Fundamentals of Geographic Knowledge Engineering for Territorial Intelligence[1]

Robert Laurini
LIRIS, INSA-Lyon, University of Lyon, France

Abstract

In a lot of applications from agriculture to zoology, from environmental planning to territorial intelligence, actual systems of artificial intelligence are not very efficient, essentially because of a naïve representation of space.

As spatial knowledge corresponds to conventional geometric and topological knowledge, geographic knowledge corresponds to knowledge about geographic features in the real world even if real features can have spatial relationships between them. In other words, spatial knowledge is based on topological, projective and distance relations; but if applied to geographic features, one must take earth rotundity and other characteristics (demography, physical geographic, economic geography, etc.) into account.

After a rapid presentation of spatial relations and their properties, this chapter will detail the 12 principles governing geographic knowledge. First emphasis will be given to various forms of geography knowledge, such as located facts, geographic clusters, flows, gradients, co-location rules and topological constraints.

Then, based on ribbon theory, spatial relations and earth rotundity, geographic relations will be defined. For instance, let us consider two features in the real world associated with a DISJOINT relation; when down-scaling, those objects can be associated with a TOUCHES relation. As a consequence, any reasoning mechanism must be transformed accordingly.

1. Introduction

Due to the difficulties of modeling geographic analysis and reasoning, knowledge engineering must be revisited in order to integrate space and computational geometry. Spatial knowledge must be seen as an extension of geometry but also must integrate topology (Laurini-Milleret-Raffort, 1989), whereas geographic knowledge can be defined as spatial knowledge applied to the earth. In other words, the main difference between spatial and geographic knowledge deals with semantics of all the features which are on the Earth.

Territorial intelligence can be defined as business intelligence applied to any territory ranging from a parcel to a country and the whole Earth. As it is common to claim that planar geometry can be applied to parcels, for larger features the rotundity of the Earth must be taken into account. Therefore, the concept of spatial granularity of interest is very important. Indeed, for a city-level politician and a nationwide politician, the interests are different. Over the centuries, this problem was solved by considering scales. Beyond the strict ratio between the map measures and the Earth measures, a lot of semantics were hidden: for instance at the scale of 1:1,000,000 usually building are not visualized because they representations were too small: the limit of 0.1 mm was generally used according to the following rule "*if the mapped size of an object is less than 0.1 mm, then that object is removed*".

What is the origin of geographic knowledge? For centuries, geographers and urban planners have constructed skills and methodologies to analyze, to assess a territory in order to propose solutions. Sometimes this knowledge can be written in reports and some of them can be accessed by Internet. By text mining some chunks can be extracted. But due to the specific characteristics of this knowledge, for instance the problem of recognizing and extracting placenames, special algorithms are procedures have been produced. And then one can run process of geographic information retrieval (GIR).

The objective of this chapter is to present the fundamentals of geographic knowledge engineering. In the next section, an analysis of geographic features will be presented, emphasizing a new model based on ribbons. Then a difference will be made between spatial relations and geographic relations. In the fifth section, the twelve principles of geographic knowledge engineering will be discussed proceeded by twelve prolegomena. Then some

---

[1] This is an excerpt with modifications, of the book "Knowlege Engineering Principles, Methods and Applications" Edited by A. Perez Gama, Nova Science Publishing, New-York, 2015. ISBN: 978-1-63463-909-5, pp. 1-56.

various ways of encoding geographic knowledge will be introduced. To conclude this section, application in territorial intelligence will be rapidly sketched.

2. Geographic objects

The objective of this section is to revisit the modeling of geographic features for knowledge engineering. A solution will be based only on two types of areas, crisp and fuzzy, but a new sort of areas called ribbon will be used to model what is common to call "linear objects".

2.1. Representation of geographic features

When modeling a geographic feature, a very important aspect, deals with mathematical representations usually taken as attributes. For years, several models for instance for storing a simple polygon exist, but standardization has opted for one of them (OGC). For a street, several models exist depending on actor's vision. Fig. 1 illustrates four families of models:

- the first model is based on graph with edges as street axes and nodes as crossroads;
- the second based on two polylines delimitating the private part and the public part (cadaster meaning);
- the third as an areal model for describing the section reserved for traffic;
- and finally a 3D model in order to integrate engineering networks.

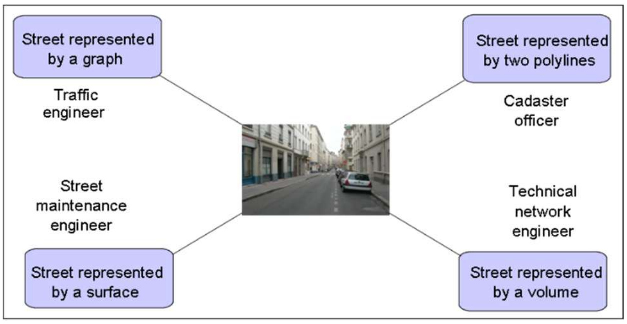

Figure 1. Multiple geometric representations of a street.

Do not forget that feature shapes are always simplified overall at smaller scales: in order to increase readability, lines are generalized i.e. some points are removed thanks for instance to the well-known algorithm designed by Douglas-Peucker (1973). In addition, depending on the context, sometimes some cartographic objects must be enlarged or slightly moved.

For decades, geometric models of geographic features have been based on points (0D), lines (1D), areas (2D) and volumes (3D). But points and lines do not exist really in the nature, since all objects are 3D and moving. Geography for its part is mainly 2D. In the majority of GIS (geographic information systems) software products, rivers and roads are modeled as lines, sometimes with a width, which is strange from a mathematical point of view. To solve this problem, the concept of ribbon will be introduced.

2.2. From lines to ribbons

Since lines do not exist in the real world, except perhaps lines such as the Equator, the meridians and the parallels, in a recent paper (Laurini, 2012), I have proposed to use ribbons to model what it is common to call linear objects such as roads and rivers.

Due to curves (circle portions, clothoids), roads are not rectilinear. So, the idea of modeling lanes by rectangles is insufficient. In order to deal with this important characteristic, a more general definition is needed. From a mathematical point of view, a ribbon can be defined as a transformation of a longish rectangle. Figure 1 illustrates this principle. Let's call R a ribbon and ρ a rectangle of length l and width w, one can state $R = H(\rho)$, in which H is a taenic[2] transformation.

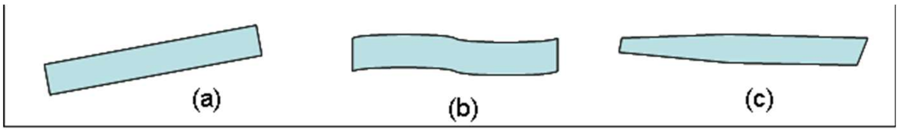

Figure 2. Various types of ribbons. (a) Rectangular ribbon. (b) Ribbon. (c) Loose ribbon.

2.2.1. Rectangular ribbons

A rectangular ribbon R is a longish rectangle. Let us call ends the two smaller extremities and sides the larger ones. The width is noted w, the length l and r_l ($r_l = l/w$) the longishness ratio which is supposed to be much greater than a positive value r_L; a possible minimum value is 5 ($r_l > r_L > 5$).

Let us call skeleton the medium line between two ends located at a distance $h = w/2$ from the sides of the rectangle. Let us note *Skel(R)*, *End1(R), End2(R), Side1(R), Side2(R),* respectively, the skeleton of R, its two ends and its two sides.

2.2.2. Ribbons and taenic transformation

A taenic transformation (Figure 2) transforms a 1D curve into a 2D area with the following properties. Let us note $y=f(x)$ a curve which is supposed to be continuous and differentiable. In each point of the curve, let us consider two points U_1 and U_2 at h distance from the curve located at the perpendicular of the first derivative. Let's call them corresponding points. The loci of those corresponding points form two curves C_1 and C_2 respectively. Knowing that $dy/dx = tg(\alpha)$, we can write respectively:

So for C_1: $\begin{Bmatrix} x_1 = x + h \times \sin(\alpha) \\ y_1 = y - h \times \cos(\alpha) \end{Bmatrix}$ and for C_2: $\begin{Bmatrix} x_2 = x - h \times \sin(\alpha) \\ y_2 = y + h \times \cos(\alpha) \end{Bmatrix}$.

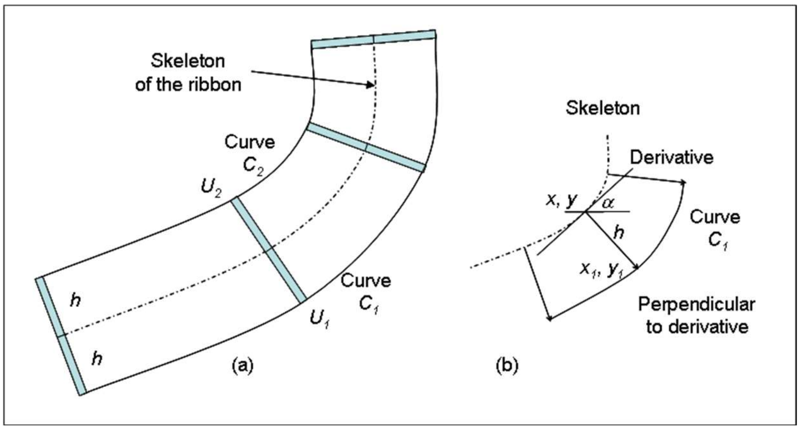

Figure 3. Construction of a regular ribbon with a taenic transformation. (a) Regular ribbon. (b) Details of the construction of a regular ribbon.

Consider the derivative at corresponding point U_1, one has:

---

[2] From Ταινία, ancient Greek for ribbon.

$$\frac{dy_1}{dx_1} = \frac{dy_1}{d\alpha} \times \frac{d\alpha}{dx_1} \quad \text{but} \quad \frac{dy_1}{d\alpha} = -h \times (-\sin(\alpha)) = h \times \sin(\alpha) \text{ and } \frac{dx_1}{d\alpha} = h \times \cos(\alpha).$$

So $\frac{dy_1}{dx_1} = \frac{dy_1}{d\alpha} \times \frac{d\alpha}{dx_1} = \frac{h \times \sin(\alpha)}{h \times \cos(\alpha)} = tg(\alpha) = \frac{dy}{dx}$.

A similar computation can be done for U_2 giving $\frac{dy_2}{dx_2} = \frac{dy}{dx}$. So, the consequence is that the derivatives of the corresponding points are equal to the derivative at the skeleton and the angle is the same.

Now, consider a ribbon end. We can write the following cross product between the tangent in U_1 and the end:

$$CP = (h \times \sin(\alpha) \times \Delta x_1 + (-h \times \cos(\alpha) \times \Delta y_1)) = \frac{h}{\cos(\alpha) \times \Delta x_1}(tg(\alpha) - \frac{\Delta y_1}{\Delta x_1})$$

So, if $\Delta x_1 \rightarrow 0$, we can see that the cross product tends also towards zero which means that the angle is right. As a conclusion, both ends are orthogonal to the sides.

2.2.3. Loose ribbons

Starting from any polygon *P*, what are the conditions to consider it as a loose ribbon? Indeed often due to measurement errors (f.i. roads) and other reasons (f.i. rivers), ribbons are not always perfect rectangles. For solving this problem, let us consider its equivalent rectangle.

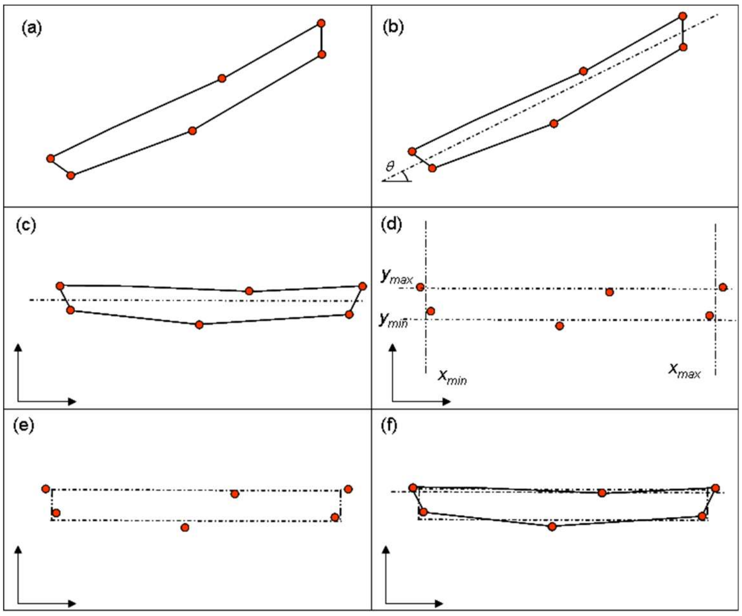

Figure 4. Loose ribbon and its equivalent rectangle. (a) Loose ribbon. (b) The regression line. (c) Rotation. (d) Determination of the two minima and the two maxima. (e) Equivalent rectangle. (f) Loose ribbon and its equivalent ribbon before the back rotation.

The first step is to consider all vertices of *P* (Fig 3a) and by the least square method to compute the regression line $y=mx+q$ (Fig3b). Let us define θ the angle so that $tg(\theta)=m$. Then we make a rotation of $-\theta$ so that the regression line is parallel to the x-axis (Fig 3c). Then, we sort all vertices according to the ascending values of respectively x and y coordinates. We determine minimum and maximum according to those orders. Along x, the mid values of the two first and the two last will determine ribbon ends; and along y, the mid values of the two first and the two last will determine sides (Fig 3d); those values will determine the equivalent rectangle of P noted *ER(P)*.

Now, let us compare the areas of P and *ER(P)*. Generally speaking, there is a small discrepancy between those values. A solution is to lightly modify l and w to reach the exact value. Let us note $A_1=Area(P)$ and $A_2=Area(ER(P))$. Generally speaking, they are not equal. To get them equal:

$$A_2 = (w+\Delta w)\times(l+\Delta l) = w\times l\times(1+\frac{\Delta w}{w})\times(1+\frac{\Delta l}{l}) = A_1\times(1+\frac{\Delta w}{w})\times(1+\frac{\Delta l}{l})$$

Let us suppose that we want to impose modification in the same proportion. So, we can write $\frac{\Delta w}{w} = \frac{\Delta l}{l} = t$ the modification ratio, so giving $A_2 = A_1\times(1+t)^2$. Its value is $t = \sqrt{\frac{A_2}{A_1}} - 1$, or $t = \frac{\sqrt{A_2}-\sqrt{A_1}}{\sqrt{A_1}}$. With this modification, both areas will be equal.

If the longishness ratio r_l of the equivalent rectangle is greater than the threshold value r_L then the polygon P is considered as a loose ribbon R =*ERR(P)*.

In order to simplify this presentation, one will only consider ribbons as rectangles, i.e. the H transformation becomes the identity transformation $R = H(\rho)$ so $R = H(\rho) = \rho$.

In Figure 5, an example of road modeled by ribbons is presented in which one can distinguish several ribbons, namely for bus lanes, bike lanes, median and so on. Immediately it can be seen that ribbons can have their own relations between them.

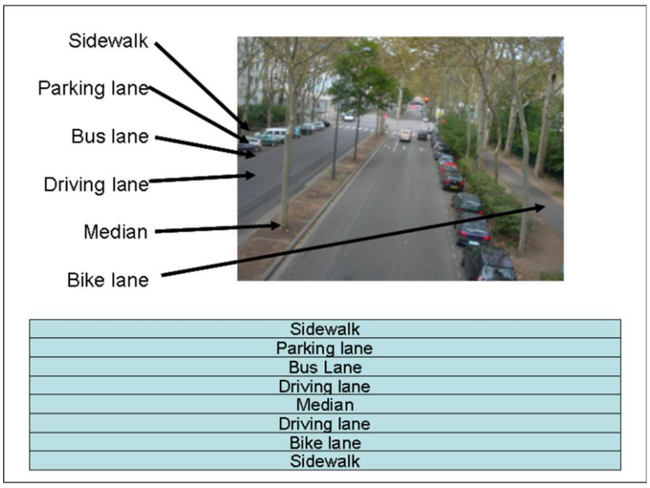

Figure 5. An urban road divided into several ribbons.

2.3. Fuzzy geographic features

Usually, two categories of features can be distinguished, crisp and fuzzy. Crisp objects must have well-defined boundaries such as administrative objects (countries, regions, provinces, natural parks, parcels, etc.) and manmade objects such as streets, buildings.

