Natural-Language Spatial Relations Between Linear and Areal Objects: The Topology and Metric of English-Language Terms*

A. Rashid B. M. Shariff
Department of Survey and Mapping Malaysia, Jalan Semarak, 50578 Kuala Lumpur, Malaysia, rashid@itc.utm.my

Max J. Egenhofer, National Center for Geographic Information and Analysis, Department of Spatial Information Science and Engineering, and Department of Computer Science, University of Maine,Orono, ME 04469-5711, max@spatial.maine.edu

and

David M. Mark, National Center for Geographic Information and Analysis and Department of Geography, State University of New York at Buffalo, Buffalo, NY 14261, geodmm@ubvms.cc.buffalo.edu

Abstract

Spatial relations are the basis for many selections users perform when they query geographic information systems (GISs). Although such query languages use natural-language-like terms, the formal definitions of those spatial relations rarely reflect the same meaning people would apply when they communicate among each other. To bridge the gap between computational models for spatial relations and people's use of spatial terms in their natural languages, a model for the geometry of spatial relations was calibrated for a set of 59 English-language spatial predicates. The model distinguishes topological and metric properties. The calibration from sketches that were drawn by 34 human subjects identifies ten groups of spatial terms with similar properties and provides a mapping from spatial terms onto significant geometric parameters and their values. The calibration's results reemphasize the importance of topological over metric properties in the selection of English-language spatial terms. The model provides a basis for high-level spatial query languages that exploit natural-language terms and serves as a model for processing such queries.

1. Introduction

Communication is paramount to people understanding each other. Among the most critical measures of the success of verbal communication is the effective capture and conveyance of the semantics of words. The better the recipient of words reconstructs the meaning the sender attaches to them, the better the recipient will understand the sender. This measure also applies to the interaction between users and geographic information systems (GISs). Although the interaction with current GISs occurs primarily through structured query languages (Egenhofer and Herring 1993), future GISs are expected to support more natural interactions with geographic data through such modalities as sketching a spatial query and simultaneously describing aspects of the sketch

through spoken language (Egenhofer 1996). Likewise, future GISs are likely to complement the presentation of query results in a graphical or tabular form through the generation of natural-language-like instructions or responses to spatial queries (Mark and Gould 1991). In order to move towards GISs with which users could interact much like they would communicate with other people, it is necessary to gain a better understanding of the semantics of the spatial terms that would play a major role in such interactions.

The focus of this work is on capturing the geometry that is associated with natural-language spatial relations. Spatial relations refer to the way people perceive spatial configurations, how they reason about such configurations, and how they describe them in a variety of languages. Based on different mathematical concepts, the GIS literature distinguishes three major types of spatial relations (Pullar and Egenhofer 1988; Worboys 1992): topological relations require the concept of neighborhood and are invariant under consistent topological transformations, such as rotation, translation, scaling; cardinal direction relations are based on the existence of a vector space and are subject to change under rotation, while they are invariant under translation and scaling of the reference frame; and distance relations express spatial properties that reflect the concept of a metric and, therefore, change under scaling, but are invariant under translation and rotation. Capturing the semantics of spatial relations has a long-standing history (Freeman 1975; Peuquet 1986), with a wave of formal models for spatial relations developed in the 1990s (Egenhofer and Herring 1990; Egenhofer and Franzosa 1991; Egenhofer and Herring 1991; Frank 1991; Hernández 1991; Frank 1992; Freksa 1992; Hazelton *et al.* 1992; Randell *et al.* 1992; Clementini *et al.* 1993; Cui *et al.* 1993; Papadias and Sellis 1993; Zimmermann 1993; Hernández 1994; Papadias and Sellis 1994; Clementini *et al.* 1995; Cohn 1995; Egenhofer and Franzosa 1995; Hernández *et al.* 1995; Hong *et al.* 1995; Nabil *et al.* 1995; Clementini and di Felice 1996; Cohn and Gotts 1996; Nabil *et al.* 1996; Papadias *et al.* 1996; Sharma 1996), these models have often stood for themselves, leading to query language extensions based on mathematically well-defined concepts (Herring 1991; de Hoop and van Oosterom 1992; Hadzilacos and Tryfona 1992; Keighan 1993), but there have been only few attempts to link these models with the way people use spatial terms in natural language. Most previous approaches to characterizing the meanings of spatial relations worked primarily with informal models and treated spatial relations case by case (Talmy 1983; Herskovits 1986). An approach in computational linguistics, based on a connectionists model, provides a framework for the definition of a set of spatial relations (Regier 1995), however, it does not lead immediately to explaining differences among relations in a high-level, visually-related domain and results in computational models (e.g., neural networks) that do not integrate well with current architectures of GISs and spatial database systems. This paper makes an effort to narrow the gap between formal models of spatial relations as developed for GISs and people's intuitive understanding of spatial relations as expressed in everyday language so that future GISs would become more natural and easier to use.

The semantics of spatial relations have many facets and aspects that may influence people's choices of words, such as the meaning of the objects, their shape, their scale, and the spatial relations among the objects, as well as the culture, education, and natural language of the individuals using the terms (Mark *et al.* 1995). Since GISs are primarily engaged in the recording of geometric and some semantic information, future GISs that could listen to and talk with users (Egenhofer 1996) require links between geometric representations and natural spatial languages; therefore, this paper focuses on the topological and metric aspects of natural-language spatial relations. We build on a model for topological relations and enhance it with metric refinements that capture details beyond topological aspects. Aspects of direction or orientation have been reserved for future investigations as possible extensions to the currently used model. The model of topological relations with metric refinements was calibrated for a set of 59 English-language spatial terms, for which 34 human subjects had generated sketches. The calibration shows ten groups with different topological and metric characteristics, and identifies for each term its significant metric parameters and their value ranges. This model enables the generation of simple sentences to

describe spatial scenes and the processing of spatial queries with natural-language spatial terms (Shariff 1996).

The remainder of this paper is structured as follows: Section 2 summarizes the topological and metric models used to specify spatial relations. Section 3 describes the experiment conducted to calibrate the model for 59 English-language terms. To analyze the effect of topology and metric, the terms were grouped into clusters with similar properties (Section 4) and within these clusters, significant parameters were identified leading to a dictionary of topological and metric parameters for the 59 terms (Section 5). Section 6 confirms the significant parameters with results from another experiment and Section 7 demonstrates the implications of this model for spatial query processing. Section 8 provides conclusions and discusses future work.

2.   Computational Model for Spatial Relations

Following the premise that *topology matters, metric refines* (Egenhofer and Mark 1995), we use a two-tier model for the analysis of natural-language spatial relations. It consists of (1) capturing the *topology* of the configuration and (2) analyzing the topological configuration according to a set of *metric* properties.

2.1   9-Intersection for Line-Region Relations

The 9-intersection model is a comprehensive model for binary topological spatial relations and applies to objects of type area, line, and point (Egenhofer and Herring 1991). It characterizes the topological relation between two point sets, A and B, by the set intersections of A's interior ($A°$), boundary (∂A), and exterior (A^-) with the interior, boundary, and exterior of B, called the *9-intersection* (Equation 1).

$$I(A, B) = \begin{pmatrix} A° \cap B° & A° \cap \partial B & A° \cap B^- \\ \partial A \cap B° & \partial A \cap \partial B & \partial A \cap B^- \\ A^- \cap B° & A^- \cap \partial B & A^- \cap B^- \end{pmatrix} \qquad (1)$$

With each of these nine intersections being empty (0) or non-empty (1), the model distinguishes 512 different topological relations between two point sets, some of which cannot be realized, depending on the dimensions of the objects and the dimensions of their embedding space. For a simple line (1-dimensional, non-branching, without self-intersections) and a region (2-dimensional, simply connected, no holes) embedded in R^2, nineteen different situations are found with the 9-intersection model (Figure 1). The nineteen relations are referred to by their line-region (LR) number, which is the conversion of the first two rows in the intersection matrix from a binary number into a decimal number. The bottom row is ignored in the LR number, because it always produces three 1's for line-region relations in R^2.

LR 11
$$\begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

LR 12
$$\begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}$$

LR 13
$$\begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

LR 22
$$\begin{pmatrix} 0 & 1 & 0 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}$$

LR 31
$$\begin{pmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

LR 32
$$\begin{pmatrix} 0 & 1 & 1 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}$$

LR 33
$$\begin{pmatrix} 0 & 1 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

LR 42
$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}$$

LR 44
$$\begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \end{pmatrix}$$

LR 46
$$\begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}$$

LR 62
$$\begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}$$

LR 64
$$\begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \end{pmatrix}$$

LR 66
$$\begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}$$

LR 71
$$\begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

LR 72
$$\begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}$$

LR 73
$$\begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

LR 74
$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \end{pmatrix}$$

LR 75
$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

LR 76
$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}$$

Figure 1: Geometric interpretations of the 19 line-region relations that can be realized from the 9-intersection model (Egenhofer and Herring 1991).

2.2 Metric Refinements for Line-Region Relations

The 9-intersection model segregates *topologically* distinct spatial terms at a coarse level, differentiating clearly, for example, between the concepts that underlie the spatial terms *inside, outside,* and *on the boundary.* Frequently, however, spatial terms require a finer level of representation. For example, "The road that exits the park" and "The road that ends just outside the park" may have the same topological configuration yet distinct differences in metrics. For this purpose, spatial terms are formalized further through the use of metric principles involving area, distance, and orientation.

The critical components of a region and a line are their interiors, boundaries, and exteriors. When the interior, boundary, and exterior of a region interact with either the boundary or interior of a line, certain metric properties can be captured about this interaction. For instance, the interior of a line can share parts with the boundary of a region such that one can measure the length of the common stretch. A purely quantitative measure would record an absolute value, like the length of the common boundary in inches. This approach would be insufficient as it does not take into consideration the relation of the associated objects; therefore, under such operations as scaling of the entire scene, a completely different value would be obtained and stronger values would be obtained if a smaller reference object were chosen. Following Talmy's (1983) observation that the objects' sizes are irrelevant for the choice of their spatial relation term, we designed a model for metric concepts that normalizes metric values for line-region relations with respect to the region's area, the line's length, and the region's perimeter.

