Improving an RCC-Derived Geospatial Approximation by OWL Axioms

Rolf Grütter, Thomas Scharrenbach, and Bettina Bauer-Messmer

Swiss Federal Research Institute WSL, An Institute of the ETH Board, Zürcherstrasse 111, CH-8903 Birmensdorf, Switzerland
`{rolf.gruetter,thomas.scharrenbach,bettina.bauer}@wsl.ch`

Abstract. An approach to improve an RCC-derived geospatial approximation is presented which makes use of concept inclusion axioms in OWL. The algorithm used to control the approximation combines hypothesis testing with consistency checking provided by a knowledge representation system based on description logics. Propositions about the consistency of the refined ABox w.r.t. the associated TBox when compared to baseline ABox and TBox are made. Formal proves of the divergent consistency results when checking either of both are provided. The application of the approach to a geospatial setting results in a roughly tenfold improved approximation when using the refined ABox and TBox. Ways to further improve the approximation and to automate the detection of falsely calculated relations are discussed.

Keywords: Geospatial approximation, Region Connection Calculus, Web Ontology Language, hypothesis testing, consistency checking.

1 Introduction

Topological relations play an important role for the description of geospatial phenomena. Accordingly, the Open GIS (OGIS) standard defines topological set operators for the retrieval of data in terms of spatial relations [1].[1] It is implemented by today's geographical information systems. However, there is currently no means to couple geometrical computations with symbolic reasoning services provided by a knowledge representation system. Such a coupling (or an alternative procedure with a similar effect) is necessary if users should be supported in constructing spatio-thematic queries which are consistent with their conceptualization of a given domain of discourse.

The coupling of geometrical computations with symbolic reasoning can be anticipated if the thematic (i.e., terminological) representation in the semantic layer of a system is complemented by a spatial representation. Ideally, the spatial representation uses topological relations which are compliant with the OGIS standard. It should also be based on a formalism which allows inferring implicit knowledge from the knowledge explicitly represented. Both requirements are fulfilled by the Region Connection Calculus (RCC) [2, 3].

---

[1] The OGIS consortium is formed by major software vendors to formulate an industry-wide standard related to GIS interoperability.

A popular method for representing the terminology of a domain together with thematic descriptions in terms of the terminology is description logic, in the context of the Semantic Web particularly the Web Ontology Language (OWL) [4]. Therefore, in order to complement the thematic representation with a spatial representation, ways to combining OWL with RCC must be explored.

The herein presented work explores how an RCC-derived geospatial approximation can be improved by OWL axioms. It builds on an approach published in [5]. The idea is to calculate or approximate geospatial settings based on attributes which can easily be queried from spatial databases such that the process can be automated. For instance, non-administrative regions, such as biotopes, may be stored in data tables together with the administrative regions (e.g., cantons), they overlap, or administrative regions, such as communes, may be stored together with the administrative regions they are externally connected to. This information about the connectedness of regions can be used as a starting point for the calculation of more complex relations such as partOf. The calculated relations allow constructing and querying complex concepts with both thematic and spatial references such as public_park_containing_a_lake $\equiv$ park $\sqcap$ public $\sqcap$ $\exists$contains.lake which is taken from [6]. The contribution of the work, however, is not limited to the geographical domain. It rather adds to the knowledge about the combination of quantitative numerical approaches with qualitative symbolic (i.e., logic) approaches in general. The presented approach is related to previous work on combining RCC with OWL [7, 8] by addressing the calculation of a geospatial world description for assertion in the ABox of a knowledge base. Reasoning with both RCC and OWL will make use of the asserted world description.

The paper is organized as follows: In section 2, related work is discussed. An introduction to RCC is provided in section 3. In section 4, a geospatial approximation is presented which is derived from RCC. In section 5, the theoretical results are applied to a geospatial setting. The approach is discussed in section 6. Section 7 concludes with an outlook on future work.

2 Related Work

Spatio-thematic reasoning with the description logic $\mathcal{ALCRP(D)}$ has been introduced in [9]. The authors define an appropriate concrete domain $\mathcal{D_P}$ for polygons. $\mathcal{RP}$ stands for role definitions based on predicates. More specifically, $\mathcal{ALCRP(D)}$ extends $\mathcal{ALC(D)}$ [10] 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 $\mathcal{ALCRP(D)}$ is provided in [11]. In order to ensure termination of the satisfiability algorithm for the concrete domain $\mathcal{D_P}$, the authors impose restrictions on the syntactic form of the set of terminological axioms which impose tight constraints on modeling spatio-thematic structures [9, 11].

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 [12]. 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 $\mathcal{ALC_{RA}}$, enforcing role disjointness on all roles R and S ($R^{\mathcal{I}} \cap S^{\mathcal{I}} = \emptyset$), is undecidable. In [12] specializations of $\mathcal{ALC_{RA}}$ which specifically consider the family of RCC related calculi are investigated. Using role axioms of the above introduced form, the author shows for both $\mathcal{ALCI_{RCC5}}$ and $\mathcal{ALCI_{RCC8}}$ 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 [13]. 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 $\mathcal{ALC(C)}$ which is $\mathcal{ALC}$ extended with two *constraint constructors*.

