Research Article

Vague Spatio-Thematic Query Processing: A Qualitative Approach to Spatial Closeness

Rolf Grütter
Swiss Federal Research Institute
WSL

Thomas Scharrenbach
Swiss Federal Research Institute
WSL

Bettina Waldvogel
Swiss Federal Research Institute
WSL

Abstract
In order to support the processing of qualitative spatial queries, spatial knowledge must be represented in a way that machines can make use of it. Ontologies typically represent thematic knowledge. Enhancing them with spatial knowledge is still a challenge. In this article, an implementation of the Region Connection Calculus (RCC) in the Web Ontology Language (OWL), augmented by DL-safe SWRL rules, is used to represent spatio-thematic knowledge. This involves partially ordered partitions, which are implemented by nominals and functional roles. Accordingly, a spatial division into administrative regions, rather than, for instance, a metric system, is used as a frame of reference for evaluating closeness. Hence, closeness is evaluated purely according to qualitative criteria. Colloquial descriptions typically involve qualitative concepts. The approach presented here is thus expected to align better with the way human beings deal with closeness than does a quantitative approach. To illustrate the approach, it is applied to the retrieval of documents from the database of the Datacenter Nature and Landscape (DNL).

1 Introduction

Knowledge representation has received considerable attention again in the last decade, fueled, most notably, by a joint initiative of research institutes and industrial organiza-

Address for correspondence: Rolf Grütter, DVM, MBA, Swiss Federal Research Institute WSL, Land Resource Assessment, Geographic Information Science, Zürcherstrasse 111, CH-8903 Birmensdorf, Switzerland. E-mail: rolf.gruetter@wsl.ch

© 2010 Blackwell Publishing Ltd
doi: 10.1111/j.1467-9671.2010.01185.x
tions to advance the Semantic Web (Berners-Lee 1998). Technically, the initiative was committed to advance Description Logics (DLs) as a means of capturing the terminological and assertional knowledge of a domain and of inferring new knowledge from existing knowledge. This kind of knowledge has been (and continues to be) made available by so-called ontologies (Gruber 1993).

Ontologies are increasingly integrated into applications to support semantic interoperability and to provide a homogeneous view of heterogeneous data (Wache et al. 2001, Fonseca et al. 2002). In Geographic Information Systems (GIS), where mereological considerations are all-important, the challenge is to enhance ontologies, typically representing thematic knowledge, with spatial knowledge (Jones et al. 2002, Bishr 2006). One way to achieve this is to combine existing logic-based approaches to spatial knowledge representation with description logics (Haarslev et al. 1998, Lutz and Milicˇic´ 2007, Grütter and Bauer-Messmer 2007, Möller and Wessel 2007).

In this article, an implementation of the Region Connection Calculus (RCC) (Randell et al. 1992, Bennett 2000) in the Web Ontology Language (OWL) (Patel-Schneider et al. 2004), augmented by DL-safe SWRL rules (Motik et al. 2004), is used to represent spatio-thematic knowledge. We show how such a representation can be operated on in order to answer queries using (possibly vague) spatial concepts. The primary goal is to demonstrate how state-of-the-art technology can be used for this purpose. While pursuing this goal, some additions to the theory of spatial knowledge representation will be made. The basic idea underlying the approach presented here is to use a spatial division into administrative regions, rather than, for instance, a metric system, as a frame of reference for evaluating closeness. Hence, closeness is evaluated purely according to qualitative criteria. Colloquial descriptions typically involve qualitative concepts. Our approach is thus expected to align better with the way human beings deal with closeness than does a quantitative approach.

To illustrate the approach, it is applied to the retrieval of documents from the database of the Datacenter Nature and Landscape (DNL), which stores a number of collections of X- and Y-coordinates for the computation of geometries and the location of protected landscapes and biotopes in Switzerland, attributive data (e.g. textual descriptions), documents (e.g. official correspondence), process data and metadata (Bauer-Messmer et al. 2009).

The article is organized as follows: Section 2 provides an overview of recent work on vague spatial concepts in geographic information science. In Section 3, the vague notion of spatial closeness is introduced into RCC and its implementation in DL (e.g. OWL DL), augmented by DL-safe SWRL rules, is outlined. In Section 4, the approach is applied to the retrieval of documents from the DNL database. Section 5 discusses the approach and Section 6 concludes with a recommendation for future work.

