the end of the adapted algorithm PerfectRef (Algorithm 1, line 35) these atoms are deleted by calling the function *drop*, whose implementation is omitted here. So the algorithm returns a classical UCQ, which can be evaluated as a SQL query on the database $DB(\mathcal{A})$.

Let us demonstrate the rewriting algorithm with a simple example [10]. Think of an engineering bureau that wants to plan additional parks in a city. It has 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, the mappings for *Park+Lake* and *Park4Playing* shall produce a (non-localized) a as instance of *Park+Lake*, *Park4Playing*, respectively. So assume further, that we have $\{Park{+}Lake(a), Park4Playing(a)\} \subseteq \mathcal{A}$. 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 following GCQ^+: $Q = Park(x) \wedge \exists hasLake \circ l, hasPlAr \circ l.(\mathcal{B}_{RCC8} \setminus \{\mathsf{ntpp}\})(x)$. Using the composition $\mathsf{tpp}; \mathsf{tppi} = \{\mathsf{dc}, \mathsf{ec}, \mathsf{po}, \mathsf{tpp}, \mathsf{tppi}, \mathsf{eq}\} \subseteq \mathcal{B}_{RCC8} \setminus \{\mathsf{ntpp}\}$, the reformulation algorithm introduced above (lines 13–15) 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 $\exists hasPlAr \circ l, l.\mathsf{tpp}$ (lines 20–21) in combination with the rewriting rule for $A_1 \sqsubseteq A_2$ (Def. 3) we get another CQ $Q'' = Park{+}Lake(x) \wedge Park4Playing(x)$. Now, Q'' captures (as desired) the object a.

That the rewriting given in Algorithm 1 is indeed correct and complete follows from Theorem 11.

Theorem 11. *Answering GCQ^+-queries w.r.t. $DL\text{-}Lite^{\sqcap}_{\mathcal{F},\mathcal{R}}(RCC8)$ ontologies that have a spatially complete ABox is FOL-rewritable.*

We give a proof sketch. The proof follows the proof of Theorem 5.15 for pure DL-Lite ontologies [3]. We adapt the chase construction to account for the RCC8 relations $r \in Rel_{RCC8}$. The main observation is that the disjunctions in r can be nearly handled as if they were predicate symbols.

Because of Prop. 7 we may assume that $\mathcal{O}$ is satisfiable. Let pr be the UCQ resulting from applying the algorithm to Q and $\mathcal{O}$. We have to show that $cert(Q, \mathcal{O}) = (pr)^{DB(\mathcal{A})}$. These can be done in two main steps of which the first will be sketched here as it contains the main chase-like construction $chase^*(\mathcal{O})$. After the construction, one has to describe what it means to answer Q with respect to $chase^*(\mathcal{O})$ is, resulting in the set $ans(Q, chase^*(\mathcal{O}))$, and then show that $ans(Q, chase^*(\mathcal{O})) = cert(Q, \mathcal{O})$. In the second step, which we leave out here (see the extended version of this paper) one has to show that $ans(Q, chase^*(\mathcal{O})) = (pr)^{DB(\mathcal{A})}$.

For the construction of $chase^*(\mathcal{O})$ one uses the chase rules of Def. 3 and the special rule (R).

Chasing Rule (R)

If $B(x) \in S_i$ and there are no y, y^*, x^* s.t. $\{R_1(x, y), l(y, y^*), l(x, x^*), r_1(y^*, x^*)\}$ is contained in S_i, then let $S_{i+1} = S_i \cup \{R_1(x, y), l(y, y^*), l(x, x^*), r_1(y^*, x^*)\}$. The constants y, y^* are completely new constants not appearing in S_i. The constant x^* is the old x^* if already in S_i, otherwise it is also a completely new constant symbol.

