Towards Spatial Reasoning in the Semantic Web: A Hybrid Knowledge Representation System Architecture

Rolf Grütter and Bettina Bauer-Messmer

Swiss Federal Institute for Forest, Snow and Landscape Research, an Institute of the ETH Board,
Zürcherstrasse 111, 8903 Birmensdorf, Switzerland
{rolf.gruetter, bettina.bauer}@wsl.ch

Abstract. Environmental databases store a wide variety of data from heterogeneous sources which are described with domain-specific terminologies and refer to distinct locations. In order to make them accessible also to non-expert users, terminological concepts and spatial relations must be represented in a way that they can be exploited for searches. In this paper we propose a hybrid knowledge representation system architecture which integrates terminological and spatial aspects of the application domain and provides support for reasoning with RCC, a well-known calculus for spatial reasoning, and the Semantic Web's ontology language OWL. Our approach is motivated by the observation that RCC cannot be expressed in OWL without a major revision of the latter. Issues of upholding consistency of the knowledge base in view of an evolving ontology and of computational complexity are discussed.

1 Introduction

Environmental databases typically store a wide variety of data from heterogeneous sources which are described with domain-specific terminologies and refer to distinct locations. Searching in these databases can be difficult, particularly for non-expert users. To give an example, the Datacenter Nature and Landscape of the Swiss Federal Institute for Forest, Snow and Landscape Research is characterized by the coexistence of several inventories of different kinds of biotopes (bogs, fens, floodplains, grasslands, amphibians spawning grounds, etc.). These biotopes cover neighboring or overlapping regions and are further related to non-inventory administrative regions such as communes or cantons. Interviews with users have shown, that they tend to search for specific locations, regardless the fact that the data they are looking for may stem from different inventories. In a later section we will outline a scenario where a user wishes to acquire information about a given commune, including both thematic information, such as population statistics, and spatial information, for instance its relation to inventory objects. This scenario implies that, in addition to terminological (and eventually also temporal) concepts, spatial relations are represented in a way that they can be exploited for searches. In order to allow for an open and intuitive search as described above we propose to add a semantic layer on top of the database of the datacenter. This semantic layer must, among others, provide a method for dealing with spatial relations.

The central component of the semantic layer is an ontology. Ontologies are embodiments of shared conceptualizations [1] and, therefore, can semantically interpret and augment user input such as search terms. The vocabularies of ontologies are typically arranged in hierarchies [2]. Topological terms, however, are intrinsically not hierarchical, in contrary, they are organized in accordance with mereological criteria. The term *mereology* refers to the discipline which is concerned with parts of any kind and their relations to wholes or to a whole (*meros* is the greek word for part). Mereology and mereotopology – the application of mereological methods to topological phenomena – represent a well-established field in geographical information science. It is addressed, for instance, by computational geometry in geographic information systems and by the Region Connection Calculus (RCC) in reasoning systems [3]. Since the semantic layer aims at enabling automatic agents (e.g., sophisticated search engines) to reason on mereological relations, this paper is focused on the latter. Even though mereology has been a research topic for years, its relationship to ontology is not well understood in practical applications. Issues of debate are how regions and mereological relations should be represented and whether or not mereological reasoning can be simulated by ontological reasoning. These issues are particularly addressed by investigating the

S. Fabrikant, M. Wachowicz (Eds.): The European Information Society: Leading the Way with Geo-information, Lecture Notes in Geoinformation and Cartography, 2007.

combination of RCC with Description Logics (DLs), the latter of which have been established as the methods of choice for constructing ontologies in the last decade.

It has been shown that there is a close relationship between modal logics and a family of RCC related calculi [4] and between DLs and modal logics [5] (with reference to possible worlds semantics [6]). Based on the assumption that a similar close relationship holds between RCC and DLs quite a number of approaches aimed at combining RC-Calculi with species of DLs have been investigated. With the introduction of the Semantic Web and its Web Ontology Language (OWL) the construction of ontologies based on DLs has become very popular. In order to enable Semantic Web agents to reason on spatial relations it is, therefore, imperative to explore whether any of the RC-Calculi can be combined with OWL. Since the existing approaches give no reason for optimism regarding the reconstruction of RCC with OWL (or extensions thereof), we present an approach which is based on a hybrid knowledge representation system architecture and which allows for reasoning with both formalisms.

The paper is organized as follows: In chapters 2 and 3 short introductions to RCC and DL are provided. In chapter 4 we review a number of recent approaches aimed at combining RCC with DLs and discuss their potential for establishing reasoning with RCC in the Semantic Web. In chapter 5 a usage scenario is outlined which involves both reasoning with RCC and reasoning in OWL. Based on the usage scenario we introduce in chapter 6 an approach which is based on an extension of the knowledge representation system architecture. The introduced approach is discussed in chapter 7. The work presented here builds on previous work on the design and implementation of a Web-based platform for visualizing, querying and analyzing environmental data [7].

