



ISTITUTO DI **A**NALISI DEI **S**ISTEMI ED **I**NFORMATICA
CONSIGLIO NAZIONALE DELLE RICERCHE

E. Pourabbas, M. Rafanelli

A PICTORIAL QUERY LANGUAGE EXTENDED
WITH CARDINAL AND POSITIONAL
OPERATORS FOR QUERYING
GEOGRAPHICAL DATABASES

R. 530 Settembre 2000

Elaheh Pourabbas - Istituto di Analisi dei Sistemi ed Informatica del CNR, viale Manzoni
30 - 00185 Roma, Italy. Email : pourabbas@iasi.rm.cnr.it.

Maurizio Rafanelli - Istituto di Analisi dei Sistemi ed Informatica del CNR, viale Manzoni
30 - 00185 Roma, Italy. Email : rafanelli@iasi.rm.cnr.it.

ISSN: 1128-3378

Abstract

In the context of geographic information systems the spatial relationships between two geographic objects play a central role at the query processing level. In this paper, we study particular spatial relationships, and we define and discuss a set of operators need to formulate queries in the context of a declarative pictorial query language which enlarge a previously proposed algebra. Such operators refer to the position of an object with respect to another one, their cardinal points, and consider specific geographic situations (e.g., water flow). We define and formalize also a set of specializations for the cardinal operators defined by prefixes *completely*, *only*, *partially*, and *also*. This language that is able to express queries on an object-oriented geographic database drawing the features which form the query is extended with respect to a previous one by cardinal and positional operators. Moreover, the main characteristics of the pictorial query language using the above mentioned operators through some query examples are shown.

Keywords: Object-Oriented Geographic Databases, Cardinal and Positional Operators, Pictorial Query Language

1. Introduction

In the context of GISs, one main task for a geographic database is the ability to handling spatial data. This requirement involves some fundamental issues like the definition and classification of spatial relationships and queries, the definition of efficient data models and structures for representing spatial data [DeM93].

The classification of spatial relations starts from classical topology [Whi57], and has been object of research in the context of GISs during the last years [Ege89, Kai90, Ege91, Wor92]. Topological relations encode the connecting structure of a map and can be further classified into adjacency, boundary relations among the basic spatial entities. Set-theoretic relations are based on concepts like inclusion, intersection, coincidence, etc., between entities. Metric relations involve the concept of distance and encode spatial proximity. Directional relations describe order in space. The topological and directional relations are considered as qualitative relations.

Recently, contributions dealing with qualitative spatial relations have appeared mainly in connection with orientation and distance relations. For instance, in [PaS94] the direction relations using representative points are defined. In [CDH97], the qualitative representation of positional information is discussed. Kainz et al. [KEG93] has modelled ordinal relations using partially ordered sets, and Frank [Fra92] proposed a method for qualitative reasoning that combines direction with distance relations.

In the context of GISs, all the relevant geometric properties of objects and the relations between them are captured through spatial *operators*. Operators that apply to all primitive geometric data types (namely, points, polylines, and regions) can be classified into topological, metric and directional operators.

One of the main topic in GISs research concerns embedding the above mentioned operators in spatial query languages. This fact arises from the need of providing the user with an interactive tool for data manipulation and data retrieval which is independent of the physical organization of data, to the enhancement of interoperability across different systems between nodes of a world-wide distributed network [CID00]. The interest of the database community in extending standard query languages goes back to the early 80's. The basic requirement, along these years in research area has been focused on development of high expressive languages to query and browse spatial data in an easier way. The contributions in this direction have been made, coming mainly from the field of pictorial databases. The MapQuery [Fra82], PSQL [RFS88], PIQUERY [JoC88], XSQL2 [Lor91], PICQUERY+ [CIT93], and Spatial SQL [Ege94] are examples of such a trend.

In this paper, we define particular relations between geographic objects, and define and discuss a set of operators need to formulate queries in a geographic context which enlarge a previously proposed algebra and discussed in [FMR99]. This algebra is based on a set of symbolic features that refer to the classic ones of geographic environment, i.e., geo-points, geo-polyline (oriented and non-oriented), and geo-region. In particular, the above mentioned operators refer to the position of an object with respect to another one, their cardinal points, and consider specific geographic situations (e.g., water flow). Such operators are defined to capture all the relevant geometry properties of the relations between two objects.

To the best of our knowledge the inclusion of cardinal and positional operators in query languages that allow user to define such type of queries in the context of GIS has not been extensively addressed. The reason behind of our proposal consists of introducing the positional and cardinal operators in the context of a pictorial query language in order to allow a user to formulate queries drawing geographic features involved in such queries.

The structure of this paper is the following. In section 2, we describe the geographic data model. In section 3, we list the operators for geographic databases. define the positional operators. In section 4, and in section 5 we define and formalize respectively the cardinal

4.

operators the positional operators. In section 6, we illustrate the main characteristics of the pictorial query language for expressing queries that invoke cardinal and positional operators through some query examples. Finally, in section 7 we conclude.

2. The Geographic Data Model

The geographical data model is based on the existence of the following sets:

- a finite set of atomic domains $D = \{D_1, \dots, D_m\}$
- an infinite numerical set $A = \{A_1, A_2, \dots, A_m\}$ of symbols called *instance variables*
- an infinite numerical set OID of *identifiers*
- the set C of all the classes defined in the database
- the set O of all the objects defined in the database
- the set M of all the methods defined in the database.

A geographical class is a set of elements called geographical objects go_1, go_2, \dots, go_n , which have the same properties A_1, A_2, \dots, A_m defined over the set of not necessarily distinct values D_1, D_2, \dots, D_m and a set of methods which describe the behaviour of the objects.

Definition: a *geographic class* (gc) is a quadruple:

$\langle n, sc, P, M \rangle$ where:

- n is the name of gc ;
- sc is the class-parent of gc ;
- P is the set of attributes which define the geographic entity properties (enclosing the set of attributes inherited from the relative superclass sc); it consists of:
 - a *geometric attribute*, is the type of geometric structure which refers to the geographic class. It can assume only one value chosen in its domain definition: $\{null, point, polyline, region\}$;
 - a set of *alphanumeric attributes* which defines the non geometric characteristics of the gc instances; each of them consists of a couple (attribute name, data type);
- M is the non empty set of methods such that:
 $M = \{inherited\ methods\ from\ sc\} \cup \{methods\ defined\ on\ gc\} \cup \{overridden\ methods\}$ \square

Example: The geographic class "river" is defined as:

$\langle river, river\ course, \{(name, string), (length, integer), (flow\ per\ second, integer), (average\ section, integer), \dots\}, polyline \}, \{speed, method-2, \dots\} \rangle$

Definition: A *geographic object* go , instance of a geographic class gc , represents a geographic entity of the real world. It consists of a quadruple:

$\langle id, gc, V, C \rangle$ where:

- $id \in OID$ (Object Instance Domain) is called identifier of a geographic class instance (go); it is unique.
- gc is its belonging class. It is also unique.
- V represents the ordered set of the alphanumeric attribute values.
- C is the ordered set of couples (latitude, longitude) which specifies the object position with respect to the geographic coordinates. \square

Definition: The *root class* (rt) is the super class of all the geographic classes. It is a quadruple $\langle n, sc, P, M \rangle$ where $sc = \Phi$ (void class) because its parent does not exist. \square

If geometric attribute = null, the geographic class is the *null geographic class* and the quadruple is defined as $\langle n, rt, P, \Phi \rangle$, where $n = "void\ class"$, $P = \langle \Phi, null \rangle$, where Φ represents the set of alphanumeric attributes which defines the non geometric characteristics

of the gc instances, and $null$ represents the type of geometric structure which refers to the geographic class.