2 Related Work

A great deal of work about vague spatial concepts has been conducted in both philosophy and geographic information science. In this article, we limit our discussion to the most recent research in geographic information science. For a comprehensive survey, particularly of approaches using fuzzy logic or contextual information, the reader is referred to Yao and Thill (2005).
Worboys (2001) describes an experiment with human subjects concerning the vague spatial relation “near” between places in environmental space. An environmental space is referred to as the space consisting of buildings, neighbourhoods and cities, without consideration of symbolic representations such as maps. In order to better understand how humans conceptualize nearness and to test the fit of formal theories to human concepts, the author tried to apply appropriate theories to the data resulting from the experiment. Amongst other insights into the conceptualization of “nearness”, the experiment shows the importance of scale factors introduced by the context of the reference place. This supports our claim made in Section 3 that scales, or more precisely, the categories in which human subjects think, are an important condition of the notion of closeness.

Brennan and Martin (2002) introduce a qualitative representation of spatial proximity that accounts for absolute binary nearness relations. The formalism is based on the notion of perceived points, called sites, in a point-based universe. Proximity concepts are determined by the parameters of distance between two sites and the weight of each of these sites. These parameters are derived from the concept of Generalized Voronoi Diagrams, i.e. Power Diagrams.

The approach introduced by Brennan and Martin (2002) and that presented here have in common that the qualitative description of nearness is based on a qualitative representation of distance: in their case Voronoi diagrams transform (quantitative) distances into a network of (qualitative) topological relations. This is different from all other approaches discussed in this article, where a mapping mechanism between qualitative and metric distance measures is established (or implied). While the authors link their concept of nearness to the topological relations equality, external connectedness and inclusion – which can also be expressed in terms of RCC – this link is established by the areas of influence of perceived points in a point-based universe. Polygons and spatial relations between polygons are not considered. While the approach is appealing because it is formally strict and provides cognitively useful models and interpretations, it does not address the issue of grounding. In particular, it is not clear how the weights \( w(p) \), every site \( p \) is associated with, are obtained from the abstracted “real world” entity.

A statistical approach to context-contingent proximity modelling is described by Yao and Thill (2005). They intend to enable metric systems (such as GIS) to translate between linguistic proximity and metric distance measures. Relevant context factors are chosen that influence, according to empirical studies, the way human beings reason about proximity. The translation mechanism presented works in one direction only: Given the corresponding metric distance measures and context information, linguistic proximity measures are “predicted”. This direction does not support the translation of local prepositions, such as “near” or “far”, used in natural language queries into distance measures processed in metric systems, although this would be very desirable. Dolbear et al. (2007) claim that the information required to bind the numerous context variables may not be available in a practical application, and hence it is difficult to see how they could be implemented on a large scale.

According to Hart and Dolbear (2006), the answer to whether something is near or not depends on the context in which the question is asked and the nature of the objects being compared. Dolbear et al. (2007) use ontologies to make explicit the vague spatial relation “near” for database querying. The algorithm used to calculate the relation “near” is kept relatively simple to make sure it is implementable in a practical system. It only uses two contextual parameters, namely Euclidean distance from a reference point.
and density of the feature class. Despite its simplicity it achieves perfect precision and recall when applied to a (small) number of test sets obtained by asking people which objects were near to each of the reference points. Unlike in the approach presented here, however, the authors base their algorithm on quantitative parameters, namely Euclidean distance and gravity (which is a measure of how objects are distributed), and not on qualitative relations. Furthermore, they reduce the objects considered to centroids and the calculation to point calculation. In the proposed approach, in contrast, the aim is to consider spatial regions (i.e. polygons).

Mata (2007) presents an approach to geographic information retrieval integrating topological, geographical and conceptual matching. For topological matching topological relations are extracted from overlaying data layers; for geographical matching constraints are obtained from dictionaries; for conceptual matching a geographic ontology is used. A constraint, provided as an example, defines two geographic objects (points or polygons) as near provided they are connected by a third object (an arc, e.g. a road), the length of which is less than 1 km. Different from the approach presented here, a metric distance measure thus is a necessary but not a sufficient condition for nearness. However, the framework seems general enough to be aligned with that presented here.

3 Representing Spatio-Thematic Knowledge

3.1 Preliminaries

The presented framework uses a number of spatial relations from different RCC sub-languages, particularly RCC-8, and a composition rule. It also uses the subsumption hierarchy of RCC relations and a sum function as introduced by Randell et al. (1992). These spatial notions are implemented in OWL DL, augmented by DL-safe SWRL rules. We thus assume that the reader is familiar with RCC (Randell et al. 1992, Bennett 2000) and description logics (Baader and Nutt 2003), particularly OWL DL (Patel-Schneider et al. 2004).

