



ELSEVIER

Contents lists available at ScienceDirect

Journal of Visual Languages and Computing

journal homepage: www.elsevier.com/locate/jvlc

A conceptual framework for geographic knowledge engineering[☆]

Robert Laurini^{*}

LIRIS–UMR 5205, INSA de Lyon, Université de Lyon, Villeurbanne, Cedex F-69621, France

ARTICLE INFO

Keywords:

Geoprocessing fundamentals
 Geographic knowledge engineering
 Spatial knowledge engineering
 Visual language
 Geographic reasoning

ABSTRACT

In many applications, the management of geographic knowledge is very important especially not only for urban and environmental planning, but also for any application in territorial intelligence. However there are several practical problems hindering the efficiency, some of them being technical and other being more conceptual. The goal of this paper is to present a tentative conceptual framework for managing practical geographic knowledge taking account of accuracy, rotundity of earth, the mobility of objects, multiple-representation, multi-scale, existence of sliver polygons, differences in classifying real features (ontologies), the many-to-many relationship of place names (gazetteers) and the necessity of interoperability. In other words, this framework must be robust against scaling, generalization and small measurement errors. Therefore, geographic objects must be distinguished into several classes of objects with different properties, namely geodetic objects, administrative objects, manmade objects and natural objects. Regarding spatial relations, in addition to conventional topological and projective relations, other relations including tessellations and ribbon topology relations are presented in order to help model geographic objects by integrating more practical semantics. Any conceptual framework is based on principles which are overall guidelines and rules; moreover, principles allow at making predictions and drawing implications and are finally the basic building blocks of theoretical models. But before identifying the principles, one needs some preliminary considerations named prolegomena. In our case, principles will be essentially rules for transforming geographic knowledge whereas prolegomena will be assertions regarding more the foundations of geographic science. Based on those considerations, 12 principles are given, preceded by 12 prolegomena. For instance, some principles deal with the transformation of spatial relationships based on visual acuity and granularity of interest, with the influence of neighboring information and cross-boundary interoperability. New categories of geographic knowledge types are presented, spatial facts, cluster of areas, flows of persons, goods, etc., topological constraints and co-location rules. To represent knowledge chunks, three styles are presented, based respectively on descriptive logics, XML and visual languages. To conclude this paper, after having defined contexts of interpretation, an example of visual language to manage geographic knowledge is proposed.

© 2013 The Author. Published by Elsevier Ltd. All rights reserved.

1. Introduction

In many applications regarding business and administration, knowledge management is very important in order to be more efficient. In geoprocessing, many knowledge engineering experiences were very disappointing essentially because of the naïve nature of modeling. Moreover, they present a lot of limitations due to several reasons among them the more important seem to be the use of geometry,

[☆] This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial-No Derivative Works License, which permits non-commercial use, distribution, and reproduction in any medium, provided the original author and source are credited.

This paper has been recommended for acceptance by Shi Kho Chang.

^{*} Tel.: +33 4 72 43 81 72; fax: +33 4 72 43 87 13.

E-mail address: Robert.Laurini@insa-lyon.fr

the importance of measurement errors and the difficulties not only to encode geographic knowledge but also to use it for geographic information retrieval and reasoning. Many years ago, a small book [27] was published on spatial knowledge engineering, but the situation now is totally different and the underlying assumptions must be largely revisited.

What is the difference between spatial knowledge and geographic knowledge? Someone deals with spatial knowledge when only geometry and topology are implied: so spatial knowledge can be seen as an extension of geometric knowledge. When one says “if A is north to B, and B is north to C then A is north to C”, this is spatial knowledge. However as soon as geographic features are used, then one deals with geographic knowledge. So when saying “London is north to Paris and Paris north to Madrid, then London is north to Madrid”, this assertion is a chunk of geographic knowledge. Similarly, one must speak about geographic data mining as a new way to extract geographic knowledge.

By extension, we can define “geographic knowledge” as information useful to solve geographic problems in various domains varying from archeology to zoology, from urban planning [24] to geostrategy, from real time sensors to crowdsourcing.

Conventional knowledge engineering is based on set theoretic, propositional, predicate, description or modal logics, in which one can distinguish declarative knowledge and procedural knowledge. Knowledge is usually decomposed as:

- facts, which are data or instance,
- concepts, which are classes of items,
- processes, which are flows of events,
- and rules, which allow to make inferences or draw implications.

However for geographic knowledge, other mathematical disciplines must be invoked. To name a few, operation research, graph and hypergraph theories, computational geometry, topology, fuzzy set theory, mathematical morphology, statistics, spatial analysis, etc. As a consequence, from a mathematical point of view, geographic knowledge modeling and engineering must include those aspects.

During centuries, cartography was the only mode to keep geographic information; but from half of a century, the difference between storing geographic information and visualization has been more and more clarified. But cartography (therefore visualization) is still the common way to represent geographic issues. And the next step will be to provide tools for reasoning and geographic Internet retrieval.

Additional difficulties to constitute geographic knowledge repositories come from several practical reasons such as the problems of:

- *links of geographic semantics and scale*: due to several practical reasons, a state politician and a city mayor do not reason similarly about a territory essentially because they do not deal with the same issues (global vision vs local vision); for centuries, in cartography, practical geographic semantics are heavily linked to scale;
- *accuracy*; indeed different apparatuses can deliver different values, for instance regarding the measurements of coordinates;
- *rotundity of earth*; even if for some practical applications, some small territories can be considered as planar;
- *moving objects*; due to continental drift all geographic objects are moving; suppose a continent going away 2 cm per year from Greenwich, do we have to correct coordinates accordingly every year?
- *multiple-representation*; indeed for instance according to applications a road can be seen as an edge in a graph, a line, a surface and a 3D object when dealing with engineering networks (see Fig. 1).
- *multi-scale*; because a real country can be modeled by a polygon with 1000 points or 100 points [4]: a nice solution is to handle multi-scale by generalization [8,14];
- *existence of sliver polygons* essentially in tessellations, for instance country-provinces;
- *differences in classifying* real features, e.g. ontologies;
- *toponymy*: several places can have the same name (Washington State, Washington DC, etc.), whereas a place can have different names, according to time (Byzantium, Constantinople, Istanbul) or according to languages (Venezia, Venice, Venise, Venedig, etc.); the role of gazetteers is to include all those variants;

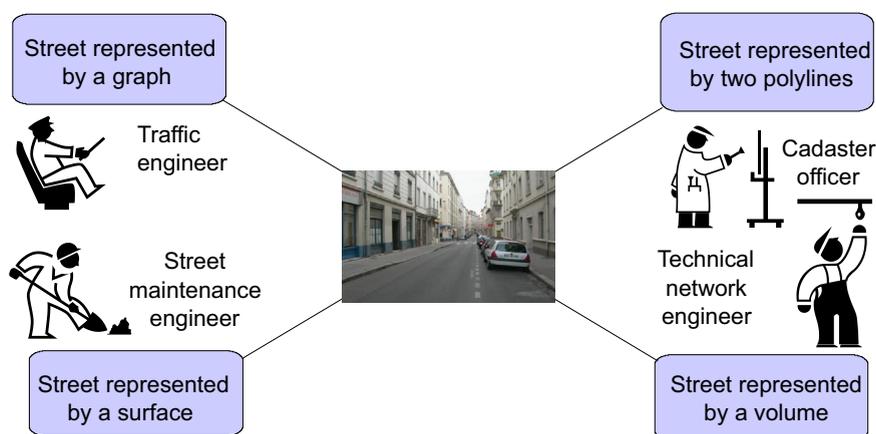


Fig. 1. Example of multiple representations of a street.

- *necessity of interoperability* between several systems belonging to different jurisdictions characterized by different ontologies and cross-boundary misalignment.

Now, for any application in territorial intelligence and smart city planning, not only new software products must be created, but overall a new conceptual framework must be set by integrating artificial intelligence, knowledge engineering, computational geometry and spatial reasoning.

Taking all those characteristics into account, the ultimate scientific question of this work is to develop a conceptual framework in order to deal with geographic knowledge management in a robust and consistent form. In other words, the goal is to list principles which can constitute a sound basis for geographic knowledge management which must be robust against scaling, generalization and errors. But before presenting those principles, key-concepts such as geographic objects and relations must be revisited. Then some preliminary notions must be stated as prolegomena.

One of the problems is the link between spatial relationships and scale. In cartography, the meaning of scale is clear, i.e. the ratio between the distance in the real world and the graphic representation (map). But in reality, semantics are hidden in scale: a cartographer does not visualize the same features at a 1:100 or at 1:1000000 scales because smaller objects disappear. Scale is not simply a multiplier (usually less than 1), but it is also an indicator about the semantics of the map. This process will be called generalization of topological relations.

Temporal and spatio-temporal knowledge is outside the scope of this paper.

The contributions of this paper are the presentation of ribbons, ribbon topology, generalization of topological relations, the construction of the foundations of a conceptual framework and a visual language for geographic knowledge management.

In this paper, many references can be cited but due to the limit of this paper, only the more seminal will be mentioned.

2. Related works

As there are many works relative to geographic data, few works have been done in order to develop a complete theory.

Maybe, the more important seems to be the work made by Goodchild et al. [20]. In essence they tried to construct a theory of geography based on a selection of concepts which can be derived from a single foundation that they term the atomic form (Geo-atom) of geographic information. This geo-atom is defined as an association between a point location in space-time and a property and can be applied to both continuous fields and discrete objects. But noting is mentioned regarding multi-representation, acquisition errors and geographic knowledge.

Golledge [18] tries to define the nature of geographic knowledge, but it is essentially revisited from a cognitive point of view. Crowther [12] examined the properties of geographic knowledge extracted from satellites images,

whereas Aldridge [1] studied geographic knowledge coming from sensors.

Several works have been done regarding geographic knowledge extracted from spatial data mining techniques such as Mennis and Peuquet [29], Ester et al. [16], etc.; but none studied the principles for the management of this knowledge.

Some other works must be mentioned based on the Sowa's seminal work [32] on knowledge represented as conceptual graphs, so to give ontologies. Regarding geographic ontologies, let's mention Karalopoulos et al. [23], the Towntology project [33], etc. Here, the authors try essentially to model features whereas few was made regarding acquisition errors and spatial reasoning.

In their famous book on knowledge representation and reasoning, Brachman and Levesque [6] do not treat the specificities of geographic knowledge. However, in their work about knowledge bases, they introduce the notation $KB| = a$, which can be read " a is a logical consequence issued from the contents of the knowledge base KB ", or " KB entails a ".

As a consequence, the development of a consistent framework for geographic knowledge management is still on the research agenda especially as a basis for automatic reasoning in urban and environmental planning and territorial intelligence.

3. Methodology

Our objective is to define a conceptual framework for the management of geographic knowledge. Any conceptual framework is based on principles which are overall guidelines and rules; moreover, principles allow to make predictions and draw implications and are finally the basic building blocks of theoretical models. But before identifying the principles, one needs some preliminary considerations named prolegomena. In our case, principles will be essentially rules for transforming geographic knowledge whereas prolegomena will be assertions regarding more the foundations of geographic science.

