ed D Abel and B C Ool, Springer Verlag, Berlin, pp 290-313, 1993

Qualitative and Topological Relationships

Z Cui 1, A G Cohn 2 and D A Randell 3

Advanced Computation Laboratory, Imperial Cancer Research Fund London WC2A 3PX, England { cui@acl.lif.icnet.uk}
 2 School of Computer Studies, University of Leeds, LS2 9JT England {agc@scs.leeds.ac.uk}
 3 School of Dentistry, The University of Birmingham Birmingham B4 6NN, England { randelda@ibm3090.bham.ac.uk}

Abstract. In this paper, we present a spatial logic which can be used to reason about topological and spatial relationships among objects in spatial databases. The main advantages of such a formalism are its rigorousness, clear semantics and sound inference mechanism. We also show how the formalism can be extended to include orientation and metrical information. Comparisons with other formalisms are discussed.

1 Introduction

A formal theory of space and time has always been an important issue in Artificial Intelligence. Recently, its importance in spatial databases has been recognized (Egenhofer 1989, 1991, Egenhofer and Herring 1990, and Pullar and Egenhofer 1988). Advances of database technology have required a database not only to store, to retrieve and to update data, but also to reason about the relationships among its data, and to have production rules and triggers. A formal theory of data models is important in securing the data consistency in such situations. A deductive database should allow its users to formulate complex queries based on simple facts (relations), otherwise it is difficult to meet the requirements of many applications.

Geographic information systems, image data bases, and CAD/CAM systems are often based upon the relationships among spatial objects. Although some query languages support queries with some spatial relationships; however, the diversity, semantics, completeness and terminology of these relationships vary dramatically (eg Egenhofer and Frank 1988, Roussopoulos, Faloutsos and Sellis 1988, Guenther and Buchmann 1990, Güting 1988). In general, the underlying basis of most existing spatial databases seems to be that of point set geometry, perhaps with some application specific ontology in addition (eg Rawlings 1985).

However, many explanations of phenomena and descriptions of the relationship between objects in informal discourse appeal to relatively high level qualitative spatial information, in particular, topological information. Much of this appears to be done unconsciously, but little reflection on our use of language

^{*} This work has been partially funded by the SERC under grant no. GR/G36852

(particularly the use given to prepositions and prepositional phrases) reveals how important this information is. While we need not assume that linguistic descriptions necessarily uncover those entities represented and exploited by the brain in all such activity (eg. in the encoding and representation of perceptual information), the design of formal spatial query languages which mirror the ontology used in informal discourse may ease the use of such a language. Thus there is a need to develop a unified theory on topological relations. Egenhofer (1991, 1989) has proposed a topological relationships based on point set combinatorical topology. However, it relies on relatively sophisticated mathematics concepts such as open and closed regions. The main purpose of this paper is to present a higher level, axiomatic approach (Randell and Cohn 1989, Randell et al 1992) to representing and reasoning about qualitative (including topological) spatial information and to discuss the application of the language to spatial databases. A distinguishing feature of our approach is that our basic ontology is that of a region, thus abstracting away from point set semantics, which may indeed be a model for our formalism but is not presupposed.

The other main advantages of this approach are its logical rigorousness and its foundation in first order predicate calculus allowing well investigated inference rules and many different theorem proving technologies to be readily used to prove theorems, make inferences and test consistency and constraints of the databases. Moreover the ontological commitments required are few: two basic primitive notions allow an arbitrarily complex taxonomy of qualitative spatial relationships and concepts to be defined (Cohn, Randell and Cui 1993).

The remainder of this paper is structured as follows: first we present the extant formalism and discuss how the calculus can be easily used to check database consistency and constraints. Then we show how the formalism can be extended to incorporate notions of orientation and direction. Related works are discussed, mainly in the context of the formalisms developed by Egenhofer (1991) and Freksa (1991); finally we mention some current and future work and summarise the paper.

2 The Extant Formalism

The basic ontological entity we consider⁴ is a region; note that boundaries, lines and points are not regions.⁵. Regions are non empty. Regions in the theory support either a spatial or temporal interpretation, though we will only consider the spatial interpretation here. Informally, these regions may be thought to be potentially infinite in number, and any degree of connection between them is

⁴ Most of the material in this section can be found in (Randell, Cui and Cohn 1992) but we review it here for convenience and to make this paper more self contained. Previously we have also considered a temporal interpretation of the formalism (Randell and Cohn 1989, 1992) but we concentrate on the spatial case here.

⁵ However we believe that from a modelling point of view at least for commonsense reasoning, such mathematical abstractions are not necessary and one can use special kinds of regions such as *skins* and *atoms* – see (Randell, Cui and Cohn 1992)

allowed in the intended model, from external contact to identity in terms of mutually shared parts. The formalism supports two or three dimensional interpretations (or higher dimensions!) and is based upon Clarke's (1981, 1985) calculus of individuals based on "connection"; it is expressed in the many sorted logic LLAMA (Cohn 1987).⁶

The basic part of the formalism assumes one primitive dyadic relation: C(x, y) read as 'x connects with y'. The relation C(x, y) is reflexive and symmetric. We can give a topological model to interpret the theory, namely that C(x, y) holds when the topological closures of regions x and y share a common point. Two axioms are introduced.

```
\begin{array}{l} (1) \ \forall x \mathcal{C}(x,x) \\ (2) \ \forall xy [\mathcal{C}(x,y) \to \mathcal{C}(y,x)] \end{array}
```

Using C(x,y), a basic set of dyadic relations are defined: 'DC(x,y)' ('x is disconnected from y'), 'P(x,y)' ('x is a part of y'), 'PP(x,y) ('x is a proper part of y'), 'x = y' ('x is identical with y'), 'O(x,y)' ('x overlaps y'), 'DR(x,y)' ('x is discrete from y') 'PO(x,y)' ('x partially overlaps y'), 'EC(x,y)' ('x is externally connected with y)', 'TPP(x,y)' ('x is a tangential proper part of y') and 'NTPP(x,y)' ('x is a nontangential proper part of y'). The relations: P,PP,TPP and NTPP being non-symmetrical support inverses. For the inverses we use the notation Φ^{-1} , where $\Phi \in \{P,PP,TPP$ and NTPP $\}$. In order to save space we will not give the definitions for any of the inverse predicates as they are all of the form $\Phi^{-1}(x,y) \equiv_{def} \Phi(y,x)$. Of the defined relations, DC,EC,PO,=,TPP,NTPP and the inverses for TPP and NTPP are provably mutually exhaustive and pairwise disjoint.

```
(3) DC(x, y) \equiv_{def} \neg C(x, y)

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

(5) PP(x, y) \equiv_{def} P(x, y) \land \neg P(y, x)

(6) x = y \equiv_{def} P(x, y) \land P(y, x)

(7) O(x, y) \equiv_{def} \exists z[P(z, x) \land P(z, y)]

(8) PO(x, y) \equiv_{def} O(x, y) \land \neg P(x, y) \land \neg P(y, x)

(9) DR(x, y) \equiv_{def} \neg O(x, y)

(3) TPP(x, y) \equiv_{def} PP(x, y) \land \exists z[EC(z, x) \land EC(z, y)]

(10) EC(x, y) \equiv_{def} C(x, y) \land \neg O(x, y)

(11) NTPP(x, y) \equiv_{def} PP(x, y) \land \neg \exists z[EC(z, x) \land EC(z, y)]
```

A pictorial representation of the relations defined above is given in Figure 1.