Other objects, for instance some natural features can be defined as crisp objects; but there are difficulties. A river at some scales can be defined as a line whereas sometimes expressions such as minor bed or major bed are used. Even some dry rivers can be without water. For seas, according to the tide levels, geometric shapes can be different. One of the more salient examples is “Mont Saint-Michel” which is roughly only 1 km^2 wide at high tide and several squared kilometer wide at low tide.

Where does a hill begin, what is the upper limit of a valley, where does a marsh begin? Those are common questions in which features can be modeled as fuzzy geographic sets.

For those objects, fuzzy sets can be used in which some membership grades can be defined (Figure 6) (Zadeh, 1995). An interesting model (Cohn-Gotts, 1996) is the egg-yolk model with two parts, the core (the yellow part) and the extension, the white part of the egg. For instance, for a river, the “yolk” represents the minor bed whereas the “egg” modeled its major bed. Another example is given in Figure 7 in which the mangrove and the jungle are modeled with the egg-yolk representation.

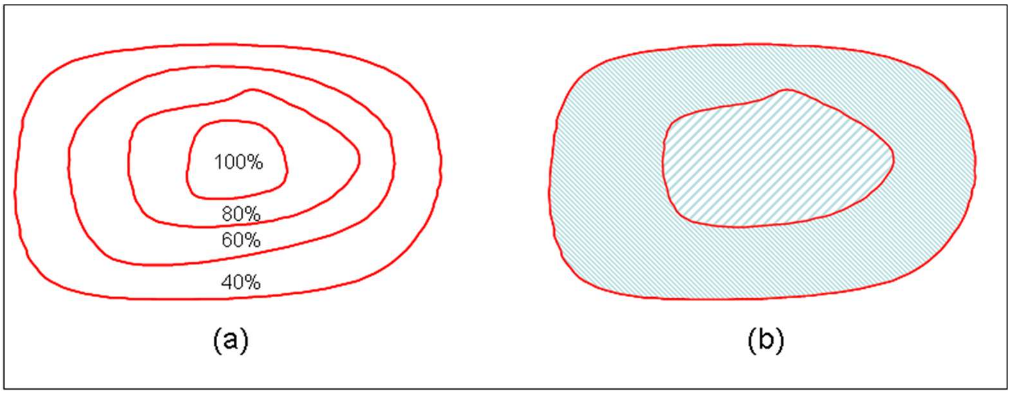

Figure 6. **Fuzzy geographic object. (a) Different membership grades. (b) The egg-yolk representation.**

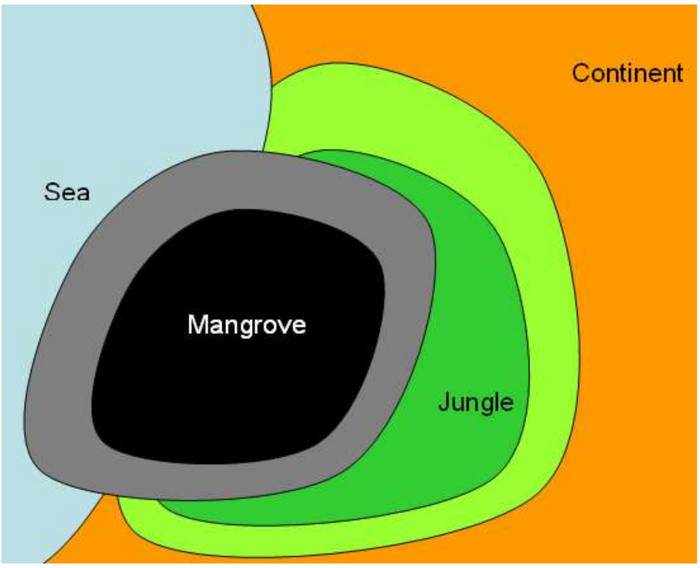

Figure 7. **Fuzzy geographic features.**

3. Spatial and geographic relations

Spatial relations describe relations concerning mathematical objects in space (Laurini-Milleret-Raffort, 1989). But before examining these relations, let us detail some aspects of spatial operations.

3.1. Spatial operations

There are a lot of spatial or geometric operations which can be used in geoprocessing. Only a few will be examined in this text, i.e., minimum bounding rectangles, centroids and buffer zones. Remember that some spatial operations include sophisticated computational geometry algorithms. For more details, refer to (Preparata-Shamos, 1985).

3.1.1. Minimum bounding rectangle

For several more sophisticated operations, it is interesting to determine the smaller rectangle encompassing an area, for instance a polygon, which sides are parallel to the coordinate axes. In the list of polygon coordinates, we extract easily x_{min}, x_{max}, y_{min} and y_{max}. Figure 8a shows a polygon *A*, and Figure 8b its minimum bounding rectangle (MBR). But in reality, due to the rotundity of the earth, the minimum/maximum longitude and latitudes, instead of lines, this MBR is limited by part of ellipsoids (Figure 8c). Let us encode it `Mbr(A)`.

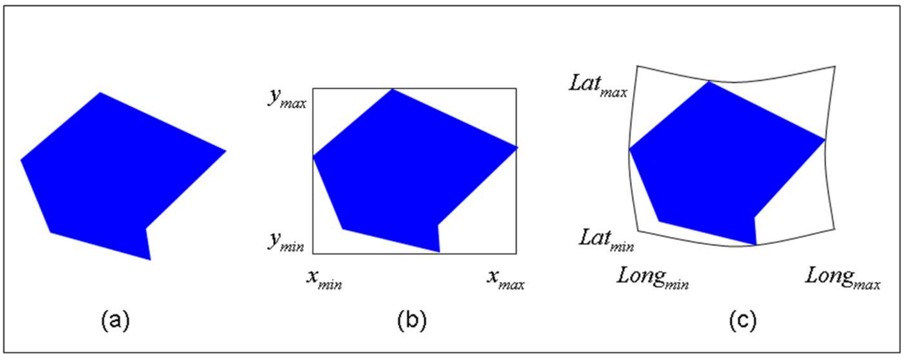

Figure 8. Minimum bounding rectangle. (a) initial polygon. (b) its MBR. (c) an MBR taking earth rotundity into account.

3.1.2. Centroid

Sometimes, it is interesting to exhibit a point representing an area: let us call in a centroid located in the center of the area. Centroids can be loosely defined as a point in the center of an area. But mathematically several possibilities (Laurini-Thompson, 1992) can be used to define a centroid, for instance (Figure 9):

- the barycenter of vertices (Figure 9a),
- the center of gravity (Figure 9b),
- the center of the minimum bounding rectangle (Figure 9c).

Usually, the last one is used because it is easier to be calculated. Let us encode it `Center(A)`. When the polygon is not connected, the center of the MBR of the bigger component is taken; for instance, for the US, only the MBR of the conterminous States is usually used.

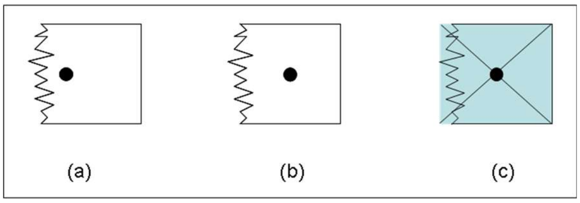

Figure 9. Several definitions of a centroid. (a) as barycenter of coordinates. (b) as center of gravity. (c) as center of the minimum bounding rectangle.

3.1.3. Buffer zones

When one wants to get the people living less than 10 km from the borders of a country, he must define a buffer zone.

More generally when considering a polygon (Figure 10a) and a distance *d*, two buffer zones can be defined, an inner buffer zone (Figure 10b) and an outer buffer zone (Figure 10c). LET us encode it `Buffer(A,d)` in which when *d* is positive, this is an outer zone and when *d* is negative an inner buffer zone.

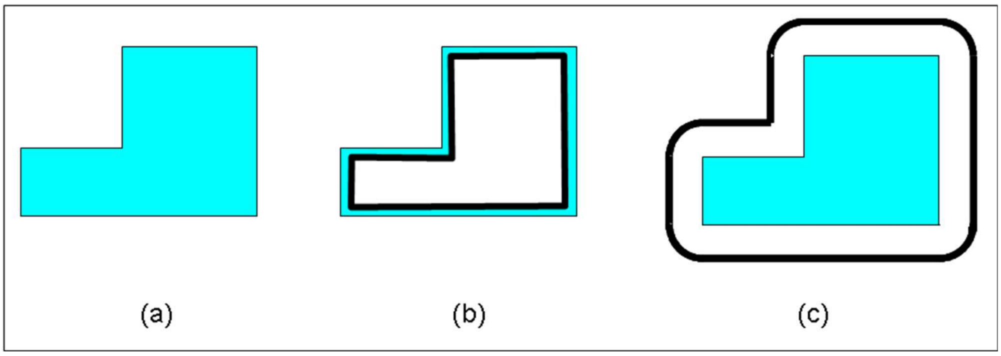

Figure 10. Several definitions of a centroid. (a) As barycenter of coordinates. (b) As center of gravity. (c) As center of the minimum bounding rectangle.

3.2. Spatial relations

First, let me say that spatial relations are hidden in coordinates. In this section, among planar spatial relations, topological and projective relations will be examined. Then, some considerations regarding tessellations will be given.

3.2.1. Topological relations

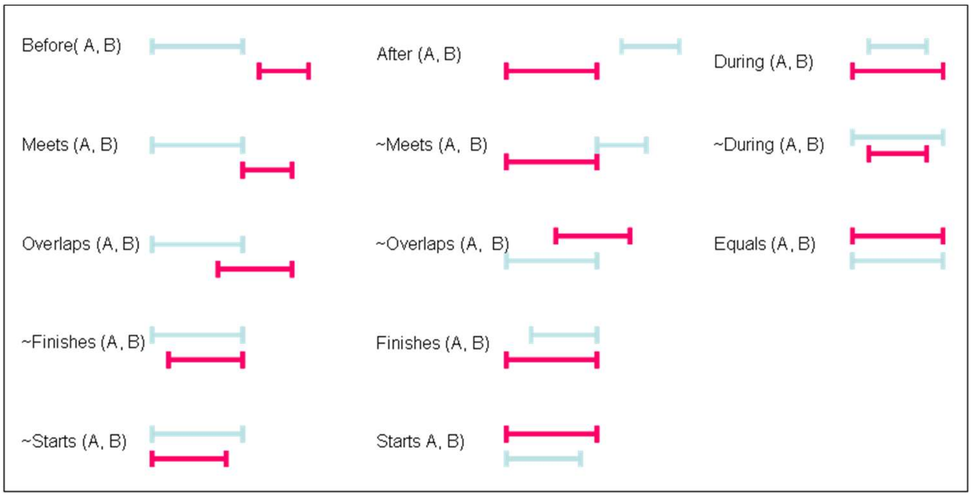

Figure 11. Topological relations at 1D (Allen 1983).

Topological relations such as at 1D, interval relations (Allen 1983) (Figure 11) and at 2D Egenhofer relations (Egenhofer, 1991, 1994]) are well known (Figure 12).

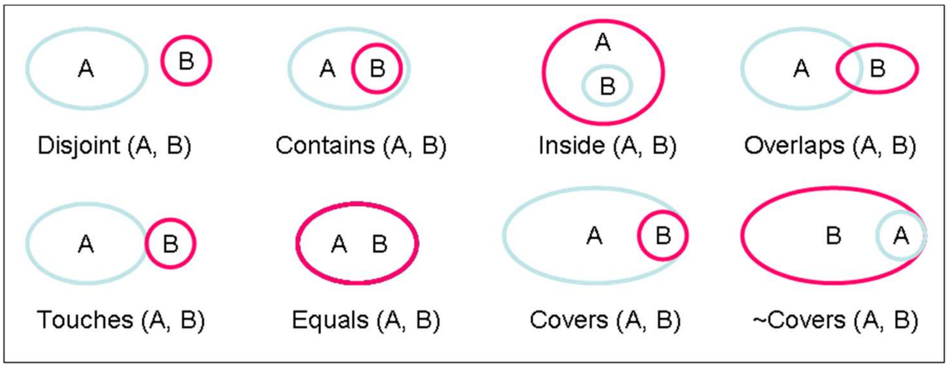

Figure 12. Topological relations at 2D (Egenhofer 1994).

To determine the topological relation between two areas, one solution (Egenhofer et al. 1992, Clementini et al. 1993) is to compute the so-called 9 intersections. Considering a polygon, let us note $A°$ the inner part, $\neg$A the outer part and ∂A its boundary (Figure 13). The answer is given by the following matrix in which the result of one intersection can be void $\varnothing$ or not void $\neg\varnothing$.

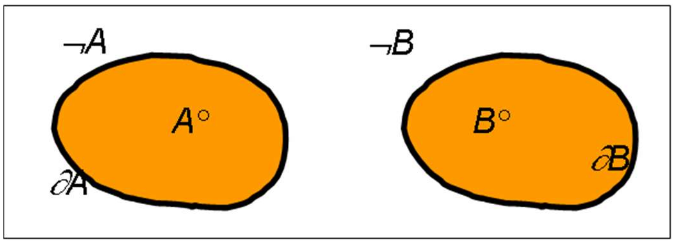

Figure 13. Determining the topological relation between two objects. A° and B° represent the inner parts, ∂A and ∂B the boundaries and ¬A and ¬B the outer parts.

$$R(A,B) = \begin{pmatrix} A° \cap B° & A° \cap \partial B & A° \cap \neg B° \\ \partial A \cap B° & \partial A \cap \partial B & \partial A \cap \neg B \\ \neg A \cap B° & \neg A \cap \partial B & \neg A \cap \neg B \end{pmatrix}$$

For instance, for *TOUCHES*, the result is as follows:

$$TOUCHES(A,B) = \begin{pmatrix} \varnothing & \varnothing & \neg\varnothing \\ \varnothing & \neg\varnothing & \neg\varnothing \\ \neg\varnothing & \neg\varnothing & \neg\varnothing \end{pmatrix}$$

Clementini et al. (1993) have developed this matrix by integrating the dimensions of the intersections (0D, 1D or 2D).

With such topological relations, one can easily define relations for geographic features. For instance, Figure 14 presents a TOUCHES relation between a river and the sea.

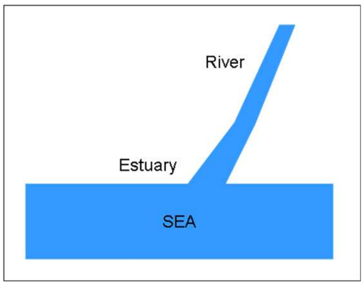

Figure 14. **There is a TOUCHES topological relation between river and sea, corresponding to the estuary.**

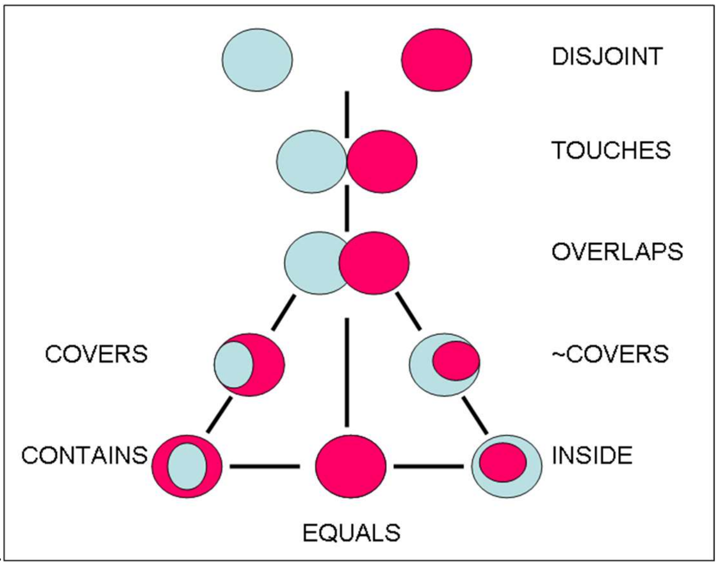

Figure 15. **Vicinity of topological relations.**

Another model for topological relations has been proposed independently in 1992, by Randell, Cui and Cohn (1992) which allowed qualitative spatial representation and consistent reasoning. This logic received the name of "Region Connection Calculus" (RCC); this acronym is also the first letter of authors' names. This model is equivalent to the Egenhofer model. The 8 relations have different names: DC (is disconnected from), EC (is externally connected with), PO (partially overlaps), TPP (is a tangential proper part of), NTPP (is a nontangential proper part of), TPPi (inverse of TPP), NTPPi (inverse of NTPP) and EQUAL.