Following a recommendation of an anonymous referee, a survey was launched in September 2013 to experts on GIS. Those persons were chosen among the scientific boards of major GIS reviews and/or among program committees of the major GIS conferences. A total of 57 experts were consulted and 17 answers were received.

4. Geographic key-concepts

In this section, mathematical modeling tools will be presented and emphasis will be given on ribbons, ribbon topology and generalization of topological relations with scale.

4.1. Modeling tools

It is common to state that there 0D (points), 1D (lines), 2D (areas) and 3D (solids) geometric objects for modeling geographic objects. But the reality is much more complex, where are located 0D objects on the Earth? Apart the

North and the South poles, there are no 0D objects. It is also common to state that streets and rivers can be modeled as lines or polylines, but in reality they are areas which specific properties so that they can be reduced to lines when needed. In this paper, the mathematical concept of ribbon will be used.

Considering two polylines A and B , what is the distance between them? An interesting definition is given by the Frechet distance which corresponds to the minimum leash between a dog and its owner, the dog walking on a line, and the owner in the other line as they walk without backtracking along their respective curves from one endpoint to the other. The definition is symmetric with respect to the two curves. By noting a , a point of A , and b of B , the Frechet Distance F is given as follows in which $dist$ is the conventional Euclidean distance [3]:

$$F = \text{Max}_{a \in A} (\text{Min}_{b \in B} (\text{dist}(a, b)))$$

But in our case, we must consider two distances, let us say, the minimum and the maximum of the leash, so giving:

$$d1 = \text{Min}_{a \in A} (\text{Min}_{b \in B} (\text{dist}(a, b))) \text{ and } d2 = \text{Max}_{a \in A} (\text{Min}_{b \in B} (\text{dist}(a, b)))$$

4.2. Geometric objects

In addition to the conventional Euclidean representations, several other geometric objects must be used, i.e. fuzzy geometric objects and ribbons.

4.2.1. Fuzzy geometric objects

As in Euclidean representation a point belongs or does not belong to a geometric object (say a polygon), in fuzzy representation a point belongs with a certain membership grade, for instance 75% (Zadeh [37]). In other words, a fuzzy geometric object is represented by a set of polylines similar to contour lines with different degrees (see Fig. 3).

But the fuzzy representation is not so easy to manipulate and Randell, Zhan and Cohn [9,31] have proposed another view based on the egg metaphor which is the so-called egg-yolk model in which an object has only two boundaries, the kernel (yolk) and its maximal extension (egg). For instance for modeling a river, one can distinguish the major and the minor beds. The yolk width will be defined by the minimum and the maximum Frechet distance.

4.2.2. Ribbon

Another interesting model for is based on ribbons which may elegantly model rivers and roads (so-called linear objects): a ribbon can be loosely defined as a line or a polyline with a width. Mathematically speaking, a ribbon is defined as a homeomorphic transformation of a longish rectangle. The width of a ribbon is defined as the Frechet distance between both longish sides. According to the ribbon definition, the axis is the core line; however based on the other definition, by considering each point of a longish side, and its Frechet distance to the point of the opposite side, the axis is the locus where the Frechet distance is 50%. See Fig. 2b for an example.

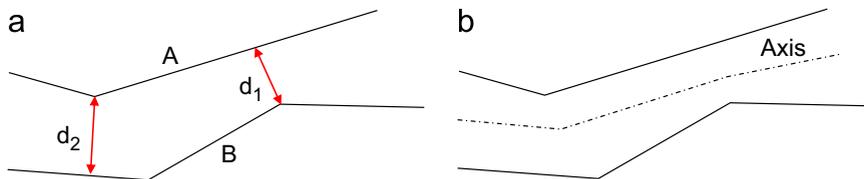


Fig. 2. Two polylines. (a) Distances. (b) Axis.

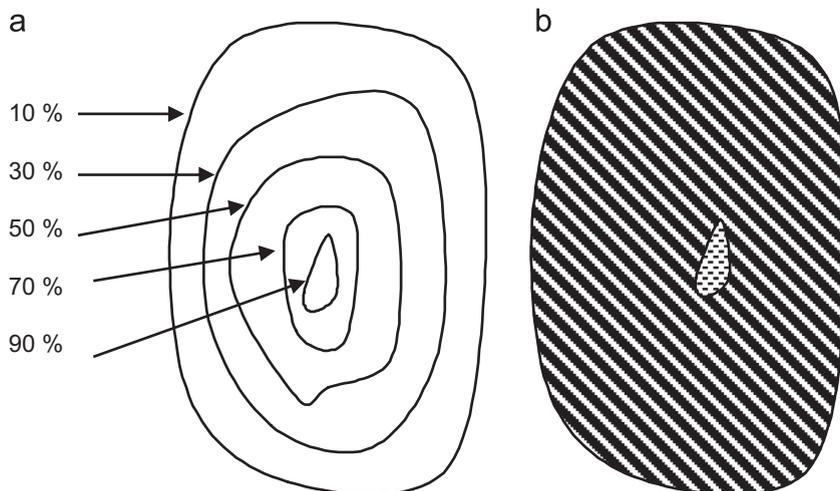


Fig. 3. Representations of geographic objects with undetermined boundaries. (a) Fuzzy set representation. (b) Egg-yolk representation.

Let us denote R a ribbon, ρ a rectangle and H an homeomorphism (Fig. 4) so that

$$R = \mathcal{H}(\rho) \quad \text{and} \quad \rho = \mathcal{H}^{-1}(R)$$

Let us note $W(R)$ and $Axis(R)$ respectively the width and the axis of a ribbon. Remember that the homeomorphism can produced holes in the ribbon. Moreover, a ribbon can be closed for example to model correctly car racing tracks.

In the sequel of this paper, to simplify the presentation, a ribbon will be represented by a longish rectangle, that is to say the homeomorphism is identity.

4.3. Geographic objects

The digital representation of geographic features is generally based on the Euclidean planar and spherical geometry. As soon as the boundaries are known, the objects can be modeled with polygons, otherwise fuzzy set geometry can also be used in some cases. Among geographic features, let us consider:

- *Geodetic objects*: they constitute the basis for geoid coordinates, i.e. equator, North and South poles, meridians and parallels. There are modeled with points and circles.
- *Administrative objects* (countries, regions, cities, parcels, parks, etc.): generally they can be described as 2D non-connected polygons. Often they form irregular tessellations and sometimes hierarchical tessellations.
- *Manmade objects* (buildings, bridges, tunnels, etc.): they are 3D objects, but in conventional GIS they are represented only at 2D. Many works try to create 3D GIS.
- *Natural objects*: generally boundaries are not determinate; for instance for a river, once has to distinguish minor beds and major beds, limits of mountains are not known, continents and seas are moving, etc. For these features, fuzzy sets can be a good candidate for modeling. In geology, sometimes the boundaries between objects can be crisp or fuzzy. For natural phenomena such as temperature, pressure and winds, models based on continuous fields can be of interest [11].

In the sequel of this paper, the word “feature” will be used for real objects whereas the expression “geographic object” will be used from a computing point of view, that is the computer representation of the feature; this representation can varied according to time, scale and precision. As a consequence topological relations between features will be modeled by topological relations between geographic objects. So, as a real bridge correspond to the crossing of a river and a road. From a computing point of view, they form the intersection of two ribbons; according to scale, those ribbons can be transformed into lines, and eventually, one or both of them can be invisible; as a consequence the topological relation will vanish.

4.4. Geographic relations

Now let us explore relations. In addition to “is-a”, “has-a”, “part-whole” relations which are common in set theory, logics and ontologies, let us explore spatial relations, first

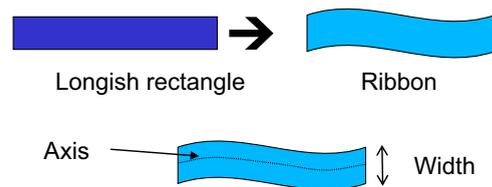


Fig. 4. A ribbon as an homeomorphism of a longish rectangle with its axes and its width.

conventional spatial relations and then more sophisticated extensions.

4.4.1. Conventional spatial relations

They are the following:

- *Topological relations*: when boundaries are known, Allen for 1D [2], Egenhofer for 2D [15] relations can be used (Fig. 5). The alternative model based on the egg-yolk metaphor can also be used. Theoretically speaking, topological relations are invariant with elastic transformations, but in geography, this is different: indeed, due to scale, a road and a lake can be disjoint or related with a TOUCHES relation; moreover, the road and/or the lake can disappear. Consequently, any realistic geographic knowledge system must integrate this aspect. See also [17].
- *Projective relations* linked to cardinal points (north, south, east, and west) which have special properties: nothing at the north of the North Pole, limited spherical transitivity relative to east and west. However, whereas projective relations are well defined for points, but for areas, the situation is different: indeed, one can say that “Canada is north of the USA” whereas some cities of Canada such as Toronto are south of Seattle. In this case, this definition must be extended.
- *Distance relations*: two objects can be neighbor or remote: for instance an airport is generally at small distance of a city.
- *Mereological relations* between non-connected parts of a polygon. For instance Alaska has a mereological relation with the conterminous states of the USA.

However in real applications, more sophisticated spatial relations or structures are necessary. Let us examine a few.

4.4.2. Ribbon relations

In a recent paper [26], ribbon relations were proposed to describe streets, roads and rivers; for instance a motorway (Fig. 6a) can be described by several ribbons corresponding to several driving lanes, emergency lanes and one median. Several other objects can be described with ribbon relations such as vegetation layers in mountains.

Four relations can be defined with ribbons as exemplified in Fig. 6b, side-by-side, end-to-end, fusion (or merging) and splitting. For describing a section of a highway, Fig. 6c is based on with a Pascal-like formalism in which the symbol “||” denotes parallel ribbon linked by a side-by-side relation.

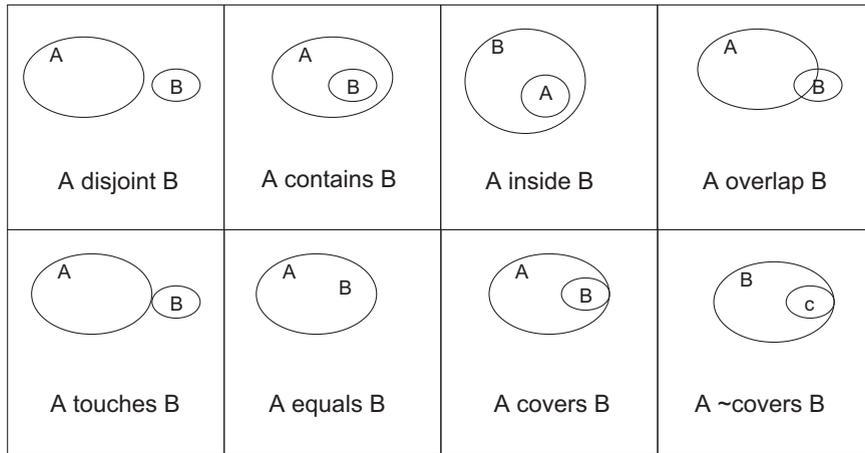


Fig. 5. Egenhofer topological relations (1994).