As mentioned, the composition rule of the framework is implemented as a DL-safe SWRL rule. DL-safe SWRL rules are function-free Horn rules with the restriction that each variable in the rule occurs in a non-DL-atom in the rule body (Motik et al. 2004). This is ensured by adding special non-DL-literals such as $\mathcal{O}(x)$ to the rule body, and by adding a fact $\mathcal{O}(a)$ for each individual $a$ to the knowledge base. While in theory DL-safe SWRL rules support complex, i.e. disjunctive, heads (or negation in the rule body) (Motik et al. 2006, Motik and Rosati 2007), there is currently no implementation that supports this feature. However, since RCC relations describe a closed world (Randell et al. 1992), it is always possible to replace a negative atom, for instance $\neg \mathrm{disconnected-From}(z, y)$, by a, possibly auxiliary (cf. Section 3.2), positive atom, for instance $\mathrm{connectsWith}(z, y)$.

3.2 Defining Closeness in RCC

A basic assumption underlying our approach is that administrative regions are social artifacts and their organization is largely, if not entirely, motivated by the property of spatial closeness. To be more precise, administrative regions are assumed to mirror how a collective perceives spatial closeness on increasing scales of social organization.
Since administrative regions are typically organized in partitions, it is necessary to introduce the notion of a partition and to reformulate it in a way that is compliant with a model-theoretic interpretation of RCC, the formalism used for expressing closeness. Compliance with this kind of interpretation is a requirement for the implementation of RCC in a DL knowledge base and in the rule base introduced in Section 3.3. Further, in order to assert closeness between individual regions, we must extend RCC and introduce closeness by an additional relation. The idea is that, given the conceptualization of a user in terms of a query and a partially ordered and typed system of partitions, closeness can be evaluated by a composition rule.

**Definition 1a (Partition).** A partition is defined as a (possibly improper) subset of the power set of a set $Y$, denoted by $(Y_i)_{i \in I} \subseteq \mathcal{P}(Y)$, for which holds

- $Y = \bigcup_{i \in I} Y_i$, where $I$ is a finite index set;
- $Y_i \cap Y_j = \emptyset$ for $i \neq j$;
- $Y_i \neq \emptyset$ for all $i \in I$.

In this definition, $Y_i$ and $Y$ refer to sets of points in a point-based universe. As mentioned, in order to be compatible with the model-theoretic semantics of DL, we use a non-standard interpretation where regions are interpreted as individuals, and not as sets, in an abstract domain. We thus reformulate the definition using the Boolean RCC function $\text{SUM}$ and the RCC relation $\text{DR}$ (i.e. “discrete from”). As is customary, we use lower case letters for variables denoting individuals.

**Definition 1b (Partition in RCC).** A family of regions $(x_i)_{i \in I}$ is a partition of a region $y$ if the following holds:

- $y = \text{SUM}_{i \in I} x_i$, where $I$ is a finite index set;
- $\forall x_i, \forall y_j : \text{DR}(x_i, y_j)$ for $i \neq j$;
- regions $(x_i)_{i \in I}$ are named for all $i \in I$.

We only consider partitions where the elements are typed by kind of administrative region, for instance, $\text{Community}(x_i)$ says that $x_i$ is of type Community. Multiple typing of regions is not allowed, that is, the concepts used for typing are mutually disjoint. Similarly, a given type is used for a single partition only. This allows distinguishing the partitions by their types.

In order to account for the different scales of social organization we define a partial order on the system of partitions in RCC by comparing partitions with regard to their granularity.

**Definition 2 (Partial Order on Typed Partitions in RCC).** Let $C(x_i)_{i \in I}$ and $D(y_j)_{j \in J}$ be partitions of the same region of types $C$ and $D$, respectively. We say that $C(x_i)_{i \in I}$ is more fine-grained than $D(y_j)_{j \in J}$, denoted by $C(x_i)_{i \in I} \preceq D(y_j)_{j \in J}$, if each element of $C(x_i)_{i \in I}$ is a (possibly improper) subset of an element of $D(y_j)_{j \in J}$. A partial order on typed partitions is reflexive, transitive and antisymmetric.

This means that each element of $D(y_j)_{j \in J}$ is partitioned by elements of $C(x_i)_{i \in I}$. For instance, $\text{Community}(x_i)_{i \in I}$ and $\text{District}(y_j)_{j \in J}$ are typed partitions of a canton and each element of $\text{District}(y_j)_{j \in J}$ is partitioned by elements of $\text{Community}(x_i)_{i \in I}$.