Every time (R) is applied to yield a new ABox S_i, the resulting constraint network in S_i is saturated by calculating the minimal labels between the new added region constants and the other region constants. The application of (R) does not constrain the RCC8-relations between the old regions and even stronger: Let (R) be applied to a TBox axiom of the form $A \sqsubseteq \exists \tilde{R}, l.r$ and $A(a) \in S_i$ resulting in the addition of $R(a, b)$, $l(b, b^*)$ and $r(b^*, a^*)$. Then it is enough to consider all $c^* \in S_i$ and all relations r_{c^*,a^*} such that $r_{c^*,a^*}(c^*, a^*) \in S_i$. The composition table gives the outcome $r_{c^*,a^*}; r = r'_{c^*,b^*}$ and one adds $r'_{c^*,b^*}(c^*, b^*)$ to S_i. After this step, which we call triangulation step, one closes the assertions up with respect to the subset relation between RCC8-relations and with respect to symmetry. I.e., if $r_1(x^*, y^*)$ is added to S_i, then one also adds $r_2(x^*, y^*)$ for all r_2 such that $r_1 \subseteq r_2$ and $r_2^{-1}(y^*, x^*)$. For different c_1^*, c_2^*, assertions of the form $r_{c_1^*,b^*}(c_1^*, b^*)$ and $r_{c_2^*,b^*}(c_2^*, b^*)$ do not constrain each other (because of the patch work property). The saturation leads to a finite set S_{i+k} (for some $k \in \mathbb{N}$) that is a superset of S_i. Let $chase^*(\mathcal{O}) = \bigcup S_i$. The set $chase^*(\mathcal{O})$ does not induce a single canonical model. But it is universal in the following sense: For every model $\mathcal{I}$ of $\mathcal{O}$ define a model $\mathcal{I}_c$ out of $chase^*(\mathcal{O})$ by taking a (consistent) configuration of the contained RCC8-network and taking the minimal model of this configuration and the thematic part of $chase^*(\mathcal{O})$. Then $\mathcal{I}_c$ maps homomorphically to $\mathcal{I}$. Now one can define that answers of GCQ^+-queries with respect to $chase^*(\mathcal{O})$ are given by homomorphic embeddings and show that these answers are exactly the certain answers w.r.t. the ontology $\mathcal{O}$.

Theorem 11 generalizes results in prior work of the authors [10, 9]. The combined logic DL-Lite(RCC8) [10] cannot be used (and was not intended to be used) as a logical tool for a fine-grained representation of geo-thematic domains as it does not offer the means to formulate sufficient conditions for geo-thematic concepts; e.g., one would like to give a sufficient condition for a park containing a lake by (among other axioms) $\tau = \exists(parkHasLake\ l)(l).\mathsf{tpp} \vee \mathsf{ntpp} \sqsubseteq ParkWithLake$. Such an axiom cannot be formulated in DL-Lite$^{\sqcap}_{\mathcal{F},\mathcal{R}}$(RCC8) either. But DL-Lite$^{\sqcap}_{\mathcal{F},\mathcal{R}}$(RCC8) provides better means for approximating τ according to the method of [6]. The general idea of the approximation methodology of [6] is to strengthen the TBox $\mathcal{T}$ to a logically stronger TBox $\mathcal{T}'$. The answer set of a query Q to $\mathcal{T}' \cup \mathcal{A}$ contains all answers of the same query to $\mathcal{T} \cup \mathcal{A}$, but it may also contain false positives. So in a post-processing step the answer set w.r.t. $\mathcal{T}' \cup \mathcal{A}$ has to be checked with a conventional prover for the approximated logic. Coming back to the example, the axiom τ could be strengthened to the logically stronger axiom $\tau' = \exists parkHasLake \sqcap \exists l \sqsubseteq ParkWithLake$ in DL-Lite$^{\sqcap}_{\mathcal{F},\mathcal{R}}$(RCC8). If one uses DL-Lite(RCC8) for approximation, the strengthening may result an even stronger axiom $\tau'' = \exists parkHasLake \sqsubseteq ParkWithLake$. Overall, the resulting strengthening $\mathcal{T}''$ w.r.t. DL-Lite(RCC8) may become stronger than the resulting strengthening w.r.t. $\mathcal{T}'$, so that $cert(Q, \mathcal{T}' \cup \mathcal{A}) \subseteq cert(Q, \mathcal{T}'' \cup \mathcal{A})$.

6 Conclusion

It is possible to define combinations of fine-grained region connection calculi (such as RCC5 or RCC8) with lightweight descriptions logics such that FOL rewritability w.r.t. to query answering is fulfilled. The result is achieved under further restrictions: Spatial completeness of the ABox; controlled interaction of the thematic and spatial component; and a restricted query language. Only for weakly expressible spatial calculi like RCC2 it seems to be possible to allow for a stronger coupling of the thematic and spatial part. Our contribution adds to the results of [3] the incorporation of a spatial component in form of RCC. Moreover, in using Ax_{RCC8} we switch the perspective of [8] from concrete domains to axiomatic theories, which can be better integrated into techniques needed for query answering. This switch could also be applied if other concrete domains (e.g., numbers) have to be accounted for in query answering.

In pushing the expressivity of the thematic part from known DL-Lite logics to DL-Lite$^{\sqcap}_{\mathcal{F},\mathcal{R}}$the application potential of the combined geo-thematic logics is increased. The approximation idea using DL-Lite$^{\sqcap}_{\mathcal{F},\mathcal{R}}$(RCC8) as an approximation aid for $\mathcal{ALC}$(RCC8) logics has still to be worked out in detail and experimentally tested on real-world ontologies. The realizability of the experiments depends on an ontology satisfiability tester for $\mathcal{ALC}$(RCC8) ontologies, which is still missing.