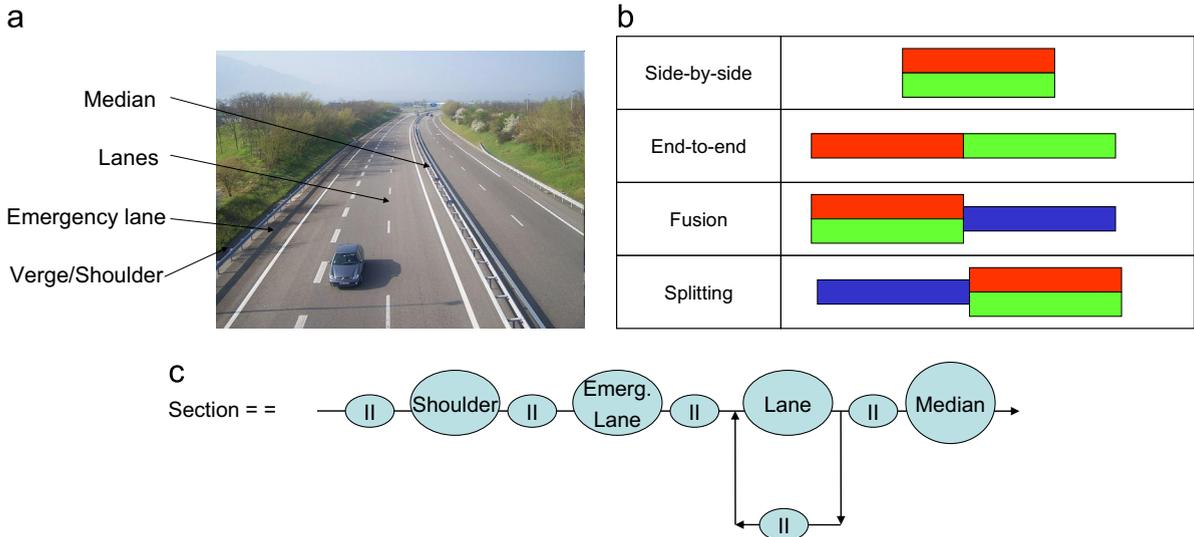


Fig. 6. Modeling a motorway by ribbon relations. (a) Example. (b) basic ribbon relations, (c) modeling of a portion of this motorway by Pascal-like formalisms in which “||” means a “side-by-side” operator.

For a real world feature (f.i. a road or a river), it can be modeled by a single composite ribbon, that is a set of ribbons linked by side-by-side and/or end-by-end relations. As scale diminishes, ribbons will be reduced to lines, for instance to their axes.

4.4.3. Irregular tessellation

By irregular tessellation (or tessellation) one means the total coverage of an area by subareas see Fig. 7a. For instance the conterminous States in the USA form a tessellation to cover the whole country. Generally speaking administrative subdivisions form tessellations, sometimes hierarchical tessellations. Let us consider a domain D and several polygons P_i ; they form a tessellation iff:

- For any point p_k , if p_k belongs to D then there exists P_j , so that p_k belongs to P_j
- For any p_k belonging to P_j , then p_k belongs to D .

Remark 1: a tessellation can be also described by Egenhofer relations applied to P_i and D .

Remark 2: in practical cases, due to measurement errors, this definition must be relaxed in order to include sliver polygons; let's call them “loose tessellations”. See Fig. 7b.

Remark 3: one must include spherical tessellations and 3D tessellations.

Remark 4: tessellation can be non-connected; for instance Italy is a good example with islands (Sicily, etc.) and Holes (San Marino and Vatican City).

Remark 5: for terrain modeling, a TIN (Triangulated irregular network) constitutes a special case of 3D triangular tessellation.

In some cases, particularly for administrative objects, a tessellation can be hierarchical, such as in the USA, country, states, and counties. That is a polygon belonging to a tessellation is itself a tessellation of other polygons.

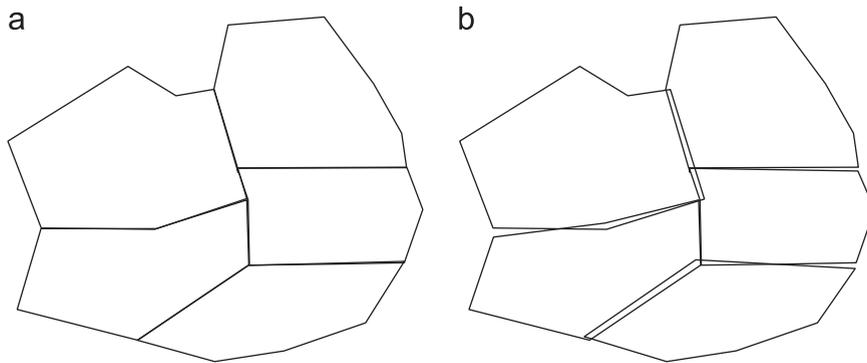


Fig. 7. Examples of irregular tessellations. (a) A mathematical good-looking tessellation. (b) A practical tessellation (loose tessellation) with sliver polygons in which errors are voluntarily exaggerated.

4.4.4. Networks

In several applications the concept of network can be of interest but it is not only the naïve extension of graph, even if graph theory can be useful to deal with networks. Indeed:

- *water supply networks* include water towers pipes, taps, etc. The first idea could be a tree-like structure. But to get a minimum number of people without water when repairs are need, a more general graph is necessary; in addition often in different municipalities, water supply networks are interconnected. A similar structure applies to gas network.
- *sewerage networks* are graphs embedded in 3D tubes or galleries; the ramp is continuous and they include manholes for repair.
- *telecommunication networks* are characterized by redundancy: indeed in order to change rapidly, some additional cables are already installed to increase the speed for new subscribers.
- *transportation networks* such as bus transportation, metro lines, boat and airways networks have different characteristics.
- etc.

Consequently, one can see that these structures are richer than conventional graphs essentially due to 3D characteristics. As a consequence, a 3D extension of ribbon will be necessary; but this consideration is out the scope of this paper. As scale diminishes, those 3D ribbons are transformed into lines, and eventually disappear.

4.5. From visual acuity to granularity of interest

As previously told according to scales, geographic objects can evolve. At the beginning, all geographic objects are 2D. When scale diminishes, they are generalized and transformed into points or lines (see Table 1). Finally they can disappear. During this process, the topological relations change. See examples of objects transformations in Fig. 8.

This process can be modeled as follows:

Step 0: original geographic objects with the major accuracy,

Step 1: as scale diminishes, small areas and ribbons will be generalized and possibly can coalesce,
 Step 2: as scale continues to diminish, areas are transformed into points and ribbons into lines,
 Step 3: as scale continues to diminish, points and lines disappear.

Let us call this process “generalization-reduction-disappearance” (GRD process).

But let us examine some particular cases.

- Suppose one wants an itinerary in France from Charles de Gaulle airport to St Tropez printed on a A4 map. Due to scale, applying strictly the previous rules, the three objects airport, village and motorways will disappear; but they should not. Let’s call it “query predominance”.
- Let consider a river such as the Danube whose width varies from 100 m to 800 m or more. At some scale, some sections may disappear. In order to ensure continuity, those sections must stay. Similarly when roads vary in width (sections with 2, 3 or more lanes), the narrow sections must stay. In other words, when dealing a composite ribbon, the widest section must be considered to pass from ribbon to line, and then from line to void. Let’s call it “graph continuity”.

Therefore the GRD process must be superseded by two rules.

Rule #1: when a feature is important in a query or in a reasoning method, it must stay even if at this level it may stay, in order to solve the first problem previously mentioned concerning query predominance.

Rule #2 (graph continuity): when a section of a composite ribbon is candidate to disappear, it may stay to ensure graph continuity (minimum vs maximum Frechet distance).

However the concept of visual acuity is too much linked to cartography. In order not to bind this concept to cartography, the notion of “granularity of interest” can

Table 1
Generalization areas and ribbons according to scale.

Higher scale	Mean scale	Lower scale
		
		

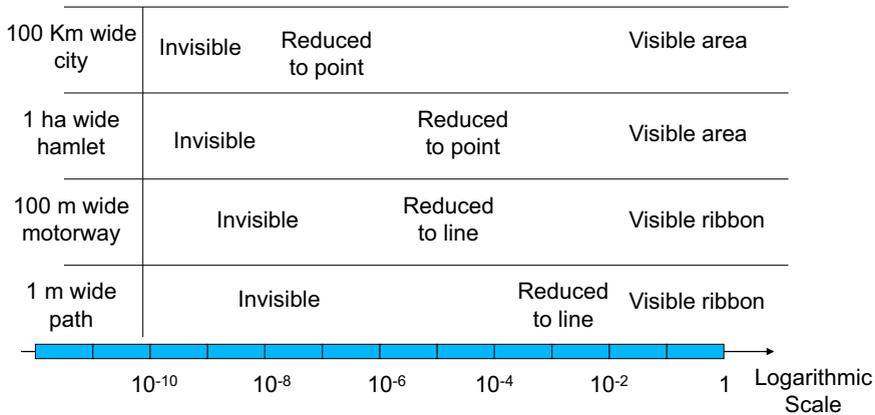


Fig. 8. Generalization of geographic objects according to scale.

extend it. Nevertheless to facilitate the understanding of this paper, visual acuity will be used instead.

Therefore two thresholds must be defined for instance based on visual acuity. Let me propose ϵ_i for invisibility threshold and ϵ_{pl} the threshold for reduction to point or line. As example, let us take $\epsilon_i = 0.1$ mm and $\epsilon_{pl} = 1$ mm.

4.6. Generalization of topological relations

There are some categories of geographic relations which can be of interest. Those spatial relations constitute the key concepts for organizing geographic ontologies [26]. Like geographic objects, relations can be generalized; the objective of this section is to give some characteristics.

Concerning topological relations, those relations will change: indeed (see Fig. 9), suppose two objects not far from each others. At the beginning they share a DISJOINT relation, then when scale diminishes, they are related by a Meet relation; after the smaller will disappear, and then both will disappear. And the same reasoning could apply to other topological relations.

5. Prolegomena

Now that the key concepts are established, we can state some prolegomena as preliminary assertions constituting the underlying foundations of principles. They are

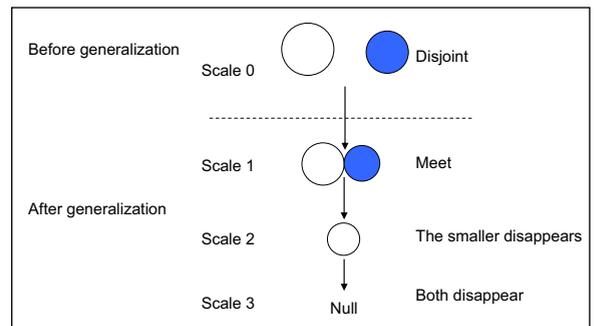


Fig. 9. Example of topological generalization according to scale, from Disjoint to Meet, and then disappearance.

organized as follows:

- the two first prolegomena state the origin of geographic data,
- the two next, particular cases of data transformation,
- the two next, updating of data,
- the five next ones, the structuring of objects and of geographic information,
- and the last one, the well-known Tobler's law.