Definition: A *method* is a triple

$\langle nm, f: S \rightarrow R, s \rangle$, where:

- nm is the name of the method.
- $f: S \rightarrow R$ is the method's interface in the form of: $f: S_1 \times S_2 \times \dots \times S_n \rightarrow R$ where S_1, S_2, \dots, S_n are the domains of the arguments, R is the domain of the result and $S_i = ist(gc)$, where $ist(gc)$ are the class instances.
- s is the set of instructions in a given language which determines the semantics of the function f (method implementation). \square

Definition: A *schema of geographical database* consists of a set of geographical classes. \square

Definition: A *geographic database* consists of a geographical object set, each of them being an instance of a schema class. \square

Definition: A *symbolic feature* (ψ) is defined by a 5-tuple:

$\psi = \langle id, objclass, objalias, S, L \rangle$ where:

- id is the identifier of the symbolic feature,
- $objclass$ is the class set iconized by ψ ;
- $objalias$ allows to create references to this ψ from inside other ψ ;
- S (set of properties) represents the attributes to which user can assign a value range; this assigning allows to carry out a selection among the objects or the classes iconized by ψ . In particular, the geometric attribute is the only attribute to which only one value (and not a range) has to be assigned;
- L is the ordered set of couples (h, v) which specifies the object position with respect to the top-left point of the query work space. By these ordered sets of couples, it is possible to determine the topologic relations among the symbolic features drawn on the work space. \square

3. Operators for Geographic Databases

In this paper we study particular relationships between geographic objects, and define and discuss a set of operators which refer either to the cardinal points (called *Cardinal Operators*) or to the relative position (called *Positional Operators*). This set enlarges an algebra previously proposed [FMR99]. We consider also some their specializations, for example, objects which are "completely" or "only" North of another object, as well as specific geographical situations (for example, water flow).

The previous algebra refers to a set of symbolic features, i.e., geo-null, geo-point, geo-polyline, and geo-region, that form the relative *alphabet* A . In this work, we extend the definition of geo-polyline including the *oriented* polyline (see Figure 1). To this finite set of spatially related symbolic features it is possible assign a specific semantics. Both the geo-polylines and the geo-regions are considered to be "discretized", in other words, each of them is represented by a set of segments and, then, a set of vertexes.

We propose a set of operators, part of which have already been discussed in [FMR99]. These last operators are: Geo-Union, Geo-Difference, Geo-Disjunction, Geo-Touching, Geo-Inclusion, Geo-Crossing, Geo-Pass-Through, Geo-Overlapping, Geo-Equality, and Geo-Distance. All of these ones have the prefix *geo* in order to avoid confusion with respect to the set theory operators, from which some differences exist.

6.

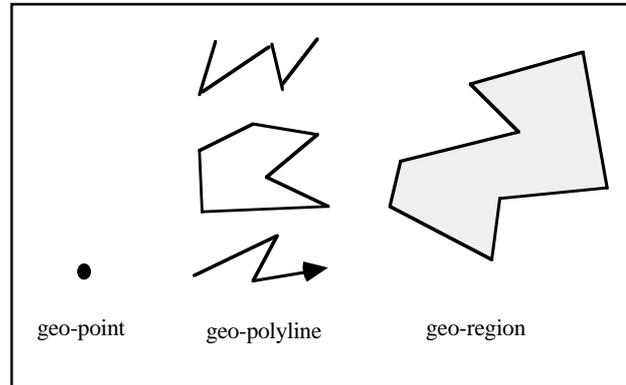


Figure 1. The symbolic features used

In fact, some of these operators are not applicable to any symbolic feature combination (that could have not sense in a generic geographic context). For example, the difference between a geo-region and a geo-polyline does not have sense because the result of this operation should be two adjacent areas without the part of boundary common to the two areas, i.e., two features which are not elements of the alphabet \mathcal{A} . For this reason the Geo-Difference operator is different from the Difference operator of the set theory, and it is applicable only to the ordered couple of operands $\langle \text{geo-polyline} \rightarrow \text{geo-region} \rangle$, and to the couples of operands $\langle \text{geo-region} \rightarrow \text{geo-region} \rangle$ and $\langle \text{geo-polyline} \rightarrow \text{geo-polyline} \rangle$, with the obvious condition that an overlapping exists between the two operands.

The proposed *cardinal* operators are: North_of, East_of, South_of, West_of, North-East_of, South-East_of, South-West_of, and North-West_of, and the set of proposed *positional* operators consists of: Inside, Left_of, Right_of, Nearest, Farest, Before, and After.

All the cardinal operators are specialized by the prefixes Completely, Only, and Partially, and by the suffix "- opposite cardinal point"; for example, North-South_of.

The operators Nearest, Farest, and Inside can be specialized using as suffix the cardinal operators, for example, Inside_North_of or Nearest_South_of. The *Farest* and *Nearest* operators are different from the *Geo-Distance* operator proposed in [FMR99]. In fact, Geo-Distance selects a set of objects which satisfy a specific built-in predicate on distance (e.g., $< 100 \text{ Km}$), while as we will see, Farest and Nearest select only one object which is "farest" or "nearest" from the reference object without using any predicate.

For instance, if the query is "Find the lake which is Nearest to the Rome city", the operator selects all the lakes which are in the considered area that includes "Rome", and measures the distance among Rome and these lakes, then selects the lake which has the minimum distance from Rome. Instead if the query is "Find all the lakes which have distance $< 100 \text{ Km}$ from Rome city", then *all* the lakes which satisfy the condition "distance $< 100 \text{ Km}$ from Rome" are selected.

The operators Left_of, and Right_of have as reference object only an oriented polyline.

Instead, the operators Before and After consider only two geo-points which are *onto* the same oriented polyline.

All these operators (as well as the set considered in [FMR99]) are dyadic, because the monadic operators (for example, length, area, boundary-box, etc.) are considered as properties of the single classes.

4. The Cardinal Operators

We distinguish eight possible cardinal relationships between two objects: (*North_of*, *South_of*, *East_of*, *West_of*, *North-East_of*, *North-West_of*, *South-East_of*, and *South-West_of*).

For each of them it is necessary to specialize their meaning using the following prefix: *Completely*, *Only*, and *Partially*. Moreover, another particular specialization, that is a pair "cardinal position-opposite cardinal position", is proposed.

In the following, for sake of brevity, we will formally define only the operator *North_of* and the relative specializations. The definition of the other operators is similar, a part the obvious changings.

Definition: A symbolic feature ψ_i is *North_of* another symbolic feature ψ_j (and we write $\psi_i \uparrow \psi_j$) if it exists at least one point $p_{\psi_i} \in \psi_i$ such that its y-coordinate $p_{\psi_i,y}$ is greater than the y-coordinate $p_{\psi_j,y}$ of all the points $p_{\psi_j} \in \psi_j$. The result of the operation is a symbolic feature ψ_h which coincides with the first operand ψ_i .

Formally we have:

Let $\psi_i, \psi_j \in A$ be two symbolic features of the alphabet A . Let $\psi_i \cap \psi_j = \emptyset$. Let $p_{\psi_i} \in \psi_i$ be a generic point of ψ_i and let $p_{\psi_j} \in \psi_j$ be a generic point of ψ_j . Let $p_{\psi_i,y}$ and $p_{\psi_j,y}$ be respectively the y-coordinate of the generic points p_{ψ_i} and p_{ψ_j} .

Then:

$\psi_i \uparrow \psi_j$ iff $\exists p_{\psi_i} \in \psi_i : p_{\psi_i,y} \geq p_{\psi_j,y} \quad \forall p_{\psi_j} \in \psi_j$
where $p_{\psi_i,y}$ is the y-coordinate of p_{ψ_i} and $p_{\psi_j,y}$ is the y-coordinate of p_{ψ_j} . \square

Before studying different *specializations* of *North_of* operator, we need to define *Bottom-point*, *Top-point*, *Left-point* and *Right-point* of a symbolic feature (geo-region or geo-polyline).

Definition: A generic point p_{ψ_i} of a symbolic feature ψ_i is the *Bottom-point* of ψ_i (called it $p_{b\psi_i}$) if its y-coordinate $p_{b\psi_i,y}$ is less than (or equal to) the y-coordinate p_{y,ψ_i} of all the other points p_{ψ_i} of ψ_i .

Formally we have:

Let $\psi_i \in A$ be a symbolic feature of the alphabet A . Let $p_{\psi_i}, p_{b\psi_i} \in \psi_i$ be generic points of ψ_i . Let $p_{\psi_i,y}$ be the y-coordinate of the generic point p_{ψ_i} and let $p_{b\psi_i,y}$ be the y-coordinate of the generic point $p_{b\psi_i}$.

Then:

$p_{b\psi_i,y} \leq p_{\psi_i,y} \quad \forall p_{\psi_i} \in \psi_i$ \square

Analogously, we can define *Top-point*, *Left-point* and *Right-point* of a symbolic feature in a similar way, with the obvious changings. In this point, we propose and discuss the above mentioned specializations.

Definition: A symbolic feature ψ_i is *Completely_North_of* another symbolic feature ψ_j (and we write $\psi_i \uparrow \psi_j$) if the y-coordinate of the bottom point of ψ_i is greater than (or equal to) the y-coordinate of the top point of ψ_j (see Figure 2).