To describe details about topological relations, we consider three metric concepts: (1) *splitting*, which determines how the region's and line's interiors, boundaries, and exteriors are cut; (2) *closeness*, which determines how far apart the region's boundary is from the parts of the line, and (3) *approximate alongness*, which combines the closeness measures and the splitting ratios.

2.2.1 Splitting

Splitting determines how a region's interior, boundary, and exterior are divided by a line's interior and boundary, and vice versa. To describe the degree of a splitting, the metric concepts of the length of a line and the area of a region are used. In the context of topological relations between lines and regions, *length* applies to the line's interior, any non-empty intersection with a line's interior, or their components; and to region boundaries, any non-empty intersection between a region's boundary and a line's exterior, or their components. *Area* applies to the interior or regions, the intersections between a line's exterior and a region's interior or exterior, and their components. Among the entries of the 9-intersection for a line and a region, there are seven intersections that can be evaluated with a length or an area (Egenhofer and Shariff 1998).

- *Interior area-splitting (IAS)* describes how the line's interior separates the region's interior (Figure 2a),
- *exterior area-splitting (EAS)* describes how the line's interior separates the region's exterior (Figure 2b),
- *interior traversal-splitting (ITS)* describes how the region's interior splits the line's interior (Figure 2c),
- *exterior traversal-splitting (ETS)* describes how the region's exterior splits the line's interior (Figure 2d),
- *perimeter alongness (PA)* describes how the line's interior splits the region's boundary (Figure 2e),
- *line alongness (LA)* describes how the region's boundary splits the line's interior (Figure 2f), and
- *region boundary splitting (RBS)* describes how the line's boundaries split the region's boundary (Figure 2g).

These concepts are formulated as ratios with respect to the region's area, its perimeter, or the line's length, therefore, forming a measure that is dimension-neutral.

(a) Interior Area Splitting

leftArea (R°∩L⁻)

rightArea (R°∩L⁻)

$$IAS = \frac{\min(leftArea(R° \cap L^-),\ rightArea(R° \cap L^-))}{area(R)}$$

(b) Exterior Area Splitting

boundedExterior (R⁻∩L⁻)

$$EAS = \frac{area(boundedExterior(R^- \cap L^-))}{area(R)}$$

(c) Interior Traversal Splitting

$L° \cap R°$

$$ITS = \frac{length(L° \cap R°)}{length(L)}$$

(d) Exterior Traversal Splitting

$L° \cap R^-$

$$ETS = \frac{length(L° \cap R^-)}{length(L)}$$

(e) Perimeter Alongness

$L° \cap \partial R$

∂R

$$PA = \frac{length(L° \cap \partial R)}{length(\partial R)}$$

(f) Line Alongness

$L° \cap \partial R$

$$LA = \frac{length(L° \cap \partial R)}{length(L)}$$

(g) Region Boundary Splitting

min (∂R∩L⁻)

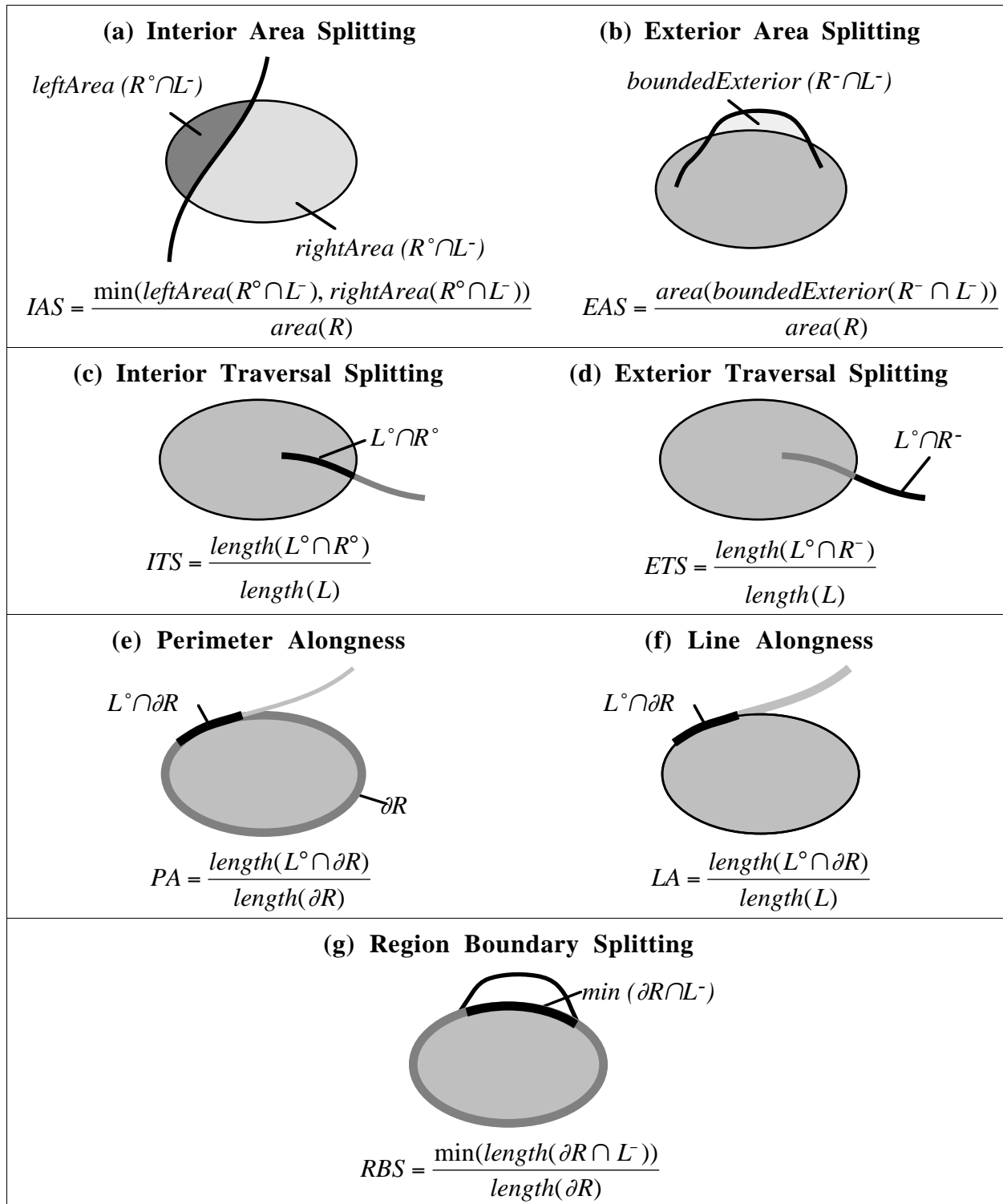$$RBS = \frac{\min(length(\partial R \cap L^-))}{length(\partial R)}$$

Figure 2: Metric refinements of topological relations: splitting ratios.

2.2.2 *Closeness*

Unlike splitting, which requires coincidence and describes how much is in common between two objects, *closeness measures* describe how far apart disjoint parts are (Egenhofer and Shariff 1998).

- *Inner closeness (IC)* captures the remoteness of the line's boundary, located in the interior of the region, from the region's boundary (Figure 3a),
- *outer closeness (OC)* describes the remoteness of the region's boundary from a boundary point of a line located in the exterior of the region (Figure 3b),
- *inner nearness (IN)* describes how far the line's interior, located in the interior of the region, is from the region's boundary (Figure 3c), and
- *outer nearness (ON)* describes how far the line's interior is from the region's boundary (Figure 3d).

These concepts are expressed in terms of the amount a region would have to grow or shrink in order to bridge the distance to the line's boundary or interior. To convert the closeness measures to dimension-neutral measures, they are expressed as ratios over the region's area.

(a) Inner Closeness

$$IC = \frac{area(\Delta_{BI}R)}{area(R)}$$

(b) Outer Closeness

$$OC = \frac{area(\Delta_{BE}R)}{area(R)}$$

(c) Inner Nearness

$$IN = \frac{area(\Delta_{II}R)}{area(R)}$$

(d) Outer Nearness

$$ON = \frac{area(\Delta_{IE}R)}{area(R)}$$
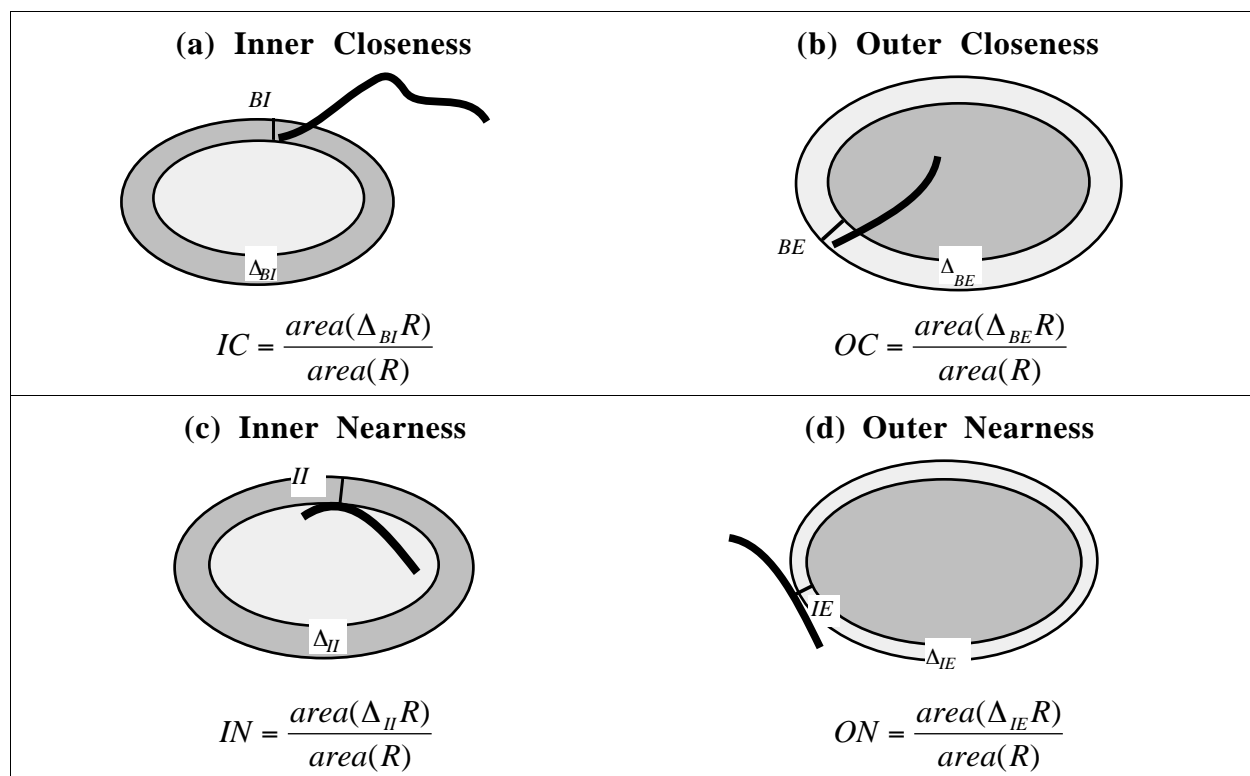
Figure 3: Metric refinements of topological relations: closeness ratios.