Prolegomenon #1 (3D + T objects): "All existing objects are tridimensional and can have temporal evolution; lower dimensions (0D, 1D and 2D) are only used for modeling (in databases) and visualization (in cartography)". Unlike

geodetic objects which were created by man, all features are 3D, can move, can change their shape and can be destroyed.

Prolegomenon #2 (Acquisition by measurements): “All basic attributes (spatial or non-spatial) are obtained by means of measuring apparatuses having some limited accuracy”. Now more and more data come from sensors; more, citizens can be seen as sensors [19]. In other, the word “apparatus” must be taken in a very wide sense, from sensors to census, etc.

In metadata, accuracy is perhaps one of the most important features, but too often real applications do not care enough about accuracy. One of the big practical difficulties is when different subsets of databases were acquired with various accuracies.

Prolegomenon #3 (Continuous fields): “Since it is not possible to store the infinite number of value points in a continuous field, some sampling points will be used to generate the whole field by interpolation”. As a consequence, special data structures must be developed in conjunction with interpolating functions to estimate value anywhere in the field. See for instance Vckovski [36], Gordillo [21] or Kang et al. [22] for more details.

Prolegomenon #4 (Raster-vector and vector-raster transformations): “Procedures transforming vector to raster data and raster to vector data must be implemented with losing less accuracy as possible”. Any geographic knowledge system must include those procedures.

Prolegomenon #5 (From Popper's falsifiability principle [30]): “When a new apparatus delivers measures with higher accuracy, these measures supersede the previous ones”. The practical consequence is that as a new generation of data comes, geographic data and knowledge basis must integrate those data and remove the previous data. But alas, due to the acquisition cost, a lot of actual systems are based on “obsolete” data.

Prolegomenon #6 (Permanent updating): “Since objects are evolving either continuously (sea, continental drift) or event-based (removing building), updating should be done permanently respectively in real-time and as soon as possible”. Remember that “updating” in computing means three different things, (i) a characteristic of an object has varied (f.i. land use in a parcel), (ii) the class of an object (so its description) has varied (a building formerly a residence is now for business), (iii) an error has been discovered in this object and then corrected (f.i. wrong coordinates or attributes). This prolegomenon implies that any procedure to check or increase data quality must be invoked.

Prolegomenon #7 (Geographic metadata): “All geographic databases or repositories must be accompanied with metadata”. The necessity to accompany data by information regarding lineage and accuracy was first observed in the GIS domain. More precisely, now the International Standard ISO 19115 “Geographic Information - Metadata” from ISO/TC 211 provides information about the identification, the extent, the quality, the spatial and temporal schema, spatial reference, and distribution of digital geographic data. Practically, many geographic databases do not implement the whole standard, but only the more important aspects, because it is very time-consuming. Moreover metadata must be also updated when necessary.

Prolegomenon #8 (Cartographic objects): “In cartography, it is common to eliminate objects, to displace or to

simplify them”. This is due to ensure a maximal readability of maps.

Prolegomenon #9 (One storing, several visualizations): “A good practice should be to store all geographic objects with the highest possible accuracy and to generate other shapes by means of generalization”. This can be seen as an extension of the well-known Douglas-Peucker's family of methods and algorithms for generalization [14,8].

Prolegomenon #10 (Place names and gazetteers): “Relationships between places and place names are many-to-many”. Mississippi is the name of a river and the name of a state. The actual city of Rome, Italy, is larger than the same Rome in Romulus's time. The main consequence is that unique feature identifiers must be defined since “popular names” are not so easy to digitally manipulate.

Prolegomenon #11 (Geographic ontologies): “All geographic object types are linked to concepts organized into a geographic ontology based on topological relations”. This comes from my own definition of geographic ontologies [26]. When necessary, raster information can be included into ontologies. For instance, roof textures can be used to identify a building, a wood texture for a wood, a corn field texture to a corn field, possibly with different level of maturity.

In the case of federation of several geographic databases, interoperability is often governed by ontologies. If ontologies of each database are different, a global ontology must be defined from the so-called local ontologies.

Prolegomenon #12 (Tobler's law [34]): “Everything is related to everything else, but near things are more related than distant things”. This statement may be seen as a key-concept also for geographic data mining.

6. Principles

Now that prolegomena are stated, principles governing geographic knowledge may be listed in order to get robust reasoning and retrieval. Let us note *GDB* a geographic database and *GKB* a geographic knowledge base. By definition, let us state $GKB_0 = GDB$.

The principles are organized as follows:

- The three first concerned the origin of geographic knowledge,
- The seven next ones, the transformation of geographic knowledge,
- The two last ones take the environment into account.

Principle #1 (Origin of geographic knowledge): “Spatial knowledge is hidden in geometry whereas geographic knowledge comes in addition from non-spatial attributes”.

In other words, spatial knowledge is implicit and the question is whether to make it explicit. We can derive from coordinates that New-York city is west of Paris, and for many cities throughout the world. The good practice is to derive knowledge on-demand when necessary.

In addition, data coming sensors will support geographic knowledge whereas any indicator will be seen as composite knowledge derived from measures.

Some geographic knowledge can be extracted from data mining techniques. Therefore:

$GKB_1 | = \text{Principle 1}(GKB_0)$

Principle #2 (Knowledge cleaning): “All geographic data, once captured, must be cleaned to remove errors and artifacts to get consistent knowledge”. This principle is directly connected with Prolegomenon #6 since all automatic acquisition system may include errors or anomalies. For instance, any airborne laser beam to capture digital data for terrain or elevation can intercept a bird: in this case, the captured data will no more be the terrain altitude, but the bird altitude. Based on this principle, all procedures to increase geographic knowledge quality must be invoked. Consequently,

$GKB_2 | = \text{Principle2}(GKB_1)$

However, in practical situations, geographic data or knowledge bases can still encompass some remaining (not yet discovered) errors, so implying often wrong results in treatment and reasoning. End-users must take care.

Principle #3 (Knowledge enumeration): “It is not necessary to enumerate all possible chunks of geographic knowledge”. For instance, if one has n object, then $(n-1)^2$ North–South relationships can be also derived accordingly. Indeed, it is truly possible to derive them automatically when reasoning.

In other words, since GKB is infinite (intensional), only implicit knowledge is stored, but other knowledge chunks can be derived when necessary.

$GKB_3 | = \text{Principle3}(GKB_2)$

Principle #4 (From geoid to plane): “On small territories, a planar representation is sufficient whereas for big territories, Earth rotundity must be taken into consideration”. But the question is “how to define a small or a big territory”? A solution can be to define a threshold, for instance a 100 km wide square. Let write O_σ , the planar map for any geographic object O at scale σ taking generalization into account: $O_\sigma = 2Dmap(O)$. As a consequence:

$GKB_4 | = \text{Principle4}(GKB_3)$

Principle #5 (Visualization and visual acuity): “Cartographic representation is linked to visual acuity”. Here again thresholds must be defined. In classical cartography, the limit ranges from 1 mm to 0.1 mm. Suppose that one takes a road and a certain scale: if the transformation gives a width more than 1 mm, this road is an area, between 1 mm and 0.1 mm a line, and less than 0.1 mm the road disappears. The same reasoning is valid for cities or small countries such as Andorra, San Marino, Monaco, etc. In these cases, the “holes” in Italy or in France disappear cartographically. With the thresholds ϵ_i , ϵ_{ip} previously defined, we can formally get (in which 2Dmap is a

function transforming a geographic object at some scale possibly with generalization)::

$$a/\forall O \in GeObject, \forall \sigma \in Scale \wedge$$

$$O_\sigma = 2Dmap(O) \wedge Area(O_\sigma) < (\epsilon_{ip})^2 \Rightarrow O_\sigma = \emptyset. b/\forall O \in GeObject, \forall \sigma \in Scale \wedge$$

$$O_\sigma = 2Dmap(O) \wedge (\epsilon_i)^2 > Area(O_\sigma) > (\epsilon_{ip})^2 \Rightarrow O_\sigma = \text{Centroid}(O).$$

But this principle must be relaxed (Rule #1) when one has to map small objects. For instance, let us consider an A4-format map showing Roman churches in France, those churches must be stayed whereas due to scale they should disappear.

The other interesting case regards loose tessellations, i.e. “tessellations” with sliver polygons: when scale diminishes, those sliver polygons will vanish due to visual acuity, and so leading to a good-standing tessellation. As a consequence:

$GKB_5 | = \text{Principle5}(GKB_4)$

As previously explained, this principle can be reformulated taking the concept of granularity of interest into account.

Principle #6 (Crispification): “At some scales every fuzzy object becomes crisp”. If the egg-yolk representation is adopted to represent of geo-object, when the egg white distance is less than a threshold, the geo-object geometry can be taken for instance where the membership grade is 50%. Fig. 10 illustrates this process. This process is similar to the reduction of a ribbon to a line.

In the case of the fuzzy set representation, the 50% membership contour line can represent the boundary of the so-transformed polygon. Therefore:

$GKB_6 | = \text{Principle6}(GKB_5)$

Principle #7 (Relativity of spatial relations): “Spatial relation varies according to scale”. Commonly, one says that a road runs along a lake. But in reality, in some place, the road does not run really along the water of the lake due to beaches, buildings, etc. At one scale, the road TOUCHES the lake, but at another scale at some places, this is a DISJOINT relation (Fig. 8). Let consider two geographic objects O^1 and O^2 and their O_σ^1 and O_σ^2 their cartographic representations, for instance the following assertion holds:

$$\forall O^1, O^2 \in GeObject \wedge \forall \sigma \in Scale \wedge$$

$$O_\sigma^1 = 2Dmap(O^1) \wedge O_\sigma^2 = 2Dmap(O^2) \wedge \text{Disjoint}(O^1, O^2) \wedge \text{Dist}(O^1, O^2) < \epsilon_2 \Rightarrow \text{Touches}(O_\sigma^1, O_\sigma^2).$$

Similar assertions could be written when CONTAINS, OVERLAP relationships. In addition, two objects in the real world with a TOUCHES relation can coalesce into a single one.

As a consequence, in reasoning what is true at one scale, can be wrong at another scale. So, any automatic system must be robust enough to deal with this issue (Fig. 11).