Formally we have:

Let $\psi_i, \psi_j \in A$ be two symbolic features of the alphabet A . Let $p_{\psi_i} \in \psi_i$ be a generic point of ψ_i and let $p_{\psi_j} \in \psi_j$ be a generic point of ψ_j . Let $p_{\psi_i,y}$ and p_{y,ψ_j} be respectively the y-coordinate of the generic points p_{ψ_i} and p_{ψ_j} . Let $p_{b\psi_i}$ be a *bottom-point* of ψ_i , and $p_{t\psi_j}$ be a *top point* of ψ_j .

Then: $\psi_i \uparrow \psi_j$ iff $p_{b\psi_i,y} \geq p_{t\psi_j,y}$ \square

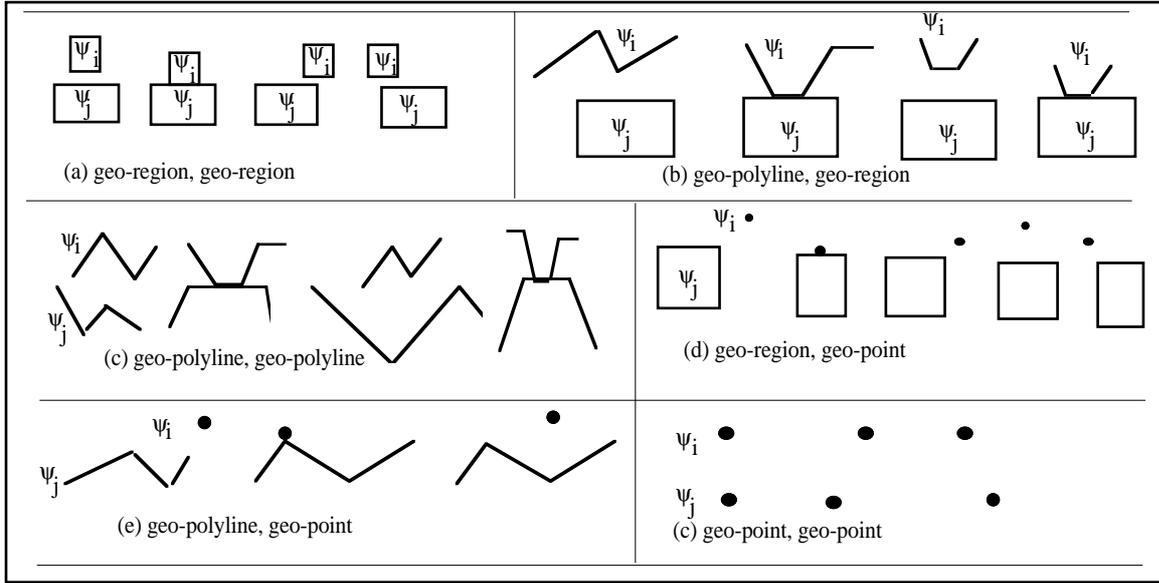


Figure 2. Cases of the *Completely_North_of* specialization

Definition: A symbolic feature ψ_i is *Only_North_of* another symbolic feature ψ_j (and we write $\psi_i \circ \uparrow \psi_j$) if all the points of ψ_i are included between the Left-point and the Right-point of ψ_j (see Figure 3).

Formally we have:

Let $\psi_i, \psi_j \in A$ be two symbolic features of the alphabet A . Let p_{lx, ψ_j} be the Left-point of ψ_j , and p_{rx, ψ_j} be the Right-point of ψ_j , and p_{b, ψ_j} be the bottom point of ψ_j . Let p_{ψ_i} be a generic point of ψ_i , and let p_{x, ψ_i} be its x-coordinate. Then,

$$\psi_i \circ \uparrow \psi_j \text{ iff } \exists p_{\psi_i} \in \psi_i : p_{l\psi_j, x} \leq p_{\psi_i, x} \leq p_{r\psi_j, x} \text{ AND } p_{\psi_i, y} \geq p_{b\psi_j, y} \quad \square$$

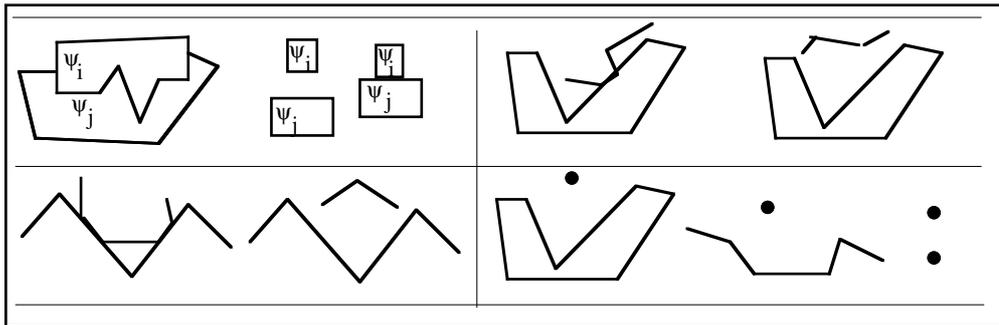


Figure 3. Cases of the *Only_at_North_of* specialization

Definition: A symbolic feature ψ_i is *Partially_North_of* another symbolic feature ψ_j (and we write $\psi_i P \uparrow \psi_j$) if at least one point of ψ_i is North_of all the points of ψ_j , and it exists at least one point of ψ_j whose y-coordinate is less than or equal to the y-coordinate of the bottom-point of ψ_i (see Figure 4).

Formally we have:

Let $\psi_i, \psi_j \in A$ be two symbolic features of the alphabet A . Let p_{ψ_i} be a generic point of ψ_i , and p_{ψ_j} be a generic point of ψ_j . Let $p_{\psi_i, y}$ and $p_{\psi_j, y}$ be respectively the y -coordinate of the generic points p_{ψ_i} and p_{ψ_j} . Let $p_{b\psi_i}$ be a *bottom-point* of ψ_i .

Then:

$$\psi_i P \uparrow \psi_j \text{ iff } \exists p_{\psi_i} \in \psi_i, p_{\psi_j} \in \psi_j: p_{\psi_i} \uparrow \psi_j \text{ AND } \exists p_{\psi_j} \in \psi_j: p_{\psi_j, y} \leq p_{b\psi_i, y} \quad \square$$

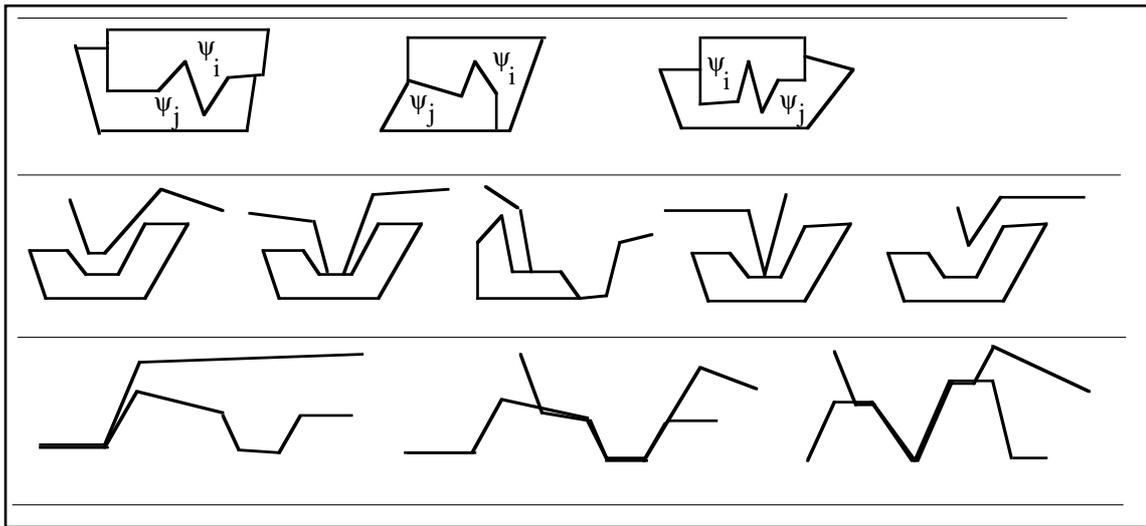


Figure 4. Cases of the *Partially_at_North_of* specialization

Definition: A symbolic feature ψ_i is *North_South_of* another symbolic feature ψ_j (and we write $\psi_i \uparrow \downarrow \psi_j$) if at least one point of ψ_i is *North_of* all the points of ψ_j , and at least one point of ψ_i is *South_of* all the points of ψ_j (see Figure 5).

