Nature of Geographic Knowledge Bases

Robert Laurini
INSA-Lyon, France

ABSTRACT

In many domains such as environmental and urban planning, experts need to make reasoning and propose solutions. However marketed GIS software products are limited to store, display geographic information together with additional tools such as in spatial analysis, but they do not offer users the real functionalities which are useful for territorial intelligence. This first step is to propose novel models to represent this kind of knowledge needing not only to integrate geographic aspects, but also be independent of data acquisition technologies (satellite images, laser, crowdsourcing, etc.) and able to be used in different languages. After the definitions of geographic ontologies (to organize geographic feature vocabulary) and gazetteers (to structure toponyms in various languages), various examples will be presented in order to extract geographic semantics. A special attention will be devoted to geographic rules.

Keywords: territorial intelligence, geographic knowledge, reasoning, geographic rules.

INTRODUCTION

In comparison with information systems in companies, gradually passing to business intelligence, it could be interesting to provide local administrations and local politicians with tools for territorial intelligence. Territorial intelligence, sometimes called geographic intelligence combines human collective intelligence with computer intelligence to reach sustainable development for any territory, a country, a region or a city. In the case of cities, the expression "smart cities" is very common.

There are several definitions for territorial intelligence. Girardot (2010) claims that "Territorial intelligence is the science having for object the sustainable development of territories and having for subjects the territorial communities" and then emphasizing the importance of Multidisciplinary knowledge. For its part, Bertacchini (2007) explains that "Territorial Intelligence can be compared with the territoriality which results from the phenomenon of appropriation of resources of a territory; it consists in know-how transmissions between categories of local actors of different cultures." Miedes Ugarte (2008) declares that there are three components in territorial intelligence, cognitive, socio-political and organizational-technological. The last includes new tool kit for analysis, monitoring and territorial communication. As a sort of synthesis, let me define territorial intelligence as the composition of collective human intelligence allied to artificial intelligence for sustainable development as illustrated Figure 1.

Figure 1. Definition of territorial intelligence

Presently, even if GIS tools are very used and efficient, few experiences have been made to use geographic knowledge for urban and regional planning. One of the reasons is the misfit of geographic data and information: they need to be revisited to able to use in automatic reasoning. Among the criteria, let me mention:

- independence from data acquisition techniques,
- independence vis-à-vis scales,
- independence from computer data representation,
- independence from human languages,

- robustness vis-à-vis errors,
- new visualization techniques more adapted to decision-makers by offering synthetic overviews in contrast with data atlases,
- etc.

Consequently, in this chapter, in order to design new tools for territorial intelligence, it is necessary to propose a generic model for geographic knowledge. But before presented a consistent and exhaustive model, the semantics of geographic objects and of geographic rules must be identified and categorized. The targeted applications deal with urban and regional planning, environmental analysis and planning, transportation and more generally all applications for which a geographic information system (GIS) is used can be revisited to see whether geographic knowledge can lead to better tools.

In other words, the goal of this chapter will be to explore the nature of geographic knowledge so to propose novel models to be efficiently used by computers.

BACKGROUND

Few research laboratories include geographic knowledge into their guidelines. People working in geographic data mining as (Mennis-Guo 2009) are only trying to extract so-called association rules in which co-localization rules are a good representative.

Several works have been in spatial knowledge, essentially based on topology (Egenhofer 1991, 1994, Clementini 1993) especially as a basis for topological reasoning. But generally, extensions to deal with geographic reasoning are very limited.

In previous papers, the author and some co-authors have examined some aspects such as the mutation of topological rules (Laurini 2014), the linguistic aspects of geographic ontologies and gazetteers (Laurini 2015).

MAIN FOCUS OF THE CHAPTER

In this chapter, geographic objects, ontologies and gazetteers will be revisited in order to be efficiently for geographic knowledge. And then after, several considerations for modeling geographic rules will be detail and for designing a new type of inference engine devoted to geographic applications. But before doing so, it looks necessary to examine a few linguistic aspects in geography.