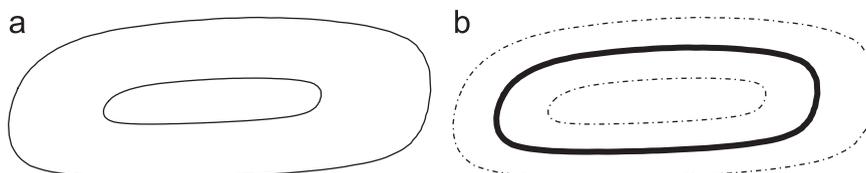


Fig. 10. Crispification of a geographic object modeled by the egg-yolk representation. (a) The original model. (b) Its reduction to a crisp object.

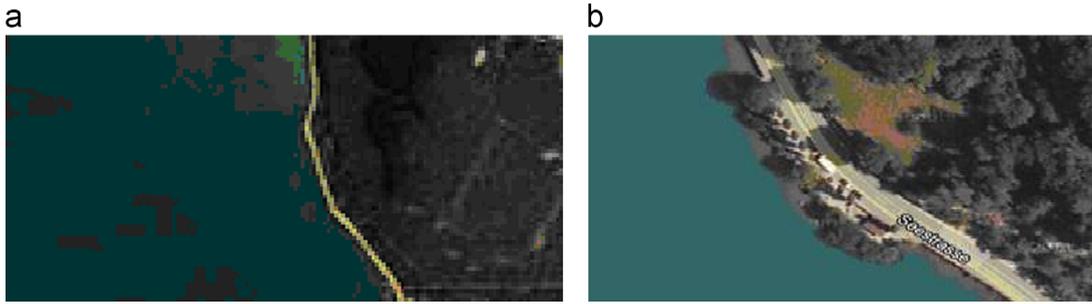


Fig. 11. According to scale, the road TOUCHes or not the lake.

As a consequence, one can write:

$$GKB_7| = \text{Principle7}(GKB_6)$$

Principle #8 (Transformation into graph): “Every set of ribbon or linear objects can be transformed into a graph”. Indeed, reasoning with graphs is often easier than to reason with computational geometry. For instance this kind of transformation can be used for roads, rivers, metrolines, sewerages, etc. Therefore:

$$GKB_8| = \text{Principle8}(GKB_7)$$

Principle #9 (From pictorial to geographic objects): “Any group of pixels having same characteristics located in a satellite image or in an aerial photo can be regrouped into a pictorial object; this pictorial object can be conferred a geographic type possibly using an ontology”. Indeed as soon as a pictorial object is recognized its type will be identified and it can be a part of a geographic object. For instance, a roof texture and adjacent garden texture can reveal a parcel. So one can write:

$$GKB_9| = \text{Principle9}(GKB_8)$$

Principle #10 (Visualization constraints): “The spatial relations between objects must hold after generalization”. In Fig. 12, an excerpt of the French Riviera coast along the Mediterranean Sea is showed from Spain to Italy. Suppose we generalize this coast by a single line between those two counties: the city of Nice will be in the middle of the sea whereas Marseilles and Montpellier will stay in the mainland. In order to enforce the topological constraints, those harbors must move so that the COVER relations hold. Same reasoning is valid for the Rhone River to link it to the sea.

Another example will be the city of Geneva which must be always outside France.

As a consequence of Prolegomena #5 and #6, when better or newer data supersede old data, topological constraints must hold on (Fig. 12).

One of the difficulties of this principle is not to follow the constraints, but to ascertain that all visualization constraints are listed. In other words, how to prove that the list is exhaustive, irredundant and consistent? Here lies a technological barrier. As a consequence

$$GKB_{10}| = \text{Principle10}(GKB_9)$$

Principle #11 (Influence of neighbors): “In geographic repositories, do not forget that objects at the vicinity (outside the jurisdiction) can have an influence”. This is a

consequence of Tobler’s law (Prolegomenon #12); however the great majority of existing GIS do not follow this law. Taking again the example of Geneva, remember that a big part of its metropolitan area is located in the French Rhône-Alpes region; any automatic social-economic reasoning must take this characteristics into consideration. Therefore a kind of out-buffer zone must be defined.

But the question is “where is the limit?” Similarly a threshold can be defined. So, by denoting NKB , the knowledge base of this buffer zone, we get:

$$GKB_{11}| = \text{Principle11}(GKB_{10}, NKB)$$

Principle #12 (Cross-boundary interoperability): “Any geographic repository must provide key-information to ensure cross-boundary interoperability”. Once solved the sliver polygons problem located at the boundary, two cases are important, graph structures and terrains. Fig. 13 illustrates a graph example in which two neighboring geographic repositories are present, obviously with geometric discrepancies [25]. Fig. 13a shows two geographic repositories before integration and 13b the situation with a magnifying glass emphasizing the discrepancies at the boundary; Fig. 13c shows the results of cartographic integration (maps look good; a successive step is not mention in the figure is object integration in which two objects (for instance, a road, a river) which were artificially cut into two pieces, fusion, i.e. same identifier. Then Fig. 13d shows the last step, graph integration: indeed before integration road graphs are not connected, but in order to allow graph reasoning, for instance minimum path algorithm across several repositories, graphs must be connected; in this case a node must be created in which a first edge belongs to the first repository, and the second edge to the second repository.

In other words, before integration there is a set of non-connected ribbons and at the end the concerned ribbons are reduced to a unique ribbon, and so a unique graph is constructed. Therefore it is compulsory to provide necessary tools for both creating cross-boundary edges, and launching graph algorithms without blockage, not only for roads, but for any kind of networks as previously mentioned (water supply, telecommunications, etc.). In addition Rule #2 must be applied when scale is diminishing.

In the case of disconnected terrains, the case is a little bit more complex for two reasons. First elevations can be defined differently essentially because the reference points (mean sea level) are different (for instance 2.34 m between Belgium and The Netherlands). And secondly, the mathematical shape of the geoid can differ. Once those discrepancies

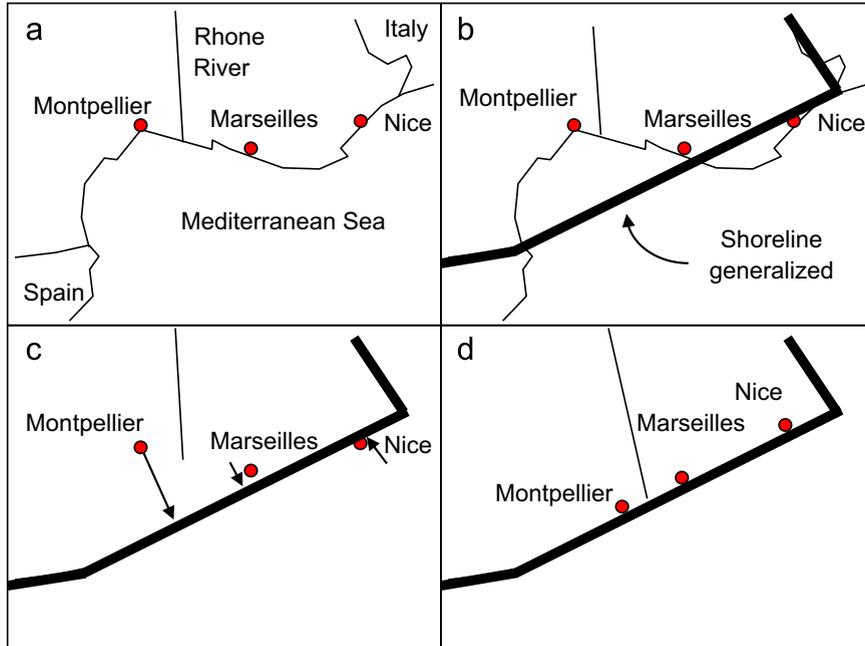


Fig. 12. Visualization constraints. (a) Before generalization; (b) generalization of the coastline, but harbors are badly located; (c) moving harbors and rivers to follow topological constraints.

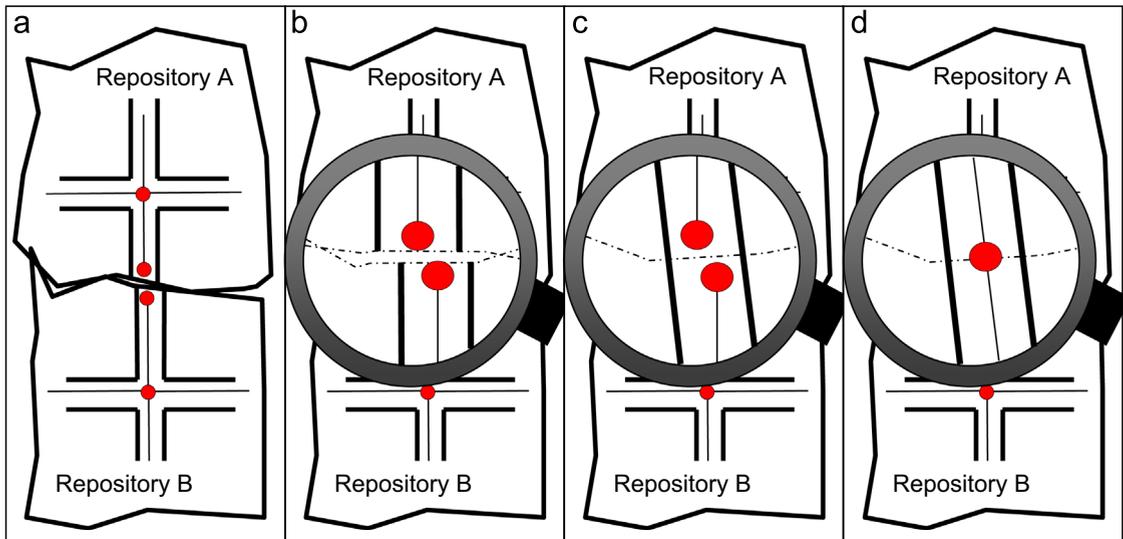


Fig. 13. Consequences of cross-boundary interoperability. (a) Before integration. (b) Cartographic integration. (c) Graph-reasoning integration.

are overcome, the integration of terrains can be launched. In Laurini [25], a solution based on triangles was proposed. See an illustration in Fig. 14.

Do not forget that by applying Principle#11, already a buffer zone is integrated in our geographic knowledge base, and some discrepancies can occur. Once they have been reconciled, as a consequence, by applying this principle on two geographic knowledge bases, namely GKB^a , GKB^b , we can get:

$$GKB_{12} = Principle12(GKB^a, GKB^b)$$

This principle drives to the design of consistent distributed geographic knowledge base systems.

Now that principles are stated, let us revisit geographic knowledge.

7. Feedbacks from the survey

As previously told, a questionnaire was sent with the 12 prolegomena and the 12 principles. Seventeen answers were received, but two of them were directly not usable. The number is low, but it can show a trend of agreement.