⁶ Although we use a sorted logic, for the most part this need not concern us here; important sortal restrictions will be mentioned as appropriate.

⁷ In Clarke's theory and in our original theory (Randell and Cohn 1989, 1992), when two regions x and y connect, they are said to share a point in common; thus the interpretation of the connects relation here and in (Randell, Cui and Cohn 1992) is weaker. Alternative models for the C relation not relying explicitly on point set semantics are that it is true when the distance between the two regions is zero, or that no other region can be 'fitted' between them.

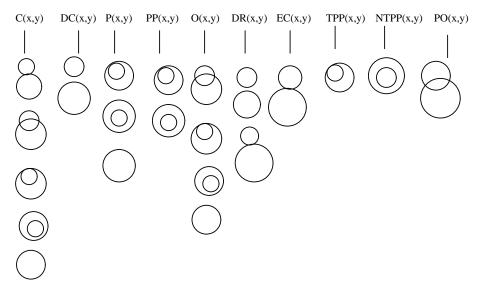


Fig. 1. A set of sample configurations (in 2D) modelling the defined relations.

The Boolean functions⁸ are: 'sum(x,y)' which is read as 'the sum of x and y', 'us' as 'the universal (spatial) region', 'compl(x)' as 'the complement of x', 'prod(x,y)' as 'the product (i.e. the intersection of x and y' and 'diff(x,y)' as 'the difference of x and y'. The functions: 'compl(x)', 'prod(x,y)' and 'diff(x,y)' are partial but are made total in the sorted logic by simply specifying sorts restrictions and by introducing a new sort called NULL. The sorts NULL and REGION are disjoint.

```
\begin{aligned} & \operatorname{sum}(x,y) =_{def} \iota y [\forall z [\operatorname{C}(z,y) \leftrightarrow [\operatorname{C}(z,x) \vee \operatorname{C}(z,y)]]] \\ & \operatorname{compl}(x) =_{def} \iota y [\forall z [[\operatorname{C}(z,y) \leftrightarrow \neg \operatorname{NTPP}(z,x)] \wedge [\operatorname{O}(z,y) \leftrightarrow \neg \operatorname{P}(z,x)]]] \\ & \operatorname{us} =_{def} \iota y [\forall z [\operatorname{C}(z,y)]] \\ & \operatorname{prod}(x,y) =_{def} \iota z [\forall u [\operatorname{C}(u,z) \leftrightarrow \exists v [\operatorname{P}(v,x) \wedge \operatorname{P}(v,y) \wedge \operatorname{C}(u,v)]]] \\ & \operatorname{diff}(x,y) =_{def} \iota w [\forall z [\operatorname{C}(z,w) \leftrightarrow \operatorname{C}(z,\operatorname{prod}(x,\operatorname{compl}(y)))]] \\ & \forall xy [\operatorname{NULL}(\operatorname{prod}(x,y)) \leftrightarrow \operatorname{DR}(x,y)] \end{aligned}
```

An additional axiom is also required which stipulates that every region has a nontangential proper part:⁹

(12) $\forall x \exists y [NTPP(y, x)]$

⁸ $\alpha(\overline{x}) =_{def} \iota y[\Phi[\alpha(\overline{y})] \text{ means } \forall \overline{x}[\Phi(\alpha(\overline{x})]]; \text{ thus, e.g., the definition for prod}(x, y) \text{ is translated out (in the object language) as: } \forall xyz[C(z, \operatorname{prod}(x, y)) \leftrightarrow \exists w[P(w, x) \land P(w, y) \land C(z, w)]].$

⁹ A consequence of this axiom is that there can be no *atomic* regions, i.e. regions which contain no subparts. For a discussion of how such regions can be introduced into the language, see Randell, Cui and Cohn (1992).

This axiom mirrors a formal property of Clarke's theory, where he stipulates that every region has a nontangential part, and thus an interior (remembering that in Clarke's theory a topological interpretation is assumed).

2.1 One piece regions

Clarke's theory supports a model where regions may topologically connected (i.e. in one piece) or disconnected (in more than one piece). This naturally arises given the above definitions for Boolean functions: the sum of two regions will be disjoint unless they are connected. Such scattered regions may be used to model, for example, a cup broken into several pieces. The definition CON(x) (x is connected one piece region) simply states that an individual region is connected if it cannot be split into parts whose union is that region, and where these parts are not connected to each other.

$$CON(x) \equiv_{def} \forall yz [sum(y, z) = x \rightarrow C(y, z)]$$

2.2 Concavity and Convexity

A primitive function 'conv(x)' ('the convex-hull of x') is defined and axiomatised.¹⁰ Informally this function generates the region of space that would arise by completely enclosing a body in a taught 'cling film' membrane. This function provides a very intuitive notion for describing objects that may be considered inside, partially inside or outside another object without forming part of that object. Figure 2 gives sample configurations. We also can define a predicate CONV(x) which is true for convex regions.

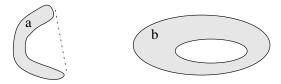


Fig. 2. Illustrations of the convex hull: a is a region (shaded area), its convex-hull is the area enclosed by outer line including the dotted line; b is the shaded area and its convex-hull is the area enclosed by the outer oval.