Formally we have:

Let $\psi_i, \psi_j \in A$ be two symbolic features, and let p_{ψ_i1}, p_{ψ_i2} two generic points of ψ_i .

Then:

$$\psi_i \uparrow \downarrow \psi_j \text{ iff } \exists p_{\psi_i1}, p_{\psi_i2} \in \psi_i: p_{\psi_i1} \uparrow \psi_j. \text{ AND } p_{\psi_i2} \downarrow \psi_j. \quad \square$$

Note that the symbol \downarrow represents the *South_of* operator, the definition of which for sake of brevity is omitted.

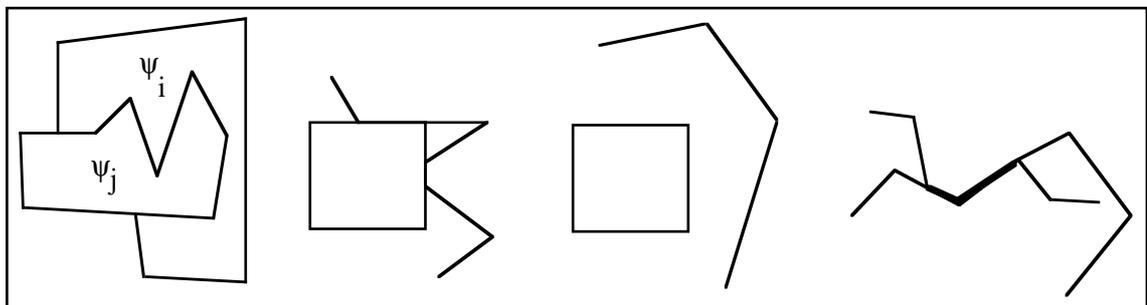


Figure 5. Cases of the *North_South_of* specialization

It is possible, subsequently, to specialize *Completely*, *Partially* and *North_South_of* specifications by adding to each of them the prefix *Weak*. that is defined as follows: if no

10.

points of the first object, ψ_i , is included in the area delimited by the left-top and the right-top points of the reference object ψ_j (see Figure 6).

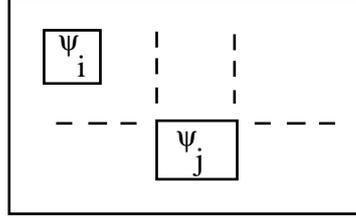


Figure 6. Case of Weak_Completely_North_of.

Definition: A symbolic feature ψ_i is *Weak_Completely_North_of* another symbolic feature ψ_j (and we write $\psi_i \text{ WC } \Uparrow \psi_j$) if the y-coordinate of the bottom point of ψ_i is greater than (or equal to) the y-coordinate of the top point of ψ_j and the x-coordinate of the Left-point (Right-point) of ψ_i is greater (less) than the x-coordinate of the Right-point (Left-point) of ψ_j .

Formally we have:

Let $\psi_i, \psi_j \in A$ be two symbolic features of the alphabet A. Let $p_{b\psi_i}, p_{r\psi_i}, p_{l\psi_i} \in \psi_i$ and $p_{t\psi_j}, p_{r\psi_j}, p_{l\psi_j} \in \psi_j$. Then:

$$\psi_i \text{ WC } \Uparrow \psi_j \text{ iff } p_{b\psi_i, y} \geq p_{t\psi_j, y} \quad \text{AND} \quad p_{l\psi_i, x} \geq p_{r\psi_j, x} \quad \square$$

Note that an important aspect of the previous specializations concern the so called *reciprocal exclusion* of this application to objects. In other words, a specific specialization of one cardinal operator indicates that, at the same time, it excludes the application of someone else. For example, *Only_North_of* specialization between two objects A and B could also indicate that A *Completely_North_of* B or A *Partially_North_of* B, but not A *North_South_of* B.

5. The Positional Operators

The proposed positional operators are: *Inside* (\sqcup), *Left_of* (\ll), *Right_of* (\gg), *Before* (\mathbf{B}), *After* (\mathbf{A}), *Nearest* (\mathbf{N}), and *Farest* (\mathbf{F}).

In the following we will give their formal definition, as well as an example of their application.

5.1. The Inside Operator

The *Inside* operator represents the particular relationship between two disjoint objects where one object is included in the "concavity" of another one. In fact, before defining the inside operator, we introduce the concept of *concavity*.

In [CCR93] the authors have identified geometric inside and topologic inside relationships. In this paper, we will consider only the later one, because we believe the topologic inside refers to an specific case of inclusion relationship where an object is included in the hole of another object.

Definition : Let p_i and p_j be two non consecutive vertexes of a geo-polyline and let $\{p_k\}$, with $1 \leq i < k < j \leq n$, be the set of the relative vertexes which are enclosed between p_i and

p_j . Let the reference Cartesian axis X be coincident with the line which joins p_i to p_j (oriented from p_i to p_j) and let p_i be the origin of the XY -axes. If it exists at least one vertex p_k of the above mentioned geo-polyline whose y -coordinate is greater than or equal to the y -coordinates of p_i and p_j , and x -coordinate is enclosed between the x -coordinates of p_i and p_j , then $\langle p_i, p_k, p_j \rangle$ defines a *concavity*, represented by $C\langle p_i, p_k, p_j \rangle$. \square

Definition : Let $\{p_1, \dots, p_n\}$ be a set of vertexes which defines a geo-polyline. If any three consecutive vertexes $\langle p_j, p_{j+1}, p_{j+2} \rangle$ (with $i = 1, \dots, n-2$) defines a concavity, then it is a *pure concavity*. \square

Obviously, if any triple vertexes $\langle p_i, p_k, p_j \rangle$ (with $1 \leq i < k < j \leq n$) of a set of vertexes is a pure concavity, then such set is a pure concavity (see Figure 7 -a).

We define *Left-vertex* (respectively, *Right-vertex*) of a geo-polyline, a vertex that the relative x -coordinate is less (respectively, greater) than to x -coordinate of all other vertexes. In a similar way the *Bottom-vertex* (respectively, *Top-vertex*) of a geo-polyline is a vertex that its y -coordinate is less (respectively, greater) than to y -coordinate of all other vertexes.

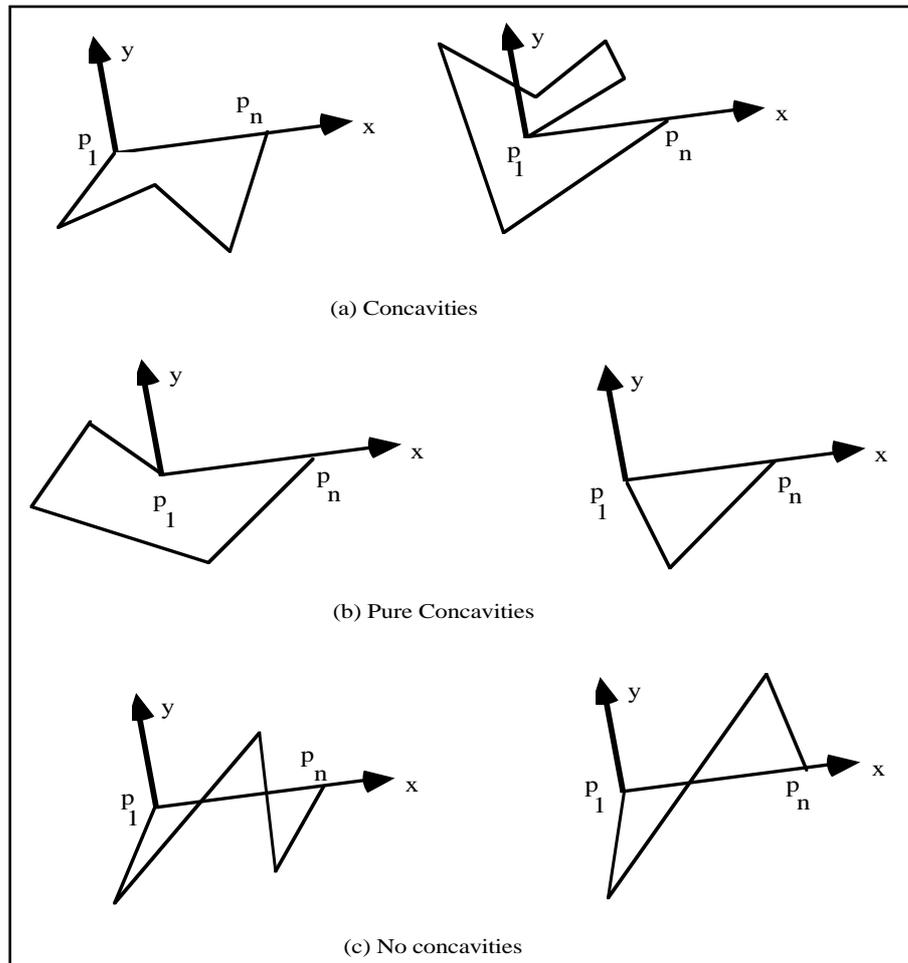


Figure 7. Examples of polylines with concavities, pure concavities, and without concavities.