2 The Region Connection Calculus

RCC is an axiomatization of certain spatial concepts and relations in first order logic [3, 4]. The basic theory assumes just one primitive dyadic relation: $\mathsf{C}(x, y)$ read as "*x* connects with *y*". Individuals (x, y) can be interpreted as denoting spatial regions. The relation $\mathsf{C}(x, y)$ is reflexive and symmetric.

Using the primitive relation $\mathsf{C}(x, y)$ a number of intuitively significant relations can be defined. Some of the most useful of these are illustrated in figure 1 and their definitions are given in table 1. The asymmetrical relations P, PP, TPP and NTPP have inverses which we write, in accordance with [4], as *R*i, where $R \in \{\mathsf{P}, \mathsf{PP}, \mathsf{TPP}, \mathsf{NTPP}\}$. These relations are defined by definitions of the form $R\mathrm{i}(x, y) \equiv_{def} R(y, x)$. Of the defined relations, DC, EC, PO, EQ, TPP, NTPP, TPPi and NTPPi have been proven to form a jointly exhaustive and pairwise disjoint set, which is known as RCC-8. Similar sets of one, two, three and five relations are known as RCC-1, RCC-2, RCC-3 and RCC-5, respectively.

RCC also incorporates a constant denoting the universal region, a sum function and partial functions giving the product of any two overlapping regions and the complement of every region except the universe [4].

According to [3], regions support either spatial or temporal interpretation. In case of spatial interpretation, there is a variety of models among which to choose. The authors provide some examples such as interpreting the relation C in terms of two regions whose closures share a common point or stating that two regions connect when the distance between them is zero.

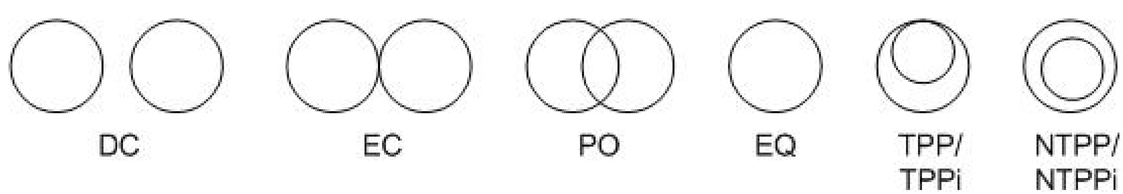

Fig. 1. RCC-8 relations (for the entire names cf. table 1)

Table 1. Basic RCC relations

C(x, y)	(primitive relation)	(x Connects with y)
DC(x, y)	$\equiv_{def}$ ¬C(x, y)	(x is DisConnected from y)
P(x, y)	$\equiv_{def}$ ∀z[C(z, x) → C(z, y)]	(x is a Part of y)
PP(x, y)	$\equiv_{def}$ P(x, y) ∧ ¬P(y, x)	(x is a Proper Part of y)
EQ(x, y)	$\equiv_{def}$ P(x, y) ∧ P(y, x)	(x is EQual to y)
O(x, y)	$\equiv_{def}$ ∃z[P(z, x) ∧ P(z, y)]	(x Overlaps y)
DR(x, y)	$\equiv_{def}$ ¬O(x, y)	(x is DiscRete from y)
PO(x, y)	$\equiv_{def}$ O(x, y) ∧ ¬P(x, y) ∧ ¬P(y, x)	(x Partially Overlaps y)
EC(x, y)	$\equiv_{def}$ C(x, y) ∧ ¬O(x, y)	(x is Externally Connected to y)
TPP(x, y)	$\equiv_{def}$ PP(x, y) ∧ ∃z[EC(z, x) ∧ EC(z, y)]	(x is a Tangential Proper Part of y)
NTTP(x, y)	$\equiv_{def}$ PP(x, y) ∧ ¬∃z[EC(z, x) ∧ EC(z, y)]	(x is a Non-Tangential Proper Part of y)

In order to check consistency of a knowledge base holding spatial relations, so-called composition tables are used (cf. the composition table for RCC-5 in table 2). The entries in these tables share a uniform inference pattern which can be formalized as composition axioms of the general form $\forall x, y, z.\ S(x, y) \wedge T(y, z) \rightarrow R_1(x, z) \vee \ldots \vee R_n(x, z)$ where S, T, and R_i are variables for relation symbols. To give an example, the composition table can check queries such as "Provided that region a partially overlaps region b, and region a is discrete from region c, is the assertion b is discrete from c valid or not?" (formally, true?[PO(a, b) ∧ DR(a, c) → DR(b, c)]).