Anyhow, these topological relations can be organized in a graph showing the vicinity of all relations (Figure 15).

3.2.2. Projective and other spatial relations

But other relations exist such as projective (or cardinal such as North/South, East/west) relations and distance (near/far) relations (Figure 16).

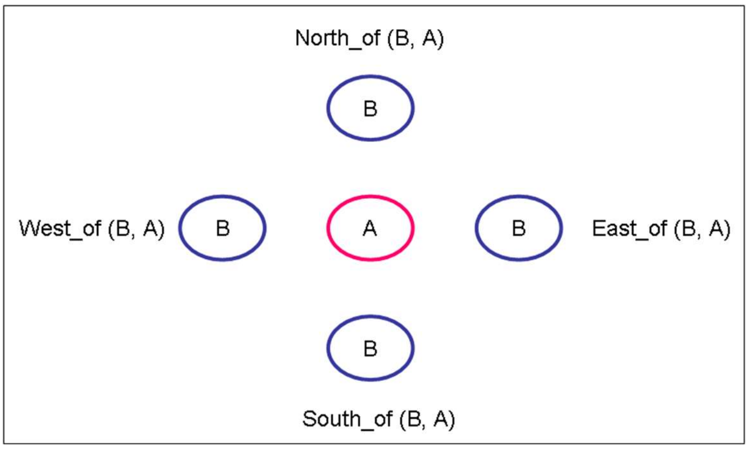

Figure 16. Projective and distance relations.

3.2.3. About tessellations

By irregular tessellation (or tessellation), one means the total coverage of an area by subareas. For instance, the conterminous States in the USA form a tessellation to cover the whole country. Generally speaking, administrative subdivisions form tessellations, sometimes organized as hierarchical tessellations. Let us consider a domain D and several polygons P_i; they form a tessellation iff (See Figure 17a):

- 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.

A tessellation can be also described by Egenhofer relations applied to P_i and D, but in practical cases, due to measurement errors, this definition must be relaxed in order to include sliver polygons (Figure 17b). Those errors are often very small, sometimes a few centimeters at scale 1. In other words, one has a tessellation from an administrative point of view, but not from a mathematical point of view. Let's call them "loose tessellation".

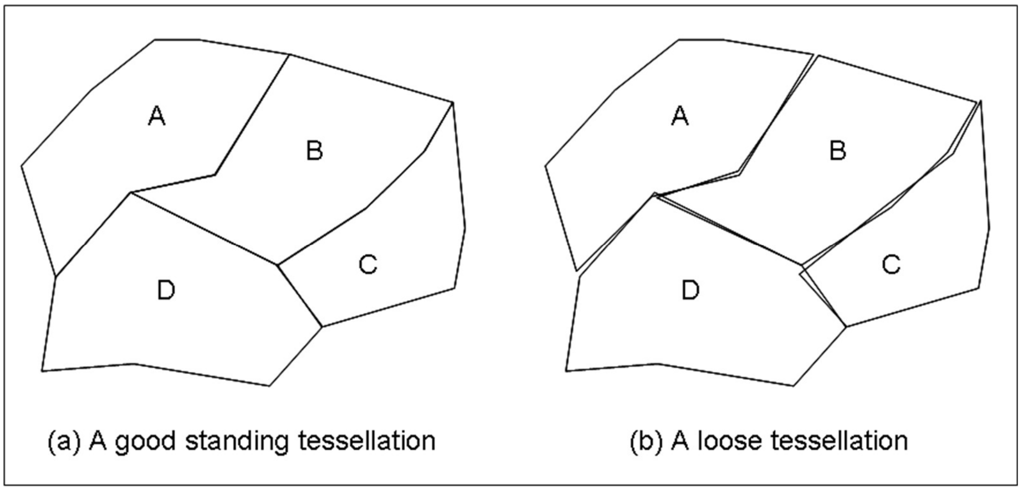

Figure 17. A tessellation with sliver polygons and a good standing tessellation.

But the Earth is not a plane, and some relations must be revisited taken rotundity into account.

3.3. Characteristics of spherical spatial relations

As example, let us consider the relation "`North_of`" and transitivity.

```
North_of (“Philadelphia”, “Boston”)
North_of (“Boston”, “Montreal”)}==> North_of (“Philadelphia”, “Montreal”)
```

But for “East_of”, this is different. Let is first consider the following:

```
East_of (“Philadelphia”, “Paris”)
East_of (“Paris”, “Berlin”)}==> East_of (“Philadelphia”, “Berlin”)
```

But for other cases, transitivity is not true:

```
East_of (“Paris”, “Beijing”)
East_of (“Beijing”, “Philadelphia”)
==> East_of (“Paris”, “Philadelphia”) // false
```

But:

```
East_of (“Philadelphia”, “Paris”) // true
```

3.3.1. East_of and West_of relations

Indeed, the real rule must integrate longitudes: when the transitivity leads to cover more than 180°, this is the converse.

```
East_of (L1, L3) = East_of (L1, L2) ➔ East_of (L2, L3) ∧ ((L3.longitude – L1.longitude) < 180°)
East_of (L3, L1) = East_of (L1, L2) ➔ East_of (L2, L3) ∧ ((L3.longitude – L1.longitude) > 180°)
```

A similar transitivity rule can be written for West_of.

3.3.2. South_of and North_of relations

There are no restrictions for transitivity. However, if one says, “what is north of North Pole?” The answer is void; and moreover, when one is at the North Pole, all directions are going south! But there is always an easternmost feature for any feature.

3.3.3. Projective relations and areas

Projective relations are easy to define for points, but they are more complex for areas. Do not forget that areas can be non-connected, such as countries with several islands. Moreover, some countries can have holes such as Italy with Vatican City and San Marino. Some are constituted of pieces of territory which are very far from the main component: for instance, France with Martinique, Guyana, New Caledonia, etc.

Let’s consider Canada, conterminous states of USA (USA for short) and Mexico. As it is easy to claim the Mexico is south of Canada, what is the exact projective relation between USA-Mexico and Canada-Mexico? It is common to claim that “Canada is north of the conterminous states of the USA”, but Canadian cities such Ottawa or Toronto are south of Seattle. A solution is to consider centroids of those areas. So, but taking this definition, we can claim:

```
North_of (“Canada”, “USA”)
North_of (“USA”, “Mexico”)
```

Moreover, one can claim the `East ("Mexico", "USA")`. But is Canada east or west of the USA? By using both centroids, there is an answer, but this answer is not totally convincing.

3.4. Spatial relations in urban space

Considering a city and spatial relations between urban objects, we can assume that the conventional 3D Cartesian is valid. Moreover, considering again streets, some observations can be made.

- they are one-way streets and sometimes two-way streets can have several lanes;
- some objects are positioned on the street (pedestrian zebras), some under such as sewerages, and some above such as traffic lights;
- some concrete concepts such as sidewalks, medians, crossroads, T-junction, runabout, road signs, curves, engineering network can be defined with the "has_a" semantic relation, but their "topological semantics " are stronger;
- some objects such as engineering networks can be under streets or under sidewalks;
- as previously told, for some actors, streets are defined by the lines with parcels whereas for others the streets are reduced to the asphalted part (Figure 1).

So those observations imply that Allen or Egenhofer relations are not sufficient to describe relationships between street objects. So, the question is "what the minimum set of useful relations could be?"

3.4.1. Binary other topological relations

```
on (street, pedonal_zebra)
under (street, sewerage)
above (street, traffic_light)
along (sidewalk, street)
on (sewerage_grid, street)
```

3.4.2. Relations between urban features and places

```
host (barrack, army)
host (hospital, health_activity).
```

3.5. Ribbon relations

As previously said, ribbons derived from longish rectangles. So, the relations between areas can be applied to ribbons. But due to their particular shapes, other interesting relations between ribbons can be detailed. First let us examine basic operations and some new relations.

3.5.1. Operations

Two operations can be defined. Considering that any ribbon can be decomposed into sub-ribbons, either longitudinally or laterally, we can define two operations, longitudinal splitting and lateral splitting (Figure 18). Of course, those operators can be recursively used.

For longitudinal splitting, one has $w1=w2=w$ and $l1+l2=l$, whereas for lateral splitting $w3+w4=w$ and $l3=l4=l$.

In order to solve the problem, Lee and Hsu (1990, 1992) proposed a table representing all spatial relations between two rectangles. They found a total of 169 types in which they number: 48 disjoint, 40 joint, 50 partial overlaps, 16 contains and 16 belongs (= inside). In our cases, disjoint, partial overlap, contains and belongs relations can be considered as outside our goal. More, due to the semantics of ribbon, a lot of them can be discarded. This is not sufficient. Suppose a road which alternates between simple and dual carriageways. In this case, we need to consider three ribbons, corresponding to dividing and merging. Finally, Figure 19 gives the more interesting ribbon relation, namely Side-by-Side, Edge-to-edge, and merging rectangular ribbons, ribbons and loose ribbons. Similarly, other relations can be defined, for instance crossing, T-junctions, etc.

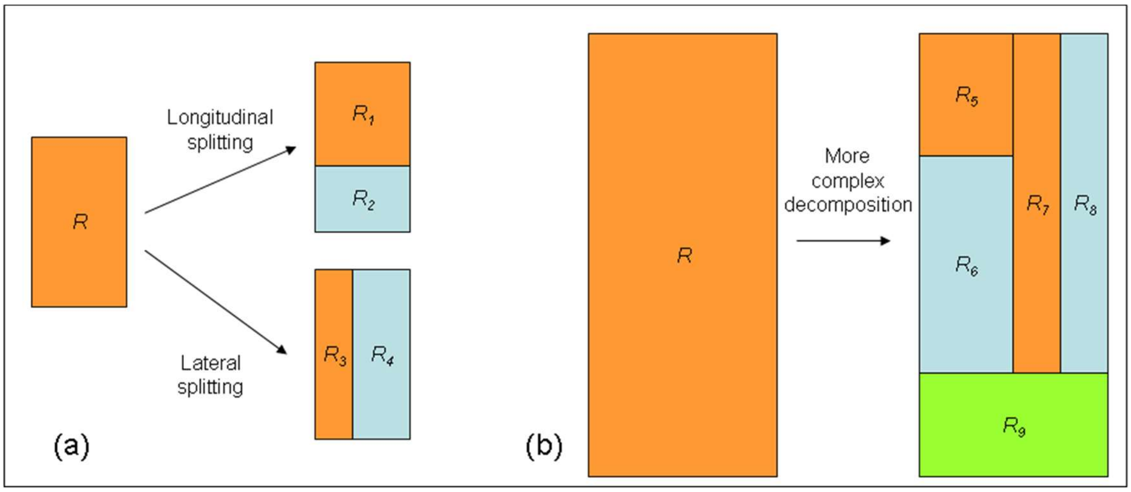

Figure 18. **Two ribbon operators, longitudinal splitting and lateral splitting. (a) Definitions. (b) A more complex decomposition.**

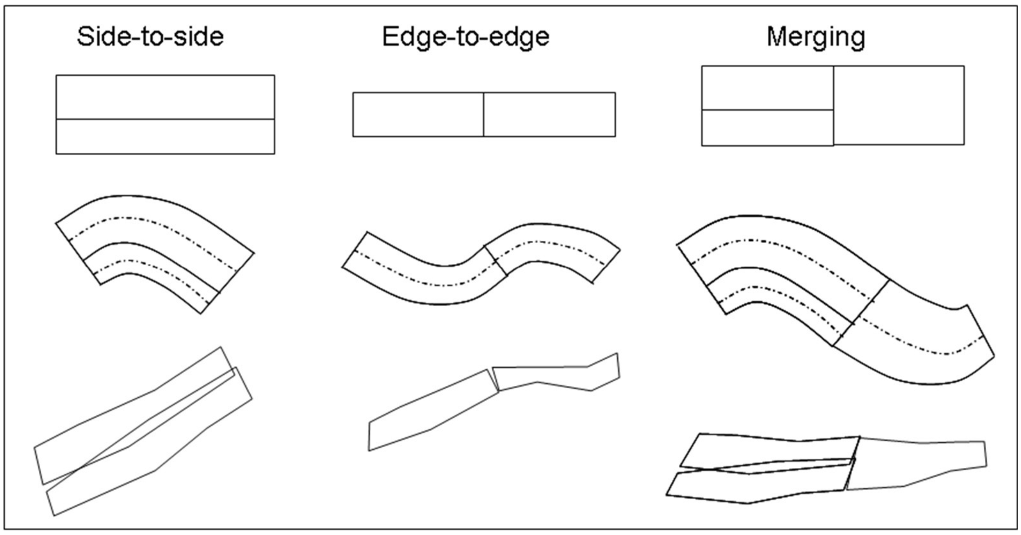

Figure 19. **The more interesting relations between ribbons.**

For instance, in transportation and along rivers, the following relations can hold:

```
SIDE-BY-SIDE (Platform, railways)
SIDE-BY-SIDE (Bus_stop, Bus_lane)
SIDE-BY-SIDE (Levee, River)
SIDE-BY-SIDE (Towpath, River).
```

Initially towpaths along rivers were made for horse-drawn boats; but more and more they are replaced by bike lanes.

3.5.2. Orientation

Since one-way or two-way streets exist, orientation can be defined whereas for some cases of ribbons orientation is not valid. For instance, in a conventional street, the decomposition into ribbons is as in Figure 20.

Pedestrian ribbon (not oriented)
Lane 🡐
Lane 🡒
Pedestrian ribbon (not oriented)

Figure 20. Ribbon orientation.

3.5.3. Chaining ribbons

To model roads and rivers, usually polylines are used to describe the axis; but sometimes two polylines can be used to model riverbanks or the limits of the road. As a consequence, those feature representations can be transformed into ribbons with different widths. For several other reasons, one can have a set of different ribbons that must be concatenated to form a chain of ribbons. Figure 21 gives an example (Fig21a) of several ribbons transformed into a chain of ribbons (Fig21b). In Figure 21c, a case is presented needing two additional curves to join the sides of two ribbons

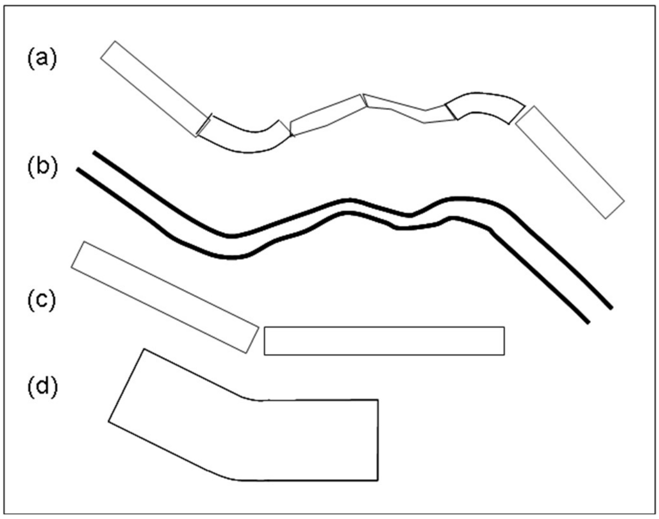

Figure 21. Chaining ribbons. (a) An example of different ribbons. (b) Chain of ribbons. (c) A case of two rectangular ribbons. (d) Additional curves to join the sides.

3.6. Mutation of topological relations according to the granularity of interest

But according to the granularity of interest, topological relations can vary. For instance, it is common to claim that a road is going along the sea, so implying a TOUCHES relation between the road and the sea. But if we consider carefully, sometimes there are small beaches between the road and the sea (Figure 22). From a cartographic point of view, the type of relation will vary: indeed at a scale of 1:1000, the relation is DISJOINT whereas at 1:100000, there is a TOUCHES. More generally, the concept of granularity of interest will enlarge the concept of scale: it is clear that the interest for the same zone for a local politician and a nation-wide politician can be different.

(a) (b)

Figure 22. According to scale, the road TOUCHes or not the sea.

So according to the granularity of interest, geographic objects can mutate according to two rules (see examples Figure 23). As scale diminishes, an area will mutate into a point and then will disappear, and a ribbon will mutate into a line and then will disappear.