**Definition 3 (Minimal Partial Order on Typed Partitions in RCC).** We say that a partial order on typed partitions is minimal with regard to a given conceptualization, denoted by...
If the conceptualization does not provide a type for any $(w_k)_{k \in K}$ such that $C(x_k)_{x_k \leq} \leq (w_k)_{w_k \leq} \leq D(y_j)_{y_j \geq}$, A minimal partial order on typed partitions is intransitive. For instance, if a given conceptualization provides the administrative types District and Community, any partial order comprising a non-typed partition of intermediate granularity is not minimal.

For asserting closeness between individual regions we extend RCC and introduce the relation $CL(x, y)$ which is read as "$x$ is close to $y". In accordance with empirical evidence (Worboys 2001), closeness is introduced as a weakly asymmetrical relation. This means that the relation is symmetrical, if $x$ and $y$ are members of the same partition, but asymmetrical, if $y$ is a member of a more fine-grained partition than $x$ or else, if $x$ is a non-administrative region.$^6$

Definition 4 (Closeness in RCC). Given a region $x_i$ of a partition used as a referent in a query, a type $C$ of a conceptualization for $x_i$ and a minimal partial order on typed partitions $C(x_i)_{x_i \leq} \leq D(y_j)_{y_j \geq}$ closeness in RCC can be inferred by the composition rule $\forall x, y (P(x_i, y_j) \land XG(z, y_j) \rightarrow CL(z, x_i))$.

In this definition, $XG(z, y_j)$, read as "exclusively connects with", is an auxiliary relation. Its main purpose is to prevent the transitive property of $P(x_i, y_j)$, which has been overridden by the definition of a minimal partial order on typed partitions, from being reintroduced through the backdoor of the composition rule. Note that since the relation is directed from $z$ to $y_j$, transitivity is excluded by removing $P(z, y_j)$ (i.e. "inverse part of") and its subrelations from $G(z, y)$.

Definition (4) shows that closeness depends on the type of region used as a referent in a query; hence, on the way a user conceptualizes a domain. This includes the scale on which spatial relations are to be evaluated. It also depends on partitions into administrative regions reflecting how a collective perceives spatial closeness on increasing scales of social organization. As a result of this dependency, $CL(z, x_i)$ is undefined unless it is related to a minimal partial order on typed partitions.

3.3 A DL Knowledge Base and Rule Base for RCC

The knowledge required to answer (possibly vague) spatio-thematic queries can be represented by a DL knowledge base $KB$ consisting of a TBox $T$ and ABox $A$, $KB = \{T, A\}$, and by a rule base $RB$ for DL-safe SWRL rules.

Among other things $T$ contains a number of concept inclusion axioms that introduce kinds of regions. It is worth recalling the definition of an ontology as an explicit, formal specification of a shared conceptualization (Gruber 1993). Accordingly, the introduced categories are not arbitrary. They are social artifacts and reflect how a collective thinks that the world (or a piece thereof) is structured. In the long run, a collectively shared conceptualization is furthermore not invariant but evolves together with the development of a society and a country.

In order to implement RCC in DL, the subsumption hierarchy of RCC relations (Randell et al. 1992) is represented as a hierarchy of binary role inclusion axioms in $T$. The RCC relation $P(x_i, y_j)$ and its subrelations are implemented as functional roles, thereby ensuring that an individual $x_i$ can be part of a single region $y_j$ only. This overrides the transitivity of the RCC relation $P(x, y)$, which prevents, for instance, communities to be related to cantons (or to countries or continents if these were represented). Partitions are represented in $T$ by (anonymous) concepts that are made up of individual names, also

© 2010 Blackwell Publishing Ltd
Transactions in GIS, 2010, 14(2)
called *nominals*, \(\{x_1, \ldots, x_n\}\). Nominals are linked to types by concept inclusion axioms of the form \(C \subseteq \{x_1, \ldots, x_n\}\) stating that the set of individuals in the interpretation of \(C\) is a (possibly improper) subset of the individuals in the interpretation of \(\{x_1, \ldots, x_n\}\). In order to disallow multiple typing the concepts used for typing are defined as mutually disjoint, \(C \subseteq \neg D\).

In order to populate \(A\), known RCC relations between individual regions are asserted as role assertions. Particularly, partitions are asserted as \(\text{partOf}(x_i, y_i)\), or any of its subrelations, for all applicable \(x_i \in \{x_1, \ldots, x_n\}\) and \(y_i \in \{y_1, \ldots, y_m\}\). In so doing, \(A\) is closed with regard to nominals denoting administrative regions.<sup>7</sup> A minimal partial order on typed partitions is implemented by asserting \(\text{partOf}(x_i, y_i)\), or any of its subrelations, exclusively for those pairs of individuals \((x_i, y_i)\) for which hold \(C(x_i) \subseteq \min D(y_i)\). \(A\) also contains facts about individual regions in terms of concept assertions.