¹⁰ The fourth of these axioms is different to our previous publication (Randell, Cui and Cohn 1992), as we have recently discovered a counterexample to the old axiom – originally the axiom was $\forall x \forall y [[P(x, \operatorname{conv}(y)) \land P(y, \operatorname{conv}(x))] \rightarrow O(x, y)]$. Also, it should be noted that whereas we previously assumed (eg Randell, Cui and Cohn 1992) that conv is only well sorted when defined on one piece regions, we have now dropped this restriction since on consideration the axioms for conv are also clearly true for non one piece regions.

```
\begin{array}{l} \forall x \mathbf{P}(x, \operatorname{conv}(x)) \\ \forall x \mathbf{P}(\operatorname{conv}(\operatorname{conv}(x)), \operatorname{conv}(x)) \\ \forall x \forall y \forall z [[\mathbf{P}(x, \operatorname{conv}(y) \land \mathbf{P}(y, \operatorname{conv}(z))] \to \mathbf{P}(x, \operatorname{conv}(z))] \\ \forall x \forall y [[\mathbf{P}(x, \operatorname{conv}(y)) \land \mathbf{P}(y, \operatorname{conv}(x))] \to \mathbf{C}(x, y)] \\ \forall x \forall y [[\mathbf{D}\mathbf{R}(x, \operatorname{conv}(y)) \land \mathbf{D}\mathbf{R}(y, \operatorname{conv}(x))] \to \mathbf{D}\mathbf{R}(\operatorname{conv}(x), \operatorname{conv}(y))] \\ \mathbf{CONV}(x) \equiv_{def} x = \operatorname{conv}(x) \end{array}
```

Note that a consequence of these axioms is that the universal region, us, is convex since us is not a proper part of any other region and thus conv(us)=us.

We use conv to define a set of relations which describe regions being inside, partially inside and outside, e.g. 'INSIDE(x, y)' (x is inside y'), 'P-INSIDE(x, y)' ('x is partially inside y') and 'OUTSIDE(x, y)' ('x is outside y'), each of which also has an inverse. Two functions¹¹ capturing the concept of the inside and the outside of a particular region are also definable (where 'inside(x)' is read as 'the inside of x', and 'outside(x)' as 'the outside of x' respectively.

```
\begin{split} & \text{INSIDE}(x,y) \equiv_{def} \text{DR}(x,y) \land \text{P}(x,\text{conv}(y)) \\ & \text{P-INSIDE}(x,y) \equiv_{def} \text{DR}(x,y) \land \text{PO}(x,\text{conv}(y)) \\ & \text{OUTSIDE}(x,y) \equiv_{def} \text{DR}(x,\text{conv}(y)) \\ & \text{inside}(x) =_{def} \iota y [\forall z [\text{C}(z,y) \leftrightarrow \exists w [\text{INSIDE}(w,x) \land \text{C}(z,w)]]] \\ & \text{outside}(x) =_{def} \iota y [\forall z [\text{C}(z,y) \leftrightarrow \exists w [\text{OUTSIDE}(w,x) \land \text{C}(z,w)]]] \end{split}
```

This particular set of relations refines DR(x, y) in the basic theory. In (Randell, Cui and Cohn 1992, Randell, Cohn and Cui 1992) we generated a pairwise disjoint and mutually exhaustive set of relations by taking the relations given above, their inverses, and the set of relations that result from non-empty intersections. The set of base relations for this particular set were then finally generated by defining an EC and DC variant for each of these relations.

A new set of base relations (using the relations defined immediately above) are constructed according to the following schema:

```
\alpha \beta \gamma(x,y) \equiv_{def} \alpha(x,y) \wedge \beta(x,y) \wedge \gamma(x,y)
```

where: $\alpha \in \{\text{INSIDE}, \text{ P-INSIDE}, \text{ OUTSIDE}\}, \beta \in \{\text{INSIDE}^{-1}, \text{ P-INSIDE}^{-1}, \text{ OUTSIDE}^{-1}\}, \text{ and } \gamma \in \{\text{EC}, \text{DC}\} \text{ excepting where } \alpha = \text{INSIDE and } \beta = \text{INSIDE}^{-1}$ This give a total of 22 base relations instead of the original 8.

2.3 Bodies v. Regions

We make an ontological distinction between physical objects (bodies) and the regions of space they occupy. Bodies and regions are represented in the formal theory as disjoint sorts. The mapping between the two is done by introducing a transfer function 'space(x,y)' read as 'the space occupied by x at y', which takes a body at a given moment in time, and maps this to the region of space it occupies. If the body does not exist at a particular moment, then it will be mapped to the sort NULL.

¹¹ Note that it does not really make much sense to define a functional analogue of P-INSIDE as this would simply be the sum of the inside and the outside, i.e. the complement of x!

3 Inferences and Constraints

Efficient inferences can be made by exploiting the structures of the formalism. The computational cost of using uncontrolled inference within computational logics for non-trivial domain problems is well known. Various hybrid representation and reasoning systems have recently gained much interest among AI research workers in an attempt to meet such difficulties (see eg. Frisch and Cohn 1990). The basic idea involves factoring out distinct ways to represent knowledge structures and assigning each "factor" to a subsystem in which specialised inference is done. It should be apparent that our representation is naturally hybrid. Although keeping (sorted) first order logic as the main language, we have knowledge about sorts, subsumption relationships (of both sorts and sets of relations) and continuity restrictions, all of which may be represented and reasoned about by special means. In (Randell and Cohn 1992) we discuss various inference mechanisms for our spatial logic. We review and discuss the most important of these here and relate them to databases.

The first point to note is that the relations naturally form a lattice structure (Figure 3) which can be exploited by any 'clash' based inference mechanism (such as a resolution based deductive engine) via theory resolution (Stickel 1990, Randell and Cohn 1992). Moreover, both facts and queries can be expressed at the most appropriate level of abstraction with respect to the hierarchy. Such a lattice can be used to derive relations efficiently. At its simplest, such a lattice can be used to answer queries such as 'is b connected to c?' (ie C(b,c)) given that the database contains the entry TPP(c,b), very efficiently (the answer being 'yes' in this case, since TPP(c,b) is equivalent to $TPP^{-1}(b,c)$ and TPP^{-1} is below C in the hierarchy. Similarly the lattice can also be used for integrity checking when entering new data. For example it may already be known that P(b,c); an update of TPP(b,c) would be consistent with this but an update of EC(c,b) would not, because the greatest lower bound in the latter case is \bot but is not \bot in the former case.

3.1 Composition Tables

For the temporal interval logic (Allen 1983), a composition table ¹³ has been used extensively for many purposes. Such a table has also been constructed for the initial eight ¹⁴ base relations in our spatial logic (Randell 1991, Randell, Cohn and Cui 1992) – see Table 1 and, independently, by Egenhofer (1991) for his closely related formalism.