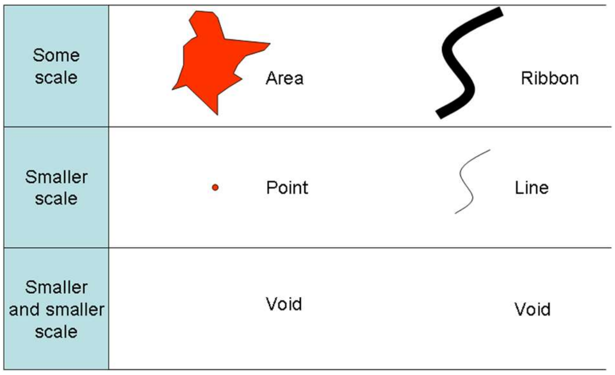

Figure 23. Mutation of geographic object representations.

In the sequel to simplify the presentation, we will continue to use the thresholds for visual acuity instead of granularity of interest.

3.6.1. General process

So, the complete process can be modeled as follows:

- Step 0: original geographic features,
- Step 1: as scale diminishes, small areas and ribbons will be generalized and possibly can coalesce,
- Step 2: as scale continues to diminish, areas mutate 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).

3.6.2. Visual acuity applied to geographic objects

It is well known that "Cartographic representation is linked to visual acuity". Thresholds must be defined. In classical cartography, the limit ranges from 1 mm to 0.1 mm. If one takes a road and a certain scale and if the transformation gives a width more thatn 1 mm, this road is an area, between 1 mm and 0.1mm, then a line and if less than 0.1mm the road disappears. The same reasoning is valid for cities or small countries such as Andorra, Liechtenstein, Monaco, etc. In these cases, the "holes" or small islands in Italy or in France disappear cartographically. With the thresholds ε_i, ε_{lp} previously defined, we can formally get (in which *2Dmap* is a function transforming a geographic object to some other scale possibly with generalization) the following:

a/ Disappearance of a geographic object (O) at scale σ:

$\forall O \in GeObject, \ \forall \sigma \in Scale \wedge O_\sigma = 2Dmap(O, \sigma) \wedge Area\ (O_\sigma) < (\varepsilon_{lp})^2 \Rightarrow O_\sigma = \varnothing$.

b/ Transformation of an area into a point (for instance the centroid of the concerned object, for instance taken as the center of the minimum bounding rectangle):

$\forall O \in GeObject, \ \forall \sigma \in Scale \wedge O_\sigma = 2Dmap(O, \sigma) \wedge \ (\varepsilon_i)^2 > Area\ (O_\sigma) > (\varepsilon_{lp})^2 \Rightarrow$

$O_\sigma = Centroid(O)$.

c/ Transformation of a ribbon R into a line (for instance its axis):

$\forall R \in Ribbon, \ \forall \sigma \in Scale \wedge R_\sigma = 2Dmap(R, \sigma) \wedge \ \varepsilon_i > Width(R_\sigma) > \varepsilon_{lp} \Rightarrow R_\sigma = (Axis(R))$.

Therefore, one can say that any spatial relation varies according to scale. As previously told, one says that a road runs along a sea; but in reality, in some places, the road does not run really along the water of the sea due to beaches, buildings, etc. As previously said, at one scale, the road TOUCHes the sea, but at another scale at some places, this is a DISJOINT relation. Let consider two geographic objects O^1 and O^2 and O_σ^1 and O_σ^2 their cartographic representations, for instance the following assertion holds:

$\forall O^1, O^2 \in GeObject \wedge \ \forall \sigma \in Scale \wedge O_\sigma^1 = 2Dmap(O^1) \wedge O_\sigma^2 = 2Dmap(O^2) \wedge Disjoint\ (O^1, O^2) \wedge Dist\ (O^1, O^2) < \varepsilon_2$

$\Rightarrow Touches(O_\sigma^1, O_\sigma^2)$.

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.

3.6.3. Example of topological mutation due to granularity of interest: DISJOINT to TOUCHES

In this section, the Egenhofer's relations (Egenhofer, 1991) are treated mainly after the generalization, the object geometries are adapted to the perceptual limits imposed by the new (smaller) scale (Laurini, 2014). As example, let us analyze the "DISJOINT" relation which can mutate into the "TOUCHES" relation according to the thresholds previously defined, the following assertion must apply (Figure 24):

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

$O_\sigma^1 = 2Dmap(O^1) \wedge O_\sigma^2 = 2Dmap(O^2) \wedge Disjoint\ (O^1, O^2) \wedge Dist\ (O^1, O^2) < \varepsilon_2$

$\Rightarrow Meet\ (O_\sigma^1, O_\sigma^2)$.

But a smaller object can disappear or be eliminated if its area is too small to be well visible. So in this case, the initial relation does not hold anymore.

$\forall O \in GeObject, \ \forall \sigma \in Scale \wedge O_\sigma = 2Dmap(O, \sigma) \wedge Area\ (O_\sigma) < (\varepsilon_1)^2 \Rightarrow O_\sigma = \varnothing$.

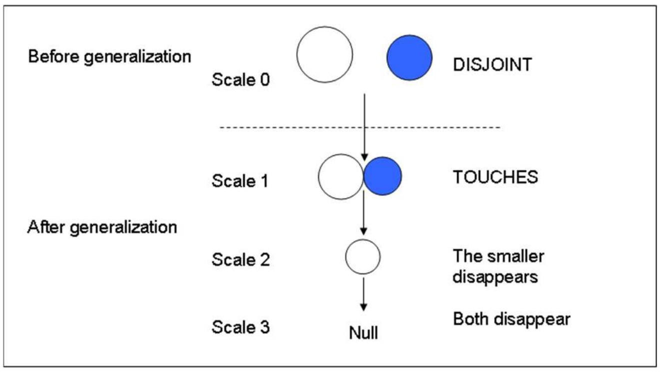

Figure 24. The mutation DISJOINT-to-TOUCHES.

4. Fundamentals of Geographic Knowledge

In addition to the importance of geometry and topology, geographic knowledge is also concerned by two types of structures, gazetteers and ontologies. As ontologies are frequently used in artificial intelligence, in addition to conventional relations, they need to integrate spatial relations. But a second structure named gazetteer is concerned by placenames, their organization and their location: indeed, several places can have the same names, and several names can be assigned to places.

But geographic knowledge is linked to cartography and visualization. From several years new types of cartography are emerging under the general umbrella of geovisualization, especially with chorems. Let us examine those issues in detail.

4.1. Geographic ontologies

In general, an ontology specifies a vocabulary of concepts together with some indication of their meanings (Gruber 1993, Guarino 1998). As discussed in Smith and Mark (2003), the term 'ontology' is used nowadays by information scientists, in a non-philosophical sense to assist in the task of specifying and clarifying the concepts employed in given domains, above all by formalizing them within the framework of some formal theory with a well-understood logical (syntactic and semantic) structure." From a computational point of view, an ontology can be seen as a network of concepts linked essentially by the following relations:

- "is a" (females and males are subtypes or subclasses of human being),
- "has a" (a paper has one or several authors),
- "part of" (a finger is a part of a hand).

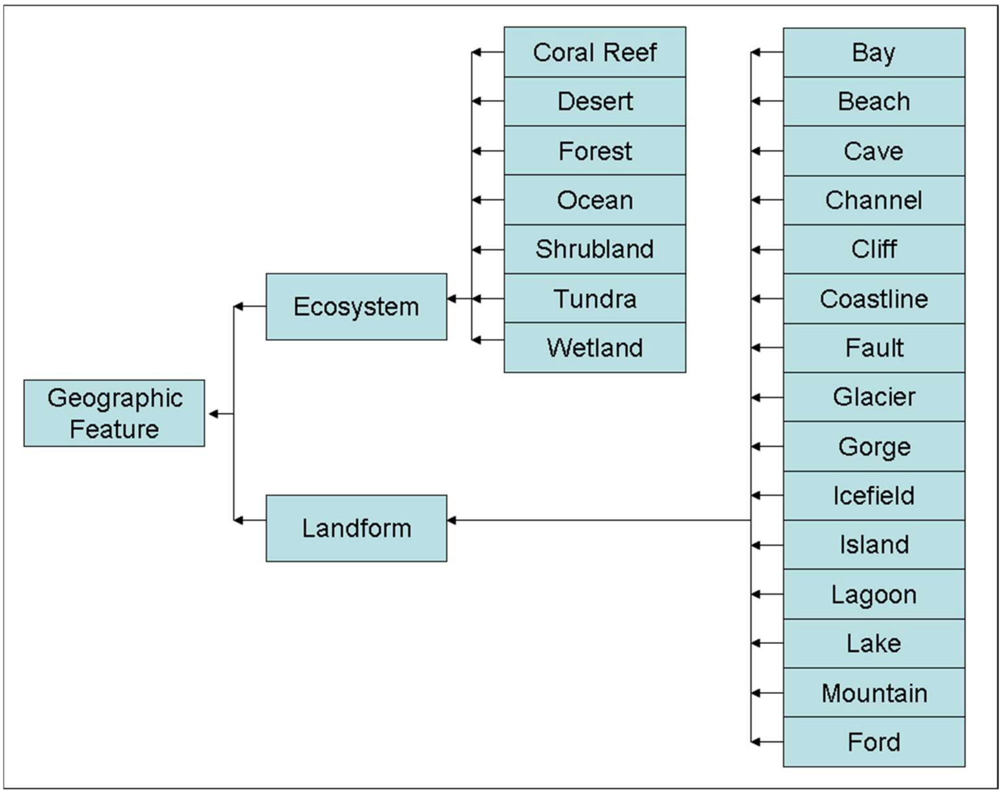

Figure 25. Example of the beginning of a geographic ontology only with is-a relations.

However, the specificity of geographic ontology does not lie only in geographic features (as illustrated in Figure 25) (Kavouras et al. (2005). But it lies overall in their geometry and in their spatial relationships (Laurini, 2012). Usually, Egenhofer or RCC relations are fully integrated in the definitions of geographic features. See Figures 26 and 27 for such examples, the first for the planet, and the second for administrative subdivisions of a country.

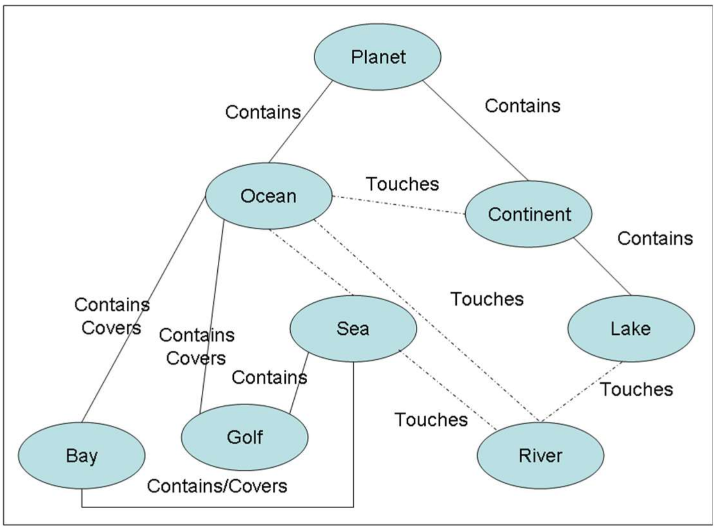

Figure 26. Example of ontology based on spatial relations.

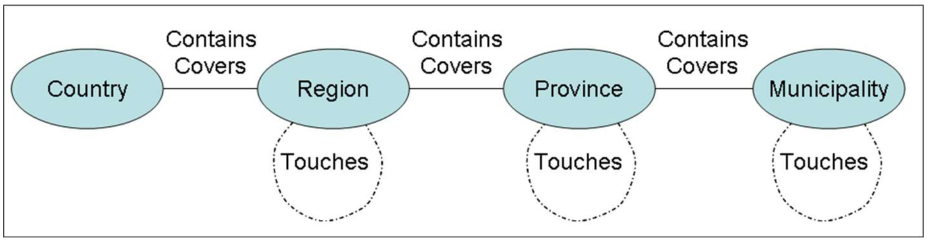

Figure 27. Example of administrative subdivisions with spatial relations.

4.2. About geographic names and gazetteers

By definition, a gazetteer is a geographic dictionary or directory (Goodchild and Hill (2008), Keßler et al. 2009). But now, more and more gazetteers become complex databases. Since they more and more include other attributes of the named features, they tend to become some toponym ontologies (Smart et al. 2010, Hećimović et al., 2013).

4.2.1. What under a name?

Under a geographic name, various objects or features can exist. In the Earth, few points have names, maybe only the North and South Poles, and only few lines such as Equator, Tropic of Cancer, Tropic of Capricorn, Greenwich Meridian, Polar Circle, etc. The majority of names are given to areas since even rivers are areas, maybe modeled as ribbons. As previously told, they must be considered as non-connected (with islands and holes) and they can be replaced by their centroid for some operations.

4.2.2. Generalities

Indeed, in addition to a pure list of placenames or toponyms, it is necessary to locate them with accuracy and to assign them some features or geographic objects. Moreover, a place can have different names in different languages and different periods of time. Let us first examine few well-known examples:

- “Mississippi” can be the name of a river or of a state,
- The city is “Venice”, Italy, is also known as “Venezia”, “Venise”, “Venedig” respectively in Italian, French and German.
- The local name of the Greek city of “Athens” is “Αθήνα” read [a'θina].
- “Istanbul” was known as “Byzantium” and “Constantinople” in the past.
- The today city of “Rome” is much bigger than in Romulus time.
- There are two Georgias, one in the United States and another one in Caucasia.
- The toponym “Milano” can correspond to the city of Milano or the province of Milano.
- Some cities have specific characteristics such as capital of a state; a river can have an estuary in the sea.
- The river “Danube” crosses several European countries; practically in each country it has a different name, “Donau” in Germany and Austria, “Dunaj” in Slovakia, “Duna” in Hungary), “Dunav” in Croatia and Serbia, “Dunav” and “Дунав” in Bulgaria, “Dunărea” in Romania and in Moldova, “Dunaj”,and Дунай” in Ukraine. It is also called “Danubio” in Italian and Spanish, “Tonava” in Finnish and “Δούναβης” in Greek. Moreover, its name in feminine in German, and masculine in other languages.
- Sometimes, names of places can be also names of something else; for instance, “Washington” can also refer to George Washington or anybody with this first name or last name.
- Some placenames are formed of two or several words; for instance, “New Orleans”, “Los Angeles”, “Antigua and Barbuda”, “Trinidad and Tobago”, “Great Britain”, “Northern Ireland”, “Tierra del Fuego”, “El Puente de Alcántara”, etc.

- Some very long names can have simplification; the well-known Welsh town “Llanfairpwllgwyngyllgogerychwyrndrobwllllantysiliogogogoch” is often simplified into “Llanfair PG” or “Llanfairpwll”.
- Some abbreviations can be common such as “L.A.” for “Los Angeles”.
- Peking became Beijing after a change of transcription to Roman alphabet; but the capital of China has not modified its name in Chinese.
- In some languages, grammatical gender is important so that placenames can be feminine or masculine; for instance in French, Italian and Spanish, names such as “Japan”, “Brazil”, “Portugal” are masculine whereas “Argentina”, “Bolivia” and “Tunisia” are feminine; in addition as the great majority of toponyms is singular, some can be plural like “The Netherlands”, “The Alps”.

As consequence, there is a complex many-to-many relationships between places and placenames (Figure 28).

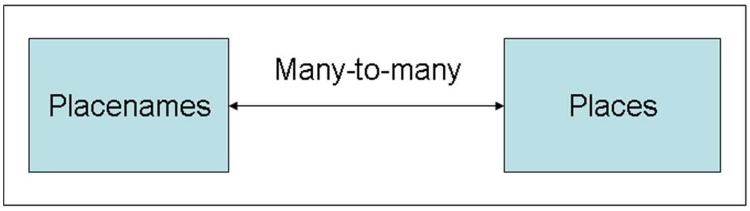

Figure 28. A very complex many-to-many relation links places and their names.