12.

Definition: Let ψ_i and ψ_j be two symbolic features of the alphabet \mathbb{A} . Let $C\langle p_i, p_k, p_j \rangle$ be a concavity in the set vertexes of ψ_j , $\{p_1, \dots, p_k, \dots, p_n\}$, with $1 \leq i < k < j \leq n$. A feature ψ_i is *Inside* $C\langle p_i, p_k, p_j \rangle$ (and denoted by $\psi_i \sqcup \psi_j$), if ψ_i is completely included in the area (V) delimited by $C\langle p_i, p_k, p_j \rangle$ itself and by the line joining p_i to p_j (see Figure 8-a).

Formally we have:

$\forall \psi_i, \psi_j \in \mathbb{A}$, if $\exists C\langle p_i, p_k, p_j \rangle \in \psi_j$, then $\psi_i \sqcup C\langle p_i, p_k, p_j \rangle \psi_j$ iff $\forall p_{\psi_i} \in \psi_i : p_{\psi_i} \in V$ \square

Definition: Let ψ_i and ψ_j be two symbolic features of the alphabet \mathbb{A} . Let $C\langle p_i, p_k, p_j \rangle$ be a concavity in the set vertexes of ψ_j , $\{p_1, \dots, p_k, \dots, p_n\}$, with $1 \leq i < k < j \leq n$. A feature ψ_i is *Partially_Inside* $C\langle p_i, p_k, p_j \rangle$ (and denoted by $\psi_i \sim \sqcup \psi_j$) if there exists a set of points $\psi_{ih} \subset \psi_i$ such that ψ_{ih} *Inside* $C\langle p_i, p_k, p_j \rangle$ and exists a set of points $\psi_{ik} \subset \psi_i$ such that it is not included in the area V (see Figure 8-b).

Formally we have:

$\forall \psi_i, \psi_j \in \mathbb{A}$, if $\exists C\langle p_i, p_k, p_j \rangle \in \psi_j$, then $\psi_i \sim \sqcup C\langle p_i, p_k, p_j \rangle \psi_j$ iff $\exists \psi_{ih}, \psi_{ik} \subset \psi_i : \psi_{ih} \sqcup C\langle p_i, p_k, p_j \rangle$ AND $\psi_{ik} \in V$ \square

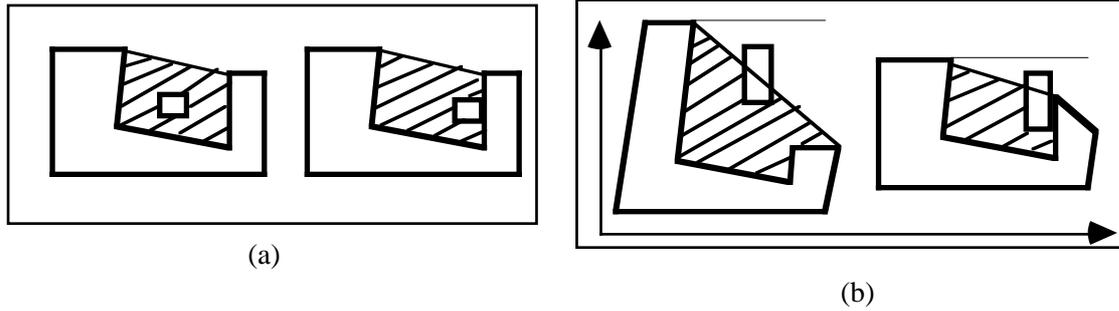


Figure 8. The cases of *Inside* and *Partially_Inside* between two geo-regions.

All the possible cases of *Inside* are: a) geo-region *Inside* geo-region; b) geo-region *Inside* geo-polyline; c) geo-polyline *Inside* geo-region; d) geo-polyline *Inside* geo-polyline; e) geo-point *Inside* geo-region; f) geo-point *Inside* geo-polyline.

The combination of the proposed cardinal operators with *Inside* gives the following operators: *Inside_North_of*, *Inside_South_of*, *Inside_East_of*, *Inside_West_of*, *Inside_North-East_of*, *Inside_North-West_of*, *Inside_South-East_of*, and *Inside_South-West_of*. We will define only the first operator (*Inside_North_of*), leaving the obvious extension of the other cases to the reader.

Definition: Let ψ_i and ψ_j be two symbolic features of the alphabet \mathbb{A} . Let $C\langle p_i, p_k, p_j \rangle$ be a concavity in the set vertexes of ψ_j , $\{p_1, \dots, p_k, \dots, p_n\}$, with $1 \leq i < k < j \leq n$. A feature ψ_i is *Inside_North_of* another feature ψ_j (and we write $\psi_i \sqcup \uparrow \psi_j$) if, translating the origin of the XY-axes and the cardinal axes onto p_1 , the Y axis is in the area defined by the two medians of respectively the angles formed by *North_West-North* and *North-North_East* axes (see Figure 9). \square

Note that the difference between *North_of* and *Inside_North_of* is that in the former it exists at least one point of ψ_i whose y-coordinate is greater than the y-coordinate of all the points

of ψ_j , while in the latter it exists at least one point of ψ_j whose y-coordinate is greater than or equal to the y-coordinate of all the points of ψ_i .

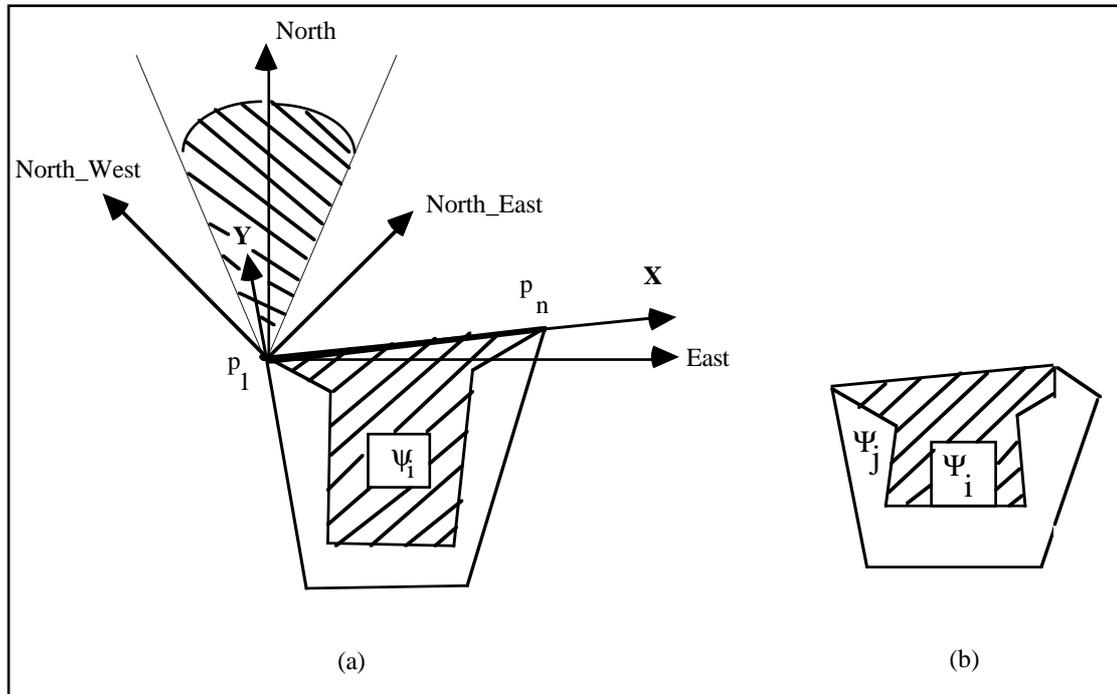


Figure 9. Cases of Inside_North operator.

5.2 The Left_of and Right_of operators

For the operators *Left_of* and *Right_of* the reference feature is always an *oriented* polyline. The definition of these operators are based on the identification of the polyline segment which has the Min-Geo_Distance from the other object that can be any symbolic feature. The orientation of this segment defines to which side the considered object belongs.