Each entry of the form $R_1(a, b)$ and $R_2(b, c)$ is mapped to a disjunctive set of base relations, corresponding to a theorem and no redundant base relations

¹² We have extended the LISP implementation of LLAMA in this way, using a bit-encoding of the lattice to allow lattice operations to be efficiently implemented – cf (Aït-Kaci 1989).

¹³ Allen originally used the term transitivity table.

¹⁴ See later for our progress in building tables for the extended set of relations.

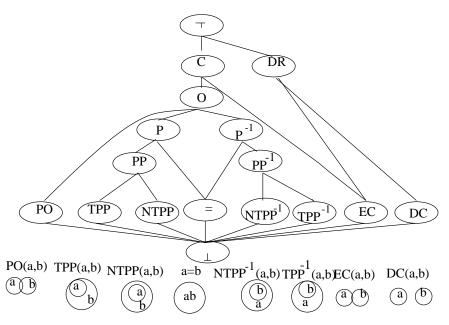


Fig. 3. A lattice defining the subsumption hierarchy of the dyadic relations defined solely in terms of the primitive relation C(x, y).

are given (ie. there is a model for each of the disjunctions). Although the composition table is only given for the base relations, non-base relations may appear in the target set (eg. PP(a,b) and PP(b,c)) so in these cases the following calculation is performed. Firstly we use the lattice L (see figure 3) to compute the set of base relations each relation covers (in this case $\{TPP(a,b)$, $NTPP(a,b)\}$ and $\{TPP(b,c), NTPP(b,c)\}$ – remembering that $\forall xy[PP(x,y) \leftrightarrow [TPP(x,y) \lor NTPP(x,y)]]$). Next we take each $R_1(a,b)$, $R_2(b,c)$ pair, where $R_1(a,b) \in \{TPP(a,b), NTPP(b,c)\}$ and $R_2(b,c) \in \{TPP(b,c), NTPP(b,c)\}$ and form the union of all disjunctive sets of base relations each $R_1(a,b)$, $R_2(b,c)$ pair yields using the composition table. In this case this would be $[TPP \lor NTPP](a,c)$ or simply PP(a,c). So given PP(a,b) and PP(b,c) we deduce PP(a,c).

An interesting open question is whether and when the composition table is sufficient to check for consistency. For example in our qualitative spatial simulation program (Cui, Cohn and Randell 1992a) a state is a conjunction of $n^2/2$ ground atoms whose predicates are the relation symbols presented here, and where there is exactly one atom R(a, b) or R(b, a) for each pair of regions a, b where n is the total number of regions in the state. Potential new states are generated by 'envisioning rules' (see below) and these must be checked for con-

R2(b,c) R1(a,b)	DC	EC	PΟ	TPP	NTPP	TPP-1	NTPP-1	=
DC	no.info	DR,PO,PP	DR,PO,PP	DR,PO,PP	DR,PO,PP	DC	DC	DC
EC	DR,PO,PP	DR,PO TPP,TP	DR,PO,PP	EC,PO,PP	PO,PP	DR	DC	EC
PO	DR,PO,PP	DR,PO,PP	no.info	PO,PP	PO,PP	DR,PO,PP	DR _A PO PP	РО
TPP	DC	DR	DR,PO,PP	PP	NTPP	DR,PO TPP,TP	DR _J PO PP	TPP
NTPP	DC	DC	DR,PO,PP	NTPP	NTPP	DR,PO,PP	no.info	NTPP
TPP-1	DR,PO,PP	EC,PO,PP	PO,PP	PO,TPP,TP	PO,PP	PP ⁻¹	NTPP	TPP-1
NTPP-1	DR,PO,PP	PO,PP	PO,PP	PO,PP	0	NTPP-1	NTPP	NTPP
=	DC	EC	PO	TPP	NTPP	TPP-1	NTPP-1	=

Table 1. Composition table for the 8 basic relations. If $R_1(a,b)$ and $R_2(b,c)$, it follows that $R_3(a,c)$ where R_3 is looked up in the table. "no info." means that no base relation is excluded. Multiple entries in a cell are interpreted as disjunctions. Note that DR stands for DC and EC, PP for TPP and NTPP, PP⁻¹ for TPP⁻¹ and NTPP⁻¹, TP⁻¹ for TPP⁻¹ and =, and O for PO, TPP, NTPP, TPP⁻¹, NTPP⁻¹, and =.

sistency with respect to our spatial theory. At present this is done simply by checking all triples of atoms of the form $R_1(a, b)$, $R_2(b, c)$, $R_3(a, c)$ in the state against the composition table. We believe that this is sufficient in that despite extensive search we cannot find a counterexample. If one restricts the intended interpretation then it is clear that such 'triangle checking' is not sufficient in general; for example if all regions are intended to be circles of the same size, then at most six distinct circles can EC with a particular circle, but this fact will not be detected by triangle checking. Similarly, also in two dimensions, the maximum number of regions that can be mutually partially inside each other (a relationship defined in a following section) is four, and again this cannot be verified by triangle checking. However in the general case, it would appear that triangle checking is sufficient, but we have been unable, despite extensive effort, to formally verify this conjecture to date. Formally stated, the conjecture is as below; readers are invited to contact any of the authors with a counterexample or proof! The conjecture is stated just for the simple case of the 8 basic pairwise disjoint and mutually exhaustive predicates, but we believe the result also holds for the extended case where the relations defined in terms of conv(x) are included.

Conjecture

 $\Gamma, \Psi \models \text{False iff } \Psi, \Pi \models \text{False where } \Gamma \text{ is the set of axioms and definitions}$ (1) to (12) in section 2 above (including the implicit definitions for the inverse relations), Ψ is a conjunction of ground atomic formulae whose predicate symbols are only drawn from the set of eight base relations and whose arguments are constants, and Π is the set of theorems expressed by table 1.

3.2 Continuity Constraints

Assuming continuous motion, there are constraints which can be imposed upon the way the base relations of L can change over consecutive moments in time for any pair of spatial regions. Intuitively, this is best illustrated by the pictorial representation in figure 4.¹⁵ These transitions are alternatively expressed as 'envisioning axioms' (Randell 1991).¹⁶ These continuity networks have been used as the basis of a qualitative spatial simulation program (Cui, Cohn and Randell 1992a), or in the present context of spatial databases could be used to perform consistency checks on the movement of regions in a spatio-temporal database. The construction of such continuity networks for the extended sets of relations is discussed in (Cohn, Randell and Cui 1993). It may also be noted that Egenhofer and Al-Taha (1992) have independently formulated topological change diagrams very similar to ours. They also consider various specialisations where further information is known about the regions in question, as does Galton (1993).