Among place names, they are street names together with the number in the street (civic number); they are not so easy to handle. This is very important, not only for the automatic processing of postal addresses, but also for all applications connecting to emergencies. The URISA association has organized many conferences on the topics (See www.urisa.org). The specificities of street names are as follows:

- some streets can have a few dozens of yards whereas other several miles;
- in some human settlements, streets have no names;
- sometimes there are variations about the way to write some street names; for instance “3rd Street”, “Third Street”, “Third St”; the words “avenues” and “boulevards” are simplified into “Ave”, or “Blvd” or “Bd”;
- in some countries, the equivalent of the words “street”, “avenue”, etc. are usually removed;
- in some places, streets can have several names; for instance in New York City, “Sixth Avenue” is also known as “Avenue of the Americas”;
- etc.

As a main consequence, the name of a place cannot be a unique ID from a computing point of view.

In order to clarify, let us give a few definitions:

- **toponym** is the general name of a geographic feature or object;
- **endonym** is a local name in the official language of the country; there may be several toponyms in countries with different official languages (Brussel in Flemish, Bruxelles in French);
- **exonym** is a name in other languages than the official languages; for instance Brussels in English;
- **archeonym** is a name which existed in the past: for instance, Byzantium for Istanbul;
- **hyperonym** and **hyponym**: names of places with subordination; hyponym is the opposite of hyperonym; for instance, Europe is a hyperonym of France whereas France is a hyponym of Europe;
- **mereonym** is a name of a part of place without subordination; “Adriatic Sea” corresponds to a mereonym of the Mediterranean Sea;
- **hydronym** is a name of a waterbody; it can be also used for seas;
- **oronym** is a name for a hill or a mountain.

Figure 29 gives the essential elements of a gazetteer, the names, the features, the dates and everything regarding geometry and georeferencing according to Jakir et al. (2011).

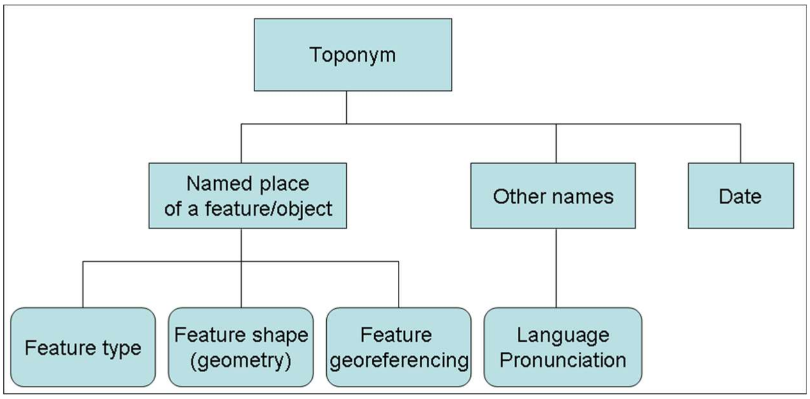

Figure 29. Essential elements of a toponym, after Jakir et al. (2011).

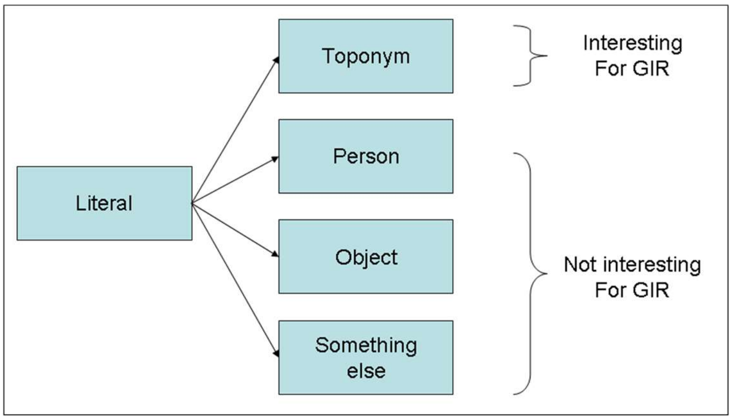

Figure 30. Disambiguation of literal to extract toponyms.

To conclude this section, in an automatic system for searching geographic information in the web (known as GIR, Geographic Information Retrieval), a preliminary phase of disambiguation is necessary since the name can correspond to something which is not geographic. Let us define as a literal a string of characters (perhaps including blank spaces and hyphens): this literal may be a toponym, the name of a person (Washington) or something else (China and porcelain) (Figure 30).

4.2.3. Examples

Generally speaking, a gazetteer is designed for a specific activity, for instance to help post offices, to assist the history of a region. As a consequence, several gazetteers can have different structures. Let us examine a few examples.

4.2.3.1. Gazetteer as an index for a map (street directory)

The starting point is a map of a certain region with a precise objective and scale with a visual vocabulary presented in the legend. In this case, the map is usually split into a crossword-like grid in which squares are located by letters and numbers. For instance, "River Street" goes from B3 to C7. The directory can have the following forms.

```
Location1 (street-name, beginning-location, ending-location)
```

In addition, an alternative could be with street names with the names of the other streets which are at the beginning and at the end.

```
Location2 (street-name, beginning-street-name, ending-street-name)
```

4.2.3.2. Gazetteer for a local post-office

For the post office, the gazetteer can have the previous forms, but in addition it can also include several important monuments, administrations and enterprises which can be stored:

```
Urban-feature (name, street-address)
```

4.2.3.3. Gazetteer for hydrology

Here, there are only names of rivers, lakes, seas, etc. Important relations are for tributaries and possible estuaries with sea.

```
Hydronym (id, onto-type, geometry)
Endonym (id, hydronym)
Exonym (id, language, hydronym)
Tributary (id1, id2, location)
Estuary (id1, id2, location)
Mereonym (id1, id2)
```

4.2.3.4. Gazetteer for the history of a place

Here, we essentially deal with ancient names. Anyway, let us start with the actual toponyms.

```
Placename (id, onto-type, geometry, beginning-date)
Archeonym (id, language, toponym, geometry, beginning-date, ending-date)
Exonym (id, language, toponym)
```

4.2.3.5. Gazetteer covering several actual countries, for instance Europe

```
Placename (id, onto-type, geometry, beginning-date)
Exonym (id, language, toponym)
Hydronym (id, onto-type, geometry)
Endonym (id, hydronym)
Exonym (id, language, hydronym)
Mereonym (id1, id2).
```

4.3. Existing systems

Concerning ontologies and gazetteers, several systems exist. Let us rapidly present two of them, GeoNames and GeoSPARQL.

4.3.1. GeoNames (http://www.geonames.org)

The GeoNames database contains over 10,000,000 geographical names corresponding to over 7,500,000 unique features. All features are categorized into one out of nine feature classes and further subcategorized into one out of 645 feature codes. Beyond places names in various languages, data stored include latitude, longitude, elevation, population, administrative subdivisions and postal codes. Among spatial relationships in use in GeoNames, let us mention:

- Children, i.e. the list of administrative divisions (first relative sublevel),
- Hierarchy, i.e. the list of toponyms higher up in the hierarchy of a place name,
- Neighbours, i.e. the list of all neighbours for a country or administrative division,
- Contains, i.e. the list of all features within the feature,
- Siblings, i.e. the list of all siblings of a toponym at same level.

For instance, here is an excerpt of the description of Sicily:

```
<geoname>
 <toponymName>Sicilia</toponymName>
 <name>Sicily</name>
 <lat>37.75</lat><lng>14.25</lng>
 <geonameId>2523119</geonameId>
 <countryCode>IT</countryCode>
 <countryName>Italy</countryName>
 <numberOfChildren>9</numberOfChildren>
</geoname>
```

4.3.2. GeoSPARQL (http://geosparql.org/)

GeoSPARQL is a standard for representation and querying geospatially linked data for the Semantic Web from the Open Geospatial Consortium (OGC). It can be seen as an extension of SPARQL. The definition of a small ontology based on well-understood OGC standards is intended to provide a standardized exchange basis for geospatial RDF data which can support both quantitative and qualitative spatial reasoning and querying with the SPARQL (http://www.w3.org/2009/sparql/wiki/Main_Page) database query language.

But with SPARQL, some simple geographic queries, i.e. without geometric information and spatial relationships, can be launched. For instance: "What are all the country capitals in Africa?"

```
PREFIX abc: <http://example.com/exampleOntology#>
SELECT ?capital ?country
WHERE {
 ?x abc:cityname ?capital ;
 abc:isCapitalOf ?y .
 ?y abc:countryname ?country ;
 abc:isInContinent abc:Africa.
}
```

But with GeoSPARQL, not only geometric attributes (shapes) but also Egenhofer/RCC topological relations can be invoked. In addition, the following functions are integrated, distance, buffer, convex hull, intersection, union, difference, etc. The general structure and an example are given in Figure 31. To get the Washington monument, one has to write a small filter as a minimum bounding rectangle:

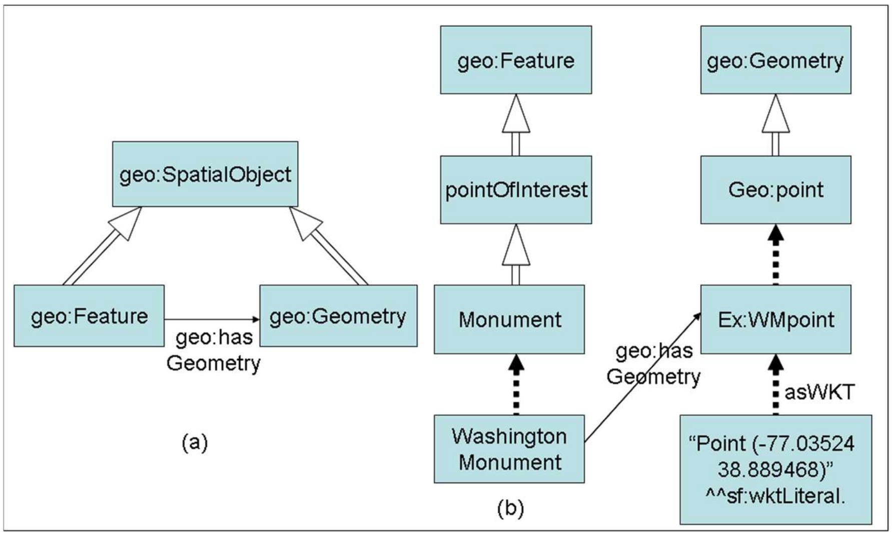

Figure 31. Example of describing geographic entities in GeoSPARQL. (a) Generic structure. (b) Example for a monument.

```
PREFIX geo: <http://www.opengis.net/ont/geosparql#>
PREFIX geof: <http://www.opengis.net/def/function/geosparql/>
PREFIX sf: <http://www.opengis.net/ont/sf#>
PREFIX ex: <http://example.org/PointOfInterest#>
SELECT ?a
WHERE {
?a geo:hasGeometry
?ageo .
?ageo geo:asWKT
?alit .
FILTER( geof:sfWithin(?alit, "Polygon((-77.089005 38.913574,-77.029953 38.913574,-77.029953 38.886321,-77.089005 38.886321,-77.089005 38.913574))"^^sf:wktLiteral)) }
```

For instance a query for getting the airports near London is as follows:

```
PREFIX co: <http://www.geonames.org/countries/#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX geo: <http://www.w3.org/2003/01/geo/wgs84_pos#>
SELECT ?link ?name ?lat ?lon
WHERE  {
   ?link gs:within(51.139725 -0.895386 51.833232 0.645447) .
   ?link gn:name ?name .
   ?link gn:featureCode gn:S.AIRP .
   ?link geo:lat ?lat .
   ?link geo:long ?lon }
```

If somebody is looking for all land parcels with some type of commercial zoning that touch some arterial street, the query is the following:

```
SELECT ?parcel ?hwy
WHERE { ?parcel rdf:type :Commercial .
 ?parcel rdf:type ogc:GeometryObject .
 ?hwy rdf:type :Arterial_Street .
```

```
?hwy rdf:type ogc:GeometryObject .
?parcel ogc:touches ?hwy }
```

4.4. Geovisualization

Conventional cartography is usually based on Bertin semiology (Bertin 1967) who has established the fundamentals of graphics. Then several other tracks were followed, especially thanks to the facilities of data processing. For instance, a cartogram is a type of graphics that depicts attributes of geographic objects as the object's area. Because a cartogram does not show geographic space, but rather changes the size of objects depending on a certain attribute, a cartogram is not a true map, but it is a good representation of the phenomenon. An example is given Figure 32 illustrating gross national products. It is argued that this kind of representation bring more knowledge than common representations.

After James J. Thomas and Kristin A. Cook (2005), visual analytics is the science of analytical reasoning facilitated by interactive visual interfaces. Applied to GIS, visual analytics is also known as geoanalytics and geovisualization. This is the way to transform data and especially digital data into drawings or maps so that anyone can access, understand and interact with to harvest rich insight from vast data sources. Visual analytics tools and techniques create an interactive view of data that reveals the patterns within it, enabling everyone to become researchers and analysts. It brings together information technology, information visualization, cognitive and perceptual sciences, interactive design, graphic design, and social sciences. For instance, cartograms are shown in the site http://www-personal.umich.edu/~mejn/cartograms/ in which all countries are presented in proportion of some phenomenon, i.e. their national gross products or the number of patients.

Remember that during millennia maps were both the visualization and the way to store geographic information. From half a century, progressively it was clear to distinguish between those two aspects. During a few decades, attention was paid on storing efficiently geographic data in databases. Now, it turns out to ameliorate visualization by trying to propose new tools which can facilitate geographic reasoning.

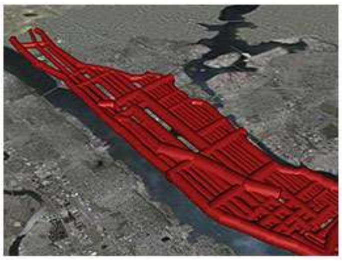

Figure 32. Novel representations: Traffic flows in New-York City, Source http://www.nadiaamoroso.com with permission.

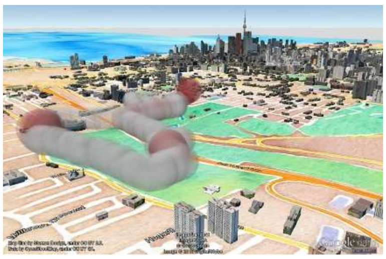

Figure 33. Novel representations: Pollution level, Source http://www.nadiaamoroso.com with permission.

In other words, visual reasoning can be seen as a competitor to conventional spatial reasoning as it is known in artificial intelligence and computational geometry. In this section, a few examples will be shown: Figure 32 for traffic flows whereas Figure 33 depicts pollution.

4.5. Chorems

For the representation of geographic knowledge, a visual solution could be based on chorems (Brunet 1980, DelFatto et al., 2007) which are schematic representations of territories. An example is given Figure 34 representing Spain and its characteristics, a desert in the middle, farming and forests in the North-Western part, tourism saturation along the coast of the Mediterranean Sea and two big cities, Madrid and Barcelona.

Chorems are essentially used to schematize information about a territory, but they can also be used as an alternative access method to geographic databases.

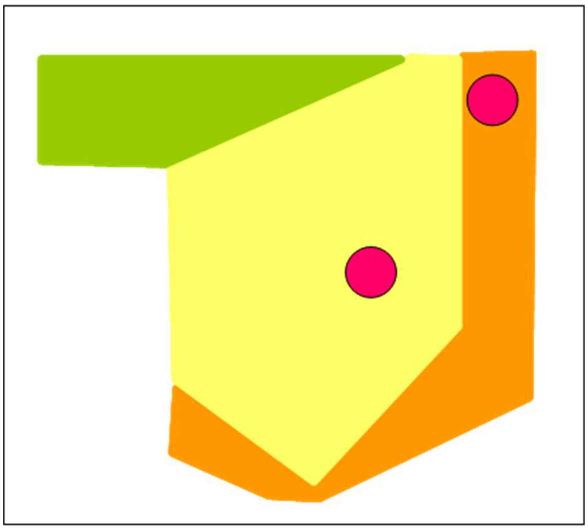

Figure 34. Chorem showing Spain and its characteristics.

5. Geographic rules

Now that the key concepts are established, we can state some prolegomena as preliminary assertions constituting the underlying foundations of principles. They are organized as follows (Laurini, 2014):

- 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. (Tobler, 1970).

5.1. Prolegomena

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 (Goodchild-Hill, 2008). In other terms, 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 (1995), Gordillo (2001) or Kang et al. (2002) 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 loosing less accuracy as possibl*e". Any geographic knowledge system must include those procedures.

Prolegomenon #5 (From Popper's falsifiability principle (Popper, 1934)): "*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 characteristics of an object has varied (f.i. landuse 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 (1973).