In future work, DL-Lite$^{\sqcap}_{\mathcal{F},\mathcal{R}}$has to be extended further to even more expressible lightweight logics guaranteeing FOL rewritability such as Datalog$^{\pm}_{0}$ [2], an extension of Datalog that allows for a (restricted) use of tuple generating and equality generating dependencies.

References

[1] Franz Baader, Diego Calvanese, Deborah L. McGuinness, Daniele Nardi, and Peter F. Patel-Schneider, editors. *The Description Logic Handbook: Theory, Implementation and Applications*. Cambridge University Press, 2003.

[2] Andrea Calì, Georg Gottlob, and Thomas Lukasiewicz. A general datalog-based framework for tractable query answering over ontologies. Technical Report CL-RR-10-21, Oxford University Computing Laboratory, November 2010.

[3] Diego Calvanese, Giuseppe De Giacomo, Domenico Lembo, Maurizio Lenzerini, Antonella Poggi, Mariano Rodríguez-Muro, and Riccardo Rosati. Ontologies and databases: The DL-Lite approach. In Sergio Tessaris and Enrico Franconi, editors, *Semantic Technologies for Informations Systems – 5th Int. Reasoning Web Summer School (RW-2009)*, volume 5689 of *LNCS*, pages 255–356. Springer, 2009.

[4] Michelangelo Grigni, Dimitris Papadias, and Christos H. Papadimitriou. Topological inference. In *IJCAI (1)*, pages 901–907, 1995.

[5] Volker Haarslev, Carsten Lutz, and Ralf Möller. A description logic with concrete domains and a role-forming predicate operator. *J. Log. Comput.*, 9(3):351–384, 1999.

[6] Alissa Kaplunova, Ralf Moeller, Sebastian Wandelt, and Michael Wessel. Towards scalable instance retrieval over ontologies. In Bi Yaxin and Williams Mary-Anne, editors, *Knowledge Science, Engineering and Management, Fourth International Conference, KSEM 2010, Proceedings*, volume 6291 of *Lecture Notes in Computer Science*. Springer, 2010.

[7] Sanjiang Li and Mingsheng Ying. Generalized region connection calculus. *Artificial intelligence*, 160(1):1–34, 2004.

[8] Carsten Lutz and Maja Miličić. A tableau algorithm for description logics with concrete domains and general TBoxes. *J. Autom. Reasoning*, 38(1-3):227–259, 2007.

[9] Özgür L. Özçep and Ralf Möller. Combining lightweight description logics with the region connection calculus. Technical report, Institute for Softwaresystems (STS), Hamburg University of Technology, 2011. Available online at `http://www.sts.tu-harburg.de/tech-reports/papers.html`.

[10] Özgür L. Özçep and Ralf Möller. Computationally feasible query answering over spatio-thematic ontologies, 2011. To appear in the Proceedings of GeoProcessing 2012.

[11] Özgür L. Özçep and Ralf Möller. 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`.

[12] David A. Randell, Zhan Cui, and Anthony G. Cohn. A spatial logic based on regions and connection. In *Proceedings of the 3rd International Conferecence on Knowledge Representation and Reasoning*, pages 165–176, 1992.

[13] Jochen Renz. *Qualitative Spatial Reasoning with Topological Information*, volume 2293 of *Lecture Notes in Computer Science*. Springer, 2002.

[14] Jochen Renz and Bernhard Nebel. Qualitative spatial reasoning using constraint calculi. In Marco Aiello, Ian Pratt-Hartmann, and Johan Benthem, editors, *Handbook of Spatial Logics*, pages 161–215. Springer Netherlands, 2007.

[15] Mariano Rodriguez-Muro and Diego Calvanese. Semantic index: Scalable query answering without forward chaining or exponential rewritings. In *Posters of the 10th Int. Semantic Web Conf. (ISWC 2011)*, 2011.

[16] Michael Wessel. 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.