ABOUT LANGUAGES AND GEOGRAPHY

Several people are saying something like "developing geographic knowledge in a unique language (say English) and after trying to translate it into other languages". But the reality is different. Consider a project about the Danube River. There will be databases in German (Germany and Austria), in Serb, in Romanian and possibly other languages. And those databases will be the basis for geographic knowledge.

It is certain that all geographic projects do not include the necessity to deal with different languages. But the idea to design geographic knowledge in one language which must be valid everywhere is not realistic. A few aspects will be examined in this section.

The well-known English term "*bank*" represents both a riverside and a financial institution. In other words, the first meaning will be translated in French and Spanish, respectively, as "*rive*", "*ribera*" and the second by "*banque*" and "*banco*".

Let us examine a special case: in the French language, the word "*quai*" defines a wharf, an embankment, a train platform or a street along a river. In Spanish, especially in Barcelona, "*rambla*" is a ravine or a special kind of broad avenue. In Venice, "*rioterà*" is a special type of pedestrian lane, whereas other denominations are used, such as *salizada*, *sottoportego*, *ramo*, *fondamenta*, *campiello*, *corte*, *calle*, *riva*, etc. As far as I know, those terms have no equivalent in English.

Therefore when speaking about geographic features and then knowledge, a possibility is to design a geographic knowledge base only in English, and then translating it into various languages. But by doing so, an impoverishment will be a natural consequence which can lead to not only difficulties of reasoning but potential errors.

ISO 639 is a set of international standards that lists short codes for language names (See http://www.iso.org/iso/home/standards/language_codes.htm?=). The following is a complete list of three-letter codes defined in part two (ISO 639-2) of the standard, including the corresponding two-letter (ISO 639-1) codes where they exist. In this chapter, we will use the three-letter codes as the prefix (ENG for English, FRE for French, ITA for Italian, SPA for Spanish, GER for German, GRE for Greek, RUS for Russian, ARA for Arabic, etc.). Therefore, for the city of Venice, we can distinguish various placenames such as: ITA.Venezia, SPA.Venecia, FRE.Venise, ENG.Venice, GER.Venedig, POL.Wenecja, GRE.Βενετία, RUS.Венеция, ARA.البندقيــــة (transliterated into Al Bundukiyya or Al Bondokia), etc.

GEOGRAPHIC FEATURES AND 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".

Representation of geographic features

A very important aspect when modeling a geographic feature 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 river, several models exist depending from actor's vision. Figure 2 illustrates four families of models totally different:

- in the first model, the river is reduced to nothing;
- the second based on one polyline corresponding to the axis of the river;
- the third as two lines describing each bank of the river;
- and finally another model giving both the minor bed and the major bed of the river.

Figure 2. Multiple geometric representations of a river

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 of 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.

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, 2014), 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 or clothoid portions), 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 need. From a mathematical point of view, a ribbon can be defined as a transformation of a longish rectangle. Figure 3 illustrates this principle.

Figure 3. Various types of ribbons. (a) Rectangular ribbon. (b) Ribbon. (c) Loose ribbon

Various types of 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 rl ($rl= l/w$) the longishness ratio which is supposed to be much greater than a positive value rL; a possible minimum value is 5 ($rl > rL > 5$).

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.

However due to error measurements, the real ribbons are not exactly rectangular. Let us denote them loose ribbon.

For some categories of features such as mountains or deserts, fuzzy models can be used to model areas. In addition, in order to represent rivers with their minor and major bed, let us define **fuzzy ribbons** when the sides vary. Considering the egg-yolk model proposed by Cohn-Gotts (1996), the ribbon yolk will model the minor bed whereas the ribbon egg, the major bed as exemplified in Figure 4. Remember that the major bed can correspond to a 100-year flood.

Figure 4. Modeling geographic objects by fuzzy sets. (a) A geographic object of a fuzzy area with membership degrees. (b) Limitation at 70% and 30%. (c) The egg-yolk representation of this area. (d) Application of the egg-yolk model to a ribbon

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.