2.2.3 *Approximate Alongness*

A third set of metric parameters assesses line alongness and perimeter alongness when the line's interior does not coincide with the region's boundary, but runs parallel to it. This set of measures is called the *approximate alongness* and distinguishes four ratios:

- Inner approximate perimeter alongness (IPA) describes how the line's interior splits a buffer zone that extends from the region's boundary into the region's interior;
- inner approximate line alongness (ILA) describes how much of the line's interior falls within a buffer zone that extends from the region's boundary into the region's interior;

- outer approximate perimeter alongness (OPA) describes how the line's interior splits a buffer zone that extends from the region's boundary into the region's exterior, and
- outer approximate line alongness (OLA) describes how much of the line's interior falls within a buffer zone that extends from the region's boundary into the region's exterior.

All approximate alongness measures are expressed in terms of a ratio with respect to the perimeter of the buffer (approximate perimeter alongness) or the length of the line (approximate line alongness).
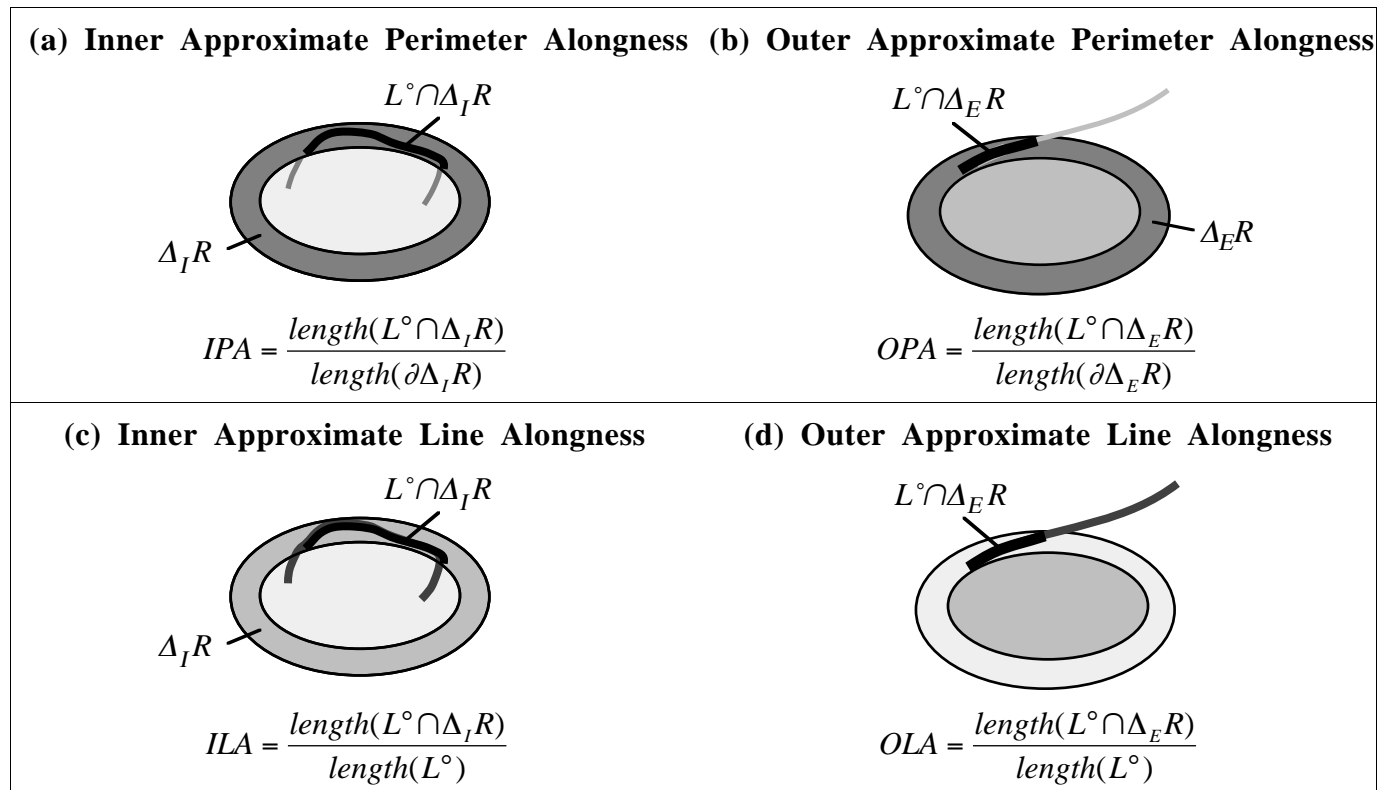
(a) Inner Approximate Perimeter Alongness

$$IPA = \frac{length(L° \cap \Delta_I R)}{length(\partial \Delta_I R)}$$

(b) Outer Approximate Perimeter Alongness

$$OPA = \frac{length(L° \cap \Delta_E R)}{length(\partial \Delta_E R)}$$

(c) Inner Approximate Line Alongness

$$ILA = \frac{length(L° \cap \Delta_I R)}{length(L°)}$$

(d) Outer Approximate Line Alongness

$$OLA = \frac{length(L° \cap \Delta_E R)}{length(L°)}$$

Figure 4: Metric refinements of topological relations: approximate alongness ratios.

2.3 Dependencies Among Topological Relations and Metric Parameters

There is a clear dependency among the topological relations and the metric parameters, because each of the 19 topological relations dictates what metric parameters are applicable. For example, if the line is completely contained in the region's boundary (LR 22), then the three metric parameters of region-boundary splitting, perimeter alongness, and line alongness describe details, while the remaining twelve metric parameters do not apply. For instance, outer closeness is not applicable to LR 22, because the line's boundary is not in the region's exterior. The mapping from the topological relations onto the metric parameters shows that there is no topological configuration for which all fifteen metric parameters would apply; nor is there a topological relation for which none of the metric parameters would be applicable. The topological relation LR 75—the line goes from the region's interior to its exterior—has the highest number of applicable metric parameters (10 out of 15 possible), whereas LR 22—the line is completely contained in the region's boudary—has the lowest count with only three pertinent metric parameters. Similarly distributed is the reverse mapping: perimeter alongness and line alongness apply to 13 out of 19 topological relations each, whereas inner and outer nearness only apply to one topological relation each (Table 1).

	IAS	EAS	ITS	ETS	RBS	PA	LA	IC	OC	IN	ON	IPA	ILA	OPA	OLA
LR 11									■		■			■	■
LR 12		■			■									■	■
LR 13			■						■					■	■
LR 22					■	■	■								
LR 31						■	■		■					■	■
LR 32		■				■	■							■	■
LR 33				■		■	■		■					■	■
LR 42	■					■						■	■		
LR 44								■		■		■	■		
LR 46			■									■	■		
LR 62	■				■	■	■					■	■		
LR 64						■	■	■				■	■		
LR 66			■			■	■					■	■		
LR 71	■								■			■	■	■	■
LR 72	■	■				■			■			■	■	■	■
LR 73	■			■					■			■	■	■	■
LR 74		■				■	■	■				■	■	■	■
LR 75		■	■			■	■	■				■	■	■	■
LR 76		■	■			■	■	■				■	■	■	■

Table 1: The dependencies among topological relations and metric parameters. Highlighted boxes mark those metric parameters that apply to a topological relation.

3. Calibration of Topological and Metric Parameters for 59 English-Language Terms

The splitting, closeness, and alongness ratios provide a framework for describing the metric properties of natural-language terms for line-region relations. The actual values of each parameter depend on the terms for which they are used. Some parameters may not be applicable at all to a specific configuration, while others may apply based on the geometry, but their values may not matter. The parameters that are of particular interest are those that are significant for a certain term. Also the value ranges of the applicable and significant parameters for each natural-language term are of interest, because this knowledge will allow us to determine the best spatial configurations for a particular term.