Appendix: Proofs

Construction of $cln(\mathcal{T})$

We give details on the construction of $cln(\mathcal{T})$. The finite closure $cln(\mathcal{T})$ of the negative inclusions axioms and the functionality axioms are (only) relevant for checking the satisfiability of the ontology which can be tested by a simple FOL query. With induction on the stepwise construction of the chase one can show that $can(\mathcal{O})$ is a model of the whole ontology $\mathcal{O}$ iff the negative inclusion axioms and functionality axioms are in accordance with the original ABox. The authors of [3] define the *negative closure* $cln(\mathcal{T})$ in order to capture all possible conflicts [3, Def. 4.7, p. 292]. We extend their definition of the negative closure to the logic DL-Lite$^{\sqcap}_{\mathcal{F},\mathcal{R}}$ and reformulate the definition in an alternative representation that fits better to the extensions $cl_{\perp}$ that we will use in the proofs of some propositions. Let $\bar{B}$ be an abbreviation for $B_1 \sqcap \cdots \sqcap B_n$. (1) All functionality axioms of $\mathcal{T}$ are in $cln(\mathcal{T})$; (2) for all negative inclusions $\bar{B} \sqsubseteq \neg B \in \mathcal{T}$ let $\bar{B} \sqcap B \sqsubseteq \perp \in cln(\mathcal{T})$; (3) if $\bar{B} \sqsubseteq B \in \mathcal{T}$ and $B \sqcap B' \sqsubseteq \perp \in cln(\mathcal{T})$, then $\bar{B} \sqcap B' \sqsubseteq \perp \in cln(\mathcal{T})$; (4) if $P_1 \sqsubseteq P_2 \in \mathcal{T}$ and $\exists P_2 \sqcap \bar{B} \sqsubseteq \perp \in cln(\mathcal{T})$, then $\exists P_1 \sqcap \bar{B} \sqsubseteq \perp \in cln(\mathcal{T})$; (5) if $P_1 \sqsubseteq P_2 \in \mathcal{T}$ and $\exists P_2^- \sqcap \bar{B} \sqsubseteq \perp \in cln(\mathcal{T})$, then $\exists P_1^- \sqcap \bar{B} \sqsubseteq \perp \in cln(\mathcal{T})$; (6) let $X := \{\exists P \sqsubseteq \neg \exists P, \exists P^- \sqsubseteq \neg \exists P^-\}$; if $X \cap cln(\mathcal{T}) \neq \emptyset$, then $X \subseteq cln(\mathcal{T})$. The FOL query for testing the satisfiability is built as the disjunction of boolean queries q_τ for every $\tau \in cln(\mathcal{T})$ in the following way: if $\tau = (\text{funct } R)$ then $q_\tau = \exists x, y, z R(x,y) \wedge R(x,z) \wedge y \neq z$; if $\tau = A \sqsubseteq \perp$, then $q_\tau = \exists x.A(x)$; for the other $\tau \in cln(\mathcal{T})$ the query q_τ is defined similarly [3, p. 296].

Proof of Proposition 4

We construct a generic TBox $\mathcal{T}_g$ that allows one to encode any RCC8 constraint network so that checking the consistency of RCC8 constraint networks is reducible to a satisfiability check of this TBox and a spatially complete ABox. Let for every $r \in Rel_{RCC8}$ be given role symbols R^1_r, R^2_r. The generic TBox $\mathcal{T}_g$ has for every $r \in Rel_{RCC8}$ 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.

$$\mathcal{T}_g = \{A_r \sqsubseteq \exists \tilde{R}^1_r, \tilde{R}^2_r.r, (\text{funct } l, R^1_r, R^2_r) \mid r \in Rel_{RCC8}\}$$

Now, let $\mathcal{N}$ be an arbitrary RCC8 constraint network which has to be tested for relational consistency. Let $\mathcal{A}_{\mathcal{N}}$ be an ABox such that for every $r(a,b)$ in $\mathcal{N}$ three new constants are introduced: x_{ab}, x_a, x_b.

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

The construction immediately implies the following fact: $\mathcal{T}_g \cup \mathcal{A}_{\mathcal{N}} \cup Ax_{RCC8}$ is satisfiable iff $\mathcal{N} \cup Ax_{RCC8}$ is satisfiable. If the data complexity of the satisfiability check for $\mathcal{L}^0_{\mathcal{F}}(RCC8)$-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 RCC8 constraint networks is NP-hard and AC$^0 \subsetneq$ NP.

Chasing Rule (R)
If $B(x) \in S_i$ and there are no y, y^*, x^* s.t. $\{R_1(x, y), l(y, y^*), l(x, x^*), r_1(y^*, x^*)\}$ is contained in S_i, then let $S_{i+1} = S_i \cup \{R_1(x, y), l(y, y^*), l(x, x^*), r_1(y^*, x^*)\}$. The constants y, y^* are completely new constants not appearing in S_i. The constant x^* is the old x^* if already in S_i, otherwise it is also a completely new constant symbol.