Topological relations

Topological relations such as at 1D, interval relations (Allen 1983) and at 2D Egenhofer relations (Egenhofer, 1991, 1994) are well known (Figure 5).

Figure 5. 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, ∂A the outer part and ¬A its boundary (Figure 6). The answer is given by the following matrix in which the result of one intersection can be void or not void.

Figure 6. 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

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

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 7 presents a TOUCHES relation between a river and the sea.

Figure 7. There is a TOUCHES topological relation between river and sea, corresponding to the estuary

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.

Other spatial relations

But other relations exist such as projective (or cardinal such as North/South, East/west) (see Figure 8) relations and distance (near/far) relations.

Figure 8. Projective relations

Ribbon relations

As previously told, 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.
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 9). Of course, those operators can be recursively used.

Figure 9. Two ribbon operators, longitudinal splitting and lateral splitting. (a) Definitions. (b) A more complex decomposition

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 10 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 10. 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.

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 11.

Figure 11. Ribbon orientation

Chaining ribbons

To model roads and rivers, usually polylines are used to describe the axis; but sometimes two polylines can be used to model river banks 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 12 gives an example (Figure 12a) of several ribbons transformed into a chain of ribbons (Figure 12b). In Figure 12c, a case is presented needing two additional curves to join the sides of two ribbons

Figure 12. 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.

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 if and only if (See Figure 13a):

- 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 13b). 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 us call them "loose tessellation".

Figure 13. A tessellation with sliver polygons and a good standing tessellation

But the Earth is not flat and all relations must be revisited taken rotundity into account.

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).

Geographic ontologies

However, the specificity of geographic ontology does not lie only on geographic features (as illustrated in Figure 14) (Kavouras et al. (2005). But it lies overall on their geometry and on their spatial relationships

(Laurini, 2012). Usually Egenhofer or RCC relations are fully integrated in the definitions of geographic features. See Figures 15 and 16 for such examples, the first for the planet, and the second for administrative subdivisions of a country.

Figure 14. Example of the beginning of a geographic ontology only with is-a relations

Figure 15. Example of ontology based on spatial relations

Figure 16. Example of administrative subdivisions with spatial relations

TOPONYMS AND GAZETTEERS

There are various solutions to identify places. The solution taken in geographic databases is to confer each place an ID which is generally unknown by users. In the case of geographic knowledge, it could be interesting to keep place names as they are used by politicians and lay-people (Goodchild-Hill 2007). Whereas this solution is very interesting for users, there are lots of difficulties to use toponyms in knowledge representation. In this section, several characteristics will be examined.

What is beneath a name?

Beneath a geographic name, various objects or features can exist. On the Earth, few points have names, perhaps only the North and South Poles, and only a few lines, such as the 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 or may be modeled as lines or ribbons (Laurini 2013). As previously mentioned, they must be considered as simply connected (with islands and holes), and they can be replaced by their centroids for some operations. In some geographic databases, for instance, the geographic object named "Italy" can include Vatican and San Marino, whereas those places do not officially belong to the country named Italy.

Generalities

Indeed, in addition to a pure list of place names, 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 a few well-known examples:

- "Mississippi" can be the name of a river or of a state.
- The city, "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 modern city of Rome is much bigger than in Romulus's 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.

- 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 is feminine in German, and masculine in some 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.
- In the U.K., there are several rivers named Avon.
- Some place names 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 simplifications; the well-known Welsh town "Llanfairpwllgwyngyllgogerychwyrndrobwllllantysiliogogogoch" is often simplified to "Llanfair PG" or "Llanfairpwll" (See Figure 17).
- Some abbreviations can be common, such as "L.A." for "Los Angeles", whereas its name at its inception was "El Pueblo de Nuestra Señora la Reina de los Angeles del Rio de la Porciúncula";
- Peking became Beijing after a change of transcription to the Roman alphabet; but the capital of China has not modified its name in Chinese.
- In some languages, grammatical gender is important, so that place names can be feminine or masculine; for instance, in French, Italian and Spanish, names such as "Japan", "Brazil" and "Portugal" are masculine, whereas "Argentina", "Bolivia" and "Tunisia" are feminine.
- In addition, as the great majority of toponyms are singular, some can be plural, like "The Alps"; but for "The Netherlands", the situation is more complex: plural in French (Les Pays-Bas), in Italian (I Paesi Bassi) and in Spanish (Los Países Bajos), whereas singular and plural are both acceptable in English (The Netherlands are, The Netherlands is);
- Some places have nicknames; e.g., Dixieland, Big Apple, City of the Lights, etc.
- Do not forget that in some languages, toponyms can have declensions, for instance for the Rhine River in German (der Rhein, des Rheins, etc.).