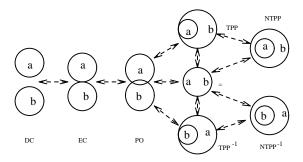


Fig. 4. A pictorial representation of the base relations and their direct topological transitions.

We now turn to consider the continuity network required when we include the predicates defined in terms of conv(x). It is easiest to specify the possible transitions using relatively high level predicates rather than in terms of the base predicates. First we will consider the transitions whose name includes OUTSIDE, P-INSIDE, INSIDE. The transition network for these predicates is displayed in figure 5.

In order to determine the possible transitions for a predicate with a multipart name (such as OUTSIDE_P-INSIDE⁻¹_EC) one simply determines the allowable transitions for each part of the name; thus in the above example, OUTSIDE can

¹⁵ These continuity networks are closely related to what Freksa (1992) calls conceptual neighbourhoods.

¹⁶ Each link in the diagram corresponds to an axiom which expresses that if R1(x,y) is true then either R1(x,y) will continue to be true for ever, or x or y will disappear (become NULL) or R2(x,y) will be the next relationship to be true of x and y in the future.

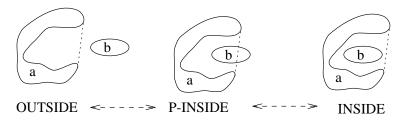


Fig. 5. The transition network for the 3 high level inside/outside relations.

only transition to P-INSIDE, P-INSIDE⁻¹ to INSIDE⁻¹ or to OUTSIDE⁻¹ and EC to PO or DC. In the case that a sub-name transitions from EC to PO, then of course the rest of the sub-names are dropped, e.g. OUTSIDE_CONT-INSIDE⁻¹_EC can transition to EC.

4 Extensions to the Expressive Power

4.1 Refining Inside and Outside

In a previous section the DR relation was specialised to cover relations describing objects being either inside, partially inside or outside other objects. However this ignores some useful distinctions that can be drawn between different cases of bodies being inside another. In this case we can separate out the case where one body is topologically inside another, and where one body is inside another but not topologically inside – this we call being geometrically inside (Randell, Cui and Cohn 1992). The important point of one body being topologically inside another is that one has to 'cut' through the surrounding body in order to reach and make contact with the contained body. In the geometrical variant this is not the case – see figure 6.

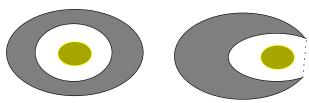


Fig. 6. The distinction between being topologically and geometrically inside. The dotted lines appearing here and in subsequent figures indicate the extent of the convex hull of the surrounding bodies.

$$\begin{split} \text{TOP-INSIDE}(x,y) &\equiv_{def} \\ \text{INSIDE}(x,y) \wedge \forall z [[\text{CON}(z) \wedge \text{C}(z,x) \wedge \text{C}(z,\text{outside}(y))] \rightarrow \text{O}(z,y)] \\ \text{GEO-INSIDE}(x,y) &\equiv_{def} \text{INSIDE}(x,y) \wedge \neg \text{TOP-INSIDE}(x,y) \end{split}$$

It is also possible to specialise the relation of being geometrically inside – in this case setting up definitions to distinguish between the pictorial representations in Figure 7.

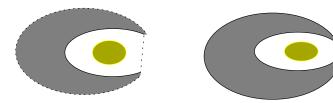


Fig. 7. Two variants of being geometrically inside. In the right hand figure the two 'arms' meet at a point.

In order to make this formal distinction we first set up a stronger case of a connected or one-piece region to that assumed above. The important part of the following definition is the P(conv(sum(v,w)),x) literal in the consequent of the definiens. This condition ensures that the connection between any two parts of a region whose sum equals that region, is not point or edge connected. That is to say it ensures a 'channel' region exists connecting any two connected parts. This notion of being connected mirrors and simplifies our previous definition of a quasi-manifold – in this case we use the concept of a convex body rather than use topological and Boolean concepts in the earlier definition – see Randell and Cohn (1989).

```
\begin{aligned} & \operatorname{CON}'(x) \equiv_{def} \operatorname{CON}(x) \wedge \\ & \forall yz [[\operatorname{sum}(y,z) = x \to \operatorname{C}(y,z)] \to \\ & \exists vw [\operatorname{P}(v,y) \wedge \operatorname{P}(w,z) \wedge \operatorname{P}(\operatorname{conv}(\operatorname{sum}(v,w)),x)]] \end{aligned}
```

Now we give the formal distinction between the two cases of being geometrical inside. In the first case a 'channel' region exists connecting the outside of the surrounding body with the contained body, in the second case the surrounding body has closed forming (in this case) a point connection. In both cases we can see how in contrast with the notion of being topologically inside, it is possible to construct a line segment that connects with both the surrounding body and the contained body without cutting through the surrounding body. Definitions distinguishing between the two cases are as follows, where the open and closed variants respectively refer to the first and second cases described above.

```
GEO-INSIDE-OPEN(x, y) \equiv_{def} \text{GEO-INSIDE}(x, y) \land
CON'(\text{sum}(\text{inside}(y), \text{outside}(y)))
GEO-INSIDE-CLOSED(x, y) \equiv_{def}
GEO\text{-INSIDE}(x, y) \land
CON(\text{sum}(\text{inside}(y), \text{outside}(y))) \land
\neg CON'(\text{sum}(\text{inside}(y), \text{outside}(y)))
```

This technique of refining a base relation into a set of finer grained mutally exhaustive and pairwise disjoint specialised relations can be continued as often as required for a particular application. In Cohn, Randell and Cui(1993) we show how the set of base relations can be expanded to some 100 relations and point to ways in which the process can be continued. We also discuss criteria to help decide when such refinements are worthwhile.

4.2 Orientations and Directions

Thus far we have only considered the relationship between two regions based on essentially topological notions. However, it is often useful to be able to express and reason about the relative or absolute orientation of two regions. In the temporal version of our calculus (Randell and Cohn 1989, 1992) we introduce an additional primitive B(x,y) which is true when the temporal region (period) x is entirely before y. Of course it is natural to introduce this to the spatial calculus as well, although we have thus far not done this. We therefore introduce three more primitives: B_1 , B_2 and B_3 . These are analogues of the three axes of the Cartesian coordinate scheme and are axiomatized below. Conventionally, we will arbitrarily associate B_1 and B_2 with the horizontal axes and B_3 with the vertical axis (though other interpretations are of course possible apart from this intended interpretation).

$$\forall x \neg B_i(x, x)$$

$$\forall x y z [B_i(x, y) \land B_i(y, z) \rightarrow B_i(x, z)]$$

$$\forall x y [B_i(x, y) \rightarrow \forall x_1 y_1 [(P(x_1, x) \land P(y_1, y)) \rightarrow B_i(x_1, y_1)]]$$

These relationships are transitive, but one often requires a non transitive relation (eg 'directly below' in the sense that a gravity vertical would intersect both regions). We can thus define $D-B_1(x,y)$, $D-B_2(x,y)$, $D-B_3(x,y)$:17

$$D-B_{i}(x,y) \equiv_{def} B_{i}(x,y) \land \neg B_{j+1}(x,y) \land \neg B_{j+1}(y,x) \land \neg B_{j+2}(x,y) \land \neg B_{j+2}(y,x)$$

These relationships are illustrated in Figure 8.