Table 2. RCC-5 composition table

T(x, z) $\equiv_{def}$ {DR(x, z), PO(x, z), EQ(x, z), PP(x, z), PPi(x, z)}

∘	DR(x, y)	PO(x, y)	EQ(x, y)	PPi(x, y)	PP(x, y)
DR(y, z)	T(x, z)	DR(x, z) PO(x, z) PPi(x, z)	DR(x, z)	DR(x, z) PO(x, z) PPi(x, z)	DR(x, z)
PO(y, z)	DR(x, z) PO(x, z) PP(x, z)	T(x, z)	PO(x, z)	PO(x, z) PPi(x, z)	DR(x, z) PO(x, z) PP(x, z)
EQ(y, z)	DR(x, z)	PO(x, z)	EQ(x, z)	PPi(x, z)	PP(x, z)
PP(y, z)	DR(x, z) PO(x, z) PP(x, z)	PO(x, z) PP(x, z)	PP(x, z)	PO(x, z) EQ(x, z) PP(x, z) PPi(x, z)	PP(x, z)
PPi(y, z)	DR(x, z)	DR(x, z) PO(x, z) PPi(x, z)	PPi(x, z)	PPi(x, z)	T(x, z)

3 Basic Description Logics

Description logics is the name for a family of knowledge representation formalisms that represent the knowledge of an application domain (the "world") by first defining the relevant concepts of the domain (its terminology), and then using these concepts to specify properties of objects and individuals occurring in the domain (the world description) [8]. One of the characteristics of these languages is that they enable systems built on them to infer implicitly represented knowledge from the knowledge that is explicitly contained in the knowledge base. DLs support inference patterns that occur in many applications of intelligent information processing systems, and which are also used by humans to structure and understand the world: classification of concepts and individuals. Classification of concepts determines subconcept/superconcept relationships (called subsumption relationships in DL) between the concepts of a given terminology, and thus allows one to structure the terminology in the form of a subsumption hierarchy. This hierarchy provides useful information on the connection between different concepts, and it can be used to speed-up other inference services. Classification of individuals (or objects) determines whether a given individual is an instance of a certain concept. These inference patterns can be exploited by reasoning services, for instance, in order to check whether the set of assertions is consistent or to retrieve the individuals satisfying a given concept description.

Similar to [8], in abstract notation, we use the letters *A* and *B* for atomic concepts (concepts denote sets of individuals), the letter *R* for atomic roles (roles denote binary relations between individuals), the letters *C* and *D* for concept descriptions, the letters *S* and *T* for role descriptions, and the letters *a* and *b* for individuals. In addition to atomic concepts and roles, which correspond to elementary descriptions and are introduced by their names, DL systems allow their users to build complex descriptions of concepts and roles using a number of *concept constructors*. Description languages are distinguished by the constructors they provide. Concept descriptions in $\mathcal{ALC}$ are formed according to the following syntax rule (first two columns) where $\top$ and $\bot$ denote the top and button concepts, respectively. Note that OWL is not based on $\mathcal{ALC}$ but on $\mathcal{SHIQ}$ which is more complex. However, since $\mathcal{ALC}$ is an ancestor of $\mathcal{SHIQ}$, the notions introduced herein also apply to $\mathcal{SHIQ}$ and OWL. For a complete specification of the OWL syntax and semantics refer to [9].

$C, D \rightarrow$	$A \mid$	(atomic concept)	$A^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$
	$\top \mid$	(universal concept)	$\Delta^{\mathcal{I}}$
	$\bot \mid$	(bottom concept)	$\emptyset$
	$\neg A \mid$	(atomic negation)	$\Delta^{\mathcal{I}} \setminus A^{\mathcal{I}}$
	$C \sqcap D \mid$	(intersection)	$C^{\mathcal{I}} \cap D^{\mathcal{I}}$
	$C \sqcup D \mid$	(union)	$C^{\mathcal{I}} \cup D^{\mathcal{I}}$
	$\forall R.C \mid$	(value restriction)	$\{a \in \Delta^{\mathcal{I}} \mid \forall b.\ (a, b) \in R^{\mathcal{I}} \rightarrow b \in C^{\mathcal{I}}\}$
	$\exists R.C$	(full existential quantification).	$\{a \in \Delta^{\mathcal{I}} \mid \exists b.\ (a, b) \in R^{\mathcal{I}} \wedge b \in C^{\mathcal{I}}\}$

In order to define a formal semantics of $\mathcal{ALC}$-concepts, we consider interpretations $\mathcal{I}$ that consist of a non-empty set $\Delta^{\mathcal{I}}$ (the domain of the interpretation) and an interpretation function, which assigns to every atomic concept *A* a set $A^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$ and to every atomic role *R* a binary relation $R^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$. The interpretation function is extended to concept descriptions as shown above (last column).

A DL knowledge base comprises two components, the TBox and the ABox [8]. The TBox introduces the *terminology*, i.e., the vocabulary of an application domain, while the ABox contains *assertions* about named individuals in terms of this vocabulary. The TBox can be used to assign names to complex descriptions.