Definition: Let ψ_i be the symbolic feature geo-polyline, and ψ_j be any other feature of the alphabet \mathbb{A} . Let $p_{\psi_i,1}$, and $p_{\psi_i,2}$ be the final points of the oriented segment of ψ_i which has the Min-Geo_Distance ψ_j . The feature ψ_j is *Right_of* ψ_i (denoted by $\psi_j \gg \psi_i$), if translating the Y-axis onto the mentioned segment such that both of them assume the same orientation, then the x-coordinates of all the points of ψ_j are positive. If ψ_j is a geo-polyline or a geo-region, then a subset of its x-coordinates could be equal to 0 (see Figure 10). \square

Analogously, the *Left_of* operator with appropriate changes in the above definition, can be defined.

5.3. The Before and After operators

Concerning *Before* and *After* operators, we consider only the case of two geo-points which are onto an *oriented* polyline.

Definition: Let ψ_i and ψ_j be two geo-points of the alphabet \mathbb{A} . Let $\psi_k \in \mathbb{A}$ be an oriented geo-polyline onto which ψ_i and ψ_j are located. Let the X-axis be coincident with ψ_k by

14.

the same orientation. The feature ψ_j is *Before* ψ_i (denoted by $\psi_j \mathbf{B} \psi_i$), if the x-coordiante of ψ_j is less than the x-coordiante of ψ_i (see Figure 11). \square

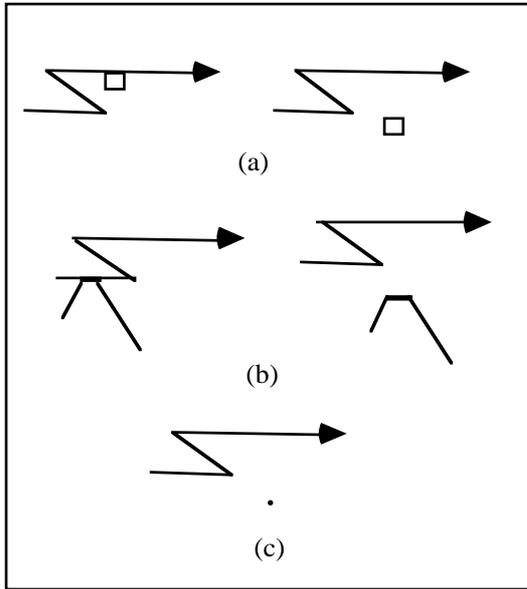


Figure 10. Examples of *Right_of* operator

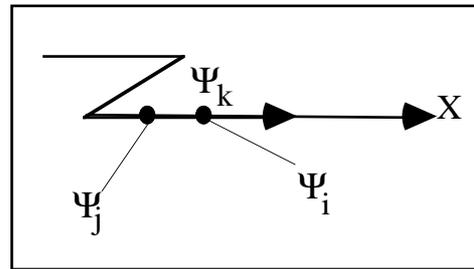


Figure 11. Example of *Before* operator

5.4. The Nearest and Farest operators

The *Nearest* and *Farest* operators are specializations of the Geo-Distance operator [FMR99]. In the case of Nearest operator, we have to measure the distance between the object of reference and all the other objects which are involved in the query, and to select the object which has the Min-Geo-Distance from the object of reference. A similar reasoning is made for the *Farest* operator, where we have to select the object which has the Max-Geo-Distance from the object of reference.

Definition: Let $\psi_i, \{\psi_{jk}\}$ with $1 \leq k \leq n$ be symbolic features of the alphabet \mathbb{A} . A symbolic feature ψ_{jk} is *Nearest* to the object of reference ψ_i (and we write $\psi_i \mathbf{N} \psi_{jk}$) if the Min-Geo-Distance between ψ_i and ψ_{jk} is the minimum value among all the Min-Geo-Distances between ψ_i and the objects of $\{\psi_{jk}\}$.

Formally we have:

$\forall \psi_i, \{\psi_{jk}\} \in \mathbb{A},$ with $1 \leq k \leq n, \psi_{jh} \in \{\psi_{jk}\}: \psi_{jh} \mathbf{N} \psi_i$ iff $\forall \text{Min-Geo-Dis}(\psi_i, \psi_{jh}) < \{\text{Min-Geo-Dis}(\psi_i, \{\psi_{jk}\} - \psi_{jh})\}$ \square

Analogously we can define the operator *Farest*, with the obvious changings. Similarly to the *Inside* operator, also for these two operators we can define the specializations *Nearest_North*, *Farest_North*, etc., leaving to the reader the relative simple definitions and formalizations.

6. The Pictorial Query Language

By the pictorial query language (PQL) complex queries can be expressed by simple and intuitive pictorial operations. In fact, the language allows the user to draw all the features

involved in the query, to instance each feature giving the selection criteria and, finally, to point the feature which represents the query target. In this way the user expresses in a declarative manner, what he wishes to know. Moreover, the user does not need to learn a query language syntax and semantics to carry out a correct query to the system.

6. 1. The queries

Using symbolic features as a representation of geographic classes and objects [DGJ95], the queries are "composed" putting graphically in evidence the spatial objects (symbolic features) which are involved in it and the topological, positional, etc. relationships which link them.

Then, the semantics of the query are expressed drawing geometric figures. Two distinct levels of semantic interpretation exist. The former is the semantics assigned to the algebra operators, the latter is due to the implementation of the class methods which customize the basic behaviour of methods instanced to solve the query. The queries which can be formulate on a geographic database express two types of geographical object properties:

- *spatial properties*, as, for example, the positional relationships or the distance; these properties are verified by the geometric attributes of the different objects which form the database.

- *descriptive properties*, as, for example, the population of a given region, or the length of a given river. These properties are verified by the alphanumeric attributes of the above mentioned objects.

Formally, we define a query as a set of seven subsets:

$$Q = \{D, F, H, E, R, S, T\} \quad \text{where:}$$

D is a non-empty set of databases which form the active domain of the current query. All the geographic object instances and classes involved in the query by the same symbolic features are searched in these databases.

F is a non-empty set of symbolic features, created directly by the user, which are involved in the query.

H is a non-empty set of type declaration which links each symbolic feature (enclosed in $S \cup F$) to a class of the class hierarchy rooted to the symbolic feature data type.

E is a set of selection expressions associated to symbolic feature fields which form the selection criterion to choose the object instances which satisfy the query target from the databases which are enclosed in *D*.

R is a set of relationships (set-type, topological-type, metric or positional-type), eventually empty, among symbolic features belonging to the sets *F* or *S*.

S is a set of symbolic features generated as a result of the relationship in *R*.

T is a non-empty set of symbolic features belonging $S \cup F$ which form the current query target.

A query carries out by the pictorial operations defined in the PQL* (dragging of a symbolic feature, selection of a symbolic feature as the target of the following pictorial operation, grouping and ungrouping of symbolic features, assignment of selection criteria to a symbolic feature, query target definition, etc.) is a combination of a sequence of these pictorial operations. It is generally terminated by the pointing out of the resulting feature, which constitutes the query target.

The creation of a symbolic feature involves not only a new object class, such as instances of the query, but also generates symbolic features to represent relationships (topological, positional, etc.) among the existing objects.

At the end of the pictorial operations the system produces a string of syntactically correct instructions, which corresponds to the graphical query.

16.

6.2. Query examples

In order to formulate and describe the following query examples, we will use to express them both by pictorial and the equivalent key-words languages. This later is a SQL-like language [Ege94] with the following four clauses:

```
SELECT  Target of query
FROM    Databases
WITH    Properties of features
WHERE   Built-in predicates on spatial relationships
```

Let us consider the following queries.

Query: "Find the Italian provinces which are situated Completely_at_North_of Viterbo province"

We generate the symbolic features representing the objects and the classes involved in the query. We suppose to express this query on the same above mentioned database. Therefore, we need two symbolic features that represent provinces A, and B. The first one, A, represents the Viterbo province, and B represents the generic instance of the *Province* class. We define the object B (the province) as the target of the query. In the same way, we don't specify any selection criteria regarding the provinces, because we are looking for all the instances which satisfy the query. We mark only the attribute "name of the province" to specify that we need only the province name. Besides, to specify the selection criteria, we need to place properly the symbolic features involved in order to specify the positional relationships existing among these objects. We put A, and B in order to obtain all provinces, represented by B, that are completely at north of Viterbo province (represented by A). This is indicated by an arrow as is shown in Figure 14.