To obtain approximate values for the parameters for a set of natural-language terms, an experiment was conducted with 34 human subjects (non-geographically trained college students, approximately 1/4 female, 3/4 male). Details of the experiment are described in Mark and Egenhofer (1995). Subjects were presented with outlines of a park and an English-language sentence printed under each, describing a particular spatial relation between a road and a park. For each outline of the park, subjects were asked to draw a road such that the resulting drawing would conform to the spatial relation described in the sentence. Each questionnaire consisted of eight pages of eight drawings per page, i.e., altogether 64 sentences were tested for each subject. The sentences were accumulated from group descriptions from a spatial relations grouping task (Mark and Egenhofer 1994b) or were listed by several native English speakers. Out of the total possible of 2,176 responses, 2,129 responses were received. The responses of the 34 subjects for the 64 terms yielded 1,801 simple line drawings. The remaining 328 drawings had either cycles or networks and are not used in the present analysis. To facilitate further analysis, the low frequency terms (i.e., terms that had seven or less returns) were dropped, reducing the number of terms to be analyzed from 64 to 59 and the total drawings to be analyzed here to 1,777 (Table 2). For these cases, the majority of the returns per term ranged between 24 and 34. Among the 59 terms, 15 out of the 19 possible topological relations occurred, however, 93% of all cases are represented by just

six groups (in decreasing frequency): LR 75 (the line goes from the region's interior to its exterior), LR 11 (the line is completely contained in the region's exterior), LR 71 (the line goes from the region's exterior through the region's interior to the region's exterior again), LR 44 (the line is completely contained in the region's interior), LR 13 (the line goes from the region's exterior up to its boundary), and LR 46 (the line goes from the region's interior up to its boundary). To investigate the correlation between topological and metric properties, we focused on the most frequent topological cases and eliminated those with small counts (i.e., less than five returns), because their analysis would not have led to statistically significant results.

Spatial Term	Topological Relation															Count
	LR 11	LR 13	LR 31	LR 33	LR 42	LR 44	LR 46	LR 62	LR 64	LR 66	LR 71	LR 73	LR 74	LR 75	LR 76	
along edge	20		7			1					1					29
avoids	31						1									32
bisects							1	2			27	1		1		32
bypasses	31					1					1					33
comes from		1				1								30		32
comes into											2	1		28		31
comes out of											2			31	1	34
comes through						1					30			1		32
connected to	5	7	7			1					5			6		31
connects					1	9	2			1	3		5	2		23
contained in edge						8	1	1	2		1					13
contained within						11	4							1		16
crosses							1				28			2		31
cuts					3	2					28			1		34
cuts across					2	1					29			2		34
cuts through							1				29	2				32
divides							1				32			1		34
enclosed by						12					1					13
encloses						15	1									16
ends at	2	14				1	1						1	12		31
ends in		1				1	1							27		30
ends just inside		1				3								27		31
ends just outside	19										5	1		9		34
ends near	31	1					1							1		34
ends outside	14	1									2	2		14		33
enters		6					1				2			24		33
entirely outside	32					1										33
exits	2	2					1				1	1		27		34
goes across					1	4	3				24	1		1		34
goes away from	17	2		1		1					2			8		31
goes by	32		1											1		34
goes into		1									1			30		32
goes out of	1	5					1				1	6		19		33
goes through					3		2				28			1		34
goes to	3	14		2							1			14		34
goes up to	12	9		1			1				1			8		32
in						19	3				1	1				24
inside						23	2				1					26
intersect		4	2	1			1				16			6		30
intersects	2	1	1			1					19			3		27
leaves		1					1				2	1		29		34
near	29		4				1									34
outside	31						1									32
passes	29		1			1										31
runs across					7		4				19	2		2		34
runs along	21		6			1			1		1					30
runs along boundary	18		4	2		1									1	26
runs into	1	1					1				6			22		31
spans	3				8	3	4	2			8	2				30
splits					2	1	1	1			27	1		1		34
starts and ends in						24				1		1	2			28
starts in						7	2							23		32
starts just inside	1					3	1		1			2		23		31
starts just outside	13	1									3			14		31
starts near	27	2				1					1			1		32
starts outside	7	1									4			18		30
transects						1					28	1				30
traverses	2					2					21			2		27
within						22	1		1							24
Total	436	76	33	7	27	182	50	6	5	2	444	26	8	473	2	1777

Table 2: Frequency of topological relations for natural-language terms.

Based on the topology of the configurations, two main categories of terms can be differentiated. One category consists of terms with a consistent topology, i.e., they tend to aggregate towards a dominant topological relation. An example of this case is the term *avoids*, which was always represented by the same topological relation, LR 11 (where the line is completely contained in the region's exterior). The second category of terms consists of terms with ambiguous topology, which are not strongly represented by a single topological relation. Terms in the latter category tend to float among different topological relations and any single topological relation by itself cannot be used as a prototype to explain that particular term. For example, the term *enters* was represented by the topological relations LR 13 (the line goes from the region's exterior up to its boundary) and LR 75 (the line goes from the region's exterior into its interior). Most terms with ambiguous topology are represented by two or three topological relations, but one term, *connected to*, is dispersed among five different topological relations. Our purpose is to analyze for each term the topology and metric of the spatial configurations; however, no single topological relation can be used to represent terms with ambiguous topology. Terms with ambiguous topology are reclassified such that the topological relation of the reclassified terms are representative of the topology of their spatial configurations. Reclassified terms were tagged with their corresponding LR value, for example, the natural-language term *goes up to* is represented by three symbols: *goes up to*-11, *goes up to*-13, and *goes up to*-75.

4. Clusters of Spatial Terms

For the investigation of the effect of topology and metric on the 59 terms, we organized the terms into groups with common parameters. This grouping was done using cluster analysis, a statistical technique for detecting natural groupings in data, which attempts to arrange members with greater commonalty into the same cluster as compared to members in other clusters. Similarity among members is measured by the distance between these members based on a given set of parameters, from which clusters are formed by minimizing the internal distances among members in the same cluster, while maximizing the distance between clusters. The method used for partitioning members of a data set into clusters was the *k-means* method (MacQueen 1967), a partitioning algorithm that produces groups minimizing the sum of squared distances to each cluster's centroid.

The cluster analysis of the natural-language spatial relations was performed for the 38 spatial terms with consistent topology as well as the 46 spatial terms that resulted from the reclassification. As partitioning variables, it used the 15 metric parameters and the topological relations as the sixteenth parameter. Since the distributions of some parameters were skewed, the median value of each parameter, rather than the mean value, was computed, because in such situations the median is a better reflector of the central tendency measure, whereas the value of the mean would become influenced by outliers. Although the metric parameters are scale-independent, they had different ranges of values. To reduce this bias when computing distances between parameters, the parameters were standardized, using for each measurement x in the data set the difference to the mean of the set of data $\bar{x}$ divided by the standard deviation s (Equation 2). Standard units indicate the number of standard deviations that a particular measured value is above or below the mean of the dataset to which it belongs; therefore, the greater the difference from the mean, the greater the significance of the parameter.

$$x_s = \frac{x - \bar{x}}{s} \tag{2}$$

As our analysis was exploratory in nature, we experimented with several numbers of clusters, ranging from five to fifteen clusters. This analysis revealed a stabilization of the significant parameters for ten clusters (Table 3). It was also at this number of clusters that the last major cluster that could intuitively be explained (Cluster 10, *goes to*) appeared.

Variables	Sum of squared distances between clusters	Sum of squared distances within each cluster	F-Ratio
topological relation	82.995	0.005	135,174
perimeter alongness (PA)	82.789	0.211	3,232
inner nearness (IN)	82.218	0.782	864
region boundary splitting (RBS)	81.963	1.037	650
line alongness (LA)	81.590	1.410	476
inner closeness (IC)	80.854	2.146	310
interior area splitting (IAS)	78.928	4.072	159
exterior traversal splitting (ETS)	73.569	9.431	64
outer approximate perimeter alongness (OPA)	73.047	9.953	60
outer nearness (ON)	69.820	13.180	44
outer approximate line alongness (OLA)	68.964	14.036	40
interior traversal splitting (ITS)	67.465	15.535	36
outer closeness (OC)	51.246	31.754	13
exterior area splitting (EAS)	83.000	0.000	*undefined*
inner approximate line alongness (ILA)	83.000	0.000	*undefined*
inner approximate perimeter alongness (IPA)	83.000	0.000	*undefined*

Table 3:   Summary statistics for ten clusters (sorted by decreasing F-ratios).

   In this clustering, parameters with higher values of the F-ratios are better discriminators among spatial terms than parameters with lower values of the F-ratios. The topological relations come out as the strongest discriminators—approximately 22 times stronger than all metric parameters combined—which confirms the underlying assumption that topology is more critical for the semantics of spatial relations than metric (Mark and Egenhofer 1994a; Egenhofer and Mark 1995). Among the metric parameters, three parameters—exterior area splitting (EAS), inner approximate line alongness (ILA), and inner approximate perimeter alongness (IPA)—could not be assessed for their relative importance, because the clusters relying on these measures have only one term each and, therefore, their deviation from the best term in the cluster is 0. Among the other clusters, the perimeter alongness (PA), inner nearness (IN), and region boundary splitting (RBS) are better discriminators among natural-language spatial terms than the outer nearness (ON), interior traversal splitting (ITS), and outer closeness (OC). A possible explanation is that such parameters as perimeter alongness are used for a narrower number of natural-language spatial terms, for instance, "The road runs along the edge of the park." On the other hand, the usage of the outer closeness is more general and applies to a wider number of terms such as *ends outside, goes through,* and *goes across.*

5.   Mapping Metric Parameters onto Individual Spatial Terms

Since topology dominates the formation of the ten clusters, the topological relation of a configuration was selected for all spatial terms as a relevant parameter. The clustering also identified what metric parameters are critical to describe the geometry of the terms in a cluster. To discriminate the influence of parameters, we defined any metric parameter with a standard score greater than one to be a *significant parameter,* because the mean of such a parameter is at least one standard deviation higher than the mean of the entire data set. Parameters with a range of values between zero and one were used as *supporting parameters,* because their values influence the outcome less strongly than the significant parameters. The other metric parameters, whose normalized values fall below zero, were not considered as relevant to describe the terms of a cluster. Table 4 shows as an example the configuration for Cluster 1 (consisting of configurations that map onto LR 71—the line goes from the region's exterior through the region's interior to its exterior again—and LR 73—the line goes from the region's exterior through the interior up to the region's boundary) with the distribution of the normalized values for each parameter. For the terms in this cluster, only the interior area splitting (IAS) is significant, whereas outer closeness (OC) is a supporting parameter.