Terminological axioms make statements about how concepts or roles are related to each other [8]. In the most general case, terminological axioms have the form $C \sqsubseteq D$ ($R \sqsubseteq S$) or $C \equiv D$ ($R \equiv S$), where *C*, *D* are concepts (and *R*, *S* are roles). Axioms of the first kind are called *inclusions*, while axioms of the second kind are called *equalities*. An equality whose left-hand side is an atomic concept is a *definition*. *Terminologies* are sets of definitions which are used to introduce *symbolic names* for complex descriptions or they are sets of inclusions. The semantics of axioms is defined as one would expect. An interpretation $\mathcal{I}$ satisfies an inclusion $C \sqsubseteq D$ if $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$, and it satisfies an equality $C \equiv D$ if $C^{\mathcal{I}} = D^{\mathcal{I}}$.

As mentioned, concepts *C* and roles *R* can be used to make assertions in an ABox [8]. These are assertions of the kinds $C(a)$ and $R(b, c)$, where *C* and *R* denote concept and role names and *a*, *b*, *c* denote individual names. By the first kind, called *concept assertions*, we state that *a* belongs to (the interpretation of) *C*, by the second kind, called *role assertions*, we state that *c* is a filler of the role *R* for *b*. An ABox is a finite set of such assertions. To give a semantics to ABoxes, interpretations are extended to individual names. Since it is assumed that distinct individual names denote distinct objects, interpretations have to respect the *unique name assumption*, that is, if *a*, *b* are distinct names, then $a^{\mathcal{I}} \neq b^{\mathcal{I}}$ (note that, different from $\mathcal{ALC}$, OWL does *not* assume unique names for individuals [10]). The interpretation $\mathcal{I}$ satisfies the concept assertion $C(a)$ if $a^{\mathcal{I}} \in C^{\mathcal{I}}$, and it satisfies the role assertion $R(b, c)$ if $(b^{\mathcal{I}}, c^{\mathcal{I}}) \in R^{\mathcal{I}}$.

4 Review of Existing Approaches

Spatioterminological reasoning with the description logic $\mathcal{ALCRP}(\mathcal{D})$ has been introduced in [11]. The authors define an appropriate concrete domain $\mathcal{D}_{\mathcal{P}}$ for polygons. $\mathcal{RP}$ stands for role definitions based on predicates. More specifically, $\mathcal{ALCRP}(\mathcal{D})$ extends $\mathcal{ALC}(\mathcal{D})$ by a *role-forming operator* which is based on concrete domain predicates. The new operator allows the definition of roles with very complex properties and provides a close coupling of roles with concrete domains. A detailed account of $\mathcal{ALCRP}(\mathcal{D})$ is provided in [12]. In order to ensure termination of the satisfiability algorithm for the concrete domain $\mathcal{D}_{\mathcal{P}}$, the authors impose restrictions on the syntactic form of the set of terminological axioms which impose tight constraints on modeling spatioterminological structures [11, 12].

With the intention to augment a description logic like $\mathcal{ALC}$ with some kind of qualitative spatial reasoning capabilities, a rich variety of extensions to $\mathcal{ALC}$ is investigated in [13]. As a basic extension the author introduces role inclusion axioms of the form $S \circ T \sqsubseteq R_1 \sqcup \ldots \sqcup R_n$ which constrain the models $\mathcal{I}$ to $S^{\mathcal{I}} \circ T^{\mathcal{I}} \subseteq R_1^{\mathcal{I}} \cup \ldots \cup R_n^{\mathcal{I}}$ ($\circ$ stands for the composition of roles). A set of these role inclusion axioms is referred to by the author as a *role box*. In previous work it has been shown that concept satisfiability in a related logic called $\mathcal{ALC}_{\mathcal{RA}}$, enforcing role disjointness on all roles R and S ($R^{\mathcal{I}} \cap S^{\mathcal{I}} = \emptyset$), is undecidable. In [13] specializations of $\mathcal{ALC}_{\mathcal{RA}}$ which specifically consider the family of RCC related calculi are investigated. Therefore, the author translates the inference represented by each entry in a RCC composition table into a corresponding role axiom of the above introduced form. Based on these translations, the author shows for both $\mathcal{ALCI}_{\mathcal{RCC}5}$ and $\mathcal{ALCI}_{\mathcal{RCC}8}$ that satisfiability of concepts quantifying over roles ($\forall R.C$) can be undecidable in a practical application.

A general property of concrete domains that is sufficient for proving decidability of DLs equipped with them and General Concept Inclusions (GCIs) is identified in [14]. The authors further present a tableau algorithm for reasoning in DLs equipped with such concrete domains. In order to obtain their first result, they concentrate on a particular kind of concrete domains which they call *constraint systems*. According to the authors, a constraint system is a concrete domain that only has binary predicates, and these predicates are interpreted as jointly exhaustive and pairwise disjoint relations. The authors show that the spatial constraint system which is based on the real plane and the RCC-8 relations has the required property and that the description logic which allows defining concepts with reference to this constraint system is decidable. As a description logic they introduce $\mathcal{ALC}(\mathcal{C})$ which is $\mathcal{ALC}$ extended with two *constraint constructors*.