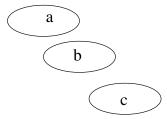


Fig. 8. The following are true in the above figure: $B_3(a,b)$, $B_3(b,c)$, $B_3(a,c)$, $D-B_3(a,b)$, $D-B_3(b,c)$, but $\neg D-B_3(a,c)$.

¹⁷ In the definition 'j+1' and 'j+2' denote modulo arithmetic, eg 3+1=1, 3+2=2.

Related to these notions are the functional terms denoting the extremities of an object in a particular direction 18 eg the top or bottom of a region. Thus we may define a set of monadic function symbols extreme, Extreme,

```
\begin{aligned} & \operatorname{Extreme}_{i}(\mathbf{y},r) \equiv_{def} P(\mathbf{y},r) \wedge \forall x [[\mathbf{P}(x,r) \wedge \neg (x=y)] \to \mathbf{B}_{i}(\mathbf{y},x)] \\ & \operatorname{extreme}_{i}(r) =_{def} \iota y [\operatorname{Extreme}_{i}(y,r) \wedge \forall x [\operatorname{Extreme}_{i}(x,r) \to \mathbf{P}(x,y)]] \\ & \operatorname{Extreme}_{i}^{-1}(\mathbf{y},r) \equiv_{def} \mathbf{P}(\mathbf{y},r) \wedge \forall x [[\mathbf{P}(x,r) \wedge \neg (x=y)] \to \mathbf{B}_{i}(\mathbf{y},x)] \\ & \operatorname{extreme}_{i}^{-1}(r) =_{def} \iota y [\operatorname{Extreme}_{i}^{-1}(y,r) \wedge \forall x [\operatorname{Extreme}_{i}^{-1}(x,r) \to \mathbf{P}(x,y)]] \end{aligned}
```

Note that it is not difficult to see that these definitions presuppose a notion of atomic region, i.e. a region which has no proper subparts. The calculus as presented thus far does not admit atomic regions because of the axiom $\forall x \exists y [\text{NTPP}(y, x)]$. However, in Randell, Cui and Cohn (1992), we discussed a number of ways to modify the calculus to allow atomic regions.

Given our intended interpretation, more natural names for extreme₃(x) and extreme₃⁻¹(x) would be top(x) and bottom(x). In a particular context one may wish to rename the other relations and functions just defined as well. For example, one might want to name B₁ as EastOf, B₁⁻¹ as WestOf, B₂ as NorthOf and B₂⁻¹ as SouthOf. Or if one assumes a particular viewpoint, then these four relations may be more naturally named LeftOf, RightOf, Behind and InFrontOf. Of course, rather than regard the viewpoint as fixed, one may want to add an extra argument to give the viewpoint and define, eg, LeftOf(x, y, z), meaning x is to the left of y when viewed by/from z. In this case we do not need to assume a fixed cartesian coordinate scheme. Interestingly, Behind(x, y, z) and InFrontof(x, y, z) are actually definable from C(x, y) and conv(x) alone, without recourse to the additional B_i primitives (see Figure 9). However, the other 3 place relative relations do seem to require the relevant additional primitives.

```
\begin{split} \operatorname{InFrontof}(x,y,z) &\equiv_{def} \operatorname{DR}(x,y) \wedge \operatorname{DR}(y,z) \wedge \operatorname{DR}(x,z) \wedge \\ &\exists w \ \operatorname{EC}(w,z) \wedge \operatorname{EC}(w,y) \wedge \operatorname{CONV}(w) \wedge \operatorname{O}(x,w) \wedge \operatorname{P}(w,\operatorname{inside}(\operatorname{sum}(z,y))) \\ \operatorname{Behind}(y,x,z) &\equiv_{def} \operatorname{InFrontof}(x,y,z) \end{split}
```

4.3 Integrating metric and scalar information

So far we have principally concentrated on developing a purely qualitative calculus. However this can never be a replacement for metric information, but rather should complement a metric representation. Indeed Forbus et al (1991) have argued that there is no purely qualitative representation of space – their so called 'poverty conjecture'. Joskowicz (1992) has also argued that there is no general

¹⁸ Remember that we allow multipiece regions and the extremity of a concave region may well be a multipiece region.

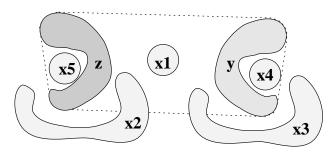


Fig. 9. Defining a relative notion of InFrontof. x1, x2 and x3 are all in front of y when viewed from z. The definition rules out the possibility that x4 and x5 are in front of y when viewed from z because w has to be convex (one piece) region touching both y and z. Arguably x2 and x3 should not be regarded as being in front of y when viewed from z; to achieve this the definition should be strengthened so that the conjunct O(x, w) is replaced by P(x, w)

purpose commonsense spatial reasoning and representation mechanism and that a hybrid representation is necessary. We do not have space here to describe our approaches to integrating metric information with our qualitative calculus. Some of our initial ideas can be found in (Randell and Cohn 1989, Randell 1991). We have also recently integrated our purely qualitative spatial simulator (Cui, Cohn and Randell 1992a) with QSIM (Kuipers 1986) which allows both region based spatial knowledge and scalar quantities (eg the qualitative distance between two regions) to be reasoned about together (Cui, Cohn and Randell 1992b).