Metric parameters	Normalized values		
	min	max	mean
IAS	0.22	2.21	1.58
EAS	-0.11	-0.11	-0.11
ITS	-0.55	-0.55	-0.55
ETS	-0.57	-0.57	-0.57
RBS	-0.15	-0.15	-0.15
PA	-0.19	-0.19	-0.19
LA	-0.19	-0.19	-0.19
IC	-0.78	-0.78	-0.78
OC	-0.91	0.51	-0.27
IN	-0.36	-0.36	-0.36
ON	-0.49	-0.49	-0.49
IPA	-0.11	-0.11	-0.11
ILA	-0.11	-0.11	-0.11
OPA	-0.32	-0.32	-0.32
OLA	-0.31	-0.31	-0.31

Table 4: The distribution the normalized values of the metric parameters in Cluster 1 consisting of the topological relations LR 71 (the line goes from the region's exterior through the region's interior to the region's exterior again) and LR 73 (the line goes from the region's exterior through the interior up to the region's boundary).

Corresponding values were obtained for the other nine clusters (Shariff 1996). All significant and supporting parameters are compatible with the topological relations of their clusters, because for each cluster the important metric parameters are a subset of those that apply based on the cluster's topological relations (i.e., Table 5 describes subsets of Table 1).

Cluster	Topological relations	IAS	EAS	ITS	ETS	RBS	PA	LA	IC	OC	IN	ON	IPA	ILA	OPA	OLA
1	LR 71, LR 73	■								░						
2	LR 74		■													
3	LR 31						■	■								
4	LR 75			■	■				■	░					■	
5	LR 11									░	■				■	■
6	LR 42	■				■										
7	LR 44								░		■		■	■		
8	LR 44								░		■					
9	LR 11								░		■					
10	LR 13								■							

Table 5: Cluster classification with significant (■) and supporting (░) metric parameters.

Since the significant and supporting parameters are critical to describing the clusters, no two clusters should have the same set of significant and supporting parameters. The summary of the significant and supporting parameters of all ten clusters (Table 5) confirms this assumption. There are, however, two pairs of clusters—Clusters 5 and 9, and Clusters 7 and 8— that differ only by

additional significant parameters (i.e., Clusters 9 and 8 provide refinements of Clusters 5 and 7, respectively).

The clustering also relates the individual spatial terms to the clusters and, therefore, to their significant and supporting parameters. Since the clustering process groups terms with similar geometric properties together, all terms within the same cluster are assumed to respond to the same metric parameters; however, the values of these parameters may differ for terms grouped in the same cluster. These differences in the values are critical to distinguish terms within the same cluster. If two different spatial terms, grouped into the same cluster, have the same values for their significant and supporting parameters, the terms would respond to the same geometry and, therefore, be synonyms. To compare parameter values, the ranges and medians of the significant and supporting parameters of each spatial terms were derived from the values obtained from the sketch experiment. Table 6 shows as an example the value ranges for the terms associated with Cluster 1. The stacked intervals of the ranges and median values of the significant and supporting parameters give a quick picture of the similarities and differences of the terms in a cluster. For example, for Cluster 1 they show that *ends just outside* and *spans* are significantly different from the other terms in this cluster due to the ranges of the interior area splitting. The distribution of the medians for the interior area splitting also shows that *runs into* differs notably from the other terms in this cluster. Corresponding values were obtained for the other nine clusters (Shariff 1996), resulting in the *Metric Table of Spatial Terms* (Table 7). It serves as a dictionary for the geometry of the 59 terms, because it identifies for each term its relevant topological relation(s), as well as the significant and supporting metric parameters with their value ranges and median values. All terms with ambiguous topology were assigned to multiple clusters, thereby enabling differentiations for metric parameters that were impossible before. Another important result it that only a small subset (approximately 12%) of all possible metric parameters are significant and an even small percentage (less than 6%) making the set of supporting parameters. This means that despite 15 potential metric parameters per spatial term, only a few—on average less than three—are necessary to link a geometric configuration with a spatial term.
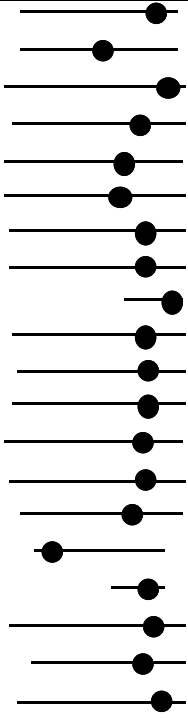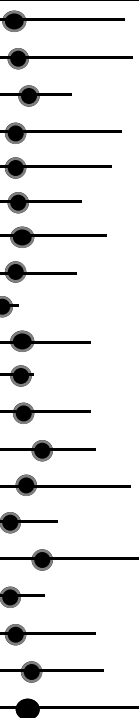
Spatial term	Topological relation	IAS min	max	median		OC min	max	median	
bisects	LR 71	0.05	0.48	0.42		0.40	8.78	1.49	
comes through	LR 71	0.05	0.48	0.27		0.51	9.29	1.83	
connected to	LR 71	0.01	0.50	0.45		0.27	5.24	2.49	
crosses	LR 71	0.03	0.50	0.37		0.29	8.62	1.63	
cuts	LR 71	0.01	0.49	0.33		0.68	7.85	1.62	
cuts through	LR 71	0.01	0.50	0.32		0.41	6.04	1.75	
cuts across	LR 71	0.02	0.50	0.39		0.19	7.62	1.58	
divides	LR 71	0.02	0.50	0.39		0.33	5.58	1.69	
ends just outside	LR 71	0.34	0.49	0.46		0.24	1.77	0.49	
goes across	LR 71	0.03	0.50	0.39		0.25	6.55	2.01	
goes out	LR 73	0.04	0.50	0.40		0.36	2.77	1.95	
goes through	LR 71	0.03	0.50	0.40		0.70	6.21	2.17	
intersect	LR 71	0.01	0.49	0.38		0.41	6.87	3.36	
intersects	LR 71	0.02	0.50	0.39		0.25	9.10	2.25	
runs across	LR 71	0.05	0.49	0.35		0.27	4.31	1.32	
runs into	LR 71	0.09	0.44	0.13		0.54	11.96	3.42	
spans	LR 71	0.30	0.43	0.40		0.54	3.58	1.30	
splits	LR 71	0.02	0.50	0.41		0.42	6.90	1.62	
transects	LR 71	0.08	0.50	0.38		0.53	7.32	2.57	
traverses	LR 71	0.04	0.50	0.43		0.56	10.03	2.38	

Table 6:   Parameter ranges for the terms in Cluster 1.

6.   Verification of the Significant Parameters

The significant parameters that have been determined are based on one experimental data set. In order to validate these findings, we compare them with another independent data set. The experimental data used for this comparison are based on results from another human-subjects experiment, the agreement tasks for English-language spatial terms (Mark and Egenhofer 1994b). In these agreement tasks, subjects were presented with a sentence in English that described a relation between a road and a park. Subjects were asked to compare the sentence with each of the given 60 diagrams and then to evaluate their agreement or disagreement on a 5-step scale. These tests give us an opportunity to identify how well the metric parameters are able to explain the semantics of the spatial terms tested in the agreement task.

For any spatial term *X*, the results of the cluster analysis provide the significant parameters. Similarly, the set of non-significant parameters for the term *X*, can be inferred from these results. This knowledge is used to select configurations from the agreement task, that have all the significant parameters, set *A*, and those that do not have the set of significant parameters, set *B*. The mean agreement ratings for the sets *A* and *B*, for the spatial term *X* is computed to determine if the set *A* has a higher agreement rating than the set *B*. In analyzing the metric parameters, a strong agreement rating for the spatial term is a good indicator of the suitability of the metric parameters in resolving the semantics of the particular spatial term. It also reflects that the metric parameters are consistent and stable in their usage for both the prototype and the agreement task.

For five spatial terms—*goes through, enters, goes along, inside,* and *outside*—the mean agreement ratings for these configurations (μ_0), the sample size (n_1), the sample mean ($\bar{x}$), and the standard deviation (s_1) were determined from the agreement task. The level of significance used for this testing was 0.5, the value of $-Z_{0.025} = -1.96$ and $Z_{0.025} = 1.96$. For all five spatial terms, the null hypothesis is μ_0 = mean of agreement ratings for all configurations with the relevant

significant parameters (assumed as population mean), while the alternative hypothesis is $\mu_0 \neq$ assumed population mean. The value of Z can be computed using the test statistic (Equation 3). If the computed value of Z is less than $-Z_{0.025}$ or greater than $Z_{0.025}$, the null hypothesis is rejected.

$$Z = \frac{(\bar{x} - \mu_0)\sqrt{n_1}}{s_1} \tag{3}$$

For all the five spatial terms tested, the computed value of Z is less than -1.96, therefore, the null hypothesis is rejected for all five cases (Table 8). As such, the presence of the significant parameters for the corresponding spatial terms results in a significant increase in the agreement ratings as compared to configurations where these corresponding parameters are not present.

Spatial term	Significant parameters	μ_0	n_1	$\bar{x}$	s_1	$Z_{computed}$
through	IAS	0.696	42	0.397	0.245	-7.910
enters	ITS, ETS, IC	0.709	44	0.513	0.214	-6.075
goes along	PA, LA	0.530	32	0.310	0.134	-9.287
inside	IN	0.885	58	0.546	0.259	-9.968
outside	ON	0.970	56	0.393	0.252	-17.134

Table 8: Results of significance test.

This finding is important as it implies that the metric parameters are found to be significant not only on an independent data set, but on a data set that was based on a different kind of a test. The agreement task accommodated a range of answers from the subjects, while the prototype test only allowed a single best example to be drawn. Based on the assumption that a person will use the best term when making a query about spatial relations in a spatial database, the results of the prototype test can be used for spatial queries. On the other hand, if the best scene descriptor about spatial relations needs to be generated from a GIS, then the results from the agreement task provide natural-language spatial descriptors on a scale graduated for human acceptability. The ramification for GIS is that the metric parameters are suitable for conducting spatial relations queries as well as generating the best natural-language spatial terms for describing spatial relations in a scene.