Figure 1: Additional chasing rule that accounts for $\exists U_1, U_2.r$-concepts

Proof of Proposition 7

Let $\mathcal{T} \cup \mathcal{A} \cup Ax_{RCC8}$ be an ontology with a spatially complete ABox $\mathcal{A}$ and the ω-admissible background theory Ax_{RCC8}. We built a simple closure $\mathcal{T}'$ of the pure DL-Lite part of $\mathcal{T}$ in the following way. Every DL-Lite axiom of $\mathcal{T}$ is in $\mathcal{T}'$. For every $B \sqsubseteq \exists \tilde{R}_1, \tilde{R}_2.r \in \mathcal{T}$ let $\{B \sqsubseteq \exists R_1, B \sqsubseteq \exists R_2\} \subseteq \mathcal{T}'$, for $B \sqsubseteq \exists \tilde{R}, l.r \in \mathcal{T}$ and $B \sqsubseteq \exists l, \tilde{R}.r \in \mathcal{T}$ let $\{B \sqsubseteq \exists R, B \sqsubseteq \exists l\} \subseteq \mathcal{T}'$. We claim that $\mathcal{T} \cup \mathcal{A} \cup Ax_{RCC8}$ is satisfiable iff the DL-Lite ontology $\mathcal{T}' \cup (\mathcal{A} \setminus \mathcal{N}_{\mathcal{A}})$ is satisfiable. Here $\mathcal{N}_{\mathcal{A}}$) denotes the RCC8-network contained in $\mathcal{A}$. The difficult direction is the one from right to left which we will prove in the following. Let $\mathcal{T}' \cup (\mathcal{A} \setminus \mathcal{N}_{\mathcal{A}})$ be satisfiable by some model $\mathcal{I}$ which can w.l.o.g. be assumed to be the canonical model for the chase $chase(\mathcal{T}'_p \cup (\mathcal{A} \setminus \mathcal{N}_{\mathcal{A}}))$. Here, $\mathcal{T}'_p$ denotes the PI axioms in $\mathcal{T}$. As $\mathcal{A}$ is spatially complete, there exists a model $\mathcal{I}' \models \mathcal{N}_{\mathcal{A}} \cup Ax_{RCC8}$. We let $X = chase(\mathcal{T}'_p \cup \mathcal{A})$ which is satisfiable by an interpretation $\mathcal{J}$ built as a merge of $\mathcal{I}$ and $\mathcal{I}'$. Now, we will extend X by further chasing steps in the following way. Different from the chase construction of [3], we start with the set X which may already be infinite. This fact poses no problems as we chose an ordering over set of all possible strings over the signature of the ontology and the chasing constants.

In addition to the chasing rules listed in Definition 3, we will use a chasing rule for axioms of the form $B \sqsubseteq \exists \tilde{R}_1, l.(r_1 \vee \cdots \vee r_k) \in \mathcal{T}$ and the other axioms of the form $B \sqsubseteq \exists U_1, U_2.R \in \mathcal{T}$ (Figure 1). Let S_i denote the sets created during the chasing process. Directly after applying this chasing rule a completion step is applied in order to make the generated constraint network have a unique model modulo isomorphism. For every node n^* appearing in S_i an atom $r_{n*}(n^*, y^*)$ with $r_{n^*} \in \mathcal{B}_{RCC8}$ being some basic relation is added. We may have infinite many lalities already in S_0 but these are not constrained in anyway. We can assume that these nodes are pairwise related by the disjointness relation dc. So, in every chasing step there can be defined two disjoint sets of localities V_i^{dc} and V_i^{fin} with the following properties: For all pairwise distinct nodes in V_i^{dc} it is the case that $\mathsf{dc}(a^*, b^*) \in S_i$ and for all nodes $a^* \in V_i^{\mathsf{dc}}$ and nodes $b^* \in V_i^{fin}$ it is the case that $\mathsf{dc}(a^*, b^*) \in S_i$ and both networks are complete and relationally consistent. Now, the complete constraint network induced by V_i^{fin} is finite and is consistent with $r_1(y^*, x^*)$. This fact follows, e.g., from the patch-work property of ω-admissible theories. We use a path-consistency algorithm or some other appropriate algorithm to find a complete and consistent set induced by $V_i^{fin} \cup \{y^*\}$ that extends the networks induced by V_i^{fin} and y^*, resp. The new node y^* is related to the nodes in V_i^{dc} by dc-edges. This step does not disturb the consistency of the whole network because every composition of some basic relation with dc results in a disjunction which again contains dc. Proceeding in this way, we finally define $\bigcup_{i=0}^{\infty} S_i$ which induces a canonical model of $\mathcal{T} \cup \mathcal{A} \cup Ax_{RCC8}$.