Prolegomenon #10 (Place names and gazetteers): "*Relationships between places and place names are many-to-many*". As previously told, 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 (Laurini, 2012). When necessary, raster information can be included in 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 levels 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 1970): "*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.

5.2. Principles

Now that prolegomena are stated, principles governing geographic knowledge may be listed in order to get robust reasoning and retrieval. The principles are organized as follows:

- The three first concern 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 same kind of relations for all cities throughout the world. A 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.

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 systems 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 longer be the terrain altitude, but the bird altitude. Based on this principle, all procedures to increase geographic knowledge quality must be invoked.

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 objects, 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 any geographic knowledge repository is infinite (intensional), only implicit knowledge is stored, but other knowledge chunks can be derived when necessary.

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_σ =2Dmap(O).

Principle #5 (Visualization and visual acuity): "*Cartographic representation is linked to visual acuity*". Here again thresholds must be defined. In classical cartography, as previously said, the limit ranges from 1 mm to 0.1 mm. Let us 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 0.1mm a line, and less than 0.1mm the road disappears. The same reasoning is valid for cities or small countries such as Andorra, etc. In these cases, the "holes" in Italy or in France disappear cartographically. With the thresholds ε_i, ε_{lp} previously defined, we can formally get:

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

b/ $\forall O \in GeObject, \forall \sigma \in Scale \wedge O_\sigma = 2Dmap(O) \wedge (\varepsilon_i)^2 > Area(O_\sigma) > (\varepsilon_{lp})^2 \Rightarrow O_\sigma = Centroid(O)$.

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 stay 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 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 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%. Figure 35 illustrates this process. This process is like the reduction of a ribbon into a line.

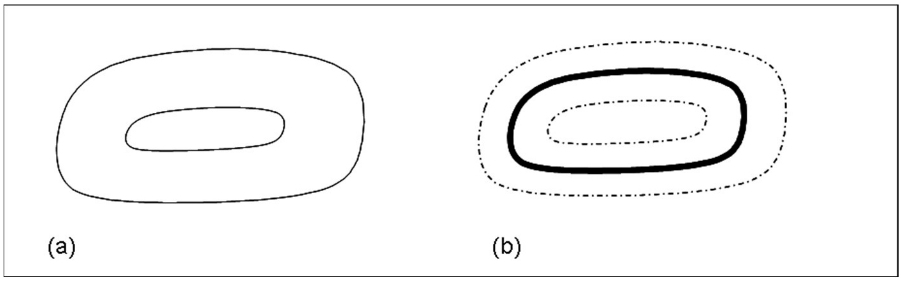

Figure 35. Crispification of a geographic object modeled by the egg-yolk representation. (a) the original model. (b) its reduction to a crisp object.

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

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 places, 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 (Figure 20). Let consider two geographic objects O1 and O2 and their Oσ1 and Oσ2 their cartographic representations, for instance the following assertion holds:

$\forall O^1, O^2 \in GeObject \wedge \forall \sigma \in Scale \wedge O_\sigma^1 = 2Dmap(O^1) \wedge O_\sigma^2 = 2Dmap(O^2) \wedge$

$Disjoint(O^1, O^2) \wedge Dist(O^1, O^2) < \varepsilon_2 \Rightarrow Touches(O_\sigma^1, O_\sigma^2)$.

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.

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.

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 an adjacent garden texture can reveal a parcel.

Principle #10 (Visualization constraints): "*The spatial relations between objects must hold after generalization*". In Figure 36, an excerpt of the English Riviera coast along the Channel is showed. Suppose we generalize the shoreline by a single line: the city of Eastbourne will be in the middle of the sea whereas Plymouth and Bournemouth 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 also valid for the rivers going to the sea. In the same spirit, some cities at the borders must stay in the proper country. See Geneva for example with the French border.

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

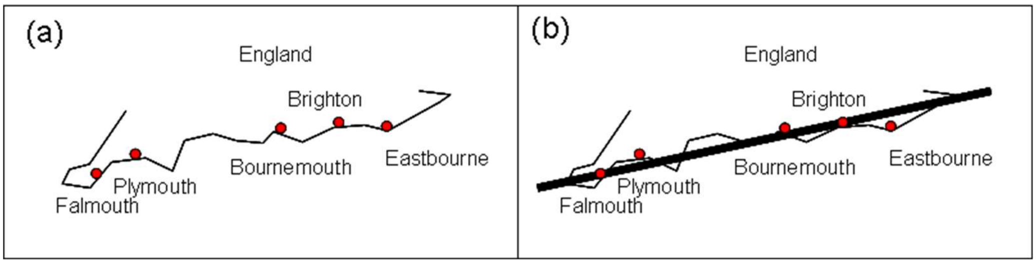

Figure 36. Visualization constraints. (a) Before generalization; (b) Generalization of the coastline, but harbors are badly located.

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.

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 in located in the French Rhône-Alpes region; any automatic social-economic reasoning must take those 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.

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. Figure 35 illustrates a graph example in which two neighboring geographic repositories are present, obviously with geometric discrepancies (Laurini, 1998). Figure 37a shows two geographic repositories before integration and 37b the situation with a magnifying glass emphasizing the discrepancies at the boundary; Figure 37c 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 Figure 37d 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.

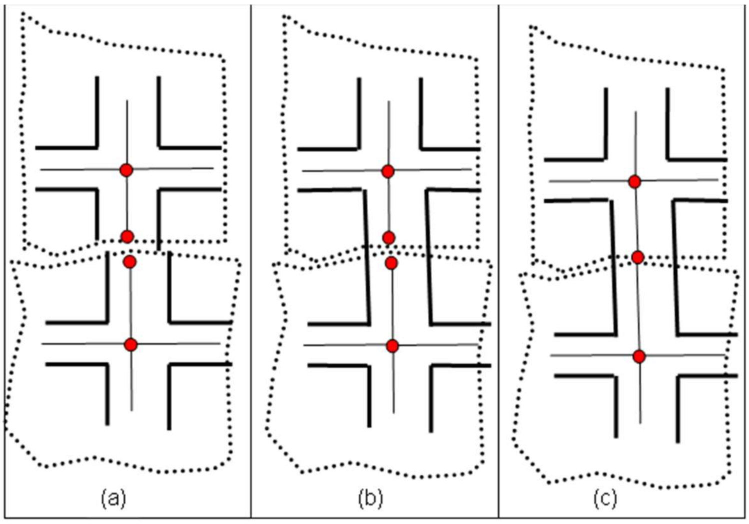

Figure 37. Consequences of cross-boundary interoperability. (a) Before integration. (b) Cartographic integration. (c) Graph-reasoning integration.

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 network 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 2.34m between Belgium and The Netherlands). And secondly, the mathematical shape of the geoid can differ. Once those discrepancies are overcome, the integration of terrains can be launched. In Laurini (1998), a solution based on triangles was proposed.

Do not forget that by applying Principle#11, already a buffer zone is integrated in our geographic knowledge base, and some discrepancies can occur.

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

6. Encoding geographic knowledge

In addition to conventional categories such as facts, concepts, processes and rules, geographic knowledge engineering needs not only to redefine them but also to include new items. First let us examine the different types of geographic knowledge and then the different ways to represent knowledge chunks.

6.1. Types of geographic knowledge

The ChorML project (Coimbra, 2008) mentions the following, geographic facts, cluster of areas, flows (persons, goods, etc.), and co-location rules. But we need to add gradients, external information and topological constraints.

6.1.1. Geographic facts

The notion of geographic facts must be revisited. In some cases, 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 4807m". 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; as previously said, a solution can be based on centroids.

6.1.2. Clusters of area

In some situations, it could be interesting to regroup areas (polygonal zones) into a single cluster according to some criteria (See Figure 38). This cluster will constitute a new tessellation perhaps with disconnected pieces or with holes.

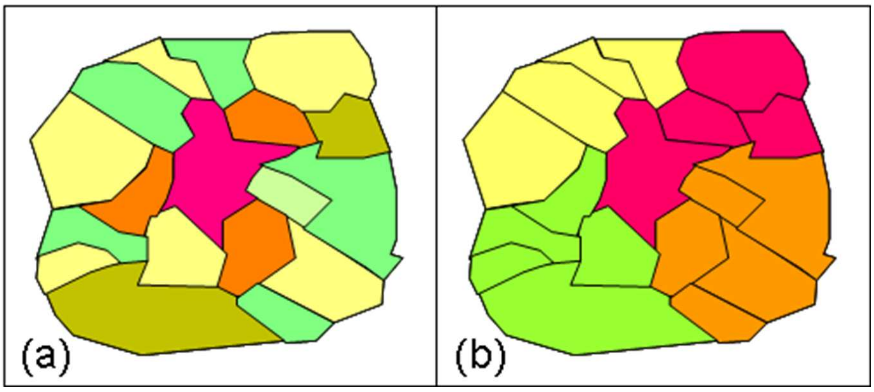

Figure 38. Example of clustering. (a) Initial configuration. (b) One clustering solution in four clusters.

For instance, in UK:

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

6.1.3. Flows

Generally two areas (seen as 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 Figure 39.

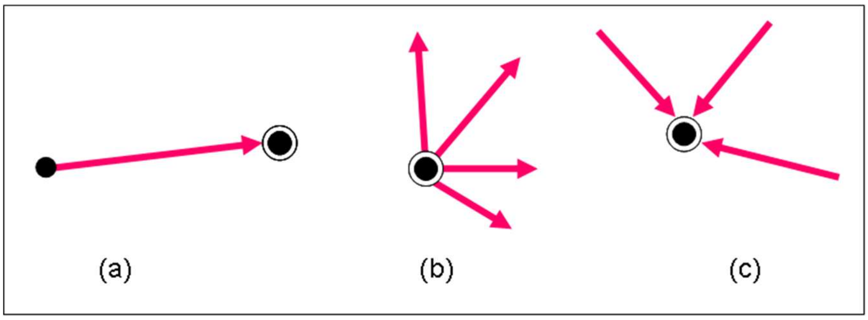

Figure 39. Examples of flows, (a) unidirectional, (b) diverging, (c) converging.

6.1.4. Co-location rules

One of the scopes of geographic data mining (Mennis-Guo, 2009) is to look for co-location rules. When two sets of features are concerned; for instance, "near a big city, there is an airport". Figure 40 shows an example of co-locations.

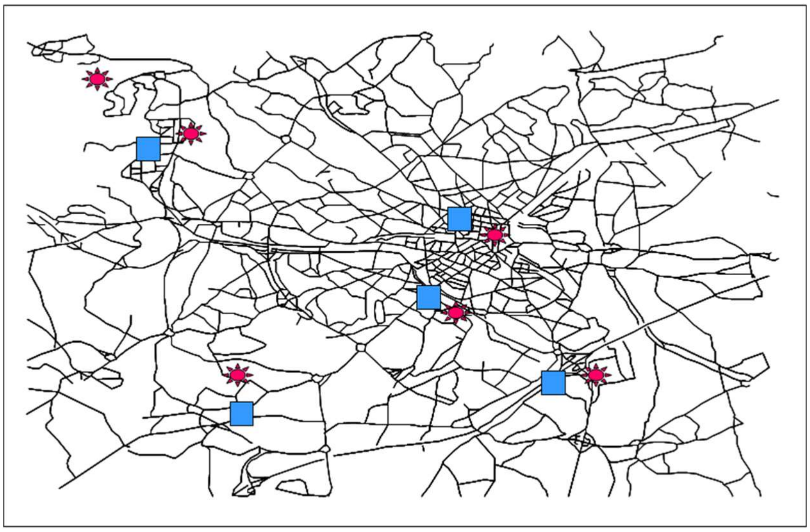

Figure 40. Example of co-locations.

6.1.5. Gradients

When a geographer reads "the more of this, the more of that", there is a high probability that some continuous fields will be concerned. For instance, let us examine the following assertion, "in a beach resort, the nearer to the beach, the higher the price of houses". Behind this assertion, one can discover the continuous field of house prices, or more exactly a smoothed continuous field. Since it is not easy to describe knowledge neither with functions nor with derivatives, a solution can be to consider two similar houses *H1* and *H2*, each at the distances *d1* and *d2* from the beach. Suppose d1 is smaller than d2, one can write:

```
When d1 < d2 Then price (H1) > price (H2).
```

6.1.6. External information and knowledge

As stated in Principle #11, information regarding the vicinity is important. By external information, we mean everything which is outside the concerned territory, but which can be of interest for geographic reasoning. An example is given Figure 41 showing external information for Spain in which two categories are mentioned:

- land borders for France, Portugal, Gibraltar and Morocco (considering Ceuta and Melilla),
- and sea borders for the Atlantic Ocean and the Mediterranean Sea.

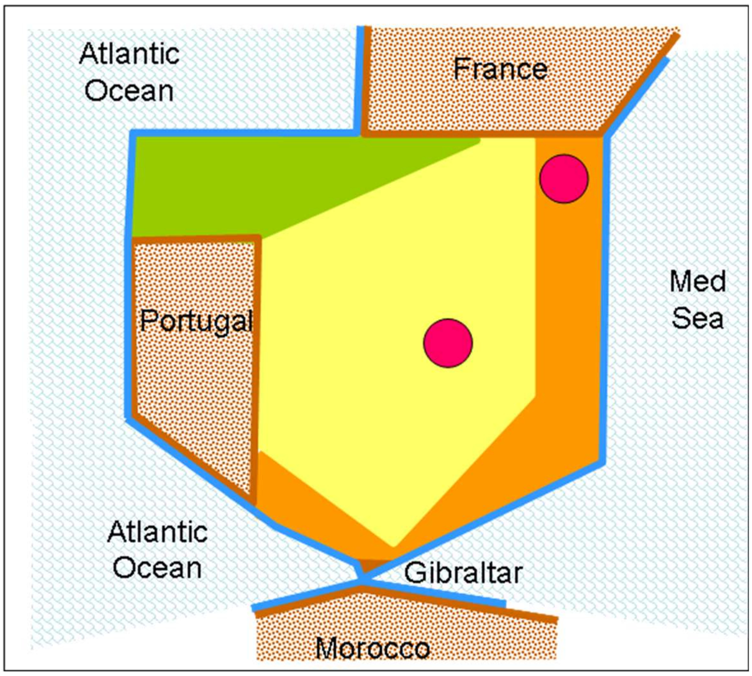

Figure 41. **Spain and its external information to support some reasoning.**

6.1.7. 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. Figure 42 gives a visual example of topological constraints in France; due to generalization, some cities can be badly located, for instance the French city of Marseilles and the Swiss city of Geneva.

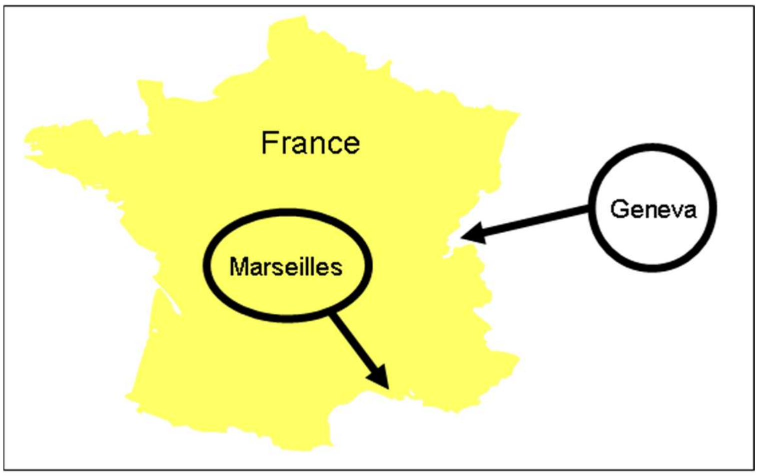

Figure 42. **Example of visual representation of constraint stating that the city of Marseilles must be always inside (`TOUCHES` relation) the French territory, and Geneva outside (`~COVERS`).**

6.2. Representing geographic knowledge

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

6.2.1. 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*".

6.2.2. Logics

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

$$\forall l \in Lake \wedge \forall s \in Street \wedge (\text{ Touches } s\ (l, s) \Rightarrow \exists r \in Restaurants \wedge (distance\ (r,l) < 100 \wedge (\text{Dist}\ (r,s) < 100.$$

Practically, everything can be encoded provided that the corresponding spatial relations can be used. Do not avoid spatial relations sometimes include sophisticated computational geometry algorithms.