In [15] the authors aim at representing qualitative spatial information in OWL DL (the DL fragment of OWL). On the basis of the (assumed) close relationship between the RCC-8 calculus and OWL DL they extend the latter with the ability to define reflexive roles. The extension of OWL DL with a reflexive property is motivated by the requirement that such a property, together with the transitive one, is needed in order to describe the accessibility relation which relates possible worlds to each of the modal operators of the logic S4. The modal logic S4 is considered, because the RCC-8 calculus can be translated into an extension of it [4]. In order to express RCC-8 knowledge bases the authors use a translation in which regions are expressed as non-empty regular closed sets. The RCC-8 relations are then translated into (sets of) concept axioms in OWL DL and the classes denoted by the introduced concepts are instantiated by asserting for each concept an individual in the ABox in order to ensure that the classes cannot be empty.

It seems to be more intuitive to define the RCC relations in terms of role descriptions than to translate them into (sets of) concept axioms. Since OWL DL does not provide constructors for role descriptions (apart from inverse), the underlying description logic has to be extended with these constructors. In [16] it is shown that the extension of $\mathcal{SHIQ}$ with complex role inclusion axioms of the form $S \circ T \sqsubseteq R$ is undecidable, even when these axioms are restricted to the forms $S \circ T \sqsubseteq S$ or $T \circ S \sqsubseteq S$, but that decidability can be regained by further restricting them to be acyclic. However, a closer look at an arbitrary composition table reveals that, in order to axiomatize the composition of RCC relations, a language must even support an extension of the unrestricted form of role inclusion axioms, namely $S \circ T \sqsubseteq R_1 \sqcup \ldots \sqcup R_n$. If decidability should be preserved, complex role inclusion axioms are, therefore, not a solution to the translation problem of RCC. Axioms describing the basic RCC relations even require additional role constructors such as intersection, union and complement. Extensions of $\mathcal{SHIQ}$ with these kinds of role constructors have, to our knowledge, not been investigated so far.

To summarize, the existing approaches give no reason for optimism regarding the effective combination of RCC with DL ontologies in the Semantic Web. Since OWL and the reviewed extensions to $\mathcal{ALC}$ are not as closely related as one might expect, approaches based on $\mathcal{ALC}$ require a major revision of the existing Web ontology language. This is not desirable as an alternate language would surely miss some of the favorable features of the existing, such as property hierarchies, which make it compatible with RDF (Resource Description Framework), the Web's description language for resources. The approaches based on $\mathcal{ALC}$ further have in common that they do not consider ABoxes. While the most recent approach would require only a minimal extension of OWL, the notion of regions as sets in the (abstract) object domain (and not in a concrete domain) prevents RCC from effectively combining with domain ontologies. The reason therefore is that OWL DL requires type separation: A class cannot also be an individual (or a property) [17]. However, in order to classify regions in a domain ontology they must be represented as individuals and not as concepts.

5 A Usage Scenario for Terminological and Spatial Reasoning

A user wishes to acquire the available thematic and spatial information about a given commune, say Birmensdorf. The spatial information particularly refers to inventory objects which are spatially related to Birmensdorf. In order to keep the scenario simple we concentrate on a single inventory, the National Inventory of Territories and Natural Monuments, while in reality the Data Center Nature and Landscape holds information about 16 different inventories (as of November 2006).

In order to initiate the search, the user enters the term "Birmensdorf" into the search form and clicks on "search". Let's assume that the (syntactic) search engine returns a site about the commune of Birmensdorf as best ranked resource which contains some thematic information about the commune but no information about spatially related inventories. However, as the site is indexed by keywords which were defined in an ontology upon creation, the user is given the option to semantically expand her query, for example by clicking on an correspondingly named button. When activated, the query expansion method consults the ontology in order to look for terms that are semantically related to the keywords of the site. Provided that one of the keywords is "Birmensdorf", the expansion method finds out that "Birmensdorf" is a commune, that "Birmensdorf" is connected with an individual named "Albiskette-Reppischtal" and that this individual is an inventory object. Based on this expansion, the user can be provided with links to resources about the National Inventory of Territories and Natural Monuments in general or about the inventory object "Albiskette-Reppischtal" in particular.

Note that the ontology in table 3 contains more information than is required for expanding the example query. For instance, it holds that the concepts "Object" and "Commune" are included in the concept "Region", that regions are spatially related to one another and that the role "connectsWith" is included in the role "spatiallyRelated". Since the role assertions of the world description are calculated using computational geometry they reflect explicit information which is implicitly also contained in the geographic information system and therefore provide a minimal redundant interface to the reasoning services based on both description logic and RCC.

Having expanded the query semantically, the method comes back to the user with the question whether she would like to expand her search also spatially. When expanded spatially, the method passes the query to a RCC reasoner which calculates the RCC-8 relation holding between the commune of Birmensdorf and the inventory object "Albiskette-Reppischtal", that is, it finds out that the commune of Birmensdorf *partially overlaps* the object "Albiskette-Reppischtal" of the National Inventory of Territories and Natural Monuments (in RCC formalism, PO(Birmensdorf, Albiskette-Reppischtal). For this calculation, the reasoner makes use of a number of connected pairs of regions in the neighborhood of Birmensdorf and Albiskette-Reppischtal which are asserted in the ABox of the knowledge base (not shown) and which must be passed to the RCC reasoner together with the query. The spatial query expansion also reveals how the commune of Birmensdorf and the inventory object "Albiskette-Reppischtal" are spatially related to other regions.