In [14] the authors aim at representing qualitative spatial information in OWL DL. 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, in addition to the transitive one, is needed in order to describe the accessibility relation. In order to represent 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. 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 concept axioms. In [15] 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. Complex role inclusion axioms of the unrestricted form are supported by the description logic $\mathcal{SROIQ}$ which serves as a logical basis for OWL 1.1 [16]. However, in order to axiomatize the composition of RCC relations, a language must 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 defining the basic RCC relations require additional role constructors such as intersection and complement. Extensions of $\mathcal{SHIQ}$ with these kinds of role constructors have, to our knowledge, not been investigated so far. $\mathcal{SROIQ}$ supports negation of roles (i.e. complement) but not intersection.

To summarize, the existing approaches show that the combination of formalisms for thematic and spatial reasoning is not straightforward. In order to uphold decidability, approaches based on $\mathcal{ALC}$ require that the resulting language is constrained. This bears on its expressivity for modeling spatio-thematic structures. $\mathcal{SHIQ}$ and $\mathcal{SROIQ}$ do not provide for the expression of role inclusion axioms of the form $S \circ T \sqsubseteq R_1 \sqcup \ldots \sqcup R_n$ which is a requirement for spatial reasoning with RCC. The approach based on $\mathcal{SHOIN(D)}$ (OWL DL) requires only a minimal extension of the language which has been considered in the draft to OWL 1.1 [17]. However, the notion of regions as sets in the abstract object domain prevents RCC from effectively combining with domain ontologies. The reason for this is that OWL DL requires type separation: a class cannot be an individual (or a property) at the same time [18]. Yet, in order to classify regions in a domain ontology they must be represented as individuals, and not as concepts.

Furthermore, in [19] it is proposed to encode spatial inferences in the Semantic Web Rule Language (SWRL) [20]. Even though not explicitly mentioned, the examples are provided in a RCC-like style. SWRL uses Horn-like rules which are combined with OWL DL (and OWL Lite). Horn rules do not allow complex heads (which refer to the expressions on the right hand side of the implication connective). However, complex heads in terms of disjunctions are required in order to formalize the RCC composition axioms.

In [6] a generic architectural framework for building ontology-based information systems is presented which covers regions in the system design space instead of isolated points. The framework introduces a graph-based *substrate data model* and a *substrate query language*. An ABox can be seen as a substrate, an ABox with an associated TBox as a substrate with a background theory. A substrate can also encode geometric or spatial structures in a geometric substrate which is called an *SBox* (Space Box). The authors propose four options to solve the spatial representation problem: (1) Use an ABox, (2) use a map substrate, (3) use a spatial ABox, (4) use an ABox + RCC substrate. The herein presented work addresses the fourth representation option.

3   The Region Connection Calculus

The Region Connection Calculus (RCC) is an axiomatization of certain spatial concepts and relations in first order logic [2, 3]. 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. The most common of these are illustrated in figure 1 and their definitions together with those of additional relations are given in table 1. The asymmetrical relations P, PP, TPP and NTPP have inverses which we write, in accordance with [3], as $R\mathsf{i}$, where $R \in \{\mathsf{P, PP, TPP, NTPP}\}$. These relations are defined by definitions of the form $R\mathsf{i}(x, y) \equiv_{def} R(y, x)$.
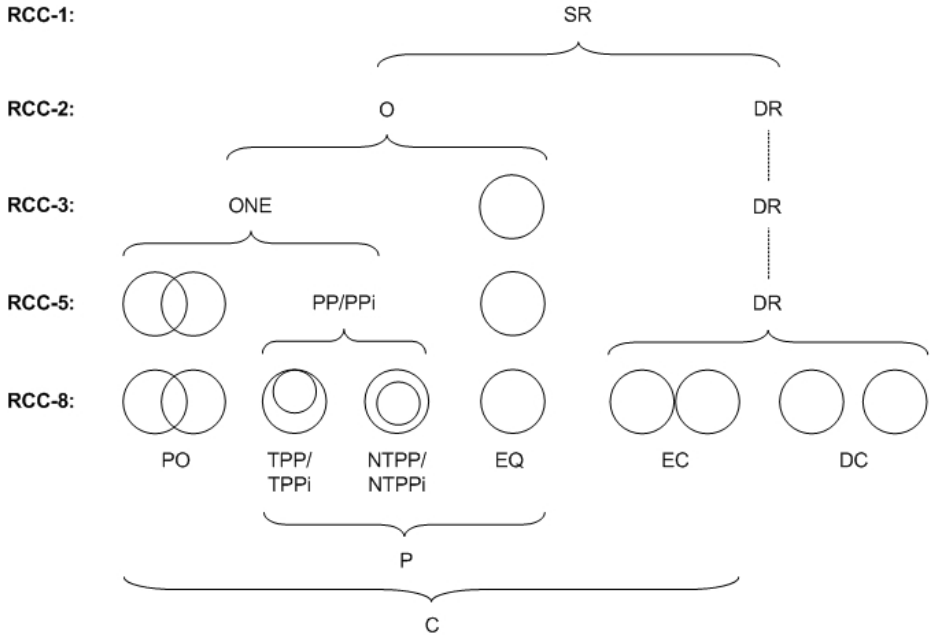
Fig. 1. RCC family tree (for the entire names of the relations cf. table 1)