In order to simplify the questionnaire, the word “prolegomenon” was not used and no further explanations were given to the experts. Table 2 presents the results

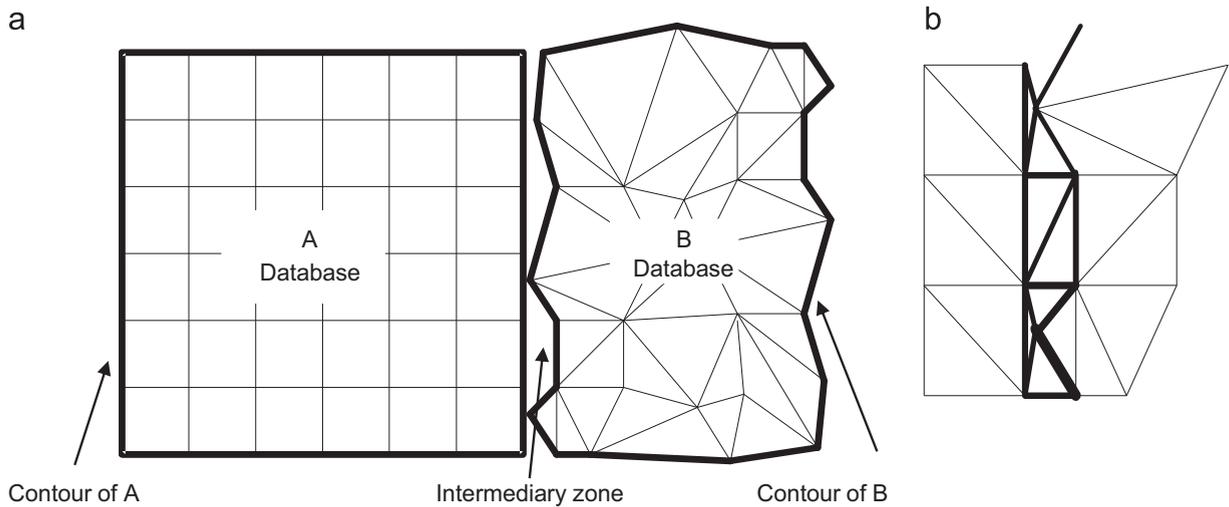


Fig. 14. Examples of terrain continuity. (a) Two adjacent terrain databases with their intermediary zone. (b) A solution to ensure continuity based on additional triangles.

Table 2

Table showing the results of experts' agreement for prolegomena and principles.

Prolegomena	1	2	3	4	5	6	7	8	9	10	11	12
Agreement rate	67	60	93	87	67	77	87	93	80	87	73	87
Principles	1	2	3	4	5	6	7	8	9	10	11	12
Agreement rate	73	87	77	100	83	60	83	70	80	73	93	90

showing a high level of consensus, the minimum being 60% and the average 80%.

The lower values (60%) correspond to the Prolegomenon #2 and Principle #6 which deals with the transformation of fuzzy objects. For the prolegomenon dealing with acquisition of data by measurement, nobody gave reasons for disagreement, likely a problem of misunderstanding; for me, along the Goodchild's motto "Citizens as sensors". As a consequence, I have slightly modified it.

Few experts explained the reasons of their disagreement. Some states that it was a problem of definition.

In the questionnaire, there was an additional open question concerning the completeness: a lot feel that the list is not complete, but none suggests additional considerations, except one who suggests something regarding decision-making.

In a presentation made in a GIS conference in September 2013 in Kuala Lumpur, Malaysia, some participants suggest to integrate tridimensional and temporal considerations.

8. About geographic knowledge categories

In addition to conventional categories such as facts, concepts, processes and rules, geographic knowledge engineering need not only to redefine them but also to include new items. The ChorML project [10] the following, geographic facts, cluster of areas, flows (persons, goods,

etc.), topological constraints and co-location rules. Let us examine all of them.

8.1. Geographic facts

The notion of geographic facts must be revisited. In some case, facts are simple to define. For instance a place must be mentioned either by a place name or identifier or by means of its coordinates; for instance "The Mont Blanc summit is located in North 45°49'59 and East 6°51'53 and its elevation is 4807 m". But when saying that there are 60 million of inhabitants in France, there is no problem. But when one says that France is located at the South of Belgium, it is a little more awkward to encode because some points of Belgium are located at the South of some French places. A solution seems to claim that the majority of Belgium points are located at the North of French points; but from a mathematical point of view, one has to compare infinity of points. Consequently, another mathematical definition must be proposed for instance based on centroids.

8.2. Clusters of area

In some situation, it could be interesting to regroup areas (polygonal zones) into a single cluster according to some criteria. This cluster will constitute a new tessellation perhaps

with disconnected pieces or with holes. Here again, Principle #5 applies when the newly created cluster is discovered as a loose tessellation. For instance:

UK = CLUSTER (England, Scotland, Wales, NorthernIreland)

8.3. Flows

Generally two areas (origin and destination) can be linked by flows of people or goods; flows can be unidirectional, bidirectional; if origin is multiple or unknown, the flow is converging (for instance speaking of immigration) then the destination zone is called sink, elsewhere if destination is multiple or unknown, the flow is diverging (the origin node is a source); flows are defined from areas which can be reduced as points at some scales. See Fig. 15.

8.4. Topological constraints

Some examples were already given in Principle #10; these constraints must be used not only for visualization but also for reasoning and retrieval.

8.5. Co-location rules

One of the scopes of geographic data mining is to look for co-location rules. When two sets of features are concerned; for instance “near a big city, there is an airport”.

9. Rapid state of the art of geographic knowledge representations

Presently, geographic knowledge can be represented by four different methods, natural language (but which is outside the scope of this paper), predicate logics, XML-encoding and visual representation by chorems. Let us examine them.

9.1. Natural language

Historically speaking, the objective of conventional geography was to exhibit geographic knowledge with natural language; but the main drawback is that this mode of representation is not very machine-treatable. Let us take an example “when there is a lake and a road going to the lake, then there is a restaurant”.

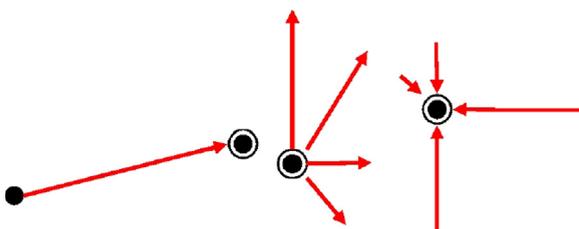


Fig. 15. Examples of flows, unidirectional, diverging, converging.

9.2. Predicate logics

Geographic knowledge can be expressed by propositional logic under the condition to include spatial relations and spatial operators. The previous statement can be encoded:

$$\begin{aligned} &\forall l \in Lake \wedge \forall s \in Street \wedge (touches(l, s) \\ &\Rightarrow \exists r \in Restaurants \wedge (distance(r, l) \\ &< 100 \wedge (distance(r, s) < 100. \end{aligned}$$

9.3. XML-encoding

XML can be the basis of geographic knowledge: for instance SpatialML (Mani et al. [28]) is a markup language for representing spatial expressions in natural language documents; its goal is to allow for better integration of text collections with resources such as databases that provide spatial information about a domain.

9.4. Chorems

Another possible track is a visual language since geography and cartography are essentially visual. A solution could be found with chorems (Brunet [7], DeFatto et al. [13]) which are schematic representations of territories. An example is given Fig. 16 representing the mobility of population in the United States [35]. In this chorematic map, the USA conterminous states are simplified into a rectangle and only major cities are illustrated. External or internal migrations are shown by arrows Fig. 17.

Chorems are essentially used to schematize information about a territory. They can be an interesting prototype to represent visually geographic knowledge.

10. Towards visual representation of geographic knowledge

To design a visual language, one has to define the vocabulary and the grammar of the visual language in order to define statements (=knowledge) and define interrogations (queries), and the context of interpretation. Among the generic characteristics, a visual language must be universal, i.e. everybody must understand it; in other words icons must come from a fully-agreed visual ontology, and object icons must be known by anyone. In addition its expressive power must be as large as possible.

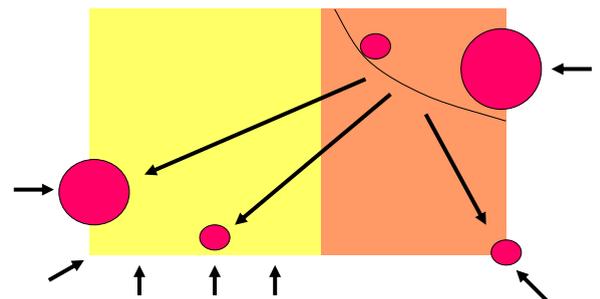


Fig. 16. Chorem showing population mobility in the United States.

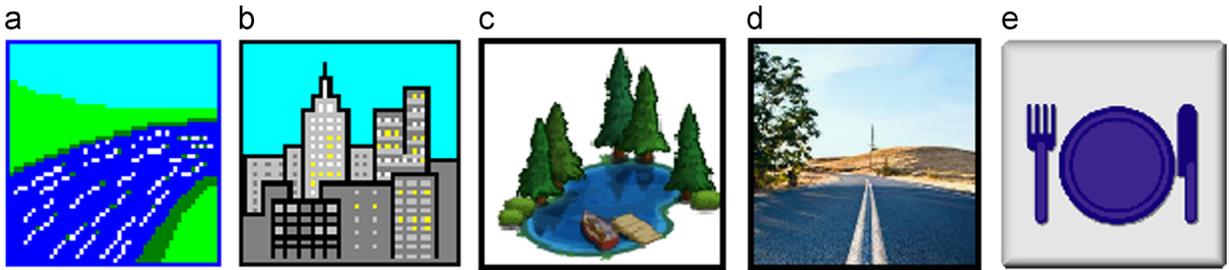


Fig. 17. Examples of geographic object type icons. (a) River; (b) City, (c) Lake, (d) Route, (e) Restaurant.

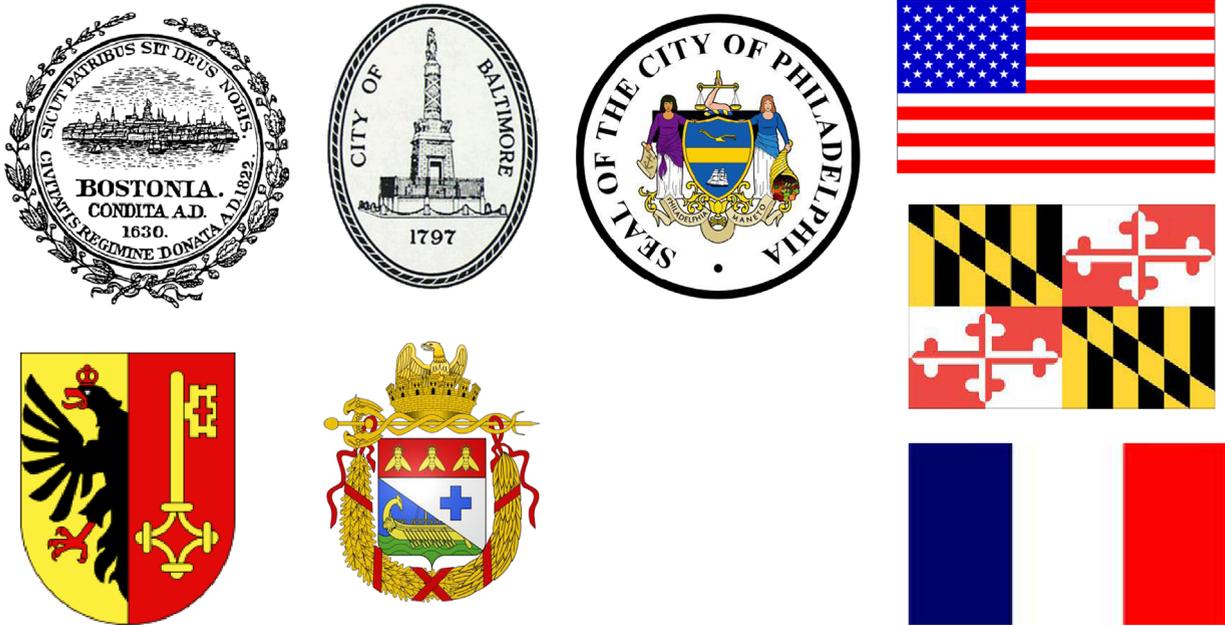


Fig. 18. Examples of geographic object icons issued from a gazetteer, seals of cities of Boston, Baltimore and Philadelphia, then flags for the USA and Maryland State.