Figure 17. What to do with such a place name in territorial intelligence?

Consider now the toponym "Granada": there are places in practically all Spanish-speaking countries bearing this name:

- a small country located in the Caribbean Islands;
- in Spain, a city capital of the eponymous province, a few other places located in Barcelona and Huelva provinces and a river in the Vizcaya province;

- in Colombia, three cities with this toponym;
- in Mexico, a city in Yucatán;
- in Nicaragua, a city capital of the eponymous department;
- in Peru, a district.

In addition, in the U.S., there are cities named Granada in California, Colorado, Kansas, Minnesota, Mississippi, etc.;

As a consequence, there is a very complex many-to-many relationship between places and place names (Figure 18).

Figure 18. Very complex many-to-many relations link places and their names.

Among place names, there are street names together with the number in the street (civic number); these are not so easy to handle. This is very important, not only for the automatic processing of postal addresses, but also for all applications connected to an emergency. The Urban and Regional Information Systems Association (URISA) association has organized many conferences on the topic (see http://www.urisa.org/). The specificities of street names are as follows:

- some streets comprise a few dozens of yards, whereas others 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 commonly simplified into "Ave" and "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";

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;
- endonym is a local name in the official language of the country or in a well-established language occurring in that area where the feature is located; there may be several toponyms in countries with different official languages (Brussel in Flemish, Bruxelles in French);
- exonym is a name in languages other than the official languages; for instance Brussels in English;
- archeonym is a name that existed in the past: for instance, Byzantium for Istanbul;
- hyperonym and hyponym are the names of places with a hierarchy; hyponym is the opposite of hyperonym; for instance, Europe is a hyperonym of France, whereas France is a hyponym
- of Europe;
- meronym is a name of a part of a place without a hierarchy; sometimes the expression partonym is used; for instance "Adriatic Sea" is a meronym of the Mediterranean Sea;
- hydronym is a name of a waterbody;

- oronym is a name for a hill or a mountain;

In Jakir *et al.* (2011), the essential elements of a gazetteer, the names, the features, the dates and everything regarding geometry and georeferencing are given.

In addition, places, such as airports, can have several names. Sometimes, their International Air Transport Association (IATA, http://www.iata.org/) codes are used: the well-known New York airport, John F. Kennedy International Airport, is often referred to as JFK airport. Zip codes or postcodes can also be considered as toponyms. However, the definition of postcodes differs according to country: In some cases, one postcode can correspond to a few houses, and in others some hundred thousand inhabitants.

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

GEOSGRAPHIC RULES

According to Graham (2006) and Morgan (2008), rules or business rules must be considered as first-class citizen in information technology. In enterprises, many applications, such business intelligence are using rules. Systems such as SAP are based on a declarative formalism: i.e. for the description of the tasks to be done, are not given algorithms, but lists of rules, such as "car insurance does not cover drivers who have been convicted of driving drunkenness in the last two years", or" when you send your monthly bills, marketing materials that match the customer profile should be included ".

In this chapter, a rule is not necessarily a juridical norm, but only an inference from things or phenomena whose origin can be physical, statistical or best practice type or from mining. Mathematically it is written as: $A \Rightarrow B$.

In fact, the SAP software is the ERP (Enterprise Resource Planning) and appears as integrated management systems. These systems are used to manage all information and knowledge relevant to a company or an organization, allowing data access management to all functions of the company in an integrated, modular and scalable.