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-1 = {SR}, RCC-2 = {O, DR}, RCC-3 = {ONE, EQ, DR}, RCC-5 = {PP, PPi, PO, EQ, DR}. 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 [3].

Table 1. RCC relations

$SR(x, y)$	$\equiv_{\text{def}} \top(x, y)$		(Spatially Related)
$C(x, y)$	(primitive relation)		(Connects with)
$DC(x, y)$	$\equiv_{\text{def}} \neg C(x, y)$		(DisConnected from)
$P(x, y)$	$\equiv_{\text{def}} \forall z[C(z, x) \rightarrow C(z, y)]$		(Part of)
$O(x, y)$	$\equiv_{\text{def}} \exists z[P(z, x) \wedge P(z, y)]$		(Overlaps)
$DR(x, y)$	$\equiv_{\text{def}} \neg O(x, y)$		(DiscRete from)
$EC(x, y)$	$\equiv_{\text{def}} C(x, y) \wedge \neg O(x, y)$		(Externally Connected to)
$EQ(x, y)$	$\equiv_{\text{def}} P(x, y) \wedge P(y, x)$		(EQual to)
$ONE(x, y)$	$\equiv_{\text{def}} O(x, y) \wedge \neg EQ(x, y)$		(Overlaps Not Equal)
$PP(x, y)$	$\equiv_{\text{def}} P(x, y) \wedge \neg P(y, x)$		(Proper Part of)
$PO(x, y)$	$\equiv_{\text{def}} O(x, y) \wedge \neg P(x, y) \wedge \neg P(y, x)$		(Partially Overlaps)
$TPP(x, y)$	$\equiv_{\text{def}} PP(x, y) \wedge \exists z[EC(z, x) \wedge EC(z, y)]$		(Tangential Proper Part of)
$NTTP(x, y)$	$\equiv_{\text{def}} PP(x, y) \wedge \neg\exists z[EC(z, x) \wedge EC(z, y)]$		(Non-Tangential Proper Part of)

According to [2], 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 ("connects with") in terms of two regions whose closures share a common point or stating that two regions connect when the distance between them is zero.

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.

A similar approach which is based on the description of topological relations between two spatial regions was introduced as the 9-intersection model in [21]. In this model, eight out of nine relations can be interpreted in the same way as we interpret the RCC-8 relations, namely as spatial relations between polygons in the integral plane [7]. Only the ninth relation is specific for the model. Since it is based on a topological framework – and not on a logical one – the 9-intersection model is harder to combine with OWL DL than RCC.

Table 2. RCC-5 composition table ($\mathsf{T}(x, z) \equiv_{def} \{\mathsf{DR}(x, z), \mathsf{PO}(x, z), \mathsf{EQ}(x, z), \mathsf{PP}(x, z), \mathsf{PPi}(x, z)\}$)

∘	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)

4   An RCC-Derived Geospatial Approximation

In order to approximate a geospatial setting we derive a hypothesis for the RCC relation $\mathsf{P}(x, y)$ from its definition in table 1 (section 4.1). For each pair of connecting regions, this hypothesis is tested against the role assertions in the ABox of a knowledge base (section 4.2). If the hypothesis is not falsified it is checked whether its assertion causes an inconsistency of the knowledge base or not. If the knowledge base remains consistent the relation $\mathsf{P}(x, y)$ is asserted in the ABox.

We compare two procedures which use different levels of knowledge. The *baseline approximation* uses baseline knowledge in terms of the concept Region which is used

for asserting individual regions in the ABox of the knowledge base. The *improved approximation* uses additional knowledge from cartographic analysis. In particular, it distinguishes communes as a special kind of regions and introduces a property restriction which constrains the values of the property to some communes. Since the approach is generic and can be applied to similar problems in other domains as well, the algorithm and the axioms of TBoxes $\mathcal{T}_1$ (baseline) and $\mathcal{T}_2$ (constrained) are introduced in an abstract way. The notation used for the DL axioms is taken from [22].