The composition rule \(\forall x \forall y \forall z [P(x, y) \land XC(z, y) \rightarrow CL(z, x)\] is implemented in \(RB\) by the DL-safe SWRL rule \(\text{partOf}(x, y) \land \text{exclusivelyConnectsWith}(z, y) \land O(x) \land O(y) \land O(z) \rightarrow \text{closeTo}(z, x)\) where \(O(x), O(y)\) and \(O(z)\) are non-DL-literals.<sup>8</sup> In order to make the rule DL-safe, a fact \(O(a)\) is asserted for each individual \(a\) in the ABox. The rule is read as “A region \(z\) is close to a region \(x\) if \(x\) is part of a region \(y\) and \(z\) exclusively connects with \(y\) where the identity of all regions is known.”

### 3.4 Processing Vague Spatio-Thematic Queries

The concepts implicitly and explicitly used in a query reveal how a user conceptualizes a domain. Thereby the user is assumed to be a member of the social collective in question. Query concepts can be used to determine the scale on which closeness is to be evaluated. They translate what is often referred to as the context of a vague concept, such as closeness, from contingencies in the real world into linguistic constraints. We assume a query of the form \(\forall z [Q(z) \land CL(z, a)]\) which is expected to return the set of those individuals of type \(Q\) that are close to a given individual \(a\) of a partition. In this query, the type of individual \(a\), for instance \(C(a)\), sets the scale for the evaluation of closeness.

**Algorithm 1** Function \text{CLOSETO} computes \((Q \sqcap \exists \text{closeTo}\{a\})(z)\) from \(KB\) and \(RB\).

**FUNCTION CLOSETO**

**INPUT:** Knowledge Base \(KB = \{T, A\}\), Rule Base \(RB\), Concept \(Q\), Individual \(a\)

**OUTPUT:** Set<Individual>

1. \(\{b\} \leftarrow \{b \mid A \models \text{partOf}(a, b)\}\)
2. \(V \leftarrow \{v_i \in I \mid A \models \text{exclusivelyConnectsWith}(v_i, b)\}\)
3. \(W \leftarrow \{w_j \in J \mid A \models Q(w_j)\}\)
4. \(Z \leftarrow V \cap W\)
5. **OUTPUT** \(Z\)

The query \(\forall z [Q(z) \land CL(z, a)]\) is implemented in DL by the concept description \(Q \sqcap \exists \text{closeTo}\{a\}\). Given an ABox \(A\) and a concept description \(Q \sqcap \exists \text{closeTo}\{a\}\), the retrieval problem is thus to find all individuals \(z\) in \(A\) such that \(A \models (Q \sqcap \exists \text{closeTo}\{a\})(z)\). Algorithm 1 shows the steps (1–5) to take when processing a query. Note that in order to process steps 1 and 2 the composition rule is required.
4 Applying the Approach to an Environmental Database

As stated in Section 3, we assume that administrative regions mirror how a collective perceives spatial closeness on increasing scales of social organization. In order to verify our assumption, it is necessary to recall the design principles underlying the organization into administrative units. For example, the following official statement clarifies the purpose of districts as administrative units:

“The administrative districts . . . perform decentralized administrative tasks of the cantons, particularly in the areas of health (district hospitals, public health), to some extent education (district schools), judiciary (district courts) and general administration (taxation, business failures, etc.). Furthermore, in several cantons the administrative districts correspond to the electoral wards.” (Bundesamt für Statistik 2008)

The statement implies that the organization of districts is largely, if not entirely, motivated by the spatial closeness of the communities. District hospitals, for instance, are decentralized entities of the health care system. They might have been established with the intention of keeping the distance between the patients (living in the communities) and the care providers short. This obviously reflects the experience of human subjects who perceive this distance as short. Similar arguments apply to district schools, district courts and electoral wards. It is, therefore, reasonable to claim that the organization of districts is motivated by the property of spatial closeness. The organization of other administrative regions can be motivated in a similar way.

The application proceeds in two steps. In a first step, queries of the form (Landscape closeTo(x))(z) are processed using a knowledge base, a rule base and the algorithm described in Sections 3.3 and 3.4, respectively. The variable x stands for the name of a community; the variable z binds landscapes that are close to x.

In a second step, the names of the returned landscapes, for instance, “Albiskette-Reppischtal” (No. 1306 in Figure 1), are used as terms for searches for textually indexed documents in the DNL database (cf. Section 1). The results are compared to those from searches in the same database using the strings <Landschaften "in der Nähe von" ?x> (i.e. landscapes close to x, where x is the name of a community). Thus, the idea is to compare the results from two conceptually (although not syntactically) identical database searches, the first without, the second with spatio-thematic query pre-processing. Table 1 shows the results of the query pre-processing.