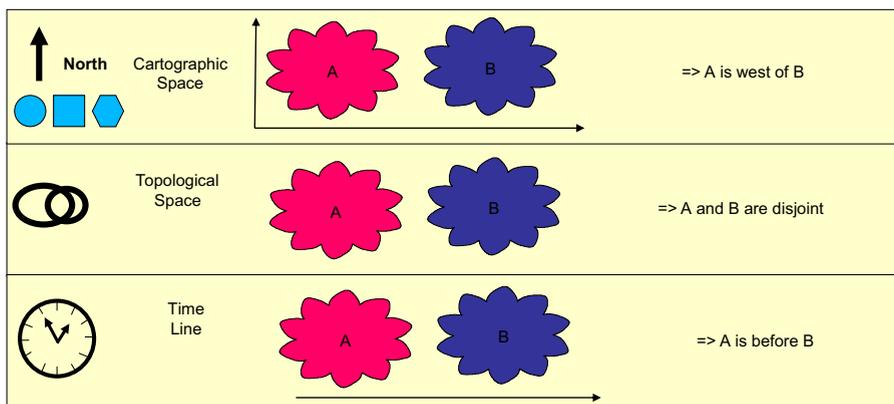


Fig. 19. The contexts of interpretation, cartographic space, topological space and time line.

10.1. Vocabulary

Not only object types must be representing by an icon but also object classes themselves. As type icons (Bonhomme et al. [5] (Fig. 17) must be defined in the ontology, object

icons (f.i. a city, a river) must be defined in the gazetteer (Fig. 18). Some cities have emblems or seals which can be used and countries have flags. As Mississippi State has a flag and a seal, Mississippi river has no official symbol (as far as we know). The consequence is that if somebody creates an

icon for the Mississippi river, a lot of people will have difficulties understanding either the statement or the queries integrating it.

The main difficulty to define geographic object icons is that in cartographic legends, they have various representations. For instance, restaurants and museums have dozens of visual representations.

Concerning mathematical vocabulary, let us propose to continue using the usual symbols.

10.2. Sentences

Three types of sentences must be defined, statements, constraints, rules and queries. Statements will correspond to geographic facts, clusters and flows. Rules must be defined by using \Rightarrow symbol. And queries by using Spanish interrogation marks (ζ ?) which can be used as parentheses.

10.3. Contexts of interpretation

In fact, three types of interpretation spaces are possible, identified by three interpretation icons (Fig. 19),

- Cartographic space which corresponds to conventional cartography with an arrow to North and a scale; the horizontal axis represents eastings; this context is identified by the North arrow icon; according to scale, it can be based on projections or being spherical; It has three alternatives, the planar one (iconized by a square), the global one (circle) and the chorem represented by a hexagon (indeed, as example France is often schematized as a hexagon);
- Topological space in which only cardinal directions have no importance, but the importance is given to the respective positioning of geographic objects; the horizontal and vertical axes have no meaning; this context is identified by an “overlap relation”;

- Time line in which the horizontal axis represents time; this context is identified by a clock icon; remind that this interpretation context is outside the scope of this paper.

10.4. Examples

Fig. 20 is taken in the cartographic context. Fig. 20a illustrates a fact that the city of Baltimore is located south of Boston, whereas Fig. 20b represents a query in order to know whether Baltimore is located south of Boston.

Fig. 21 illustrates a more complex topological query in order to get rivers crossing the State of Maryland. Since the scale is mentioned, it means that one is only interested by rivers wider than 100 m. Fig. 22 represents a query in the chorem space: what is the migration flow between two clusters, one for the provinces of the Northern part of Italy, and the cluster of Sicilian provinces. Fig. 23 represents

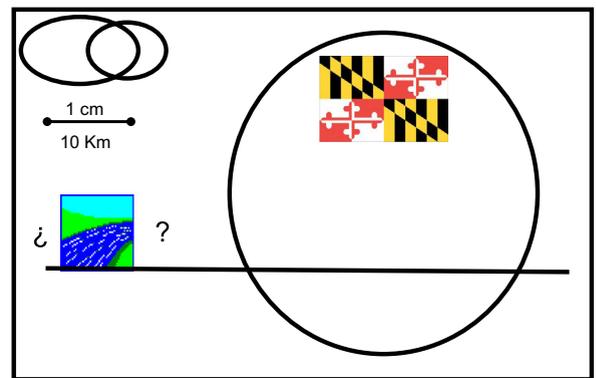


Fig. 21. A topological query to get the list of rivers crossing the State of Maryland and wider than 100 m.

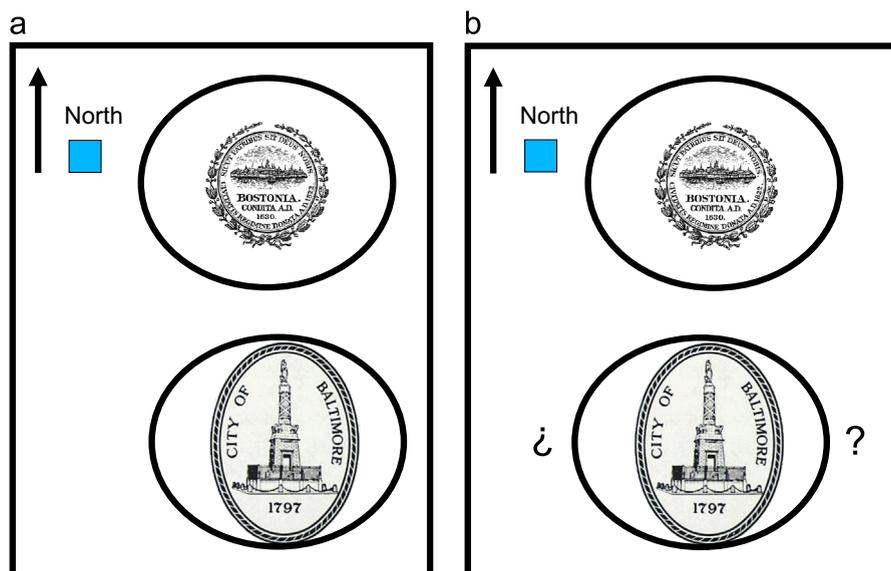


Fig. 20. A fact and a query in a geometric context. (a) representing the fact that Baltimore is south of Boston. (b) representing a query in order to know whether Baltimore is located south of Boston.

a topological constraint concerning the cities of Marseilles and Geneva with France. Finally, Fig. 24 represents the co-location rule (Lake, Road) → (Lake, Road, Restaurant).

11. Conclusion

The ultimate goal of this work is to develop a conceptual framework for geographic knowledge management

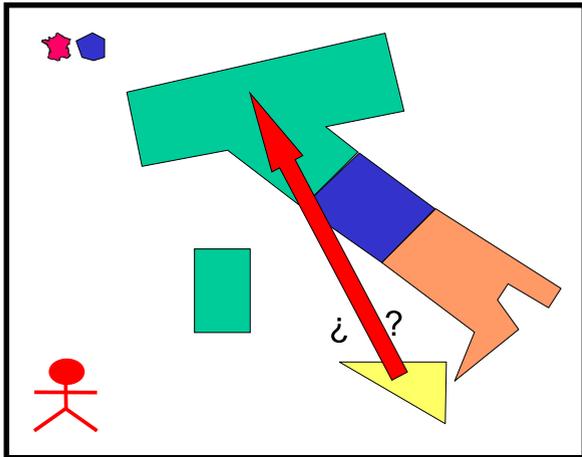


Fig. 22. Example of visual representation of a query in the chorem space: what is the migration flow between a cluster of Northern regions and the cluster of Sicily provinces in Italy?

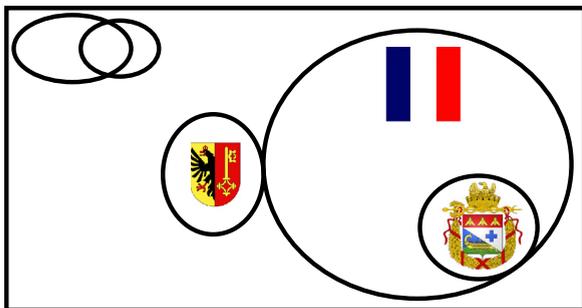


Fig. 23. Example of visual representation of constraint stating that the city of Marseilles must be always inside (TOUCHES relation) the French territory, and Geneva outside (\sim COVERS).

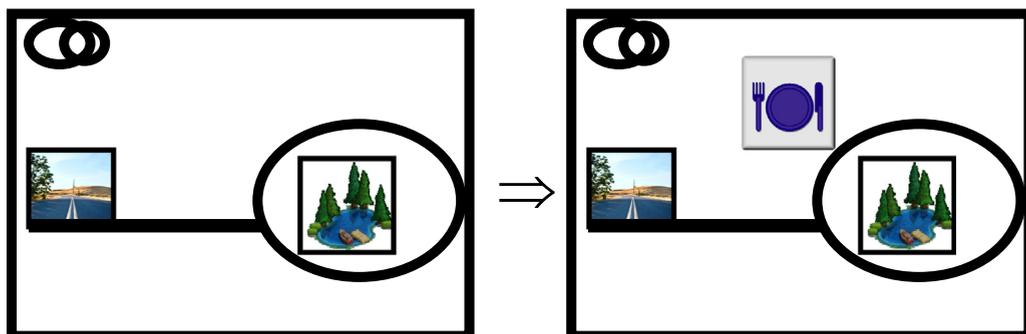


Fig. 24. Example of visual representation of co-location rule “Lake, Road and Restaurant” in a topological space.

essentially because the conventional framework in artificial intelligence is not totally adequate for novel applications in territorial intelligence. In presenting key concepts, emphasis was given to the generalization of spatial relations that constitute the basement for the foundations of geographic knowledge modeling and engineering in connection with scale. For that purpose, the ribbon theory was proposed as a kind of intermediary between a longish area and a line. With this concept, generalized topological relations can be defined in order to elegantly represent the transformation of spatial relationship between two areal objects in connection with scale. Towards this objective, the process of generalization-reduction-disappearance was the key to govern the generalization of topological relations.