The ESRI Company has decided to connect GIS systems with systems such as SAP. But their purpose is only to consider administrative rules related to the area, such as the licensing of building public works. Figure 19 shows the possible connections between a GIS and a SAP system.

Figure 19. Connecting a GIS to a SAP system. (Source: http://www.esri.com/news/arcnews/spring09articles/integrating-gis.html)

But, beyond such rules in Government, there are other rules that govern a territory. Consider some examples:

- in the United Kingdom, driving on the left,
- in Canada, the majority of the population lives along the border with the United States,

- each capital city has an international airport nearby,
- between the two capitals, generally there are direct flights,
- the more one proceeds northward, the colder (but locally is not true),
- the higher you climb a mountain, the colder,
- large rain upstream, flooding downstream,
- mosques are oriented towards Mecca,
- if a zone is a swamp, it is necessary to prohibit construction,
- if there is unemployment, the creation of enterprises and/or industrial areas must increase ,
- if some land is adjacent to an airport, there is a need to limit the height of the building,
- do not open a new pharmacy in less than 500 meters from another existing,
- a best practice in America is to use a bus to go from one city to another city.
- in the US, when 3 or more people are in a car, they can use the HOV lane (High Occupancy Vehicle) in several highways.

In the domain of geographical knowledge (Laurini 2014, 2015), it is not only important to refer to the study of objects and their spatial relationships, but one must also consider the rules to build geographic intelligence systems which can be useful to local decision-makers.

Although there are several works on business rules, virtually nothing has been done about geographic rules. We only know jobs as those of the book edited by Kim et al. (1980) on expert systems for urban planning in which they were presented some spatial rules: but during that time, the domain was not totally mature. Without forgetting the Golledge's paper (Golledge 2002) about the nature of geographic knowledge.

However, there are many jobs on the accumulation of knowledge management, especially for the business. The paper by Ross (2011) explores the IF-THEN form to represent rules with two alternatives analysis IF-THEN-FACT and IF-THEN-ACTION. If the knowledge base is composed only of facts and rules, then with the IF-THEN-FACT form, you can derive new facts; but, in general, knowledge does not only reduce to facts. For the other form, the idea is to propose actions, i.e. when a certain context holds then one or several actions can or must be launched. Sometimes the before-issue is call “antecedent” whereas the after-issue, “consequent”.

Mathematical tools

The mathematical modeling of the classical rules of business intelligence is based on first-order logic, descriptive logic and predicates. But for geographic rules, we must consider and then integrate other fields of mathematics such as flat and spherical geometry, topology, statistics. The consequences include:

- the creation of mathematical objects that can extend logic,
- but above all creating inference engines able to run these functions.

Fuzzy logic can also be of interest to model geographic rules. Three extensions may be defined:

1 – when dealing with fuzzy geographic objects and classical logic, the antecedent and consequent are fuzzy;
2 – when dealing with crisp objects, the implication is fuzzy, style (IF-OFTEN THEN);

3 – both objects and implication are fuzzy.

Mathematical formulas applied to some geographic problems can also be considered as rules.

Analysis of certain geographic rules

Now, let's investigate some rules in various areas.

Geodetic rules
Rules of this type are valid everywhere on the globe; consider the cardinal points. If you are North of B and if B lies to the North of C, then A is in the North of C. So transitivity of this rule applied; but there is nothing in the North of the North Pole, then the transitivity is limited. A similar rule can be written for the South. But in the East and the West directions, transitivity is partial because of the rotundity of the Earth. If Rome is to the East of Los Angeles, and if Los Angeles is East of Peking, the consequence is Rome is in the East of Beijing, but to truth is towards West. The rule can be written like this:

```
IF East (A, B) and East (B,C) and Longitud (A,C) < 180°
THEN East (A,C)
ELSE West (A,C)
```

Rules of physical geography
In this domain, the rules should represent natural phenomena and their implications. For example as a result of the tsunami, volcanic eruptions, storms and heavy rains, you may consider some automatic consequences. But in addition, you must consider more recently the preventive, protection or mitigation, as well as the existence of effective monitoring systems in real time.