5 Related Work

We have already mentioned Clarke's calculus of individuals, our earlier work on which this present theory is based, and Allen's work on interval logics. The other work of which we are aware, that uses Clarke's theory for describing space, is Augnague (1991) and Vieu (1991). Other work on the description of space using a body rather than a point based ontology, can be found in Laguna (1922), Tarski (1965) and Whitehead (1978). There have been some attempts in the qualitative spatial reasoning literature to employ Allen's interval logic for describing space, see for example Freska (1990), Mukerjee and Joe (1990) and Hernandez(1990,1992), but here a stronger primitive relation is used, which does not allow the full range of topological relationships to be formally described as given in both Clarke's and our original and new theories. Apart from the question raised by adding atoms to the theory, we are currently working on the question as to what a decidable subset the new theory supports. We have already indicated some extensions to this new logic above. For other extensions to the spatial theory itself, work described in Randell (1991) can also be included. For example, we could add a metric extension to the theory, using a ternary relation (along the lines of Van Benthem 1982, appendix A) that gives comparative distances between objects.

Egenhofer's work (1991) on topological relations has similar relations to ours, though we have an extended set owing to the relations defined in terms of conv(x). However, we draw distinction between physical objects and regions; intuitively, a region of space may be likened to a space that could be conceivably occupied by a physical body. We restrict our interpretation of time and space so that time is treated as a one-dimensional region, and space a three-dimensional region. In Egenhofer's work, the usual concepts of point-set topology with open and closed sets are assumed. For the corresponding part of his theory, we only use one primitive C. Open, closed, etc concepts can be defined in our theory (see Randell et al 1989, 1992) although in the theory presented above, we assumed no distinctions between open and closed regions because we believe that applications, and reasoning about physical objects and commonsense reasoning in particular, do not need to differentiate between open and closed regions. Another difference is that we do not have a notion of points in the theory, ¹⁹ nor do we view an arbitrary set of points as a region. Finally we introduce the convex hull operator allowing many more base relations to be defined.

In both theories, eight disjoint base relations are defined. There is a one to one correspondence between the base relations: DC (disjoint), EC (meet), = (equal), TPP (coverBy), PO (overlap), NTPP (inside) TPP⁻¹ (covers) and NTPP⁻¹ (contain). One minor difference is between EC and meet. In our theory, if the sum of two regions equals the universe (assuming universe is continuous), the two regions EC, whereas the two regions do not meet in Egenhofer's (as they did not in our original theory which distinguished between open and closed regions).

The composition table in both theories is the same although we used a combination of theorem proving and model building (Randell, Cohn and Cui 1992b) to derive the composition table whilst Egenhofer used exhaustive search by PROLOG. Egenhofer used matrix representations of the base relations in the theory and this obviously provides a model for his theory. We have investigated a linear bitmap representation in calculating and constructing the composition tables (Randell, Cohn and Cui 1992b). In Randell, Cohn and Cui (1992), we proposed calculating an extended composition table. The task is formidable. We have in fact constructed the table for 22×22 case. This took some 2 days of cpu time on a Sparc IPC. For an even larger table (Cohn, Randell and Cui 1993), this becomes more difficult. It would be very interesting to investigate whether an extension of Egenhofer's matrix method could be used.

Perhaps the most important work in qualitative spatial orientation is that of Freksa(1991), who also surveys the existing work in this area. The principal tenets of his approach are that he just treats points rather than regions²⁰,

¹⁹ In the earliest version we did include points (Randell and Cohn 1989) but have since abandoned them as pragmatically unnecessary (Randell et al 1992) though they could be reintroduced if necessary (Randell, Cui and Cohn 1992a).

²⁰ Freksa does remark that regions may be treated as points at a certain level of abstraction.

orientations are relative and are terms in his language rather than relations as outlined in our approach above; he then uses qualitative relations to compare orientations (eg he would say same(ab,cd) meaning that the directions from a to b and from c to d are identical). Other relations include additional exact directions (such as 'opposite') and inexact relations (such as 'left', which cover a segment of directions. Further refinements of these relations allow a notion of relative qualitative distance to be expressed as well. He also defines a composition table on his qualitative directions.

The inferential power of his approach relies crucially on having some exact directions (as compared eg to Hernadez (1990, 1992)). An interesting question arises as to whether it is possible to integrate Freksa's calculus with a region based approach (whilst retaining Freksa's inferential power). We speculate that this may be possible using either atomic regions (Randell, Cui and Cohn 1992a) or by defining 'strong' versions of D-B_i. The relations defined in this paper are 'weak' in the sense that, for example, D-B₃(b, c) is true if some part of b is directly above some part of c. A strong version would only be true if every part of b was directly above some part of c (or, perhaps, some part of b was above every part of c). These ideas need further investigation but should allow a sufficient notion of transitivity to be gained (the transitivity of exact directions are most important to Freksa's calculus). Also worthy of investigation is the extension of Freksa's calculus to 3D (at present he only considers the 2D case).

We conclude this discussion of qualitative orientation by noting that the purpose of the exposition of qualitative orientation in the body of this paper was not to give a definitive axiomatisation, but rather to show how such information might be axiomatised, defined and related to the rest of the calculus.

6 Conclusion

We have presented a spatial logic based on a primitive notion of connection. Eight mutually exclusive and pairwise disjoint relations were defined giving an alternative formulation of Egenhofer's calculus. By introducing a further primitive, more expressive calculi can be defined. A number of inference techniques, dealing with both static and dynamic spatial configurations were outlined which may be efficiently implemented and their relevance to databases was discussed. Of course many questions remain unanswered and deserve further research. We have mentioned some of these above.

References

- H. Aït-Kaci, et al: "Efficient Implementation of Lattice Operations," ACM Trans. on Programming Languages and Systems, 11(1), 1989
- 2. J. F. Allen: "Maintaining Knowledge about Temporal Intervals," Comm. ACM26(11), 1983.
- 3. M. Aurnague: "Contribution a l'etude de la semantique formelle de l'espace et du raisonnement spatial: la localisation interne en francais, semantique et structures inferentielles", PhD thesis, l'Universite Paul Sabatier de Toulouse, 1991.