7. Processing Spatial Queries with Natural-Language Spatial Terms

The calibration of the model for spatial terms and its verification showed that configurations people commonly use as prototypes for representing spatial terms in the English language are also those configurations that they can readily agree upon to reflect a particular spatial term. The testing of prototypes implies that in the execution of natural-language queries, people are more likely to use the prototypical terms as descriptors for their queries. As such, the results of the prototypical test provide us with a foundation for designing natural-language queries in GISs.

In order to integrate such natural-language spatial relations into query languages, several extensions to current query languages are necessary to incorporate the terminology into the syntax, map spatial terms onto executable database queries, and process the results of such queries. For the integration of user-defined spatial terms, latest extensions of standard spatial query languages, such as those proposed for SQL3/Multimedia (ISO 1996), provide a foundation with explicit topological relations. The following three steps are necessary to process a spatial query with natural-language spatial relations as constraints:
- based on the spatial term of the query, one must determine the relevant topological and metric parameters and their values from the Metric Table of Spatial Terms (Table 7);
- the values of these parameters must be translated into an SQL query that selects all configurations fulfilling the query constraints; and
- the query result must be sorted to distinguish better from not-so-good matches.

Since supporting parameters are less important for capturing the semantics of spatial terms than significant parameters, supporting parameters are used only in the sorting of query results, but not in the selection of candidate configurations. This ensures that configurations with a very good match with respect to the significant parameters are retrieved even if they fall outside the ranges of the supporting parameters.

The following example illustrates this process. Suppose a user asks the query, "Show all roads that *go to* the park." The query formulation may be through a form-based user interface or some other high-level query language. To process this query, the term *go to* must be mapped onto its corresponding geometric representations and translated into an executable database query. The first step in processing the user's request is to select for the term *goes to* the applicable parameters from the Metric Table of Spatial Terms. For *goes to*, there are two topological relations, LR 13 and LR 75. For the first case, the only significant metric parameter is the outer closeness, which must be between 1.16 and 16.74. For the second configuration, LR 75, three metric parameters are significant: interior traversal splitting (between 0.04 and 0.57), exterior traversal splitting (between 0.44 and 0.96), and inner closeness (between 0.18 and 0.99). These parameters get incorporated into the WHERE clause of an SQL-like query.

```
SELECT   road.geometry, park.geometry
FROM     road, park
WHERE    (topological_relation (road, park) = LR_75 AND
             0.04 ≤ ITS (road, park) ≤ 0.57 AND
             0.44 ≤ ETS (road, park) ≤ 0.96 AND
             0.18 ≤ IC (road, park) ≤ 0.99)
         OR
         (topological_relation (road, park) = LR_13 AND
             1.16 ≤ OC (road, park) ≤ 16.74)
```

Terms such as `topological_relation`, `ITS`, `ETS`, `IC`, and `OC` must be part of the syntax of the extended spatial query language. The result of this query is a set of road-park tuples. Some of these configurations provide a better match for the query than others. Since the query acts as a filter, it is further necessary to sort through the query result during a subsequent *query result prioritization* and to rank the configurations retrieved according to their similarity with the query. To assess the best match, the query results are prioritized by least deviations from the medians of the significant and the supporting parameters.

8.   Conclusions

This paper defined an approach that refines the 9-intersection model to capture the semantics of natural-language spatial terms based on their geometry. It built upon the foundations for topological properties of spatial-relation terms (Mark and Egenhofer 1994a; Mark and Egenhofer 1994b) and defined a complimentary metric approach, which consists of 15 metric parameters. The testing of these metric parameters on data obtained through human-subject testing revealed that all the parameters were relevant in defining the semantics of natural-language spatial terms, though no single term responded to all parameters. The clustering of natural-language terms, the determination of their ranges, and median values, all based on these parameters, also reveal the inherent metric dependencies of the natural-language terms being tested.

The major conclusions about the nature of natural-language spatial relations are:
- Topology is a more important influence for a large set of spatial-relation terms than metric.
- Many spatial-relation terms fall under the same topology, but have different metric parameters; therefore, the metric parameters are critical to distinguish between such similar configurations.

- Several spatial-relation terms were found to have similar values for the topological and metrical parameters—although no pair of the 59 terms tested had exactly the same values—which indicates that multiple terms may be used interchangeably to describe a spatial configuration.

While the model for the natural-language spatial relations contributes primarily to a better understanding of the semantics of natural-language spatial relations, it has several practical implications for future GIS design. The focus of current GISs on presenting every spatial configuration in a map-like representation leaves users with the task of interpreting the result displayed. Speech as a complementary interaction modality may provide complementary information. (Egenhofer and Kuhn 1998). The combined topological and metric formalism has been shown to be able to answer queries or describe scenes using natural-language spatial terms. As all the parameters can be easily determined with computational-geometry algorithms built into current GISs, the model is a step toward such innovative, futuristic GIS user interfaces as a "Talking GIS." In a simlar vein, future generations of car navigation systems may be another beneficiary of this formalism. Because current voice interaction mechanisms in car navigation systems are static, as the verbal instructions are pre-recorded, the metric parameters allow for a dynamic interaction with scenes. As each spatial term generated or query answered is based on the analysis of that particular scene in real time, the response can be based on spatial terms that have been calibrated to mimic human thinking.

In order to work as the basis for a comprehensive GIS query language, several extensions of the model are required. While the model addresses the influences of topology and metric, it currently lacks any considerations of orientation; therefore, no distinction can be made between such pairs of terms as *ends outside* and *starts outside*. Through the addition of an orientation parameter, such differences could be covered. Line-region relations have been thoroughly studied and similar investigations for region-region relations and line-line relations are necessary. Of particular interest should be studies of how spatial terms apply across different representations (e.g., when the geometry of the objects changes with higher resolutions and increased detail).

The efficient implementation of such high-level operators in spatial database systems requires appropriate indices. Current indices for spatial databases are tailored for window queries and neighborhood searches. They are insufficient to guarantee reasonable performance for the types of queries discussed here; therefore, indices over topological relations and, if necessary, certain metric parameters need to be developed and tested.

A theory of the dependencies between object classes and spatial relations is another area for future research. Up to now, the work on natural-language spatial relations was based on Talmy's (1983) assumption that spatial predicates are used independent of size and material; however, examples can be constructed that question parts of this generalization. For example, there is a significant difference in the semantics of "we went *through* Canada" and "the ant went *through* my hand." If one substitutes, however, *through* with *across*, the differences are much more subtle.

9. References

CLEMENTINI, E., AND DI FELICE, P., 1996, An Algebraic Model for Spatial Objects with Undetermined Boundaries. In *Geographic Objects with Indeterminate Boundaries*, edited by P. Burrough and A. Frank, (London: Taylor & Francis), pp. 155-169.

CLEMENTINI, E., DI FELICE, P., AND CALIFANO, G., 1995, Composite Regions in Topological Queries. *Information Systems*, **20**, 579-594.

CLEMENTINI, E., DI FELICE, P., AND VAN OOSTEROM, P., 1993, A Small Set of Formal Topological Relationships Suitable for End-User Interaction. In *Third International Symposium on Large Spatial Databases, SSD '93*, edited by D. Abel and B. C. Ooi, *Lecture Notes in Computer Science* 692 (Berlin: Springer-Verlag), pp. 277-295.

COHN, A., 1995, A Hierarchical Representation of Qualitative Shape Based on Connection and Convexity. In *Spatial Information Theory—A Theoretical Basis for GIS, International Conference COSIT '95, Semmering, Austria*, edited by A. Frank and W. Kuhn, *Lecture Notes in Computer Science* 988 (Berlin: Springer-Verlag), pp. 311-327.

COHN, A., AND GOTTS, N., 1996, The "Egg-Yolk" Representation of Regions with Indeterminate Boundaries. In *Geographic Objects with Indeterminate Boundaries*, edited by P. Burrough and A. Frank, (London: Taylor & Francis), pp. 171-187.

CUI, Z., COHN, A., AND RANDELL, D., 1993, Qualitative and Topological Relationships in Spatial Databases. In *Third International Symposium on Large Spatial Databases*, edited by D. Abel and B. Ooi, *Lecture Notes in Computer Science* 692 (Berlin: Springer-Verlag), pp. 296-315.

DE HOOP, S., AND VAN OOSTEROM, P., 1992, Storage and Manipulation of Topology in Postgres. In *Proceedings of Third European Conference on Geographical Information Systems, EGIS '92*, pp. 1324-1336.

EGENHOFER, M., 1996, Multi-Modal Spatial Querying. In *Proceedings of Seventh International Symposium on Spatial Data Handling*, edited by M.-J. Kraak and M. Molenaar (London: Taylor & Francis), pp. 785-799.

EGENHOFER, M., AND FRANZOSA, R., 1991, Point-Set Topological Spatial Relations. *International Journal of Geographical Information Systems*, **5**, 161-174.

EGENHOFER, M., AND FRANZOSA, R., 1995, On the Equivalence of Topological Relations. *International Journal of Geographical Information Systems*, **9**, 133-152.

EGENHOFER, M., AND HERRING, J., 1990, A Mathematical Framework for the Definition of Topological Relationships. In *Proceedings of Fourth International Symposium on Spatial Data Handling*, edited by K. Brassel and H. Kishimoto, pp. 803-813.

EGENHOFER, M., AND HERRING, J., 1991, Categorizing Binary Topological Relationships Between Regions, Lines, and Points in Geographic Databases. In *A Framework for the Definition of Topological Relationships and an Algebraic Approach to Spatial Reasoning within this Framework, NCGIA Technical Report 91-7*, edited by M. Egenhofer, J. Herring, T. Smith, and K. Park, (Santa Barbara, CA: National Center for Geographic Information and Analysis).