Now we define the closure $cl_{\perp}(\mathcal{T})$ by extending the rules (1)–(6) for $cln(\mathcal{T})$ with the rules: (i) If $B_1 \sqsubseteq \exists l, l.r$ for $\mathsf{eq} \notin r \in \mathcal{T}$, then $B_1 \sqsubseteq \bot \in cl_\bot$; (ii) if $B_1 \sqsubseteq \exists l, l.r_1 \in \mathcal{T}$ and $B_2 \sqsubseteq \exists l, l.r_2 \in \mathcal{T}$ such that $\mathsf{eq} \notin r_1 \cap r_2$, then $B_1 \sqcap B_2 \sqsubseteq \bot$. Using $cl_\bot(\mathcal{T})$ one can define a FOL query that is false in $DB(\mathcal{A})$ iff $\mathcal{T} \cup \mathcal{A}$ is satisfiable.

Proof of Proposition 9

We argue for DL-Lite$^{\sqcap}_{\mathcal{F},\mathcal{R}}$(RCC2). Similar ideas can be used for proving that DL-Lite$^{\sqcap,+}_{\mathcal{F},\mathcal{R}}$(RCC3) allows for FOL rewritability. In case of RCC3 only eq may cause conflicts. But these can be tested by a FOL query, i.e., checking the consistency of RCC3 networks is in AC^0.

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 $\mathcal{A}$. But the conflicts can be calculated without looking at the implicitly constructed RCC constraint networks because these are always satisfiable unless there are regions x^*, y^* for which $\mathsf{dr}(x^*, y^*)$ or $\emptyset(x^*, y^*)$. We must consider this case when defining a closure $cl_\bot(\mathcal{T})$ of the $\mathcal{T}$ with respect to all concepts that are equivalent with $\bot$ w.r.t. $\mathcal{T}$. $cl_\bot(\mathcal{T})$ is defined by the rules for the negative closure $cln(\mathcal{T})$ and the following additional rules: If $B \sqsubseteq \exists l, l.\mathsf{dr} \in \mathcal{T}$, then $B \sqsubseteq \bot \in cl_\bot(\mathcal{T})$. If $\{B \sqsubseteq \exists \tilde{R}, \tilde{R}.\mathsf{dr}, (\mathrm{funct}\, R)\} \subseteq \mathcal{T}$, $B \sqsubseteq \bot \in cl_\bot(\mathcal{T})$. If $B \sqsubseteq \exists U_1, U_2.r \in \mathcal{T}$ and U_1 and U_2 contain only functional roles, then $B \sqcap \exists U_1, U_2.(B_{RCC2} \setminus r) \in cl_\bot(\mathcal{T})$ and $B \sqcap \exists U_2, U_1.(B_{RCC2} \setminus r)^{-1} \in cl_\bot(\mathcal{T})$. If $B_2 \sqcap \cdots \sqcap B_n \sqcap \exists \tilde{R}_1, l.r \sqsubseteq \bot \in cl_\bot(\mathcal{T})$ and $R_2 \sqsubseteq R_1 \in \mathcal{T}$ or $R_2^- \sqsubseteq R_1^- \in \mathcal{T}$, then $B_2 \sqcap \cdots \sqcap B_n \sqcap \exists \tilde{R}_2, l.r \sqsubseteq \bot \in cl_\bot(\mathcal{T})$. (Similar rules are applied for other instances of $\exists U_1, U_2.r$). If $B_1 \sqsubseteq \exists U_1^1, U_2^1.r_1 \in \mathcal{T}$, $B_2 \sqcap \cdots \sqcap B_n \sqcap \exists U_1^2, U_2^2.r_2 \sqsubseteq \bot \in cl_\bot(\mathcal{T})$ and either (i) $U_1^1 = U_1^2, U_2^1 = U_2^2$ and $r_1 \cap r_2 = \emptyset$ or (ii) $U_1^1 = U_2^2, U_2^1 = U_1^2$ and $r_1^{-1} \cap r_2 = \emptyset$, then $B_1 \sqcap B_2 \sqcap \cdots \sqcap B_n \sqsubseteq \bot \in cl_\bot(\mathcal{T})$. Using $cl_\bot(\mathcal{T})$ one can define a FOL query that is false in $DB(\mathcal{A})$ iff $\mathcal{T} \cup \mathcal{A}$ is satisfiable.

Proof of Theorem 11

The proof follows the proof of Theorem 5.15 for pure DL-Lite ontologies [3]. We adapt the chase construction to account for the RCC8 relations $r \in Rel_{RCC8}$. The main observation is that the disjunctions in r can be nearly handled as if they were predicate symbols.