But because of the local topography, certain rules can be invalidated. In the northern hemisphere, the more you go north, the colder. To a certain scale, this type of rule is valid, but there are places where this reasoning is no longer valid. So we have to distinguish between local and global rules. Here the local rule supersedes the global rule. Same assertions hold, for example, also the height above sea level.

The rules of the spatial distribution of flora and fauna, hydrology, etc. derived from climatology, meteorology fall into this category. Figure 20 gives an example in the Alps.

Figure 20. Spatial distribution of flora in the Alps. (Source: https://www.jardinalpindulautaret.fr/garden/exceptional-natural-environment/zoning-plant-life-mountains

Suppose that we are on the South side of the mountain. Two possibilities to write a rule:

• first case: `IF birches THEN 1100 < altitude < 1800`;
• second case: `IF 1100 < altitude < 1800 1100 THEN birches`.

The first case would be in a study that links a kind of trees and the altitude likely where they are to while the second indicates that, since we are at some level of the sea, which are the trees that we can meet.

Rules derived from laws
Each country has its own rules, not only administrative but also geolocated. Analyzing aerial photos of traffic is important to know that in the United Kingdom one drives on the left. Concepts such as language, local currency can impact geographic rules.

Generally in all countries, there are a Constitution and many laws governing geographic aspects as for urban and regional planning. Let us take the example of pharmacies. Figure 21 gives an illustration of this rule.

Figure 21. Example of administrative rule: "It is prohibited to open a new pharmacy within 500 meters of another existing one".

Although international standards exist for the Highway Code, every country has its peculiarities such as precedence in intersections, roundabouts, etc. See HOV above-mentioned example.

Socio-economic rules
Demographics with its great inertia generate rules. Example: "more children, more schools."
A majority of countries have devised rules for the organization of the economy and enterprises. These rules have a significant impact on land use. Figure 22 gives an illustration of the rule "On the seashores, the larger the distance from the sea, the lower housing prices".

Figure 22. Example of economic rule: "On the seashores, the larger the distance from the sea, the lower housing prices"

Rules of urban planning
Generally, in each country there are laws governing urban planning. We will take a small sample for building permits, as presented in Figure 23 where one can see a building that must adhere to certain conditions.

In this Figure 23, condition#1 gives the minimum distance from the road, condition#2, the maximum height of the building, condition#3 the volume of the building, condition#4 the distance on the back of the lot, and condition#5 the distance from neighbors. Here, the conditions will be treated as constraints, i.e. to be valid (in the sense of the rule), the building must comply with these constraints.

Figure 23. Examples of rules of town planning that limit the volume of buildings

From an explanatory point of view, it can be interesting to split those conditions into separate rules. Indeed, if a construction is not compliant to some condition, the user will be informed about the exact condition (f.i. height condition).

Cross-border rules
A geographic knowledge base system must not only store knowledge within the jurisdiction of its owner, but also outside. Indeed, in some case, the cause of a phenomenon can be outside the jurisdiction. Take for instance the Swiss city of Geneva in which a lot of people working in Geneva are living in France; indeed, wages are higher in Switzerland, but living conditions are more interesting in France. In some other cases, some goods are cheaper across the borders for several reasons among which taxes.

More generally, cross-border rules can hold in places such as harbors, airports and any passing points for entering a territory.

Association rules extracted from spatial data mining
In the recent years, a great evolution is related to the potential of data mining in databases: the aim is by using the image of miners, to extract not gold but frequent items, i.e. things that often go together. These frequent items between various things can be seen as rules, but with a particular semantics. For example, "in the center of the village, there are church and town hall", "near a big city, there is an airport".

These links are known as association rules generally qualified with two measures, the support and confidence. Considering a database table T, the association rule $X \Rightarrow Y$ has as support s if s% transactions contain $X \cup Y$. In addition, this rule $X \Rightarrow Y$ has as confidence c if c% of transactions in D containing X, also contain Y. Rules that have a greater confidence of a user-specified threshold c is said to have the slightest confidence.