4.1   Deriving a Hypothesis from RCC

The relation $\mathsf{P}(x, y)$ ("x is a part of y") plays a key role in the definitions of the RCC relations (cf. table 1): It is directly defined by the primitive relation and, conversely, a number of relations are defined in terms of $\mathsf{P}(x, y)$. For this reason, we use the definition of the relation $\mathsf{P}(x, y)$ as a starting point for our geospatial approximation. Since $\mathsf{P}(x, y)$ is directly defined by the primitive relation it will be sufficient to assert the primitive relation in an input representation. The theory defines the relation $\mathsf{P}(x, y)$ as follows:

$$\mathsf{P}(x, y) \equiv_{def} \forall z[\mathsf{C}(z, x) \rightarrow \mathsf{C}(z, y)].$$

In accordance with the Semantic Web philosophy, this definition assumes an open world: x is a part of y if and only if for any imaginable region z the following holds: If z connects with x then z also connects with y. In order to adapt the definition to the closed world of a practical application we replace the universal quantifier by a conjunction ranging over all regions represented. In a closed world the condition on the right hand side of the expression is no longer sufficient – but still necessary (cf. below) – for the relation $\mathsf{P}(x, y)$. Accordingly, we replace the equality sign by an inclusion sign.

$$\mathsf{P}(x, y) \sqsubseteq \wedge_{z_i}[\mathsf{C}(z_i, x) \rightarrow \mathsf{C}(z_i, y)] \tag{1}$$

with $1 \leq i \leq n$, n the number of regions represented.

In the minimum case (i = 1) the region x connects only with itself (remember that the relation C is by definition reflexive) and it holds that $x = y = z$. In the maximum case (i = n) all regions, including x (y, respectively) connect with x (y, respectively). Intuitively, the calculation of $\mathsf{P}(x, y)$ is expected to be more precise with a high number of regions z_i represented.

Note that from an epistemic viewpoint the shift from an open world to a closed world limits the range of the proposition – which is originally formalized as a definition – from the partially unknown universe to a known subset thereof. Both the universal proposition and the middle range proposition cannot be empirically verified but only falsified.[2] The first cannot be verified as a matter of principle: It is not possible to test for the infinite number of all imaginable regions z connecting with x whether they also connect with y. The second cannot be verified because the condition on the right hand side of the expression is not sufficient. However, both can be falsified: A single observation of a region z connecting with x but not with y is sufficient to falsify the hypothesis that x is a part of y. Following this line of argumentation a calculus for $\mathsf{P}(x, y)$ is not expected to be sound. Instead, the question is whether it is complete or

---

[2] The expression *middle range* is adopted from the Middle Range Theory [23].

not. In a practical application it is further of interest how good a calculus approximates the geospatial setting and how the approximation can be controlled.

The question whether a calculus using formula (1) is complete or not can be answered by referring to the reflexive and symmetric properties of the primitive relation $C(x, y)$. If $z = y$ and x connects with y the formula $C(y, x) \rightarrow C(y, y) \equiv C(x, y) \rightarrow C(y, y)$ evaluates to true for $P(x, y)$. Thus, the condition that x connects with y is sufficient for hypothesizing that x is a part of y. This means that a calculus using formula (1) is expected to be complete in a practical application. $P(x, y)$ is generalized to the hypothesis $T(a, b)$ in the next section.

4.2   Combining Hypothesis Testing with Consistency Checking

The algorithm used to control the RCC-derived geospatial approximation combines hypothesis testing with consistency checking provided by a knowledge representation system based on description logics. Generically, it proceeds as follows:

FUNCTION	*geospatialApproximation*
Input:	hypothesis $T(., .)$, knowledge base $KB = \{\text{ABox}, \text{TBox}\}$
Output:	*counter*

```
0     counter ← 0
1     M ← {(a, b)} ⊆ ABox     /* the set of pairs of individuals in ABox */
2     WHILE M is NOT empty
3              SELECT (a, b) ∈ M
4              IF T(a, b) is NOT falsified in ABox THEN
5                      ABox ← ABox ∪ T(a, b)
6                      IF KB is consistent THEN
7                              counter ← counter + 1
8                      ELSE
9                              ABox ← ABox \ T(a, b)
10                     ENDIF
11             ENDIF
12             M ← M \ (a, b)
13    ENDWHILE
```

For the consistency check in step 6 two different TBoxes are alternatively applied: TBox $\mathcal{T}_1$ consisting of axioms 0–1 and 4–7 or TBox $\mathcal{T}_2$ consisting of axioms 0–7:

0	$C \sqsubseteq \top$	$C^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$
1	R	$R^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$
2	$D \sqsubseteq C$	$D^{\mathcal{I}} \subseteq C^{\mathcal{I}}$
3	$D \sqsubseteq \neg(\exists S.D)$	$D^{\mathcal{I}} \subseteq \setminus \{a \in \Delta^{\mathcal{I}} \mid \exists b. (a, b) \in S^{\mathcal{I}} \wedge b \in D^{\mathcal{I}}\}$
4	$\exists R.\top \sqsubseteq C$	$\{a \in \Delta^{\mathcal{I}} \mid \exists b. (a, b) \in R^{\mathcal{I}}\} \subseteq C^{\mathcal{I}}$
5	$\top \sqsubseteq \forall R.C$	$\Delta^{\mathcal{I}} \subseteq \{a \in \Delta^{\mathcal{I}} \mid \forall b. (a, b) \in R^{\mathcal{I}} \rightarrow b \in C^{\mathcal{I}}\}$
6	$S \sqsubseteq R$	$S^{\mathcal{I}} \subseteq R^{\mathcal{I}}$
7	$T \sqsubseteq S$	$T^{\mathcal{I}} \subseteq S^{\mathcal{I}}$