Let Q be a n-ary GCQ^+-query. If $\mathcal{T} \cup \mathcal{A}$ is not satisfiable, $cert(Q, \mathcal{T} \cup \mathcal{A})$ is the set of all n-ary tuples of constants of the ontology $\mathcal{O}$. But Proposition 7 states that satisfiability is FOL-rewritable. So we can assume, that $\mathcal{O}$ is satisfiable.

Let pr be the UCQ resulting from applying the algorithm to Q and $\mathcal{O}$. We have to show that $cert(Q, \mathcal{O}) = (pr)^{DB(\mathcal{A})}$. We proceed in two main steps. First, we construct a chase-like set $chase^*(\mathcal{O})$, declare what it means to answer Q with respect to $chase^*(\mathcal{O})$, resulting in the set $ans(Q, chase^*(\mathcal{O}))$, and then show that $ans(Q, chase^*(\mathcal{O})) = cert(Q, \mathcal{O})$. In the second step, we show that $ans(Q, chase^*(\mathcal{O})) = (pr)^{DB(\mathcal{A})}$.

First Step We construct a chase-like set $chase^*(\mathcal{O})$ that will be the basis for proving the correctness and completeness of our algorithm. We use the chase rules of Definition 3 and the special rule (R) of Figure 1 to build $chase^*(\mathcal{O})$. Every time (R) is applied to yield a new ABox S_i, the resulting constraint network in S_i is saturated by calculating the minimal labels between the new added region constants and the other region constants. The application of (R) does not constrain the RCC8-relations between the old regions and even stronger: Let (R) be applied to a TBox axiom of the form $A \sqsubseteq \exists \tilde{R}, l.r$ and $A(a) \in S_i$ resulting in the addition of $R(a, b)$, $l(b, b^*)$ and $r(b^*, a^*)$. Then it is enough to consider all $c^* \in S_i$ and all relations r_{c^*,a^*} such that $r_{c^*,a^*}(c^*, a^*) \in S_i$. The composition table gives the outcome

$r_{c^*,a^*}; r = r'_{c^*,b^*}$ and one adds $r'_{c^*,b^*}(c^*, b^*)$ to S_i. This step will be called the step of triangle completion. After the triangle completion step one closes the assertions up with respect to the subset relation between RCC8-relations and with respect to symmetry. I.e., if $r_1(x^*, y^*)$ is added to S_i, then one also adds $r_2(x^*, y^*)$ for all r_2 such that $r_1 \subseteq r_2$ and $r_2^{-1}(y^*, x^*)$. For different c_1^*, c_2^*, assertions of the form $r_{c_1^*,b^*}(c_1^*, b^*)$ and $r_{c_2^*,b^*}(c_2^*, b^*)$ do not constrain each other (because of the patch-work property). As the number of regions is finite and we excluded the non-atomicity axiom, the saturation leads to a finite set S_{i+k} (for some $k \in \mathbb{N}$) that is a superset of S_i. Let $chase^*(\mathcal{O}) = \bigcup S_i$ be the union of all ABoxes constructed in this way. The set $chase^*(\mathcal{O})$ does not induce a single canonical model. But it is universal in the following sense:

(*) For every model $\mathcal{I}$ of $\mathcal{O}$ define a model $\mathcal{I}_c$ out of $chase^*(\mathcal{O})$ by taking a (consistent) configuration of the contained RCC8-network and taking the minimal model of this configuration and the thematic part of $chase^*(\mathcal{O})$. Then $\mathcal{I}_c$ maps homomorphically to $\mathcal{I}$.

The claim (*) holds because $\exists U_1, U_2.r$-constructs do not appear on the left-hand side of the PI axioms; hence new information on the spatial side cannot be used during the chasing process to produce new information on the thematic part.

We explain what it means to answer a GCQ^+-query with respect to $chase^*(\mathcal{O})$. We transform Q into a CQ $\tau_1(Q)$ where the relations $r \in Rel_{RCC8}$ are considered as predicate symbols. E.g., let $Q = \exists R.A(x) \wedge \exists \tilde{R}_1, l.(\mathsf{tpp} \vee \mathsf{ntpp})(y)$ and $X = \tau_1(Q, \{\exists \tilde{R}_1, l.\mathsf{tpp}(y)\})$ for short. Then

$$\begin{aligned} X \quad = \quad & \exists z, z^*, x^*.\exists R.A(x) \wedge R(x, z) \wedge l(z, z^*) \wedge \\ & l(x, x^*) \wedge (\mathsf{tpp} \vee \mathsf{ntpp})(z^*, x^*) \end{aligned}$$