Also was presented the concept of granularity of interest as an extension of visual acuity, concept which is too much bound to cartography. With this concept, artificial intelligence applied to geographic reasoning can be revisited.

But before presenting the principles as a sound basis for the conceptual framework, some preliminary considerations (prolegomena) were necessary to set the foundations of this framework. Indeed, any conceptual framework is based on principles which are overall guidelines and rules; moreover, those principles allow at making predictions and drawing implications and are finally the basic building blocks of theoretical models.

In order to present a final reference framework, several conditions must be met:

- to be comprehensive; maybe some key-issues are still missing and must be integrated,
- to be consistent, i.e. no contradictions must exist between principles,
- to be robust, i.e. by taking errors into account, because actual geographic databases are error-prone implying sometimes geometric inconsistencies,
- and to be minimal, i.e. the necessity to avoid direct or indirect redundancies.

In addition, this framework can be extended by including tridimensional and temporal considerations.

In the last part of this paper, languages for representing geographic knowledge were presented and emphasis was given on visual languages based on chorems. Taking

interpretation context into account, several examples were shown for presenting or querying facts, constraints and co-location rules.

Among the perspective, we can mention the necessity to assess this framework and to extend the visual language to increase its expressive power. Let us also mention the necessity to design an exhaustive visual gazetteer and to claim for adding class icons in ontologies.

Finally, those principles can represent the key-element of the design of geographic knowledge systems which will be the basis of information technology for territorial intelligence.

My acknowledgements are due to all people who have helped me in this research, my former students and my colleagues in research in various countries, especially Dr. Milleret-Raffort, Dr. Sol-Martinez, and Dr. Boulmakoul. I want also to thank the anonymous referee who stimulates me for the validation of the framework and the experts who give their opinion to help me construct this framework.

References

- [1] C. Aldridge, Geographic Knowledge: Bringing Geographers to Their Sensors, in: Proceedings of Thirteenth Annual Colloquium of the Spatial Information Research Centre. P.A. Whigham (Ed.), 2–5 Dec, Dunedin, New Zealand. University of Otago, ISBN 1877139416. 2001, pp. 195–210.
- [2] J.F. Allen, Maintaining knowledge about temporal intervals, Communications of the ACM. 26 November 1983, ACM Press, 1983, 832–843 (ISSN 0001-0782).
- [3] H. Alt, M. Godau, Computing the Fréchet distance between two polygonal curves, Int. J. Comput. Geom. Appl. 05 (75) (1995) 75–91.
- [4] M. Bertolotto, L. De Floriani, P. Marzano, A unifying framework for multilevel description of spatial data, in: Proceedings COSIT95 (Spatial Information Theory—A Theoretical Basis for GIS), Lecture Notes in Computer Science, N. 988, A.U. Frank, W. Kuhn (Editors), Springer-Verlag, 1995.
- [5] C. Bonhomme, C. Trepied, M.A. Aufaure, R. Laurini, A visual language for querying spatio-temporal databases, in: Proceedings of the Seventh International Symposium on GIS, ACMGIS'99, Kansas City, Nov. 5–6, 1999, edited by C. Bauzer-Medeiros, ACM-Press, 1999, pp. 34–39.
- [6] R. Brachman, H. Levesque, Knowledge Representation and Reasoning The Morgan Kaufmann Series in Artificial Intelligence, San Francisco, 2004, 340 pages.
- [7] R. Brunet, La Composition des modèles dans l'analyse spatiale", in L'Espace géographique, no. 4, 1980. 1980.
- [8] B. Buttenfield, R. McMaster, Map Generalization: Making Rules for Knowledge Representation, Longman, London, 1991, 1991.
- [9] A. Cohn, N. Gotts, The "Egg-Yolk" representation of regions with indeterminate boundaries, in: P. Burrough, A. Frank (Eds.), Geographic Objects with Indeterminate Boundaries, Taylor & Francis, 1996, pp. 171–187.
- [10] A. Coimbra, ChorML: XML Extension for Modeling Visual Summaries of Geographic Databases Based on Choresms, Master Dissertation, INSA-Lyon. 2008.
- [11] H. Couclelis, People manipulate objects (but cultivate fields): beyond the raster-vector debate in GIS, in: A. U. Frank, I. Campari, U. Formentini (Eds.), Theories and methods of spatio-temporal reasoning in geographic space, Berlin: Springer Verlag, (1992) 65–77.
- [12] P. Crowther, A visual geographic knowledge classification and its relationship to the KADS model, in: Norman Fo (Ed.), Advanced Topics in Artificial Intelligence, vol. 1747, Lecture Notes in Computer Science, 1999, pp. 478–479.
- [13] V. Del Fatto, R. Laurini, K. Lopez, R. Loreto, F. Milleret-Raffort, M. Sebillo, D. Sol-Martinez, G. Vitiello, Potentialities of choresms as visual summaries of spatial databases contents, VISUAL 2007, in: Ninth International Conference on Visual Information Systems, Shanghai, China, 28–29 June 2007. (Eds.), Qiu, G., Leung, C., Xue, X.-Y., Laurini, R., Springer Verlag LNCS, vol. 4781 Advances in Visual Information Systems, 2007, pp. 537–548.
- [14] D. Douglas, T. Peucker, Algorithms for the reduction of the number of points required to represent a digitized line or its caricature, Can. Cartographer 10 (2) (1973) 112–122.
- [15] M. Egenhofer, Deriving the composition of binary topological relations, J. Visual Lang. Comput. 5 (2) (1994) 133–149.
- [16] M. Ester, H.-P. Kriegel, J. Sander, Spatial data mining: a database approach, In: Proceedings of the Fifth International Symposium on Large Spatial Databases (SSD '97), Berlin, Germany, Lecture Notes in Computer Science Vol. 1262, Springer, 1997, pp 47–66, 1997. 1997.
- [17] C. Freksa, T. Barkowsky, On the relation between spatial concepts and geographic objects, in: P. Burrough, A. Frank (Eds.), Geographic Objects with Undetermined Boundaries, Taylor & Francis, London, 1995, p. 1995.
- [18] R.G. Golledge, The nature of geographic knowledge, Ann. Assoc. Am. Geographers 92 (Issue 1) (2002) 1–14. (March 2002).
- [19] M.F. Goodchild, Citizens as sensors: the world of volunteered geography. GeoJournal 69(4): 211–221. Reprinted in M. Dodge, R. Kitchin, and C. Perkins, editors, The Map Reader: Theories of Mapping Practice and Cartographic Representation, pp. 370–378. Hoboken, NJ: Wiley, 2007.
- [20] M.F. Goodchild, M. Yuan, T.J. Cova, Towards a general theory of geographic representation in GIS, Int. J. Geog. Inf. Sci. 21 (No. 3) (2007) 239–260. (March 2007).
- [21] S. Gordillo, Modélisation et Manipulation de phénomènes Continus Spatio-temporels. Ph.D. Dissertation, October 12, 2001. Université Claude Bernard Lyon I.
- [22] M.A. Kang, C. Faloutsos R. Laurini, S. Servigne, Indexing values in continuous field databases. in: Proceedings of the International Conference on Extending Database Technology (EDBT 2002), March 2002, Prague. ISBN:3-540-4332. 2002.
- [23] A. Karalopoulos, M. Kokla, M. Kavouras, Comparing representations of geographic knowledge expressed as conceptual graphs, in: (Eds.), in: A. Rodríguez, I. Cruz, S. Levashkin, M.J. Egenhofer (Eds.), GeoSpatial Semantics, vol. 3799, Lecture Notes in Computer Science, 2005, pp. 1–14in: (Eds.), in: A. Rodríguez, I. Cruz, S. Levashkin, M.J. Egenhofer (Eds.), GeoSpatial Semantics, vol. 3799, Lecture Notes in Computer Science, 2005, pp. 1–14.
- [24] T.J. Kim, L. Wiggins, J.R. Wright (Eds.), Expert Systems, Applications to Urban Planning", Springer Verlag, 1990.
- [25] R. Laurini, Spatial multidatabase topological continuity and indexing: a step towards seamless GIS data interoperability, Int. J. Geog. Inf. Sci. 12 (1998) 373–402. (4, June 1998).
- [26] R. Laurini, Importance of spatial relationships for geographic ontologies, in: Planning Support Tools: Policy Analysis, Implementation and Evaluation Proceedings of the Seventh International Conference on Informatics and Urban and Regional Planning INPUT 2012 ((Eds.): M. Campagna, A. De Montis, F. Isola, S. Lai, C. Pira, C. Zoppi) (ISBN code: 9788856875973), 2012, pp. 122–134.
- [27] R. Laurini, F. Milleret-Raffort, L'ingénierie des connaissances spatiales. Paris, Hermès, Collection Technologies de pointe, no. 30, 64 p. 1989.
- [28] I. Mani, C. Doran, D. Harris, J. Hitzeman, R. Quimby, J. Richer, B. Wellner, S. Mardis, S. Clancy, SpatialML: Annotation Scheme, Resources and Evaluation. The MITRE Corporation 202 Burlington Road Bedford, MA 01730, USA. (http://www.mitre.org/work/tech_papers/tech_papers_09/09_3827/09_3827.pdf). 2010.
- [29] J. Mennis, D.J. Peuquet, The role of knowledge representation in geographic knowledge discovery: a case study, Trans. GIS 7 (3) (2003) 371–391.
- [30] K. Popper, The Logic of Scientific Discovery, (Logik der Forschung, English trans. 1959), ISBN 0-415-27844-9. 1934.
- [31] D.A. Randell, C. Zhan, A.G. Cohn, A spatial logic based on regions and connection, in: Proceedings of Third International Conference on Knowledge Representation and Reasoning, Cambridge, MA (Morgan Kaufmann), 1992, pp. 165–176.
- [32] J.F. Sowa, Knowledge Representation: Logical, Philosophical, and Computational Foundations, Brooks Cole Publishing Co., Pacific Grove, CA, 594 pages; ISBN 0-534-94965-7. 2000.
- [33] J. Teller, J. Lee, C. Roussey, Ontologies for urban development, Studies in Computational Intelligence, Springer Verlag, 2006, 61 (212 pages).
- [34] W. Tobler, A computer movie simulating urban growth in the Detroit region, Econ. Geogr. 46 (2) (1970) 234–240.
- [35] USAmaps, WebSite (http://lettres-histoire.info/lhg/geo/geo_usa/carte_s_usa_geo.htm); Accessed October 2013.
- [36] A. Vckovski, Representation of continuous fields, in: Proceedings of 12th AutoCarto, vol. 4, 1995, pp. 127–136, 1995.
- [37] L. Zadeh, Fuzzy sets, Inf. Control 8 (1995) 338–353.