The concept description $\exists S.D$ introduced by axiom 3 is interpreted as the set of those individuals each of which is in relationship $S^{\mathcal{I}}$ to some individuals of the set $D^{\mathcal{I}}$. The concept D is interpreted as a possibly improper subset of the complement of this set. Together with a refinement of the model in the ABox – a subset of the individuals in $C^{\mathcal{I}}$ are also members of the interpretation of the included concept D – axiom 3 is responsible for the divergent result of the consistency check when using two different TBoxes $\mathcal{T}_1$ or $\mathcal{T}_2$.

Proposition 1. Asserting the role $T(a, b)$ in the ABox $\mathcal{A}_1 = \{C(a), C(b)\}$ results in an ABox $\mathcal{A}_1' = \{C(a), C(b), T(a, b)\}$ which is consistent w.r.t. $\mathcal{T}_1$.

Proof

$(T \sqsubseteq S) \wedge T(a, b) \rightarrow S(a, b)$	Axiom 7
$(S \sqsubseteq R) \wedge S(a, b) \rightarrow R(a, b)$	Axiom 6
$(\exists R.\top \sqsubseteq C) \wedge R(a, b) \rightarrow C(a)$	Axiom 4
$(C \sqsubseteq \top) \wedge C(a) \rightarrow \top(a)$	Axiom 0
$(\top \sqsubseteq \forall R.C) \wedge R(a, b) \rightarrow C(b)$	Axiom 5
$(C \sqsubseteq \top) \wedge C(b) \rightarrow \top(b)$	Axiom 0

Proposition 2. Asserting the role $T(a, b)$ in the ABox $\mathcal{A}_2 = \{D(a), D(b)\}$ results in an ABox $\mathcal{A}_2' = \{D(a), D(b), T(a, b)\}$ which is *inconsistent* w.r.t. $\mathcal{T}_2$.

Proof

$(T \sqsubseteq S) \wedge T(a, b) \rightarrow S(a, b)$	Axiom 7
$(D \sqsubseteq \neg(\exists S.D)) \wedge S(a, b) \rightarrow (\neg D)(a)$	Axiom 3
$(\neg D)(a) \wedge D(a) \rightarrow \bot(a)$	

5   Applying the Approach to a Geospatial Setting

In order to demonstrate the approach, we use a sample of 44 two-dimensional spatial regions (polygons) from four layers of a productive GIS. The spatial regions refer to districts, communes, biotopes and are located in the canton of Zurich, Switzerland (cf. figure 2). The regions are asserted as individuals in the ABox of the knowledge base. For inference with TBox $\mathcal{T}_2$ 38 out of 44 regions are asserted as communes. The connections between regions – which were identified by cartographic analysis – are asserted as role assertions of type connectsWith. Overall, there are 262 relations asserted in our sample. The knowledge base is created from an OWL ontology using the reasoner Pellet (version 1.4). Most descriptions in the ontology are written in OWL DL. Only for the description of the role partOf the OWL 1.1 feature irreflexiveProperty is used. The description logic expressivity of the ontology is $\mathcal{ALHI}$ for $\mathcal{T}_1$ and $\mathcal{ALCHI}$ for $\mathcal{T}_2$. The algorithm used to compute formula (1) and to control the approximation as described in section 4.2 (steps 0–13) is programmed in Java. It accesses the knowledge base at the WonderWeb OWL API.[3]
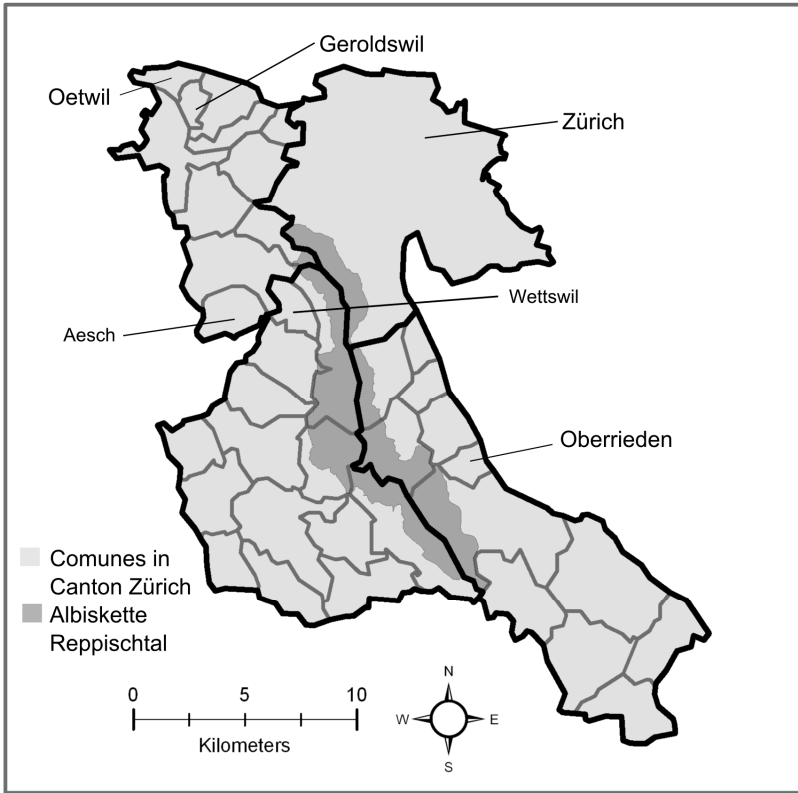
---