The knowledge required to answer the queries is represented in a consistent DL knowledge base $\mathcal{KB} = \{\mathcal{T}, \mathcal{A}\}$ and a DL-safe SWRL rule base $\mathcal{RB}$. The description language used for the $\mathcal{KB}$ is OWL DL (Patel-Schneider et al. 2004); the DL expressivity is $\text{ALCHOIF}$. It is clear that the representation of relevant knowledge is not sufficient. In order to answer the queries, an engine must be able to make use of such knowledge. We used Pellet 2.0 in order to process the queries. Pellet 2.0 is a DL reasoner that integrates a SPARQL query engine and a rule engine for the processing of DL-safe SWRL rules (see http://clarkparisa.com/pellet for additional details). It can handle all aspects of the introduced framework.

Searches in the DNL database make use of Oracle Text (version 10.2.0.3). The database stores a total of 31,645 documents. The total numbers of relevant documents in the database were counted by manually sorting through the documents of the collect-
Table 1  Landscapes close to a community in each of the twelve districts of the Canton of Zurich (results from query pre-processing)

| Community | District          | Landscapes (Nos.) |
|-----------|------------------|-------------------|
| Adliswil  | Bezirk Horgen    | 1306, 1307        |
| Affoltern | Bezirk Affoltern | 1305, 1306        |
| Bachs     | Bezirk Dielsdorf | 1011, 1404        |
| Dietikon  | Bezirk Dietikon  | 1306               |
| Dorf      | Bezirk Andelfingen | 1403, 1410, 1411, 1412 |
| Egg       | Bezirk Uster     | 1408               |
| Meilen    | Bezirk Meilen    | 1417, 1419        |
| Pfäffikon | Bezirk Pfäffikon | 1409, 1420        |
| Wald      | Bezirk Hinwil    | 1401, 1409, 1420  |
| Winkel    | Bezirk Bülach    | 1404, 1410, 1411  |
| Winterthur| Bezirk Winterthur| 1410, 1420        |
| Zürich    | Bezirk Zürich    | 1306, 1407        |

Figure 1  Landscapes (shaded and numbered) in Canton of Zurich (bold borderline)
in some cases, none are considered relevant when checking them manually (column “Control” in the table). Accordingly, recall and precision of these searches are either zero or undefined. Conversely, searches using the results from knowledge/rule base queries such as <Albiskette-Reppischtal> as inputs return relevant matches in all cases but one (column “Test” in the table). Recall ranges between 0 and 0.10 (mean 0.06) and precision between 0 and 1 (mean 0.75). Nine out of 12 searches are located in the quadrant of the recall × precision matrix that is far from the recall axis and close to the precision axis (not shown). According to Salton and McGill (1983), this characterizes narrow searches put in specific terms.

### 5 Discussion

Closeness is a vague concept, in the sense that borderline cases exist for which it is difficult to decide whether they are covered by the concept or not (Worboys 2001). While our approach takes a Boolean decision on closeness, it still accounts for borderline cases by using a qualitative formalism. Whether a region is close to another region or not depends on the size and shape of the administrative units serving as a frame of reference. When comparing the evaluation of closeness, even on the same scale of social organization, a given metric distance may in one case be interpreted as close and in another case as not close. Our claim is that the size and shape of administrative regions are not arbitrary but reflect how a collective perceives spatial closeness on increasing scales of social organization. Whether our claim is empirically well founded or not remains to be seen.

The concept of closeness evolves over time. What is perceived as close by the members of a social collective (at least in the industrialized countries) has been subject to change for decades. Similarly, at an institutional level, the concept of closeness evolves.
In recent years, several cantons in Switzerland, for instance, have revised their administrative structures or established a legal basis for future revisions. One result of these revisions is a reduction in the number of districts. It is important to note that societal change precedes institutional change. If this was not the case, proposals for structural revisions would not obtain a majority of popular votes. Since our approach evaluates closeness within the frame of administrative structures and institutional change lags behind societal change, it tends to underestimate closeness. While our approach takes into account evolution of closeness, it does so with a view to the slow pace of the institutions and not to the fast pace of the social collective.

Recall of the database searches is low, even for the searches with query pre-processing (column “Test” in Table 2). The reason for this is that the database stores a lot of scanned documents, which are not indexed. Recall should be improved by making these scans accessible using Optical Character Recognition (OCR) methods and re-creating the text index. The reason why the precision of three searches with query pre-processing is also low is that some landscape names (e.g. “Pfäffikersee”, No. 1409 in Figure 1) are not unique. They are shared by different kinds of objects (e.g. moorlands). Precision should thus be improved by making the kind of objects searched for explicit.