EGENHOFER, M., AND HERRING, J., 1993, Querying a Geographical Information System. In *Human Factors in Geographical Information Systems*, edited by D. Medyckyj-Scott and H. Hearnshaw, (London: Belhaven Press), pp. 124-136.

EGENHOFER, M., AND KUHN, W., 1998, Interacting with Geographic Information Systems. In *Geographical Information Systems: Principles, Techniques, Management and Applications*, edited by P. Longley, M. Goodchild, D. Maguire, and D. Rhind, (London: GeoInformation International), (in press).

EGENHOFER, M., AND MARK, D., 1995, Naive Geography. In *Spatial Information Theory—A Theoretical Basis for GIS, International Conference COSIT '95, Semmering, Austria*, edited by A. Frank and W. Kuhn, *Lecture Notes in Computer Science* 988 (Berlin: Springer-Verlag), pp. 1-15.

EGENHOFER, M., AND SHARIFF, R., 1998, Metric Details for Natural-Language Spatial Relations. *ACM Transactions on Information Systems*, **16**, (in press).

FRANK, A., 1991, Qualitative Spatial Reasoning about Cardinal Directions. In *Proceedings of Autocarto 10*, edited by D. Mark and D. White, pp. 148-167.

FRANK, A., 1992, Qualitative Spatial Reasoning about Distances and Directions in Geographic Space. *Journal of Visual Languages and Computing*, **3**, 343-371.

FREEMAN, J., 1975, The Modelling of Spatial Relations. *Computer Graphics and Image Processing*, **4**, 156-171.

FREKSA, C., 1992, Using Orientation Information for Qualitative Spatial Reasoning. In *Theories and Methods of Spatio-Temporal Reasoning in Geographic Space*, edited by A. Frank, I. Campari, and U. Formentini, *Lecture Notes in Computer Science* 639 (Berlin: Springer-Verlag), pp. 162-178.

HADZILACOS, T., AND TRYFONA, N., 1992, A Model for Expressing Topological Integrity Constraints in Geographic Databases. In *Theories and Methods of Spatio-Temporal Reasoning in Geographic Space*, edited by A. Frank, I. Campari, and U. Formentini, *Lecture Notes in Computer Science* 639 (Berlin: Springer-Verlag), pp. 252-268.

HAZELTON, N. W., BENNETT, L., AND MASEL, J., 1992, Topological Structures for 4-Dimensional Geographic Information Systems. *Computers, Environment, and Urban Systems*, **16**, 227-237.

HERNÁNDEZ, D., 1991, Relative Representation of Spatial Knowledge: The 2-D Case. In *Cognitive and Linguistic Aspects of Geographic Space*, edited by D. Mark and A. Frank, (Dordrecht: Kluwer Academic Publishers), pp. 373-385.

HERNÁNDEZ, D., 1994, *Qualitative Representation of Spatial Knowledge*, *Lecture Notes in Articifical Intelligence* 804 (Berlin: Springer-Verlag).

HERNÁNDEZ, D., CLEMENTINI, E., AND DI FELICE, P., 1995, Qualitative Distances. In *Spatial Information Theory—A Theoretical Basis for GIS, International Conference COSIT '95, Semmering, Austria*, edited by A. Frank and W. Kuhn, *Lecture Notes in Computer Science* 988 (Berlin: Springer-Verlag), pp. 45-57.

HERRING, J., 1991, The Mathematical Modeling of Spatial and Non-Spatial Information in Geographic Information Systems. In *Cognitive and Linguistic Aspects of Geographic Space*, edited by D. Mark and A. Frank, (Dordrecht: Kluwer Academic Publishers), pp. 313-350.

HERSKOVITS, A., 1986, *Language and Spatial Cognition—An Interdisciplinary Study of the Prepositions in English*, (Cambridge, MA: Cambridge University Press).

HONG, J.-H., EGENHOFER, M., AND FRANK, A., 1995, On the Robustness of Qualitative Distance- and Direction Reasoning. In *Proceedings of Autocarto 12*, edited by D. Peuquet (Bethesda, MD: American Society for Photogrammetry and Remote Sensing and American Congress on Surveying and Mapping), pp. 301-310.

ISO, 1996, *SQL Multimedia and Application Packages—Part 3: Spatial.* ISO/IEC JTC 1/SC 21 N 10441,

KEIGHAN, E., 1993, *Managing Spatial Data within the Framework of the Relational Model.* Technical Paper, Oracle Corporation, Canada.

MACQUEEN, J., 1967, Some Methods for Classification and Analysis of Multivariate Observations. In *Fifth Berkeley Symposium on Mathematical Statistics and Probability*, edited by L. Le Cam and J. Neyman, 1 (Berkeley, CA: University of California Press), pp. 281-297.

MARK, D., COMAS, D., EGENHOFER, M., FREUNDSCHUH, S., GOULD, M., AND NUNES, J., 1995, Evaluating and Refining Computational Models of Spatial Relations Through Cross-Linguistic Human-Subject Testing. In *Spatial Information Theory—A Theoretical Basis for GIS, International Conference COSIT '95, Semmering, Austria*, edited by A. Frank and W. Kuhn, *Lecture Notes in Computer Science* 988 (Berlin: Springer-Verlag), pp. 553-568.

MARK, D., AND EGENHOFER, M., 1994a, Calibrating the Meanings of Spatial Predicates from Natural Language: Line-Region Relations. In *Proceedings of Sixth International Symposium on Spatial Data Handling*, edited by T. Waugh and R. Healey, pp. 538-553.

MARK, D., AND EGENHOFER, M., 1994b, Modeling Spatial Relations Between Lines and Regions: Combining Formal Mathematical Models and Human Subjects Testing. *Cartography and Geographic Information Systems*, **21**, 195-212.

MARK, D., AND EGENHOFER, M., 1995, Topology of Prototypical Spatial Relations Between Lines and Regions in English and Spanish. In *Proceedings of Autocarto 12*, edited by D. Peuquet (Bethesda, MD: American Society for Photogrammetry and Remote Sensing and American Congress on Surveying and Mapping), pp. 245-254.

MARK, D., AND GOULD, M., 1991, Interaction with Geographic Information: A Commentary. *Photogrammetric Engineering & Remote Sensing*, **57**, 1427-1430.

NABIL, M., SHEPHERD, J., AND NGU, A., 1995, 2D Projection Interval Relationships: A Symbolic Representation of Spatial Relationships. In *Advances in Spatial Databases—4th International Symposium, SSD '95, Portland, ME*, edited by M. Egenhofer and J. Herring, *Lecture Notes in Computer Science* 951 (Berlin: Springer-Verlag), pp. 292-309.

NABIL, N., NGU, A., AND SHEPHERD, J., 1996, Picture Similarity Retrieval Using the 2D Projection Interval Representation. *IEEE Transactions on Knowledge and Data Engineering*, **8**, 533-539.

PAPADIAS, D., EGENHOFER, M., AND SHARMA, J., 1996, Hierarchical Reasoning about Direction Relations. In *Proceedings of Fourth ACM Workshop on Advances in Geographic Information Systems*, edited by S. Shekhar and P. Bergougnoux (New York: ACM Press), pp. 107-114.

PAPADIAS, D., AND SELLIS, T., 1993, The Semantics of Relations in 2D Space Using Representative Points: Spatial Indexes. In *Spatial Information Theory, European Conference COSIT '93, Marciana Marina, Elba Island, Italy*, edited by A. Frank and I. Campari, 716 (Berlin: Springer-Verlag), pp. 234-247.

PAPADIAS, D., AND SELLIS, T., 1994, Qualitative Representation of Spatial Knowledge in Two-Dimensional Space. *VLDB Journal*, **3**, 479-516.

PEUQUET, D., 1986, The Use of Spatial Relationships to Aid Spatial Database Retrieval. In *Proceedings of Second International Symposium on Spatial Data Handling*, edited by D. Marble, pp. 459-471.

PULLAR, D., AND EGENHOFER, M., 1988, Towards Formal Definitions of Topological Relations Among Spatial Objects. In *Third International Symposium on Spatial Data Handling*, edited by D. Marble, pp. 225-242.

RANDELL, D., CUI, Z., AND COHN, A., 1992, A Spatial Logic Based on Regions and Connection. In *Proceedings of Principles of Knowledge Representation and Reasoning, KR '92*, edited by B. Nebel, C. Rich, and W. Swartout, pp. 165-176.

REGIER, T., 1995, A Model of the Human Capacity for Categorizing Spatial Relations. *Cognitive Linguistics*, **6**, 63-88.

SHARIFF, R., 1996, *Natural Language Spatial Relations: Metric Refinements of Topological Properties*. Ph.D. Thesis, Department of Spatial Information Science and Engineering, University of Maine, Orono, ME.

SHARMA, J., 1996, Integrated Topology- and Direction-Reasoning in GIS. In *ESF-NSF Young Scholars Summer Institute*, edited by M. Craglia and H. Onsrud, (London: Taylor & Francis), (in press).

TALMY, L., 1983, How Language Structures Space. In *Spatial Orientation: Theory, Research, and Application*, edited by H. Pick and L. Acredolo, (New York: Plenum Press), pp. 225-282.

WORBOYS, M., 1992, A Geometric Model for Planar Geographical Objects. *International Journal of Geographical Information Systems*, **6**, 353-372.

ZIMMERMANN, K., 1993, Enhancing Qualitative Spatial Reasoning—Combining Orientation and Distance. In *Spatial Information Theory, European Conference COSIT '93, Marciana Marina, Elba Island, Italy*, edited by A. Frank and I. Campari, *Lecture Notes in Computer Science* 716 (Berlin: Springer-Verlag), pp. 69-76.