Table 3. Axioms and assertions in our example OWL DL knowledge base (note that only a small part is presented)

Terminology:

	DL Syntax	Semantics
(1)	$\top$	$\Delta^{\mathcal{I}}$
(2)	$\text{Region} \sqsubseteq \top$	$\text{Region}^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$
(3)	spatiallyRelated	$\text{spatiallyRelated}^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$
(4)	$\exists\text{spatiallyRelated}.\top \sqsubseteq \text{Region}$	$\{a \in \Delta^{\mathcal{I}} \mid \exists b.\ (a, b) \in \text{spatiallyRelated}^{\mathcal{I}}\} \subseteq \text{Region}^{\mathcal{I}}$
(5)	$\top \sqsubseteq \forall\text{spatiallyRelated.Region}$	$\Delta^{\mathcal{I}} \subseteq \{a \in \Delta^{\mathcal{I}} \mid \forall b.\ (a, b) \in \text{spatiallyRelated}^{\mathcal{I}} \rightarrow b \in \text{Region}^{\mathcal{I}}\}$
(6)	$\text{Object} \sqsubseteq \text{Region}$	$\text{Object}^{\mathcal{I}} \subseteq \text{Region}^{\mathcal{I}}$
(7)	$\text{Commune} \sqsubseteq \text{Region}$	$\text{Commune}^{\mathcal{I}} \subseteq \text{Region}^{\mathcal{I}}$
(8)	$\text{connectsWith} \sqsubseteq \text{spatiallyRelated}$	$\text{connectsWith}^{\mathcal{I}} \subseteq \text{spatiallyRelated}^{\mathcal{I}}$

World description:

	DL Syntax	Semantics
(9)	Object(Albiskette-Reppischtal)	$\text{Albiskette-Reppischtal}^{\mathcal{I}} \in \text{Object}^{\mathcal{I}}$
(10)	Commune(Birmensdorf)	$\text{Birmensdorf}^{\mathcal{I}} \in \text{Commune}^{\mathcal{I}}$
(11)	connectsWith(Birmensdorf, Albiskette-Reppischtal)	$(\text{Birmensdorf}^{\mathcal{I}}, \text{Albiskette-Reppischtal}^{\mathcal{I}}) \in \text{connectsWith}^{\mathcal{I}}$

Terminology and world description in OWL:

(1)	<owl:Class rdf:about="http://www.w3.org/2002/07/owl#Thing"/>
(2)	<owl:Class rdf:ID="Region"> <rdfs:subClassOf rdf:resource="http://www.w3.org/2002/07/owl#Thing"/> </owl:Class>
(3) (4) (5)	<owl:ObjectProperty rdf:ID="spatiallyRelated"> <rdfs:domain rdf:resource="#Region"/> <rdfs:range rdf:resource="#Region"/> </owl:ObjectProperty>
(6)	<owl:Class rdf:ID="Object"> <rdfs:subClassOf rdf:resource="#Region"/> </owl:Class>
(7)	<owl:Class rdf:ID="Commune"> <rdfs:subClassOf rdf:resource="#Region"/> </owl:Class>
(8)	<owl:ObjectProperty rdf:ID="connectsWith"> <rdfs:subPropertyOf rdf:resource="#spatiallyRelated"/> </owl:ObjectProperty>
(9)	<Object rdf:ID="Albiskette-Reppischtal"/>
(10) (11)	<Commune rdf:ID="Birmensdorf"> <connectsWith rdf:resource="#Albiskette-Reppischtal"/> </Commune>

6 A Hybrid Knowledge Representation System Architecture

Regions are usually interpreted in terms of classical point-set topology as (non-empty) regular closed sets of points [4]. Motivated by the way how regions are represented in geographic information systems we introduce an interpretation which deviates from the common one. Our interpretation is further motivated by the observation that interpreting regions as sets easily leads to misinterpretation when combining RCC with DLs. This point has been elaborated in section 4.

We say that a region is interpreted as a *polygon in the integral plane* and that two regions are connected if at least one of the following holds (the described properties can be calculated using computational geometry):

- One of the vertices of a polygon lies inside the other polygon;
- any two arcs (i.e., edges), one of either polygon, intersect;
- both polygons have an arc or a fragment of an arc (which can be minimal) in common.

Note that a similar approach is adopted in [11]: There the authors restrict the predicates for the spatial domain to the description of polygons and compute the base relation between each pair of concrete polygons with the help of standard algorithms from computational geometry. However, different from [11] we calculate only the primitive relation “connects with” in order to provide for a minimal interface to the terminological component and to keep the knowledge base of the knowledge representation system small.