6 Conclusions

In this article, an implementation of RCC in OWL DL, augmented by DL-safe SWRL rules, is used to represent spatio-thematic knowledge. We show how such a representation can be operated on in order to answer queries using (possibly vague) spatial concepts. Accordingly, a spatial division into administrative regions rather than, for instance, a metric system is used as a frame of reference for evaluating closeness. Hence, closeness is evaluated purely according to qualitative criteria. This is expected to align better with the way human beings deal with closeness than does a quantitative approach. The approach is applied to document retrieval from a database on protected landscapes and biotopes in Switzerland.

So far the approach presented here supports the evaluation of closeness of regions with regard to an administrative region. An evaluation of closeness between arbitrary regions would be desirable. Exploring whether and how the frame of reference can be leveraged to support evaluation of closeness between arbitrary regions is left to future work. Likewise, the scalability of the implementation and, possibly, alternative implementation strategies remain to be explored.

The article only considers the concept of closeness. There are a number of additional vague spatial concepts such as “near”, “next to”, “a short distance outside”, “a long way off”, and “far away from”. It would be interesting to formalize these concepts in a way similar to that demonstrated for “close to”. Such a formalization might result in a theory of vague spatial concepts in RCC, which could be implemented, for instance, in OWL DL, augmented by DL-safe SWRL rules.

Acknowledgments

The authors sincerely thank Jürg Schenker (FOEN) and Martin Hägeli (WSL) for the fruitful discussions and leadership that made this research possible. This research has been funded by and conducted in cooperation with the Swiss Federal Office for the Environment (FOEN).
Endnotes

1 In the literature the term “near” is often used to denote the same concept as “close to”.
2 Cf. “In the discussion of proximity perception and cognition, the term ‘scale’ refers to the size of the frame of reference in a perceiver’s mental map, i.e. the spatial extent of the area that is considered when a distance is assessed” (Yao and Thill 2005, p. 162).
3 An approach similar to RCC which is based on the description of topological relations between two spatial regions was introduced as the 9-intersection model by Egenhofer and Franzosa (1991).
4 $\sum_{i \in X} x_i$ is defined as $\forall z [C(z, y) \leftrightarrow \bigvee_{i \in X} C(z, x_i)]$ (Randell et al., 1992). $C$ stands for “connects with”.
5 This implies $\forall x_i P(x_i, y)$. $P$ stands for “part of”.
6 Worboys (2001) argues that for nearness the subject-referent dichotomy plays a dominant role in that the referent creates the scale in which the relation has context.
7 Note that we use partOf$(x_i, y_j)$ and its subrelations only for asserting partitions into administrative regions.
8 In DL-safe SWRL rules, all variables are universally quantified.
9 Further queries that are successfully processed include communities, districts or landscapes adjacent to a given community or district; communities close to a given community; districts close to a given community or district; landscapes close to a given district.
10 Cf. http://www.wsl.ch/personal_homepages/scharren/download/tgis_2009.owl; please contact the authors if you wish to access the Web page and it is no longer maintained.
11 This might be a peculiarity of the Swiss political system.

References

Baader F and Nutt W 2003 Basic description logics. In Baader F, Calvanese D, McGuinness D L, Nardi D, and Patel-Schneider P F (eds) The Description Logic Handbook. New York, Cambridge University Press: 47-100

Bauer-Messmer B, Wotruba L, Müller K, Bischof S, Grütter R, Scharrenbach T, Meile R, Hägeli M, and Schenker J 2009 The Data Centre Nature and Landscape (DNL): Service oriented architecture, metadata standards and semantic technologies in an environmental information system. In Wohlgemuth V, Page B, and Voigt K (eds) Environmental Informatics and Industrial Environmental Protection: Concepts, Methods and Tools: Proceedings of the Twenty-third International Conference on Informatics for Environmental Protection (EnviroInfo 2009). Aachen, Shaker Verlag: 101-12

Bennett B 2000 Logics for topological reasoning. In Proceedings of the Twelfth European Summer School in Logic, Language and Information, Birmingham, United Kingdom (available at http://www.scs.leeds.ac.uk/brandon/ESLLI-2000-notes.ps)

Berners-Lee T 1998 What the semantic web can represent. WWW document, http://www.w3.org/DesignIssues/RDFnot.html

Bishr Y 2006 Geospatial semantic web. In Rana S and Sharma J (eds) Frontiers of Geographic Information Technology. Berlin, Springer-Verlag: 139-54