	LR	IAS	EAS	ITS	ETS	RBS	PA	LA	IC	OC	IN	ON	IPA	ILA	OPA	OLA
along edge	11									0.32 11.34 (1.23)		0.08 1.18 (0.36)			0.20 1.00 (0.40)	0.20 1.00 (0.80)
along edge	31						0.01 0.20 (0.17)	0.01 0.59 (0.40)								
avoids	11									1.02 17.13 (8.34)		0.26 2.96 (0.73)			0.10 0.80 (0.40)	0.10 0.70 (0.40)
bisects	71	0.05 0.48 (0.42)								0.40 8.78 (1.49)						
bypasses	11									0.67 18.31 (4.01)		0.28 1.36 (0.76)			0.10 0.70 (0.40)	0.20 0.80 (0.50)
comes from	75			0.03 0.78 (0.41)	0.22 0.97 (0.60)				0.21 0.99 (0.81)	1.75 21.53 (5.54)						
comes into	75			0.09 0.74 (0.40)	0.26 0.91 (0.60)				0.31 0.99 (0.86)	0.84 14.89 (4.96)						
comes out of	75			0.08 0.91 (0.47)	0.09 0.92 (0.54)				0.31 0.99 (0.89)	0.50 15.30 (3.57)						
comes through	71	0.05 0.48 (0.27)								0.51 9.29 (1.83)						
connected to	11									0.03 0.23 (0.03)		0.03 0.23 (0.03)				
connected to	13									3.30 9.87 (5.51)						
connected to	31						0.15 0.54 (0.46)	0.09 0.19 (0.19)								

	LR	IAS	EAS	ITS	ETS	RBS	PA	LA	IC	OC	IN	ON	IPA	ILA	OPA	OLA
connected to	71	0.01 0.50 (0.45)								0.27 5.24 (2.49)						
connected to	75			0.09 0.55 (0.36)	0.45 0.91 (0.64)				0.29 0.96 (0.73)	1.78 8.34 (3.90)						
connects	44								0.25 0.62 (0.51)		0.25 0.61 (0.47)					
connects	74		0.04 0.14 (0.11)													
contained in edge	44								0.16 0.51 (0.28)		0.15 0.51 (0.28)		0.20 0.70 (0.40)	1.00 1.00 (1.00)		
contained within	44								0.21 0.68 (0.49)		0.19 0.55 (0.46)					
crosses	71	0.03 0.50 (0.37)								0.29 8.62 (1.63)						
cuts	71	0.01 0.49 (0.33)								0.68 7.85 (1.62)						
cuts across	71	0.02 0.50 (0.39)								0.19 7.62 (1.58)						
cuts through	71	0.01 0.50 (0.32)								0.41 6.04 (1.75)						
divides	71	0.02 0.50 (0.39)								0.33 5.58 (1.69)						
enclosed by	44								0.19 0.79 (0.57)		0.19 0.79 (0.43)					

	LR	IAS	EAS	ITS	ETS	RBS	PA	LA	IC	OC	IN	ON	IPA	ILA	OPA	OLA
encloses	44								0.25 0.86 (0.51)		0.17 0.66 (0.46)					
ends at	13									1.95 15.59 (7.19)						
ends at	75			0.02 0.65 (0.24)	0.36 0.98 (0.76)				0.19 0.98 (0.79)	2.09 20.73 (7.69)						
ends in	75			0.17 0.91 (0.51)	0.09 0.83 (0.49)				0.18 0.99 (0.81)	0.59 9.94 (3.66)						
ends just inside	75			0.05 0.71 (0.16)	0.29 0.95 (0.84)				0.05 0.89 (0.55)	0.86 10.27 (3.95)						
ends just outside	11									0.21 2.81 (0.57)		0.21 1.71 (0.57)				
ends just outside	71	0.34 0.49 (0.46)								0.24 1.77 (0.49)						
ends just outside	75			0.46 0.93 (0.88)	0.07 0.54 (0.12)				0.25 0.96 (0.74)	0.29 4.93 (1.12)						
ends near	11									0.22 6.73 (0.78)		0.05 2.23 (0.57)				
ends outside	11									0.04 1.82 (0.89)		0.04 1.82 (0.87)				
ends outside	75			0.47 0.90 (0.67)	0.10 0.53 (0.33)				0.07 0.90 (0.59)	0.63 6.79 (1.83)						
enters	13									3.28 11.27 (5.21)						

	LR	IAS	EAS	ITS	ETS	RBS	PA	LA	IC	OC	IN	ON	IPA	ILA	OPA	OLA
enters	75			0.18 0.83 (0.46)	0.17 0.82 (0.54)				0.42 0.99 (0.86)	1.06 9.50 (2.47)						
entirely outside	11									0.40 10.80 (2.46)		0.25 2.28 (0.85)			0.10 0.90 (0.30)	0.10 1.00 (0.60)
exits	75			0.13 0.84 (0.56)	0.16 0.87 (0.44)				0.33 1.00 (0.84)	0.14 14.07 (2.65)						
goes across	71	0.03 0.50 (0.39)								0.25 6.55 (2.01)						
goes away from	11									0.28 10.43 (0.71)		0.28 1.90 (0.65)				
goes away from	75			0.12 0.53 (0.33)	0.48 0.88 (0.67)				0.39 0.99 (0.78)	2.42 20.14 (6.83)						
goes by	11									0.21 24.33 (4.19)		0.17 2.37 (0.69)			0.10 0.50 (0.20)	0.10 1.00 (0.40)
goes into	75			0.18 0.76 (0.40)	0.24 0.82 (0.60)				0.29 0.99 (0.81)	1.25 9.29 (3.78)						
goes out of	13									2.46 8.42 (3.87)						
goes out of	73	0.04 0.50 (0.40)								0.36 2.77 (1.95)						
goes out of	75			0.18 0.75 (0.48)	0.25 0.82 (0.52)				0.38 0.99 (0.80)	0.64 10.86 (2.22)						
goes through	71	0.03 0.50 (0.40)								0.70 8.21 (2.17)						

	LR	IAS	EAS	ITS	ETS	RBS	PA	LA	IC	OC	IN	ON	IPA	ILA	OPA	OLA
goes to	13									1.16 16.74 (6.99)						
goes to	75			0.04 0.57 (0.20)	0.44 0.96 (0.81)				0.18 0.99 (0.75)	1.63 12.04 (7.37)						
goes up to	11									0.03 4.79 (0.39)		0.03 0.75 (0.33)				
goes up to	13									3.29 17.68 (4.77)						
goes up to	75			0.06 0.97 (0.30)	0.03 0.94 (0.70)				0.12 0.97 (0.55)	0.37 16.25 (3.05)						
in	44								0.03 0.78 (0.42)		0.03 0.69 (0.41)					
inside	44								0.02 0.87 (0.46)		0.02 0.77 (0.45)					
intersect	71	0.01 0.49 (0.38)								0.41 6.87 (3.36)						
intersect	75			0.09 0.40 (0.29)	0.60 0.91 (0.72)				0.35 0.99 (0.63)	1.56 8.51 (3.22)						
intersects	71	0.02 0.50 (0.39)								0.25 9.10 (2.25)						
leaves	75			0.05 0.76 (0.55)	0.24 0.96 (0.45)				0.30 0.99 (0.81)	1.41 18.69 (3.03)						
near	11									0.49 10.09 (1.46)		0.26 2.10 (0.74)			0.10 0.90 (0.20)	0.30 1.00 (0.80)

	LR	IAS	EAS	ITS	ETS	RBS	PA	LA	IC	OC	IN	ON	IPA	ILA	OPA	OLA
outside	11									0.64 10.23 (2.65)		0.14 2.43 (0.84)				
passes	11									0.96 14.32 (3.84)		0.18 1.84 (0.64)			0.10 0.60 (0.20)	0.10 0.90 (0.40)
runs across	42	0.31 0.50 (0.48)				0.40 0.50 (0.50)										
runs across	71	0.05 0.49 (0.35)								0.27 4.31 (1.32)						
runs along	11									0.30 7.13 (1.06)		0.16 1.29 (0.33)			0.10 0.90 (0.40)	0.10 1.00 (0.90)
runs along	31						0.11 0.20 (0.19)	0.35 0.60 (0.55)								
runs along boundary	11									0.22 11.21 (1.38)		0.15 0.87 (0.33)			0.10 1.00 (0.50)	0.10 1.00 (0.80)
runs into	71	0.09 0.44 (0.13)		0.06 0.59 (0.30)	0.41 0.94 (0.70)					0.54 11.96 (3.42)						
runs into	75								0.29 0.97 (0.70)	1.32 17.15 (5.20)						
spans	42	0.29 0.47 (0.40)				0.30 0.50 (0.40)										
spans	71	0.30 0.43 (0.40)								0.54 3.58 (1.30)						
splits	71	0.02 0.50 (0.41)								0.42 6.90 (1.62)						

	LR	IAS	EAS	ITS	ETS	RBS	PA	LA	IC	OC	IN	ON	IPA	ILA	OPA	OLA
starts and ends in	44								0.23 0.71 (0.46)		0.22 0.71 (0.41)					
starts in	44								0.16 0.76 (0.60)		0.16 0.76 (0.44)					
starts in	75			0.12 0.85 (0.47)	0.15 0.88 (0.53)				0.25 0.99 (0.83)	0.98 17.13 (3.42)						
starts just inside	75			0.09 0.63 (0.22)	0.37 0.91 (0.78)				0.24 0.99 (0.60)	0.81 20.88 (4.79)						
starts just outside	11									0.19 2.50 (0.81)		0.19 1.38 (0.51)				
starts just outside	75			0.27 0.94 (0.75)	0.06 0.73 (0.25)				0.49 0.98 (0.66)	0.28 5.43 (1.14)						
starts near	11									0.32 17.98 (0.84)		0.30 1.63 (0.67)				
starts outside	11									0.33 3.89 (1.65)		0.26 1.87 (0.68)				
starts outside	75			0.17 0.86 (0.46)	0.14 0.83 (0.54)				0.40 0.99 (0.72)	1.23 18.54 (4.50)						
transects	71	0.08 0.50 (0.38)								0.53 7.32 (2.57)						
traverses	71	0.04 0.50 (0.43)								0.56 10.03 (2.38)						
within	44								0.14 0.76 (0.53)		0.14 0.76 (0.42)					