6.2.3. XML-encoding

XML can be the basis of geographic knowledge: for instance, SpatialML (Mani et al. 2010) 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. Here is an example for the phrase "a building 5 miles east of Fengshan":

```
a <PLACE id="1" type="FAC" form="NOM">building</PLACE>
<SIGNAL id="2">5 miles</SIGNAL>
<SIGNAL id="3">east</SIGNAL> of
<PLACE id="4" type="PPL" country="TW" form="NAM" latLong="22∘37′N
120∘21′E">Fengshan</PLACE>
<PATH id="5" source="4" destination="1" distance="5:mi" direction="E"
signals="2 3"/>
```

The used spatial relations based on the RCC model are given in Table 1

Spatial relations	Example
IN (tangential and non-tangential proper parts)	[Paris], [Texas]
EC (extended connection)	the border between [Lebanon] and [Israel]
NR (near)	visited [Belmont], near [San Mateo]
DC (discrete connection)	the [well] outside the [house]
PO (partial overlap)	[Russia] and [Asia]
EQ (equality)	[Rochester] and [382044N, 0874941W]

Table 1. Spatial relations in SpatialML

6.2.4. Visual representations

Another possible track is a visual language since geography and cartography are essentially visual. Let us very rapidly present a few visual concepts. But first let us detail contexts of interpretation.

6.2.4.1 Contexts of interpretation

In fact, by considering drawings, four types of interpretation spaces are possible, identified by four interpretational icons (Figure 43):

- **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 planar projections or spherical; It has two alternatives, the planar one (iconized by a square), the global one (circle)
- **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";
- **Timeline** 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 chapter
- **Chorematic space** represented by a hexagon (indeed, as example France is often schematized as a hexagon).

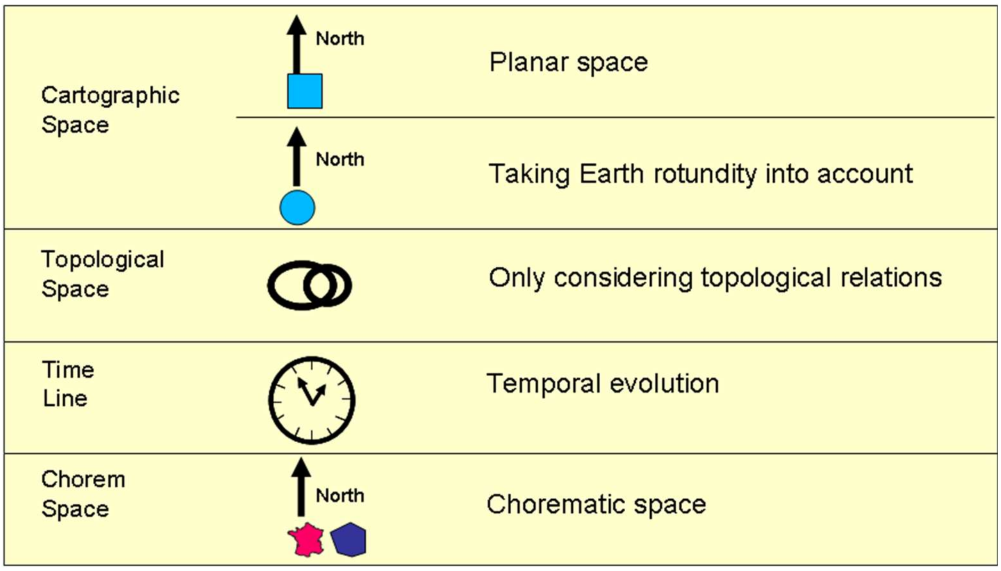

Figure 43. The contexts of interpretation, cartographic space, topological space, timeline and chorematic space.

6.2.4.1. Examples

In those examples, only three elements will be introduced, presentation of facts, presentation of a query and presentation of a co-location rule. Some other examples can be found in (Laurini, 2014). But first let us give some pieces of visual vocabulary in Figure 44 in which in the first row some entity icons are presented whereas in the second row there are emblems of two cities and a flag.

Figure 44. **Excerpt of visual vocabulary.**

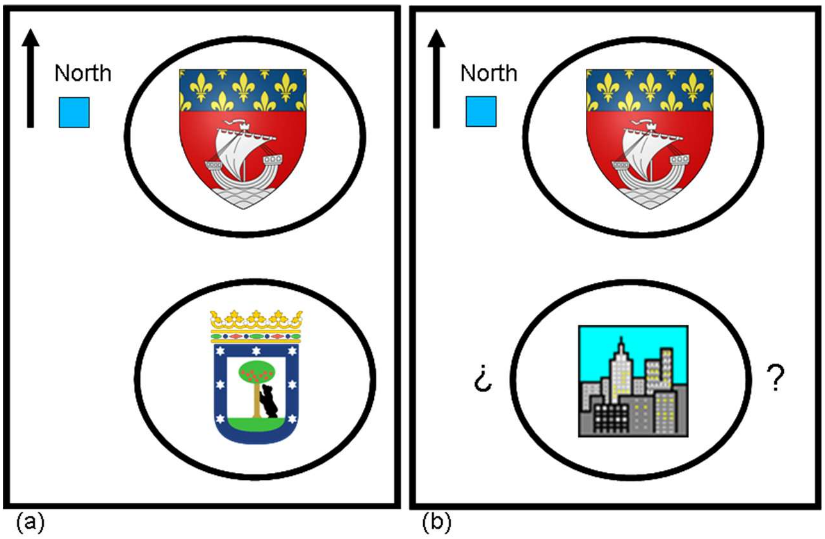

Figure 45. **Example of visual representation in the cartographic space. (a) Fact stating that London is north of Madrid. (b) Query asking for cities located south of London.**

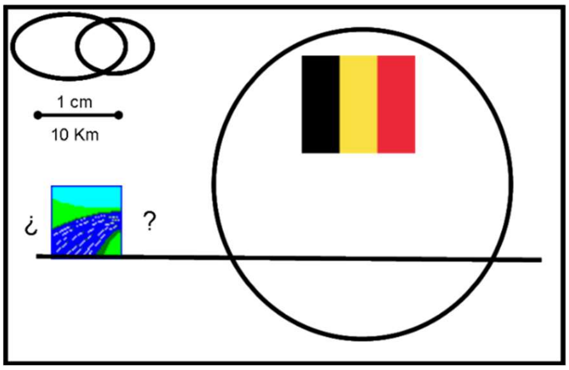

Figure 46. **A topological query to get the list of rivers crossing Belgium.**

7. Discovering geographic knowledge

As stated in Principle#1, geographic knowledge comes from data, not only from vector databases, but also from other various sources such as data mining, Internet, existing maps which are scanned, aerial photos, satellite images, sensors and people. The role of this section is not to detail all those aspects, but only to give some important issues.

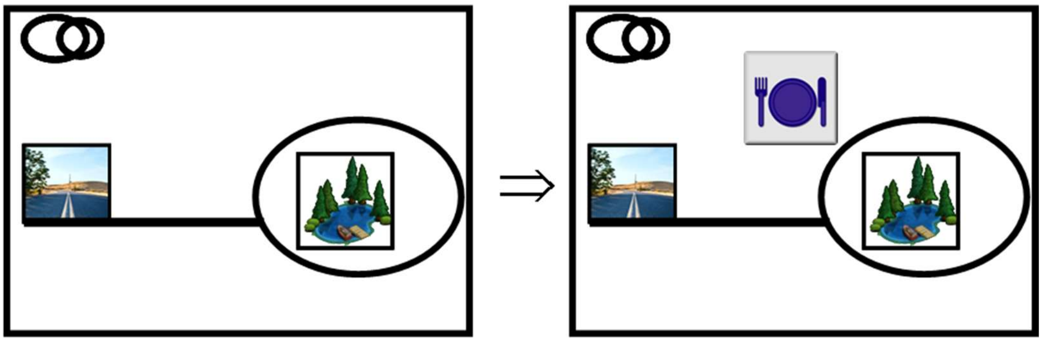

Figure 47. Example of visual representation of co-location rule "*Lake, Road and Restaurant*" in a topological space.

7.1. Geographic data mining

In information technology, the scope of data mining is to analyze database contents so to find frequent itemsets. In other words, this is to extract unknown knowledge from data. The prototype example is taken from a relational database in which it can be shown that if for instance A and B have certain value then C has a precise value according to two indicators, confidence and support. Data mining can be used for anomaly detection, associative rules modeling, clustering, classification and regression.

In our domain, (Mennis-Guo, 2009) data mining can be used for instance for public health services searching for explanations of disease clustering, for environmental agencies assessing the impact of changing land-use patterns on climate change, for police analyzing criminality, etc.

But the main application is called co-location rule discovery in which two neighboring features are usually linked, for example in a lot of places, the town hall is located in the main square.

An interesting direction is the use of sensors. More and more sensors for temperature, noise, etc. are placed in cities. Those sensors send regularly measures to a database against which data mining queries can be addressed in real time against logs.

7.2. Geographic Information Retrieval in Internet

During decades, the key-GIS problem was to search information in geographic databases, rapidly and efficiently. Now the main query is to be launched not against one or several databases, but against the whole Internet. Suppose somebody wants to retrieve some geographic information about Prussia, a country which does not exist anymore. In Internet, various information pieces can be retrieved in various languages, predominantly in German in this case. In addition to textual information, maps and paintings can be retrieved.

This is the role of GIR (Geographic Information Retrieval) to develop methodology able not only to retrieve this kind of information based essentially on gazetteers and ontologies, but also to correctly synthesize information coming from different media.

7.3. Knowledge from scanned maps

In libraries, there are a lot of historical maps which were made for different purposes. Generally, they are accompanied by a legend describing the visual language which is used for the map. Sometimes, there are also dates and other information which can be considered as metadata.

The first step is the analysis of this legend in order to extract patterns in the legend such as illustrated in Figure 47. Based on this vocabulary, the results can be easily encoded to be processed.

In other words, this is a kind of image processing problem driven by well-defined textures. So, several features can be recognized, for instance cities and their sizes, rivers and their beds, roads, etc. For that task, historical gazetteers would be a key-element. As feedback, those gazetteers can be enriched by old placenames which were not already included.

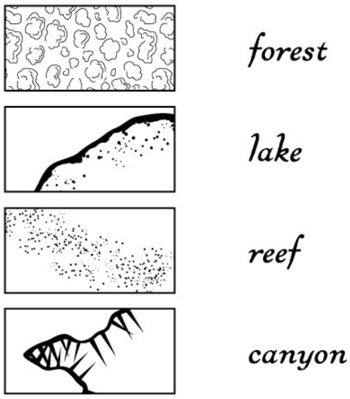

Figure 48. Example of symbols in an old map.
Source: http://a-patel0710-dc.blogspot.fr/2009/12/map-legends-compass.html

7.4. Damaged map recovery

Alas, often the old maps have not resisted the wear of time, they are damaged and they need some recovery. Often, badly made restorations have also deteriorated the quality of information (Figure 48). To recover those maps, the first preliminary steps can be as follows:

- scanning the maps with high density of pixels,
- determining the type of earth projection,
- choosing four points for which we know exactly longitudes and latitudes,
- establishing the formula for rubber-sheeting so that latitude lines be parallel to the x-axis,
- applying rubber sheeting to all pixels
- possibly filling some gaps

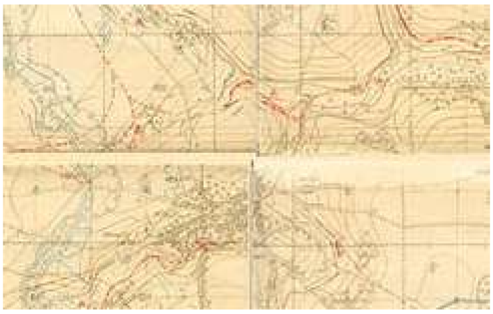

Figure 49. Example of an old, damaged map

In addition, sometimes maps were crumpled, torn and badly glued with Scotch tape: the roads, rivers, etc. are no longer aligned. In this case, a special algorithm must be launched to join all features which were torn or damaged.

After those initial steps, the recognition of features can be launched based on the visual vocabulary.

7.5. Cartography from textual information

Suppose on radio you hear the weather forecast for the whole country: you try to imagine what the meteorological conditions in your city could be. By doing that you are constructing a map from text. Cartography of textual information is based on:

- existing basic maps,
- gazetteers,
- and existing symbology for weather phenomena (rain, snow, temperature, wind, etc.).

In GIR, this is common to try to transform a text either into a map or into some textual knowledge representations.

7.6. From experts and from citizens

In the 80-90ies, the buzzword "expert systems" was mainstream meaning that some so-called experts can deliver rules of their disciplines. But in reality, those experts had difficulties expressing their knowledge as IF-THEN rules. Moreover, it was really difficult to manage an important number of rules especially for checking inconsistencies and completeness.

More recently, citizens were seen as possible sensors, i.e. as providers of geographic information (Goodchild, 2008). With the creation of VGIS (Volunteered GIS), any citizen can bring fresh information and novel knowledge chunks. But this information must often be cross-checked before being stored. Another interesting track is provided by crowdsourcing (Papadapolou-Giaoutzi, 2014) which is a new way to acquire geographic knowledge. Indeed, crowdsourcing is mainly based on the idea of an open-call publication of a problem, requesting the response of the crowd to reach the most appropriate solution.

8. Territorial business intelligence

But the main application of geographic reasoning is targeted to decision-makers for territorial intelligence or more exactly for all those who have some territory to plan and manage. Essentially those decision-makers involve politicians, urban and environmental managers. But also, companies can be interested in geomarketing needs like demarcating territories with best sales. Territorial intelligence can have various facets such as:

- Site selection for commercial or industrial establishments,
- Urban area housing permit and re-zoning decisions,
- Path-finding,
- Emergency vehicle dispatching,
- Delivery truck routing and scheduling,
- Hydrological research like modeling chemical discharges,
- Territorial waters delineation,
- Understanding climatological processes,
- Hazard and disaster prevention, risk management,
- Pollution fighting,
- Disease distribution studies,
- Inferring the geometric shape of strata from borings,
- Proximity to social services design,
- Crime prediction and prevention,
- Positioning of antennas for mobile phones,
- Positioning of sensors and video-cameras,
- Etc.

In this section, not all facets will be presented but only some small applications in territorial intelligence.

8.1. Urban planning

The design of the city of the future is a huge challenge. Urban planners are trying to do their best for centuries and new tools in information technology (Laurini, 2002) are key-elements not only to plan a city, but also as cornerstone of smart cities. In this chapter, only a few examples will be presented as rules.

In strategic urban planning, let us take some examples:

```
Rule 1:  IF a zone is a marshland or floodable
        THEN prohibit construction.
Rule 2:  IF there is unemployment
        THEN raise up enterprise creation
        AND create industrial areas.
Rule 3:  IF a parcel is closed to an airport
        THEN limit building height.
Rule 4:   IF a parcel is near to a fireman station
        THEN prohibit hospital construction.
Rule 5:  IF a building has a good architecture
        AND IF it is more than 100 year old
        AND IF the building state is mediocre
        AND IF the owner does agree
        THEN suggest restoration.
Rule 6:  IF a building falls into ruins
        AND IF nobody dwells in it
        THEN demolish it.
Rule 7:  IF a building has a poor architecture
        AND IF inner rooms are degrading
        AND IF money is raised
        THEN suggest rehabilitation.
```

8.2. Districting for elections

Among the many facets of geographic reasoning, districting is an important real world task. We use the term zoning for the division of a continuous space of land into a tessellation, and the term districting to refer to any regrouping of basic zones into another tessellation. Sometimes the word clustering is used. Generally speaking, districting is undertaken for a special purpose, according to certain rules and criteria to be optimized.

Drawing boundaries is an element common to the situations of a country modifying its electoral precincts, or a municipality creating new school regions. Among the different instances, we examine only the case of electoral districting. In some countries like the USA, boundaries are redrawn every so often in order to match more closely population distributions (Laurini-Thompson, 1993).

We set out now the nature of a spatial reasoning approach to the task of districting an hypothetical state with three hundred elementary zones into ten electoral districts in order to elect ten representatives. For each zone we have data for the number of registered voters, the number of votes cast in prior elections for each party, and the geometric description (Figure 49). We utilize a process beginning with some seed zones, adding neighbors to form the districts and stopping the process when some conditions are met. Assuming that the protection of incumbents is important (to reduce their competition one against another), then we will initiate the process from the zones of incumbent representatives, or a spatially random selection if there is no incumbent.