Brennan J and Martin E 2002 Foundations for a formalism of nearness. In McKay R and Slaney J (eds) AI 2002: Advances in Artificial Intelligence. Berlin, Springer-Verlag Lecture Notes in Computer Science Vol. 2557: 71–82

Bundesamt für Statistik 2008 Institutionelle Gliederungen der Schweiz: Schweiz – die Bezirke. WWW document, http://www.bfs.admin.ch/bfs/portal/de/index/regionen/11/geo/institutionelle_gliederungen/01a.html

Dolbear C, Hart G, and Goodwin J 2007 From theory to query: Using ontologies to make explicit imprecise spatial relationships for database querying. WWW document, http://www.ordnancesurvey.co.uk/oswebsite/partnerships/research/publications/docs/2007/COSTT_07.pdf

Egenhofer M and Franzosa R 1991 Point-set topological spatial relations. International Journal of Geographical Information Systems 5: 161-74
Fonseca F, Egenhofer M, Agouris P, and Câmara, C 2002 Using ontologies for integrated information systems. *Transactions in GIS* 3: 231-57

Gruber T R 1993 A translation approach to portable ontology specifications. *Knowledge Acquisition* 5: 199-220

Grüter R and Bauer-Messmer B 2007 Towards spatial reasoning in the semantic web: A hybrid knowledge representation system architecture. In Fabrikant S and Wachowicz M (eds) *The European Information Society: Leading the Way with Geo-information*. Berlin, Springer-Verlag Lecture Notes in Geoinformation and Cartography: 349-64

Haarslev V, Lutz C, and Möller R 1998 Foundations of spatio-terminological reasoning with Description Logics. In Cohn A G, Schubert L K, and Shapiro S C (eds) *Proceedings of the Sixth International Conference on Principles of Knowledge Representation and Reasoning (KR'98)*. San Mateo, CA, Morgan Kaufmann: 112–23

Hart G and Dolbear C 2006 What’s so special about spatial? In Scharl A and Tochtermann K (eds) *The Geospatial Web: How Geobrowsers, Social Software and the Web 2.0 are Shaping the Network Society*. London, Springer-Verlag: 39-44

Jones C B, Purves R, Ruas A, Sanderson M, Sester M, van Kreveld M, and Weibel R 2002 Spatial information retrieval and geographical ontologies: An overview of the SPIRIT project. In *Proceedings of the Twenty-fifth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. Tampere, Finland: 387-8

Lutz C and Miličić M 2007 A Tableau algorithm for Description Logics with concrete domains and General TBoxes. *Journal of Automated Reasoning* 38: 227-59

Mata F 2007 Geographic information retrieval by topological, geographical, and conceptual matching. In Fonseca F, Rodríguez M A, and Levashkin S (eds) *Geospatial Semantics*. Berlin, Springer-Verlag Lecture Notes in Computer Science Vol. 4853: 98-113

Möller R and Wessel M 2007 Flexible software architectures for ontology-based information systems. *Journal of Applied Logic* 7: 75-99

Motik B and Rosati R 2007 A faithful integration of Description Logics with Logic Programming. In Veloso M M (ed) *Proceedings of the Twentieth International Joint Conference on Artificial Intelligence (IJCAI 2007)*. San Mateo, CA, Morgan Kaufmann: 477–82

Patel-Schneider P F, Hayes P, and Horrocks I 2004 Query answering for OWL-DL with rules. In McIlraith S A, Plexousakis D, and van Harmelen F (eds) *The SemanticWeb: ISWC 2004*. Berlin, Springer-Verlag Lecture Notes in Computer Science Vol. 3298: 549-63

Randell D A, Cui Z, and Cohen A G 1992 A Spatial Logic based on Regions and Connections. In Nebel B, Rich C, and Swartout W (eds) *Proceedings of the Third International Conference on Principles of Knowledge Representation and Reasoning (KR’92)*. San Mateo, CA Morgan Kaufmann: 165-76

Salton G and McGill M J 1983 *Introduction to Modern Information Retrieval*. New York, McGraw-Hill

Wache H, Vögele T, Visser U, Stuckenschmidt H, Schuster G, Neumann H, and Hübner S 2001 Ontology-based integration of information: A survey of existing approaches. In *Proceedings of the IJCAI-01 Workshop on Ontologies and Information Sharing*. Seattle, Washington: 108-18

Worboys M F 2001 Nearness relations in environmental space. *International Journal of Geographical Information Science* 15: 633-51

Yao X and Thill J-C 2005 How far is too far?: A statistical approach to context-contingent proximity modeling. *Transactions in GIS* 9: 157-78