[3] http://wonderweb.semanticweb.org/

Fig. 2. Regions in the canton of Zurich. The dark grey shaded region depicts Albiskette-Reppischtal, a biotope of national interest. Regions with bold borderlines depict districts. Regions with regular borderlines depict communes. Note that the district of Zurich and the commune of Zurich share the same geometry, in terms of RCC: EQ (Bezirk_Zürich, Zürich).

The TBoxes $\mathcal{T}_1$ (axioms 0–1 and 4–7) and $\mathcal{T}_2$ (axioms 0–7) are instantiated as follows. Intuitively, axiom 3 says that communes must not overlap each other. The intuition behind axioms 4 and 5 is that regions are spatially related to each other.

0	Region $\sqsubseteq \top$	Region$^\mathcal{I} \subseteq \Delta^\mathcal{I}$
1	spatiallyRelated	spatiallyRelated$^\mathcal{I} \subseteq \Delta^\mathcal{I} \times \Delta^\mathcal{I}$
2	Commune $\sqsubseteq$ Region	Commune$^\mathcal{I} \subseteq$ Region$^\mathcal{I}$
3	Commune $\sqsubseteq \neg(\exists$overlaps.Commune$)$	Commune$^\mathcal{I} \subseteq \setminus \{a \in \Delta^\mathcal{I} \mid \exists b. (a, b) \in$ overlaps$^\mathcal{I} \wedge b \in$ Commune$^\mathcal{I}\}$
4	$\exists$spatiallyRelated.$\top \sqsubseteq$ Region	$\{a \in \Delta^\mathcal{I} \mid \exists b. (a, b) \in$ spatiallyRelated$^\mathcal{I}\}$ $\subseteq$ Region$^\mathcal{I}$
5	$\top \sqsubseteq \forall$spatiallyRelated.Region	$\Delta^\mathcal{I} \subseteq \{a \in \Delta^\mathcal{I} \mid \forall b. (a, b) \in$ spatiallyRelated$^\mathcal{I} \rightarrow b \in$ Region$^\mathcal{I}\}$
6	overlaps $\sqsubseteq$ spatiallyRelated	overlaps$^\mathcal{I} \subseteq$ spatiallyRelated$^\mathcal{I}$
7	partOf $\sqsubseteq$ overlaps	partOf$^\mathcal{I} \subseteq$ overlaps$^\mathcal{I}$

Using TBox $\mathcal{T}_1$ the algorithm calculates 109 relations of type $P(x, y)$. The cartographic evaluation results in 27 relations being falsely calculated as $P(x, y)$ whereas they are relations of type $EC(x, y)$. Using TBox $\mathcal{T}_2$ instead of $\mathcal{T}_1$ the algorithm calculates 85 relations of type $P(x, y)$. In this case only three relations of type $EC(x, y)$ are falsely calculated as $P(x, y)$. In both cases all relations of type $P(x, y)$ verified by cartography are identified as such. As expected, the calculations are complete but not sound in our sample. The approximation with $\mathcal{T}_2$ is roughly ten times better than that with $\mathcal{T}_1$.

To give an example, one of the relations of type $EC(x, y)$ which is falsely calculated as $P(x, y)$ using TBox $\mathcal{T}_1$ refers to the relation between Geroldswil and Oetwil (cf. figure 2). Since all regions connecting with Geroldswil also connect with Oetwil the relation between them is (falsely!) assumed to be of type $P(x, y)$. Since TBox $\mathcal{T}_2$ introduces the restriction that Geroldswil and Oetwil – which are both communes – must not overlap, partOf(Geroldswil, Oetwil) is removed from the ABox of the knowledge base in step 9 of the algorithm (cf. section 4).

At http://webgis.wsl.ch/rcc-webclient/faces/rcc-client.jspx the presented approach can be tested against the constraint (TBox $\mathcal{T}_2$) and the unconstraint (TBox $\mathcal{T}_1$) ontology, respectively.[4] The reader is encouraged to perform tests on their own by uploading a custom ontology. For being processible a custom ontology has to be consistent either with $\mathcal{T}_1$ or with $\mathcal{T}_2$.