The set of answers $ans(chase^*(\mathcal{O}), Q)$ is defined by homomorphisms of the atoms of $\tau_1(Q)$ into $chase^*(\mathcal{O})$. Let $\vec{x} = (x_1, \ldots, x_n)$ for $Q = \psi(\vec{x})$. $(a_1, \ldots, a_n) \in ans(chase^*(\mathcal{O}), Q)$ iff there is a homomorphism h from $\tau_1(Q)$ into $chase^*(\mathcal{O})$ with $h(x_i) = a_i$ (for $i \in \{1, \ldots, n\}$). The homomorphic image of $\psi(\vec{x})$ in $chase^*(\mathcal{O})$ is called a witness of ψ w.r.t. $\vec{a}$ in $chase^*(\mathcal{O})$. Clearly, if $\mathcal{I} \models chase^*(\mathcal{O})$ and $\vec{a} \in ans(chase^*(\mathcal{O}), Q)$, then $\mathcal{I} \models \psi[\vec{x}/\vec{a}]$.

We now prove $ans(Q, chase^*(\mathcal{O})) = cert(Q, \mathcal{O})$.

$\subseteq$-direction: Let $\vec{a} \in ans(Q, chase^*(\mathcal{O}))$. Let $\mathcal{I} \models \mathcal{O}$ and I_c be the model according to claim (*). Because $\mathcal{I}_c \models \psi[\vec{x}/\vec{a}]$, it follows that $\mathcal{I} \models \psi[\vec{x}/\vec{a}]$.

$\supseteq$-direction: Let $\vec{a} \in cert(Q, \mathcal{O})$. For every $\mathcal{I} \models \mathcal{O}$ consider $\mathcal{I}_c$. All these $\mathcal{I}_c$ differ at most on the interpretations of the RCC8-Relations which are assigned to regions x^*, y^*. Consider for all x^*, y^* the assertion $r(x^*, y^*)$, $r \in Rel_{RCC8}$ where $r_i \in r$ iff there is $\mathcal{I}_c$ such that $\mathcal{I}_c \models r_i(x^*, y^*)$. Then $r(x^*, y^*)$ is in $chase^*(Q, \mathcal{O})$. Therefore we find a homomorphism h from $\psi(\vec{x})$ onto $chase(Q, \mathcal{O})$ with $h(\vec{x}) = \vec{a}$.

Second Step Let pr be the outcome of the algorithm applied to Q. We show $pr^{DB(\mathcal{A})} = ans(chase^*(\mathcal{O}), Q)$.

$\subseteq$-direction: Let $q \in pr$ be a conjunctive n-ary query. We have to show $q^{DB(\mathcal{A})} \subseteq ans(chase^*(\mathcal{O}), Q)$. This can be done by induction over the number of steps that are needed in order to construct q in the PerfectRef algorithm.

$\supseteq$-direction: This is the direction showing the completeness of the algorithm. Let $\vec{a} \in ans(Q, chase^*(\mathcal{O}))$. So there is a witness of $\vec{a}$ w.r.t. $\tau_1(Q)$ in $chase^*(\mathcal{O})$. This witness lies in some S_k of the chase $chase^*(\mathcal{O})$ and shall be denoted $\mathcal{G}_k$. We have to find a $q \in pr$ such that it has a witness in the ABox $\mathcal{A}$. This can be proved by considering the pre-witness of $\vec{a}$ with respect to Q in all S_i for $i \leq k$. The pre-witness of $\vec{a}$ with respect to Q in S_i is defined by:

$$\begin{aligned}\mathcal{G}_i \quad = \quad \bigcup_{\beta' \in \mathcal{G}_k} \{\beta \in S_i \mid & \ \beta \text{ is an ancestor of } \beta' \text{ in } S_k \\ & \text{and there exists no successor of } \beta \text{ in } S_i \\ & \text{that is an ancestor of } \beta' \text{ in } S_k\}\end{aligned}$$

By induction on i ($i \in \{0, \ldots, k\}$) one can find a $q \in pr$ such that the pre-witness of $\vec{a}$ with respect to Q in S_{k-i} is a witness for q. By induction assumption there is $q' \in pr$ such that $\mathcal{G}_{k-i+1}$ is a witness of $\vec{a}$ w.r.t. q' in S_{k-i+1}. If S_{k-i+1} results from S_{k-i} by application of one of the chase rules in Definition 3, then the argument proceeds in the same manner as in the proof of Lemma 5.13 in the paper of [3]. Otherwise, S_{k-i+1} is constructed by applying rule (R) or one of the saturation steps (triangle completion, upward closure, symmetry closure, resp.) following the application of rule (R). But all these steps have a corresponding case in the algorithm.