The query generated by the system is expressed in a SQL-like syntax as follows:

```
SELECT    ALL B.name
FROM      database := Italy
WITH      B.type = province
          A.type = province
          A. property.name ="Viterbo"
WHERE     B Completely_at_North_of A
```

The result of query is shown in Figure 15. Let us consider the following query.

Query: "Find the Italian regions whose territory has an altitude > 2,000 feet for an extension >30% of its total extension, which are situated Only_at_North_of the Viterbo province".

We describe the query formulation step by step as follows. First of all, we generate the symbolic features representing the objects and the classes involved in the query. The thematisms and the generic Italian region are modelled by geo-regions. We assume that in the considering database only data regarding Italy are stored. Then, we need four features A, B, and C. The first one, A, represents the generic instance of the "region" class. B represents the generic instance of the two thematisms. Finally, feature C represents the province of Viterbo (instance "Viterbo" of the class "Province"). We will remember that the symbol \uparrow which represents graphically the operator Only_at_North_of.

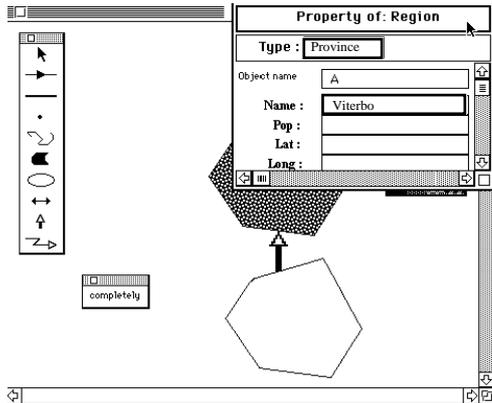


Figure 14. A pictorial query composition

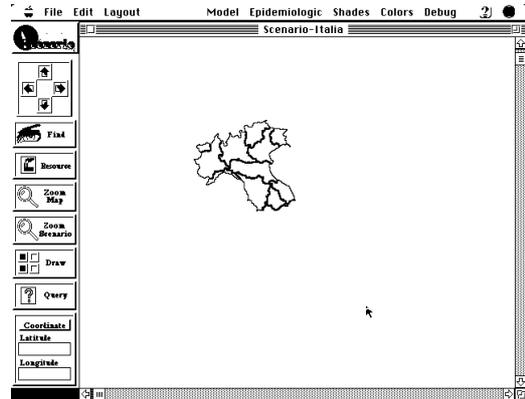


Figure 15. Result of the query

To specify the selection criteria, we need to place properly the symbolic features involved in order to specify the topological relationships existing among these features.

Note that we put A, B and C in order for avoiding the generation of a common overlapping ABC. In this way we obtain two distinct overlapping areas AB and AC.

Figure 12 shows the pictorial query formulation.

We don't specify any selection criteria regarding the regions, because we are looking for all the instances which satisfy the query. On the features representing the overlapping areas AB we set, as selection criteria, the areas (regions) with altitude > 2,000 feet for an extension > 30 %.

To set these criteria we use an alias reference (the property extension of the feature A). Now we can define the feature A (the region) as the target of the query, the target is the highlighted feature.

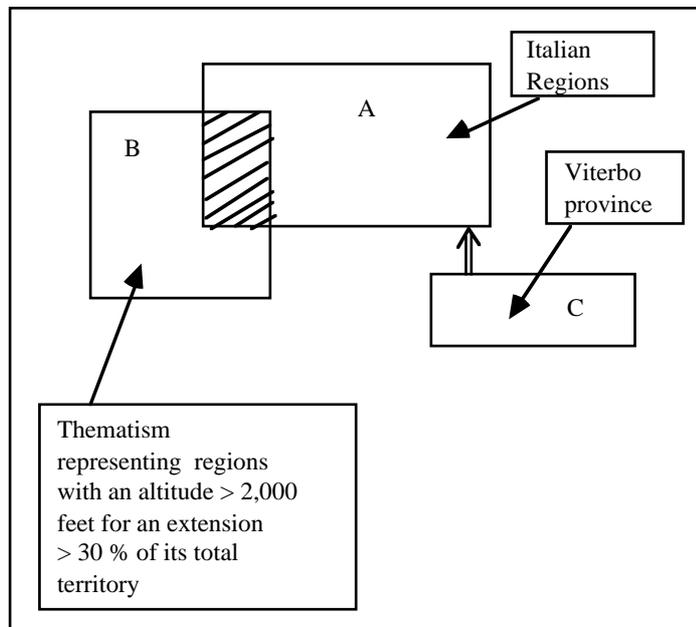


Figure 12. The pictorial query formulation

The execution of the pictorial query, that is, the target definition, is obtained as shown in Figure 13.

18.

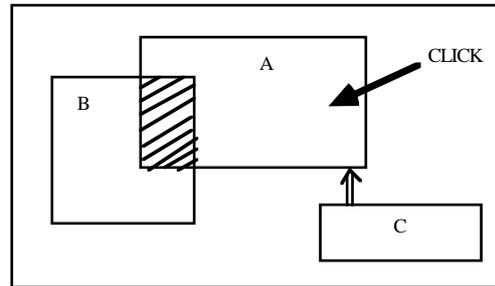


Figure 13. The pictorial query

The system, after the "click" on a well defined point of the picture representing the pictorial query, proposes all the possible, correct interpretations of the picture. In our case it is: $(A \text{ Ovl } B) \text{ Inc } F$.

The query expressed in a language SQL-like generated by the system is the following:

```
SELECT      ALL B.name
FROM        database := Italy
WITH        A.type = region
            B.type = Thematic map
            AB.type=region
            C.type = province
            C. property.name ="Viterbo"
WHERE       A Only_at_North_of C
```

7. Conclusions

The cardinal and positional operators discussed in this paper propose an enlargement and a refinement of other set of operators proposed in literature. Often such proposals are integrated in the context of new query languages. The definition of spatial query languages are among the relevant research topics. In fact, recent proposals to extend the SQL92 language with spatial operators is the proof of this trend of research [The98]. Spatial operators are used to capture the most relevant geometric properties of all the primitive geometric data types (namely points, lines, and regions). The OpenGIS Consortium, which proposed the most noticeable of the above mentioned extension of SQL, introduces several spatial operators that are extracted from previous researchers on spatial relations, also if more theoretical researches are needed to define a complete set of operators. The OpenGIS proposal defines a set of operators on the class Geometry classified in three groups: basic, topological and spatial analysis operators. In this proposal neither cardinal nor positional relations are considered. Our proposal goes in this direction.

Recently, in [CID00] several tasks, that involve the implementation of spatial operators in the context of query languages, are listed. Among them the most important are: small set of operators, expressiveness [CID95], consistency [EgF91a, CDC95, CID96], generality [GüS95], hierarchical structuring, especially with reference to topological and distance relations [CDK00, CDF94, CDH97], qualitiveness, where recent contributions deal with qualitative spatial (orientation and distance) relations [PaS94, CDH97, GoE01]. The last problem is considered from different point of views. For instance, in [PES96] the authors deal with reasoning about direction relations (East, NorthEast, NorthWest, RestrictedNorth, RestrictedWest, SamePosition) in spatial hierarchies. They identify 169 primitive relations between regions, extending to 2D space the 13 mutually exclusive relations between intervals

in 1D space discussed in [All83]. The main goal of their paper is to propose algorithms for the inference of relations between objects in different regions.

In [PaS94] the authors, as main goal, outline the computational tasks involved in relation-based information processing and describe symbolic spatial indexes in order to use them in a modelling space and to solve problems involving route planning, composition of spatial relations, and update operations. They also discuss the direction relations and their representation in spatial indexes.

Our work extends the set of positional and directional operators proposed in the above mentioned papers, defining them for two objects in 2D-space. For each of these operators, we give a detailed description and formalization, and we emphasize their main characteristics. They are defined mainly in order to facilitate the query processing and answering in geographic databases for a pictorial query language.

One of the main peculiarities which characterises this PQL with regard to other proposals (for example, [LCh95], [AP95], [Ege94], [KwM93], [Mai94], [Ege97] and [AuB99]) is the possibility to do query at an extensional level, that is, query on the database schema. The power of this query language to formulate easily queries that invoke the proposed operators through some examples are shown.