6 Discussion

The three relations which are falsely calculated as $P(x, y)$ in our sample – even with the improved algorithm – are connectsWith(Bezirk_Affoltern, Aesch), connectsWith (Bezirk_Dietikon, Wettswil), connectsWith(Albiskette-Reppischtal, Oberrieden). As can be seen from figure 2 the communes Aesch, Oberrieden and Wettswil share the property of being adjacent to the surrounding area. The question, therefore, is whether the assertion of additional connections between the border areas and the surrounding area in the ABox of the knowledge base would further improve the approximation. In our sample also the districts Bezirk_Affoltern and Bezirk_Dietikon and the biotope of national interest Albiskette-Reppischtal connect with the surrounding area. For this reason we do not expect an improvement in this case. However, an improvement can be expected in other cases.

Another question is whether the detection of the falsely calculated relations can be automated. This would substantially reduce the effort of a manual cartographic verification. To a certain degree this is indeed possible by introducing further axioms in the TBox of the knowledge base and by further refining the model in the ABox. With respect to the first, *qualified number restrictions* can be introduced, for instance in order to encode the restriction that communes may be part only of a single district. With respect to the last, those regions which are districts can be asserted as such. Qualified number restrictions are not supported by OWL DL. However, they are considered in the draft to OWL 1.1 (there a qualified number restriction is called an objectMinCardinality) [17].

---

[4] Please contact the authors if you wish to access the service and it is no longer maintained.

| 8 | District $\sqsubseteq$ Region | District$^\mathcal{I}$ $\subseteq$ Region$^\mathcal{I}$ |
| 9 | Commune $\sqsubseteq$ $\leq$ 1 partOf.District | Commune$^\mathcal{I}$ $\subseteq$ $\{a \in \Delta^\mathcal{I} \mid \mid\{b \mid (a, b) \in$ partOf$^\mathcal{I} \wedge b$ $\in$ District$^\mathcal{I}\}\mid \leq 1\}$ |

In our sample axioms 8 an 9 (together with a further refined model) are expected to cause an inconsistency of the knowledge base when asserting the relations part-Of(Aesch, Bezirk_Affoltern) and partOf(Wettswil, Bezirk_Dietikon) in the ABox (cf. step 5 of the algorithm introduced in section 4). This because the relations partOf(Aesch, Bezirk_Dietikon) and partOf(Wettswil, Bezirk_Affoltern) are (truly!) asserted in the ABox. Conversely, the falsely calculated partOf-relation between Oberrieden and Albiskette-Reppischtal would not be detected.

7   Conclusion and Outlook

We presented an approach to improve an RCC-derived geospatial approximation which makes use of concept inclusion axioms in OWL. The algorithm used to control the approximation combines hypothesis testing with consistency checking provided by a knowledge representation system based on description logics. We made propositions about the consistency of the refined ABox w.r.t. the associated TBox when compared to baseline ABox and TBox and provided formal proves of the divergent consistency results when checking either of both. The application of the approach to a sample of 44 two-dimensional regions which are related to each other through 262 spatial relations resulted in a roughly tenfold improved approximation when using the refined ABox and TBox.

Since we expect the approximation to be more precise with a high number of regions represented a next step will be to evaluate the approach on different scales. In addition to precision of the approximation the impact of scalability on performance will also be of interest.

For productive use it would be desirable to assess the degree of imprecision and to provide a measure of confidence along with the approximation. This measure could then be put into relationship with the inherent imprecision of the data. Real data are error-prone. For instance, the shape of the biotope of national interest Albiskette-Reppischtal which is geometrically computed in figure 2 is different from the shape which is reconstructed based on the textual description of the biotope (which is held in the same database!): According to the textual description, Albiskette-Reppischtal should also overlap the commune Wettswil which it does not geometrically.

Provided the relations of type P(x, y) can be sufficiently well approximated for a given spatial setting, a next step is to calculate the RCC-5 relations by using their definitions in table 1 and the relationships between RCC species in figure 1. The relations between individual regions can then be asserted in terms of RCC-5 in the ABox of a knowledge base. Based on this, a full-fledged spatio-terminological reasoning service as outlined in [7, 8] can be developed.

Since the presented approach is generic in nature it can be applied to other domains as well. It will be interesting to identify relevant problems in these domains.