Starting from this basis, a spatial reasoning mechanism can be invoked to create different sets of districts. These alternative plans will be evaluated against three kinds of criterion. First of all, we employ a **population fairness** criterion. For this element we state that the population of all districts should be approximately the same, possibly within a range of 50,000 plus or minus 5 per cent. Secondly, we require district shapes not to be suspiciously strange. By **shapefairness** we mean that the districts shall not have holes or isolated pieces, and no long tentacles. Thirdly, we require **political fairness**, meaning that no political party is drastically over- or under-represented.

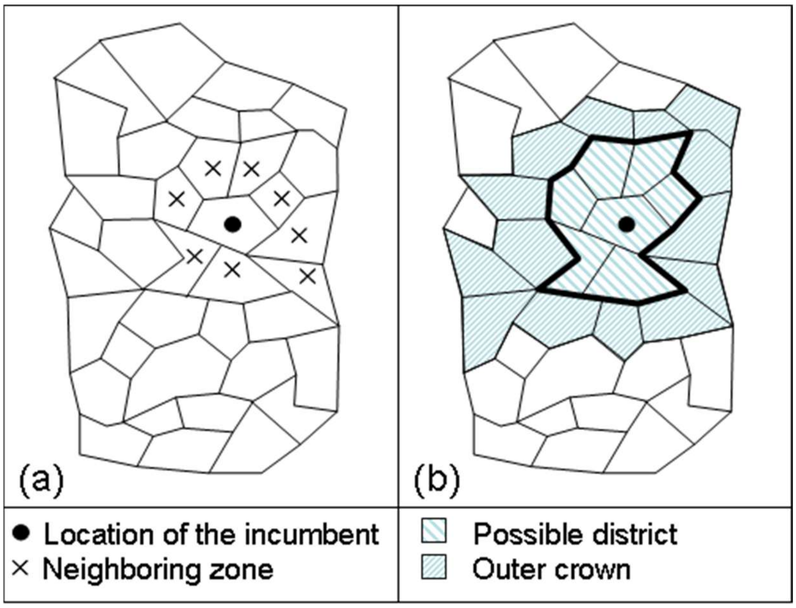

Figure 50. An example for illustrating the districting reasoning procedure. (a) Map emphasizing a seed zone and its neighbors, (b) Map presenting the inner crown zones and the outer crown zones of a possible district.

The population fairness criterion, for one example, can be based on the standard deviation of the number of voters per district.

For the criterion of shapefairness we can optimize any index which favors the shape of a circle. Among several, we use a simple measure comparing the ratio of the squared perimeter to area, which is at a minimum for a circle.

Regarding the political fairness criterion, one possibility is to aim for a result in which for each political party the ratio of elected representatives to the total number will be approximately the same as the proportion of voters by party preference. So, we can define $REPR(d, p)$ as an integer variable, having the value of 1 when the result is that the district gets a representative of the party p, and 0 in the other case. The political fairness criterion may be measured by the difference (or ratio) between the percentage of people voting for each party and the percentage of elected representatives for each party.

That is, with reference to Figure 50, we recognize an inner crown (Figure 50b), a set of zones bordering a currently active grouping as legitimate candidates to be added. There also exists a set of zones bordering a completed district, called here the outer crown. Such zones are candidates for swapping with one or more inner crown zones as adjustments are made to a district to make its population total closer to the desired number. While the map shows the possibility of swapping neighbors in the inner and outer crowns, the process can be set up to allow swapping of non-adjacent zones. For optimizing the political fairness criterion, we must determine globally the state of under- or over-representation for one or more political parties.

Thus, when a solution is reached, it corresponds to a new plan, which can then be compared with alternatives in order to make a selection.

8.3. Crime analysis and prevention

In several cities, criminality is a main problem. Police usually record criminality events, and statistics tend to show repetitions of those events. So, both criminal analysis and prevention play an important role in those cities.

Analysis is not only based on mapping on recent crimes (see Figure 51), but overall, on the discovery of causes leading to an increase of criminality. Geographic data mining can help a lot.

Prevention can have to facets. In one hand people can secure their homes and shops and create some community watch groups. In the other hand police organize patrols and police beats which can be more or less frequent.

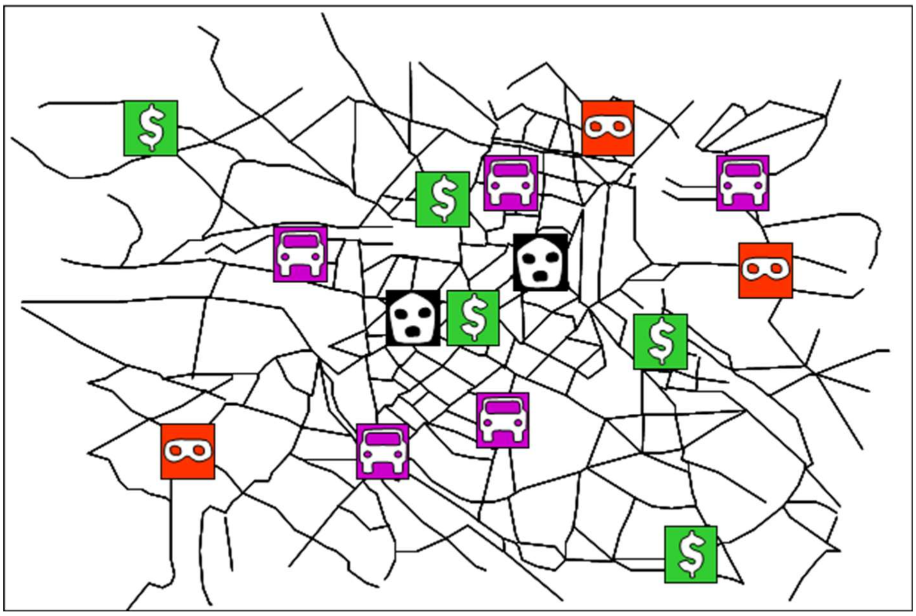

Figure 48. Example of a map for crime analysis.

The organization of those beats can be based on geographic knowledge (location of recent crimes, new alerts, etc.), time knowledge (night/day) and other knowledge essentially based on the availability of police manpower.

9. Conclusions and perspectives

Our objective was to give not only the fundamentals of geographic knowledge engineering, but also to show that it stands at the crossroads of several disciplines including computational geometry, topology, social sciences, cognitive sciences and so on.

Territorial intelligence must mix knowledge from people and automatic reasoning in order to plan the city of the future.

Good tools to tackle this kind of knowledge must not only model reality with a nice precision but overall be robust against measuring errors and scaling effects.

By the theory of ribbons, a kind of bridge has been done between lines and areas allowing a more efficient concept for modeling roads and rivers. Moreover, this concept could be of interest for a sort of grammar for urban features.

Thanks to twelve prolegomena and twelve principles, the fundamentals of geographic knowledge engineering have been proposed to build the foundations of a general framework robust to errors and scaling.

Several applications were rapidly presented in order to show the panorama of possibilities for geographic knowledge engineering.

But for designing software products able to deal efficiently with automatic geographic reasoning, this is a long way to go. Anyhow, the following issues must be clarified:

- How to introduce easily computational geometry and spatial analysis in such a framework?
- How to organize a geographic knowledge base in order to guarantee rapid access to geographic information? Will spatial indexing (for instance based on r-trees) be useful to achieve efficiency?
- What kind of intelligence for smart cities and how to integrate human intelligence and artificial intelligence?
- How to fusion knowledge from experts, knowledge from citizens and knowledge issued from geographic data mining?
- Will geographic reasoning increase or diminish citizen empowerment in a context of territorial intelligence?
- How to introduce criteria such as for sustainable development?
- How to introduce visual representations and visual interfaces?
- How to combine geographic automatic reasoning and visual reasoning from geovisualization?
- How to integrate such consideration in cloud computing?
- How to ensure interoperability of geographic knowledge base?
- Etc.

To conclude this chapter, let me thanks Dr. Françoise Milleret-Raffort for her help in discussing various aspects of this research.

References

[1] Allen J.F. (1983) Maintaining knowledge about temporal intervals. In: *Communications of the ACM*. 26 November 1983. ACM Press. pp. 832–843, ISSN 0001-0782.

[2] Bertin J. (1967) *Sémiologie Graphique. Les diagrammes, les réseaux, les cartes*. Paris: Gauthier-Villars. (Translation 1983. Semiology of Graphics by William J. Berg).

[3] Brunet R. (1980) La composition des modèles dans l'analyse spatiale, in *L'Espace géographique*, n° 4, 1980.

[4] Clementini E., Di Felice P. & van Oosterom P. (1993). A small set of formal topological relationships suitable for end-user interaction. In Abel, David & Ooi, Beng Chin. Advances in Spatial Databases: Third International Symposium, SSD '93 Singapore, June 23–25, 1993 Proceedings. *Lecture Notes in Computer Science*. 692/1993. Springer. pp. 277–295.

[5] Cohn A. & N. Gotts. (1996) The "Egg-Yolk" Representation of Regions with Indeterminate Boundaries. In: P. Burrough and A. Frank (Eds.) *Geographic Objects with Indeterminate Boundaries*, pp. 171-187. Taylor & Francis.

[6] Coimbra A. (2008). *ChorML: XML Extension for Modeling Visual Summaries of Geographic Databases Based on Chorems*, Master Dissertation, INSA-Lyon.

[7] Del Fatto V., Laurini R., Lopez K., Loreto R., Milleret-Raffort F., Sebillo M., Sol-Martinez D. & Vitiello G. (2007) Potentialities of Chorems as Visual Summaries of Spatial Databases Contents, VISUAL 2007, 9th Int'l Conf. on Visual Information Systems, Shanghai, China, 28-29 June 2007. Edited by Qiu, G., Leung, C., Xue, X.-Y. & Laurini, R., Springer Verlag LNCS, Volume 4781 *Advances in Visual Information Systems*, pp. 537-548.

[8] Douglas, D.H. & Peucker, T.K. (1973) Algorithms for the reduction of the number of points required to represent a digitised line or its caricature, *The Canadian Cartographer* 10 (2) pp. 112-122.

[9] Egenhofer M. & R.D.Franzosa R.D. (1991) Point-set topological spatial relations", *International Journal of GIS*, vol.5, no.2, pp. 161-174.

[10] Egenhofer M. (1994) Deriving the Composition of Binary Topological Relations. *Journal of Visual Languages and Computing* 5(2), pp. 133-149.

[11] Goodchild, M.F. (2007). Citizens as sensors: the world of volunteered geography. *GeoJournal* 69 (4) pp. 211–221.

[12] Goodchild M. F. & Hill L. L. (2008). Introduction to digital gazetteer research. *International Journal of Geographical Information Science,* 22(10), pp. 1039-1044.

[13] Gordillo S. (2001) *Modélisation et manipulation de phénomènes continus spatio-temporels*. PhD dissertation, October 12, 2001. Université Claude Bernard Lyon I.

[14] Gruber T. R. (1993) A translation approach to portable ontologies. *Knowledge Acquisition,* 5(2) pp. 199-220.

[15] Guarino, N. (1998). Formal Ontology and Information Systems. In N. Guarino (Ed.), *Formal Ontology in Information Systems* (pp. 3-15). Amsterdam, Netherlands: IOS Press.

[16] Hećimović Z. & Ciceli T. (2013) Spatial Intelligence and Toponyms, *Proceedings of the 26th International Cartographic Conference*, Dresden, Germany, 25–30 August 2013, Edited by Manfred F. Buchroithner, ISBN: 978-1-907075-06-3..

[17] Jakir Ž, Hećimović Ž, & Štefan Z. (2011) Place Names Ontologies. In Ruas A (Ed.) Advances in Cartography. *Lecture Notes in Geoinformation and Cartography*, Springer Heidelberg, pp. 331-349.

[18] Kang M.A., Faloutsos C. Laurini R. & Servigne S. (2002) 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

[19] Kavouras M., Kokla M. & Tomai E. (2005) Comparing categories among geographic ontologies, *Computers & Geosciences* (2005) Volume: 31, Issue: 2, Pages: 145-154 ISSN: 00983004.

[20] Keßler C., Janowicz K. & Bishr M. (2009) An agenda for the next generation gazetteer: geographic information contribution and retrieval. Proceedings of the 17th ACM SIGSPATIAL International *Conference on Advances in Geographic Information Systems*, pp. 91-100, ACM New York, ISBN: 978-1-60558-649-6.

[21] Laurini R. (1998) Spatial Multidabase Topological Continuity and Indexing: a Step towards Seamless GIS Data Interoperability. *International Journal of Geographical Information Sciences*. Vol. 12, 4, June 1998, pp. 373-402.

[22] Laurini R. (2012) Importance of spatial relationships for geographic ontologies. In *Planning Support Tools: Policy Analysis, Implementation and Evaluation* Proceedings of the Seventh Int'al Conf. on Informatics and Urban and Regional Planning INPUT 2012 (Editors: M. Campagna, A. De Montis, F. Isola, S. Lai, C. Pira & C. Zoppi) (ISBN code: 9788856875973), pp. 122-134.

[23] Laurini R. (2014) A conceptual framework for geographic knowledge engineering, *Journal of Visual Languages and Computing* 25 (2014) pp. 2–19.

[24] Laurini R. & Milleret-Raffort F. (1989) *L'ingénierie des connaissances spatiale*s. Paris, Hermès, Collection Technologies de pointe, n° 30, 64 p.

[25] Laurini R. & Thompson D. (1993) *Fundamentals of Spatial Information System*s. Academic Press, (A.P.I.C. Series, No 37), March 1993.

[26] Lee S.-Y. & Hsu F.-J., (1990) 2D C-string: A new Spatial Knowledge Representation for Image Database Systems, *Pattern Recognition*, vol. 23, no 10, pp. 1077-1087.

[27] Lee S.-Y. & Hsu F.-J., (1992) Spatial Reasoning and Similarity Retrieval of Images Using 2D C-string Knowledge Representation, *Pattern Recognition*, Vol. 25, No 3, pp. 305-318.

[28] Mani I., Doran C., Harris D., HitzemanJ., Quimby R., Richer J., Wellner B., Mardis S. & Clancy S. (2010) 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

[29] Mennis J. & Guo D. (2009) Spatial data mining and geographic knowledge discovery – An introduction, *Computers, Environment and Urban Systems*, Volume 33 (6), November 2009, pp. 403-408.

[30] Papadopoulou C.A. and Maria Giaoutzi M. (2014) Crowdsourcing as a Tool for Knowledge Acquisition in Spatial Planning. *Future Internet* 2014, 6, 109-125; doi:10.3390/fi6010109

[31] Popper K. (1934) *The Logic of Scientific Discovery*, (Logik der Forschung, English trans. 1959), ISBN 0-415-27844-9.

[32] Preparata F. & Shamos M. (1985) *Computational geometry: an introduction*. Springer-Verlag 1985, ISBN: 0-387-96131-3.

[33] Randell D.A., Zhan C. & Cohn A.G. (1992) A Spatial Logic based on Regions and Connection, In Proceedings 3rd Int'l *Conference on Knowledge Representation and Reasoning,* Cambridge, MA (Morgan Kaufmann), pp. 165-176.

[34] Smart, P.D., Jones C.B. & Twaroch F.A. (2010) Multi-Source Toponym Data Integration and Mediation for a Meta-Gazetteer Service, In *Proceedings of GIScience* 2010, Zurich, Switzerland, Lecture Notes In Computer Science 6292, pp. 234-248.

[35] Smith B. & Mark D. Do Mountains Exist? Towards an Ontology of Landforms. *Environment and Planning B*, 2003, 30(3), pp. 411-427.

[36] Thomas J.J. & Cook K. A. (2005) *Illuminating the Path: The R&D Agenda for Visual Analytics*. US National Visualization and Analytics Centre, 2005.

[37] Tobler W. (1970) A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography*, 46(2) pp. 234-240.

[38] Vckovski A. (1995) Representation of Continuous Fields, *Proceedings of 12th AutoCarto,* Vol.4, pp. 127-136, 1995.

[39] Zadeh L. (1995) Fuzzy Sets, *Information and Control*, vol. 8, pp. 338-353.