References

- [All83] J.F. Allen. Maintaining knowledge about temporal intervals. in *Communications of the ACM*, Vol. 26, N. 11, 1983, pp. 832-843.
- [AP95] M. A. Aufaure-Portier. A High Level Interface Language for GIS. *Journal of Visual Languages and Computing*, Vol. 6, 1995, pp. 167-182.
- [AuB99] M. A. Aufaure-Portier, C. Bonhomme. A High Level Visual Language for Spatial Data Management. in *Visual Information and Information Systems - Third Intern. Confer. Visual '99, Amsterdam, The Netherlands, June 1999, Lecture Notes in Computer Science*, N.1614, Springer-Verlag, 1999, pp. 325-332.
- [CCR93] Z. Cui, A.G. Cohn, D.A. Randell. Qualitative and Topological Relationships in Spatial Databases. in the *Third International Symposium on Advances in Spatial Databases, SSD'93, Singapore, June 1993, Lecture Notes in Computer Science*, N. 692, Springer-Verlag, 1993, pp. 296-315.
- [CDC95] E. Clementini, P. Di Felice, G. Califano. Composite regions in topological queries. in *Journal of Information Sciences*, Vol. 20, N. 7, 1995, pp. 579-594.
- [CDF94] E. Clementini, P. Di Felice. Topology in Object-Oriented GIS. *Second National Congress, Sistemi Evoluti per Basi di Dati, Rimini, Italy, 6-8 June 1994*.
- [CDH97] E. Clementini, P. Di Felice, D. Hernández. Qualitative representation of positional information. in *Journal of Artificial Intelligence*, Vol. 95, 1997, pp. 317-356.
- [CDK00] E. Clementini, P. Di Felice, K. Koperski. Mining multiple-level spatial association rules for objects with broad boundaries. in *Journal of Data and Knowledge Engineering*, Vol. 34, N. 3, 2000, pp. 251-270.
- [CID95] E. Clementini, P. Di Felice. A comparison of methods for representing topological relationships. in *Journal of Information Sciences*, Vol. 20, N. 3, 1995, pp. 149-178.
- [CID96] E. Clementini, P. Di Felice. A model for representing topological relationships between complex geometric features in spatial databases. in *Journal of Information Sciences*, Vol. 21, N. 1-4, 1996, pp. 121-136.
- [CID00] E. Clementini, P. Di Felice. Spatial Operators, in *ACM SIGMOD Record*, Vol. 29, N. 3, Sept. 2000, pp. 31-38.

- [CIT93] A.F. Cardenas, I.T. Jeong, R.K. Taira, R. Barker, C.M. Breant.: Knowledge-Based Object-Oriented PICQUERY+ Language. in *Journal of IEEE Transactions on Knowledge and Data Engineering*, Vol. 5, N. 4, 1993, pp. 644-657.
- [DGJ95] S. Dar, H. Gehani, V. Jagadish, J. Srinivasan. Queries in an object-oriented graphical interface, *Journal of Visual Languages and Computing*, 6(1): 27-52, 1995.
- [DeM93] L. De Floriani, P. Marzano. Spatial Queries and Data Models. in *Spatial Information Theory, A Theoretical Basis for GIS, European Conference, COSIT'93, Lecture Notes in Computer Science*, N. 716, Springer Verlag, 1993, pp. 113-138.
- [Ege89] M.J. Egenhofer. A formal definition of binary topological relationships. in *Lectures Notes in Computer Science*, N. 367, Springer Verlag, 1989, pp. 457-473.
- [Ege91] M.J. Egenhofer. Reasoning about Binary Topological Relations. in the *Second International Symposium on Advances in Spatial Databases, SSD'91, Zürich, Switzerland, August 28-30, 1991, Proceedings. Lecture Notes in Computer Science*, Vol. 525, Springer Verlag, 1991, pp. 143-160.
- [EgF91a] M.J. Egenhofer, R. Franzosa. Point-set topological spatial relations. in *International Journal of Geographic Information Systems*, Vol. 5, N. 2, 1991, pp. 161-174.
- [Ege94] M.J. Egenhofer. Spatial SQL: A Query and Presentation Language. in *Journal of IEEE Transactions on Knowledge and Data Engineering*, Vol. 6, N. 1, 1994, pp. 86-95.
- [Ege97] M. J. Egenhofer. Query processing in spatial-query-by-sketch. *Journal of Visual Languages and Computing*, 8, 403-424, 1997.
- [FMR99] F. Ferri , F. Massari, M. Rafanelli. PQL: A Pictorial Query Language for Geographic Features in an Object-Oriented Environment. in *Journal of Visual Languages and Computing*, Vol. 10, N. 6, 1999, pp. 641-671.
- [Fra82] A.U. Frank. MAPQUERY: Data Base Query Language for Retrieval of Geometric Data and their Graphical Representation. in *Journal of ACM Computer Graphics*, Vol. 16, N. 3, 1982, pp. 199-207.
- [Fra92] A.U. Frank. Qualitative Reasoning about Distances and Directions in Geographic Space. in *Journal of Visual Languages and Computing*, Vol. 3, N. 4, 1992, pp. 343-371.
- [GoE01] R.K. Goyal, M.J. Egenhofer. Cardinal directions between extended spatial objects. To appear, 2001.
- [GüS95] R. H. Güting, M. Schneider. Realm-Based Spatial Data Types: The ROSE Algebra. in *VLDB Journal*, Vol. 4, N. 2, 1995, pp. 243-286.
- [JoC88] T. Joseph, A. Cardenas. PIQUERY: A High Level Query Language for Pictorial Database Management., in *Journal of IEEE Transaction on Software Engineering*, Vol. 14, N. 5, 1988, pp. 630-638, (reprinted in *Research Foundations in OO and Semantic DBS*).
- [Kai90] W. Kainz. Spatial relationships-Topology versus order. in the *Second International Symposium on Advances in Spatial Databases, SSD'91, Zürich, Switzerland, August 28-30, 1991, Proceedings. Lecture Notes in Computer Science*, Vol. 525, Springer Verlag, 1991, pp. 814-819
- [KwM93] J. C. Kwak, S. Moon. Object Query diagram: An extended query graph for object-oriented databases. In. *Proc. IEEE Workshop on Visual Languages*, pp. 44-48, 1993.

- [KEG93] W. Kainz, M. Egenhofer, I. Greasley. Modelling spatial relations and operations with partially ordered sets. in *International Journal of Geographic Information Systems*, Vol. 7, N. 3, 1993, pp. 215-229.
- [Lor91] R.Lorie. The use of a complex object language in geographic data management. in the *Second International Symposium on Advances in Spatial Databases, SSD'91*, Zürich, Switzerland, August 28-30, 1991, Proceedings. *Lecture Notes in Computer Science*, Vol. 525, Springer Verlag, 1991, pp. 320-337.
- [LCh95] Y. C. Lee, F. L. Chin. An Iconic Query Language for Topological Relationship in GIS. *IJGIS* 9(1): 25-46, 1995.
- [Lor91] R. Lorie. The use of a complex object language in geographic data management. in the *Second International Symposium on Advances in Spatial Databases, SSD'91*, Zürich, Switzerland, August 28-30, 1991, Proceedings. *Lecture Notes in Computer Science*, Vol. 525, Springer Verlag, 1991, pp. 320-337.
- [Mai94] M. Mainguenaud, CIGALES: A Visual Query Language for Geographical Information System: The User Interface, *International Journal of Visual Languages and Computing*, Academic Press, 5, 113-126, 1994.
- [PaS94] D. Papadias, T.K. Sellis. Qualitative Representaion of Spatial Knowledge in Two-Dimensional Space. *VLDB Journal*, Ralf Hartmut Gutting Editor, Vol. 3, 1994, pp. 479-516.
- [PES96] D. Papadias, M.J. Egenhofer, J. Sharma. Hierarchical reasoning about direction relations. in the *4th ACM International Workshop on Advances in Geographic Information Systems*, November 15-16, 1996, Rockville, Maryland, USA, 1996, pp. 105-112.
- [RFS88] N. Roussopoulos, C. Faloutsos, and T.K. Sellis. An efficient pictorial database system for PSQL. in *Journal of IEEE Transactions on Software Engineering*, Vol. 14, N. 5, 1988, pp. 639-650.
- [The98] The OpenGIS Consortium. OpenGIS simple features specification for SQL. Technical report, <http://www.opengis.org/>, March 1998.
- [Whi57] H. Whitney. *Geometric Integration Theory*. Princeton University Press, 1957.
- [Wor92] M.F. Worboys. A generic model for planar geographic objects. in *International Journal of Geographical Information Systems*, Vol. 4, N. 4, pp. 369-383, 1990.