The combination of RCC with OWL ontologies as sketched in the usage scenario implies that the architecture of a knowledge representation system based on DL has been extended with RCC specific components. Figure 2 shows the architecture of a hybrid system in its simplest form.

In figure 2, the shortcut KB denotes the knowledge base. The label *RCCBox* stands for *Region Connection Calculus Box*, a term which is inspired by the role box in [13]. The RCCBox contains the definitions of the RCC relations and the composition tables for RCC-1, RCC-2, RCC-3, RCC-5 and RCC8 (cf. section 2). The reasoner uses these definitions together with selected role assertions in the ABox in order to calculate the spatial relations holding between individual regions, and it uses the composition tables in order to check spatial consistency of the ABox. The dotted arrow pointing from the RCCBox back to the ABox indicates, that the calculated relations can be asserted in the ABox, in the above case as partiallyOverlaps(Birmensdorf, Albiskette-Reppischtal), in order to speed up the processing of similar queries in the future. This implies that the names of the relations have previously been introduced as role names in the TBox, for instance as partiallyOverlaps $\sqsubseteq$ spatiallyRelated.

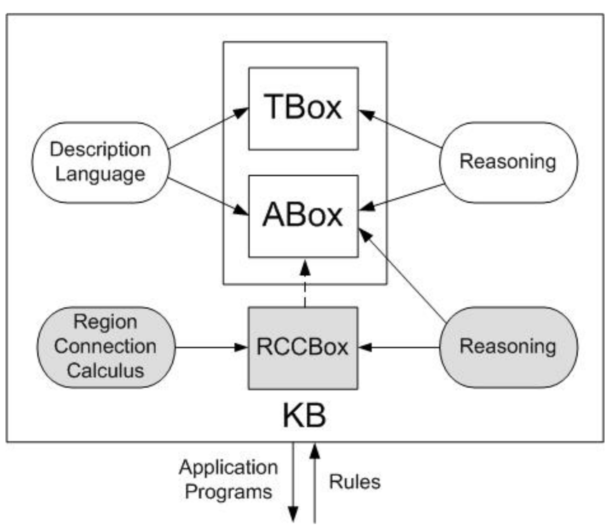

Fig. 2. Architecture of a hybrid knowledge representation system (adapted from [8])

7 Conclusion

In order to allow for an open and intuitive search in environmental databases, terminological concepts and spatial relations must be represented in a way that they can be exploited for searches. In the Datacenter Nature and Landscape we address this by introducing a semantic layer as the underlying data structure which integrates terminological and spatial aspects of the application domain. Since it is not possible to express RCC in OWL without a major revision of the latter our approach is based on an extension of the knowledge representation system architecture and not of the description language. By asserting the primitive relation “connects with” for all pairs of regions it applies to in the ABox of the knowledge base, we provide for a minimal interface between the terminological and spatial component and enable reasoning with both formalisms.

Considering the overall intention to make spatial relations between regions accessible to logic formalism and to qualitative reasoning, it can be objected that, in the domain of geography, it might be more convenient to use the cartographic user interfaces provided by up-to-date geographic information systems for answering spatial queries. However, this is the case only if the result of a region connection query is consumed by the (human) user. One can easily imagine sophisticated queries where the information about spatial relations is merely an intermediate result which is further processed. For instance, the usage scenario could be extended in such a way that the described spatial expansion is followed by a second semantic expansion which retrieves thematic information about any of the communes adjacent to Birmensdorf. In this case a formal representation is required. The same applies to RCC applications where regions are interpreted in terms of time rather than space.

We argued that with the proposed knowledge representation system architecture the calculated basic relations such as partiallyOverlaps can be asserted in the ABox in order to speed up the processing of similar queries in the future. However, since decision problems in RCC belong to the NP (Non-deterministic Polynomial time) class of computational complexity [18] (e.g., the decision whether partiallyOverlaps(Birmensdorf, Albiskette-Reppischtal) is a valid assertion or not), the question arises whether it is favorable to assert the calculated relations. The answer to this question depends on the size of the knowledge base, more specifically, on the number of regions considered, and will be given when the system is evaluated. Note that in the case where the calculated relations are asserted, ABox consistency is no longer an issue of reasoning in description logics alone but also of reasoning with RCC. ABox consistency with respect to RCC relations is an issue of practical relevance: The ontology for our environmental database will evolve over time: Some individual regions will be removed (e.g. due to the fusion of communes), others will be added (e.g. in the course of the revision of an inventory). Without a means to check for consistency, the knowledge base will soon be inconsistent.

Currently, we are working on the design and implementation of the ontology. The key concepts identified so far include inventory, object, ordinance, document, (protected or endangered) species, canton, commune; in addition to the RCC relations, the identified relations include objects (inventory $\times$ object), protection_goal (inventory $\times$ ordinance), object_list (inventory $\times$ document), object_sheet (object $\times$ document), species_pl (object $\times$ species), perimeter (object $\times$ int), area (object $\times$ int), both concept and relation names in German and French [19]. The results of first tests with different OWL reasoners are encouraging: It appears that an existing component can be embedded in the system without modification. Since existing RCC reasoners do not provide support for the calculation of the base relations, it is an open question whether they are useful for our purposes. Further research is planned on how to store the calculated RCC relations, on the representation of temporal aspects and on the integration of the knowledge representations system into the virtual data base project [20].