References

1. Shekhar, S., Chawla, S.: Spatial Databases: A Tour. Pearson Education. Upper Saddle River, New Jersey (2003)

2. Randell, D.A., Cui, Z., Cohn, A.G.: A Spatial Logic based on Regions and Connections. In: Nebel, B., Rich, C., Swartout, W. (eds.) Principles of Knowledge Representation and Reasoning, pp. 165–176. Morgan Kaufmann, San Mateo (1992)

3. Bennett, B.: Logics for Topological Reasoning. In: 12th European Summer School in Logic, Language and Information (ESSLLI-MM), Birmingham, UK, August 2000, pp. 6–18 (2000)

4. Patel-Schneider, P.F., Hayes, P., Horrocks, I.: OWL Web Ontology Language: Semantics and Abstract Syntax. W3C Recommendation, February 10, 2004. World Wide Web Consortium (2004)

5. Grütter, R., Bauer-Messmer, B., Hägeli, M.: Extending an Ontology-based Search with a Formalism for Spatial Reasoning. In: Proceedings of the 23rd Annual ACM Symposium on Applied Computing (ACM SAC 2008), pp. 2266–2270. ACM, New York (2008)

6. Wessel, M., Möller, R.: Flexible Software Architectures for Ontology-Based Information Systems. Journal of Applied Logic, Special Issue on Emperically Sucessful Systems (2007)

7. Grütter, R., Bauer-Messmer, B.: Towards Spatial Reasoning in the Semantic Web: A Hybrid Knowledge Representation System Architecture. In: Fabrikant, S.I., Wachowicz, M. (eds.) The European Information Society: Leading the Way with Geo-information. LNGC, pp. 349–364. Springer, Heidelberg (2007)

8. Grütter, R., Bauer-Messmer, B.: Combining OWL with RCC for Spatioterminological Reasoning on Environmental Data. In: Golbreich, C., Kalyanpur, A., Parsia, B. (eds.) OWL: Experiences and Directions (OWLED). CEUR Workshop Proceedings, vol. 258 (2007)

9. Haarslev, V., Lutz, C., Möller, R.: Foundations of Spatioterminological Reasoning with Description Logics. In: Cohn, A.G., Schubert, L.K., Shapiro, S.C. (eds.) Principles of Knowledge Representation and Reasoning, pp. 112–123. Morgan-Kaufmann, San Mateo (1998)

10. Baader, F., Hanschke, P.: A scheme for integrating concrete domains into concept languages. In: Twelfth International Joint Conference on Artificial Intelligence, August 1991, pp. 452–457 (1991)

11. Haarslev, V., Lutz, C., Möller, R.A.: Description Logic with Concrete Domains and a Role-Forming Predicate Operator. Journal of Logic and Computation 9(3), 351–384 (1999)

12. Wessel, M.: On Spatial Reasoning with Description Logics. Position Paper. In: Horrocks, I., Tessaris, S. (eds.) Proceedings of the International Workshop on Description Logics (DL 2002), CEUR Workshop Proceedings, Toulouse, France, April 19–21, 2002, vol. 53, pp. 156–163 (2002)

13. Lutz, C., Miličić, M.: A Tableau Algorithm for Description Logics with Concrete Domains and General TBoxes. Journal of Automated Reasoning 38(1-3), 227–259 (2007)
14. Katz, Y., Grau, B.C.: Representing Qualitative Spatial Information in OWL-DL. In: Proceedings of OWL: Experiences and Directions, CEUR Workshop Proceedings, Galway, Ireland, vol. 188 (2005)
15. Horrocks, I., Sattler, U.: Decidability of SHIQ with Complex Role Inclusion Axioms. In: Proceedings of the 18th International Joint Conference on Artificial Intelligence (IJCAI 2003), pp. 343–348. Morgan Kaufmann, Los Altos (2003)
16. Horrocks, I., Kutz, O., Sattler, U.: The Even More Irresistible SROIQ. In: Proceedings of the 10th International Conference of Knowledge Representation and Reasoning (KR 2006), Lake District, United Kingdom (2006)
17. Grau, B.C., Motik, B.: OWL 1.1 Web Ontology Language: Model-Theoretic Semantics. Editor's Draft of (April 6, 2007), http://www.webont.org/owl/1.1/semantics.html
18. Patel-Schneider, P.F., Hayes, P., Horrocks, I.: OWL Web Ontology Language: Semantics and Abstract Syntax. W3C Recommendation February 10, 2004. World Wide Web Consortium (2004)
19. Bishr, Y.: Geospatial Semantic Web. In: Rana, S., Sharma, J. (eds.) Frontiers of Geographic Information Technology, pp. 139–154. Springer, Heidelberg (2006)
20. Horrocks, I., Patel-Schneider, P.F., Boley, H., Tabet, S., Grosof, B., Dean, M.: SWRL: A Semantic Web Rule Language. Combining OWL and RuleML. W3C Member Submission May 21, 2004. World Wide Web Consortium (2004)
21. Egenhofer, M., Franzosa, R.: Point-Set Topological Spatial Relations. International Journal of Geographical Information Systems 5(2), 161–174 (1991)
22. Baader, F., Nutt, W.: Basic Description Logics. In: Baader, F., Calvanese, D., McGuinness, D.L., Nardi, D., Patel-Schneider, P.F. (eds.) The Description Logic Handbook, pp. 47–100. Cambridge University Press, Cambridge (2003)
23. Merton, R.K.: Social Theory and Social Structure (Enlarged Edition). The Free Press, New York (1968)