- 4. Van Benthem: The Logic of Time, Synthese Library vol 156, Reidel, London, 1982
- B. L. Clarke: "A Calculus of Individuals Based on Connection," Notre Dame Journal of Formal LOGIC, vol. 2, No. 3, 1981.
- B. L. Clarke: "Individuals and Points", Notre Dame Journal of Formal Logic Vol. 26, No. 1, 1985.
- A. G. Cohn: "A More Expressive Formulation of Many Sorted Logic," J. Autom. Reasoning., Vol. 3(2), pp. 113-200, 1987
- 8. A. G. Cohn, D. A. Randell and Z. Cui: "A Taxonomy of Logically Defined Qualitative Spatial Relations," Proc International Workshop on Formal Ontology in Conceptual Analysis and Knowledge Representation, Padova, Ladseb-CNR Internal Report 01/93, ed N Guarino and R Poli, pp 149-158, 1993
- 9. Z. Cui, A. G. Cohn and D. A. Randell: "Qualitative Simulation Based On A Logical Formalism Of Space And Time", Proc AAAI, 1992a.
- 10. Z. Cui, A. G. Cohn and D. A. Randell: "Qualitative Simulation Based On A Logic Of Space And Time", Proc. of QR92, 1992b.
- M. Egenhofer: "A Formal Definition of Binary Topological Relationships," In: W Litwin and H J Schek, editors, Third International Conference on Foundations of Data Organization and Algorithms, Paris, France, Lecture Notes in Computer Science, Vol. 367, pages 457-472, Springer-Verlag, New York, NY, June 1989
- M. Egenhofer: "Reasoning about Binary Topological Relations," In: Proceedings of the Second Symposium on Large Spatial Databases, SSD'91 (Zurich, Switzerland, 1991), O. Gunther and H. J. Schek, Eds. Lecture Notes in Computer Science 525, pp 143-160.
- M. Egenhofer and J. A. Herring: "A Mathematical Framework for the Definition of Topological Relationships," In: Proceedings of Fourth International Symposium on Spatial Data Handling (Zurich, Switzerland, 1990), K. Brassel and H. Kishimoto, Eds.
- 14. M. Egenhofer and A. Frank: "Towards a Spatial Query Language: user interface considerations," In: Proceedings of 14th International Conference on Very Large Data Bases (Los Angeles, CA, 1988) D. DeWitt and F. Bancilhon, Eds.
- M. Egenhofer and K K Al-Taha: "Reasoning about Gradual Changes of Topological Relationships", in Theoreies and Methods of Spatio Temporal Reasoning in Geographic Space, LNCS 639 Springer Verlag, 1992
- K. D. Forbus, P. Nielsen and B. Faltings: "Qualitative spatial reasoning: the CLOCK project," Art. Int. 51, pp. 417-471, Elsevier, 1991
- C. Freksa: "Using Orientation Information for Qualitative Spatial Reasoning", Berick Nr 11, Kognitionswissenschaft, Univ. Hamburg, 1992a.
- C. Freksa: "Temporal Reasoning based on Semi-intervals," Artificial Intelligence, 54, Elsevier, 1992b
- 19. A. M. Frisch and A. G. Cohn: "Thought and Afterthoughts on the 1988 Workshop on Principles of Hybrid Reasoning," AI Magazine 1990
- A Galton, Towards an integrated logic of space, time, and motion, to appear in Proc IJCAI93, Morgan Kaufmann, 1993.
- D. Guenther and A. Buchmann: "Research Issues in Spatial Databases," SIGMOD RECORD 19 (4), 1990
- 22. R. Güting: "Geo-Relational Algebra: A Model and Query Language for Geometric Database Systems," Advances in Database Technology, 1988.
- 23. D. Hernandez: "Using Comparative Relations to represent Spatial Knowledge", Workshop RAUM, Univ. of Koblenz, 1990.

- 24. D. Hernandez: "Qualitative Representation of Spatial Knowledge", PhD Dissertation, Fakultaet fuer Informatik, Technischen Universitaet Muenchen, Germann, 1992
- L. Joskowicz: "Commonsense Reasoning about Moving Objects: an Elusive Goal", IBM T J Watson Research Center, Yorktown Heights, 1992.
- 26. B. Kuipers: "Qualitative Simulation," Artificial Intelligence 29: 298-338, 1986
- T. de Laguna: "Point, Line and Surface as sets of Solids" The Journal of Philosophy., Vol 19., pp. 449-461, 1922.
- 28. A. Mukerjee and G. Joe: "A Qualitative Model for Space", Proc AAAI, 1990.
- D. Pullar and M. Egenhofer: "Towards Formal Definitions of Topological Relations among Spatial Objects," In: proceedings of Third International Symposium on Spatial Data Handling (Sydney, Australia, 1988), D. Marble, Ed.
- D. Randell, A. G. Cohn and Z. Cui: "Naive Topology: modeling the force pump," in Recent Advances in Qualitative Reasoning, ed B Faltings and P Struss, MIT Press, in press, 1992a.
- D. A. Randell: "Analysing the Familiar: Reasoning about space and time in the everyday world," PhD Thesis, Univ. of Warwick, UK 1991.
- 32. D. A. Randell, A. G. Cohn and Z. Cui: "Computing Transitivity Tables: a Challenge for Automated Theorem Provers," Proc CADE11, 1992b.
- D. Randell and A. G. Cohn: "Modelling Topological and Metrical Properties in Physical Processes," in *Principles of Knowledge Representation and Reasoning*, ed. R. J. Brachman, H. Levesque and R. Reiter, Morgan Kaufmann, Los Altos. 1989.
- 34. D. A. Randell and A. G. Cohn: "Exploiting Lattice in a Theory of Space and Time," Computers and Mathematics with Applications, 1992.
- D. A. Randell, Z. Cui and A. G. Cohn: "A Spatial Logic based on Regions and Connection", Proc 3rd Int Conf on the Principles of Knowledge Representation and Reasoning, Morgan Kaufmann, 1992a.
- D. Randell, Z. Cui and A. G. Cohn: "An Interval Logic for Space based on "Connection"," in proc. of ECAI92, 1992b
- C. J. Rawlings, W. R. Taylor, J. Nyakairu, J. Fox and M. J. E. Sternberg: "Reasoning about Protein Topology using the Logic Programming Language PROLOG,"

 J. Mol. Graphics 3(4), 1985
- 38. N. Roussopoulos, C. Faloutsos and T. Sellis: "An Efficient Pictorial Database System for PSQL," IEEE Transactions on Software Engineering 15 (5), 1988
- M. Stickel: "Automated Deduction by Theory Resolution," J. Automated Reasoning, 1, 1985
- 40. A. Tarski: "Foundations of the geometry of solids," in Logic, Semantics, Metamathematics, trans. J. H. Woodger, Oxford Uni. Press, 1956
- 41. L. Vieu: "Semantique des relations spatials et inference spatio-temporelles: Une contribution a l'étude des structures formelles de l'espace en Language Naturel.", PhD thesis, l'Universite Paul Sabatier de Toulouse, 1991.
- 42. A. N. Whitehead: "Process and Reality: Corrected Edition," eds. D.R. Griffin and D.W. Sherburne, The Free Press, Macmillan Pub. Co., New York, 1978

This article was processed using the $\Pi_{\hbox{\footnotesize{E}}}X$ macro package with LLNCS style