Acknowledgements. The authors sincerely thank Jürg Schenker, Martin Hägeli and Martin Brändli for the fruitful discussions and leadership that made this research possible. This research has been funded and conducted in cooperation with the Swiss Federal Office for the Environment (FOEN). Related research was funded by the European Commission and by the Swiss Federal Office for Education and Science within the 6th-Framework Programme project REWERSE number-506779 (cf. http://rewerse.net).

References

1. Gruber TR (1993) A translation approach to portable ontology specifications. Knowledge Acquisition 5:199–220
2. McGuiness DL (2005) Ontologies come of age. In: Fensel D, Hendler J, Lieberman H, Wahlster W (eds) Spinning the Semantic Web. The MIT Press, pp 171–194
3. Randell DA, Cui Z, Cohn AG (1992) A spatial logic based on regions and connections. In: Nebel B, Rich C, Swartout W (eds) Principles of knowledge representation and reasoning. Morgan Kaufmann, San Mateo, CA, pp 165–176
4. Bennett B (2000, August 6–18) Logics for topological reasoning. In: 12th European Summer School in Logic, Language and Information (ESSLLI-MM). Birmingham, UK
5. Sattler U, Calvanese D, Molitor R (2003) Relationship with other formalisms. In: Baader F, Calvanese D, McGuinness D, Nardi D, Patel-Schneider P (eds) The description logic handbook: Theory, implementation and applications. Cambridge University Press, pp 137–177
6. Grütter R (2006) Software-Agenten im Semantic Web. Informatik-Spektrum 29(1):3–13
7. Baltensweiler A, Brändli M (2004) Web-based exploration of environmental data and corresponding metadata, in particular lineage information. In: Scharl A (ed): Environmental online communication. Springer-Verlag, London, pp 127–132
8. Baader F, Nutt W (2003) Basic description logics. In: Baader F, Calvanese D, McGuinness D, Nardi D, Patel-Schneider P (eds) The description logic handbook: Theory, implementation and applications. Cambridge University Press, pp 47–100
9. Patel-Schneider PF, Hayes P, Horrocks I (2004) OWL Web Ontology Language: Semantics and Abstract Syntax. W3C Recommendation 10 February 2004. World Wide Web Consortium
10. Bechhofer S, van Harmelen F, Hendler J, Horrocks I, McGuinness DL, Patel-Schneider PF, Stein LA (2004) OWL Web Ontology Language Reference. W3C Recommendation 10 February 2004. World Wide Web Consortium
11. Haarslev V, Lutz C, Möller R (1998) Foundations of spatioterminological reasoning with description logics. In: Cohn AG, Schubert LK, Shapiro SC (eds) Principles of knowledge representation and reasoning. Morgan-Kaufmann, San Mateo, CA, pp 112–123
12. Haarslev V, Lutz C, Möller R (1999, June) A description logic with concrete domains and a role-forming predicate operator. Journal of Logic and Computation 9(3):351–384
13. Wessel M (2002, April 19–21) On spatial reasoning with description logics. Position paper. In: Horrocks I, Tessaris S. (eds) Proceedings of the International Workshop on Description Logics (DL2002), Toulouse, France. CEUR Workshop Proceedings 53, pp 156–163
14. Lutz C, Miličić M (2005) A tableau algorithm for description logics with concrete domains and GCIs. In: Proceedings of the International Workshop on Description Logics (DL2005). CEUR Workshop Proceedings 147
15. Katz Y, Grau BC (2005) Representing qualitative spatial information in OWL-DL. In: Proceedings of OWL: Experiences and directions, Galway, Ireland
16. Horrocks I, Sattler U (2003) Decidability of SHIQ with Complex Role Inclusion Axioms. In: Proceedings of the 18th International Joint Conference on Artificial Intelligence (IJCAI 2003). Morgan Kaufmann, Los Altos, pp 343–348
17. McGuinness DL, van Harmelen F (2004) OWL Web Ontology Language: Overview. W3C Recommendation 10 February 2004. World Wide Web Consortium
18. Grigni M, Papadias D, Papadimitriou C (1995) Topological inference. In: Mellish CS (ed) Proceedings of the 14th International Joint Conference on Artificial Intelligence. Morgan Kaufmann Publishers, San Mateo, CA, pp 901-906
19. Bauer-Messmer B, Grütter R (2007) Designing a bilingual eco-ontology for open and intuitive search. In: Proceedings of the Information Technologies in Environmental Engineering (ITEE), Oldenburg, Germany
20. Frehner M, Brändli M (2006) Virtual database: Spatial analysis in a Web-based data management system for distributed ecological data. Environmental Modelling & Software 21, 1544–1554