Figure 24 depicts an example of a rule coming from spatial data mining (Faiz 1999).

Figure 24. Example of rule issued from spatial data mining: "when there is a lake and a road going to the lake, there is a restaurant"

In spatial databases, it is easy to calculate confidence, but in certain cases where many relationships are implicit, assessing the support is difficult.

Superseding rules
As we have already seen, certain global rules can be replaced locally by other rules.

Technical geographic rules
In addition to those rules, there are rules coming from the techniques used in geoprocessing. Here in this chapter, only a few group will be examined, namely linked with data acquisition procedures, with the way scales are used and in pattern recognition.

Rules for data acquisition
Geographic data are coming from measures which are not free of any error. It is therefore necessary to assess the quality of any set of data in order to be sure of the validity of reasoning. Among them, let us mention:

- rules to remove artifacts and outlets;
- rules to assess geometric data and correct them (see for instance (Servigne et al. 2000)).

Another example can be taken from airborne laser covering a wood. Some beams can be reflected by the canopy, and other can penetrate touching either a leaf, a branch or ultimately the soil(Figure 25). In this case, filters can be implemented, a high-pass filter to constitute the elevation model and a low-pass filter the terrain model.

Figure 25. An example from air-borne laser. (a) Raw data. (b) Removing an artifact and low- and high-pass filters.

Mutation of topological rules
Consider a route along a lake. At certain scale, the lake and the route are linked by a DISJOINT relation. But at smaller scale, the relation become TOUCHES. In Figure 26, some other cases are presented. Let call this effect, mutation of topological rules. In Lejdel et al. (2015), the exact mathematical models are given.

Figure 26. Examples of mutation rule of topological relations

Linguistic rules

In (Laurini, 2014) some linguistic rules were given when considering two knowledge bases in two different languages. In this cases, two ontologies and two gazetteers must be considered each in different languages. By means of a homology relation noted ₪, three rules were developed and explained:

- when two places have homologous names and homologous types, therefore they have homologous geometries (see Figure 27);
- when two places have homologous geometries, therefore they have homologous names and homologous types; as a consequence, the concerned ontologies and gazetteers can be updated if necessary;
- when two pairs of places have are linked by relations, those relations are homologous.

Figure 27. Example of linguistic geographic rule with the homologous relations ₪

Rules for pattern recognition
Pattern recognition is very important in geoprocessing when processing satellite images and aerial photos. In those cases, since same features appear very differently according to the used wave lengths, it is necessary to get different images for the learning and the recognition process. The idea is to create an ontology having those images as elements; in other words, in this ontology, the main concepts will be literal whereas terminal leaves will be images. Figure 28 illustrates the beginning of such a visual ontology.

Figure 28. Excerpt from a visual ontology

Metarule
A metarule is a rule that defines a regulatory framework. For example, all local plans must be congruent to laws dealing with urban planning. In other words, metarules define a set of rules that will be valid only when we will refer to this metarule. They can also define new concepts: so a metarule may modify an ontology.

Non-geographic rules
Of course, not all rules are geography-based, and "normal" business rules must also be integrated. Among them, rules for ranking various alternatives are important, mono-criterion or multi-criterion decision-making rules.

In the context of urban and environmental planning, since several shareholders are involved, decision-making procedures must also be multi-actors.

Consequences

Another aspect that corresponds to the status of a rule, i.e. modeling should specify the semantics of the rule:

- the implication is automatic (truth of nature), "if it rains, you get wet";
- the implication can be viewed as a recommendation, "if it is raining, take your umbrella";
- a law decides an inference, i.e. there exists a legal obligation; otherwise, sanctions can generally be applied;

- a law may open possibilities; in this case, the rule provides several open alternatives or options , but none is required;
- some rules can be regarded as a constraint; and as the limits on the height of buildings;
- a local plan generates rules as part of a more general law; so a rule can supersede another rule; exceptions may exist; in some cases, illegal must also be considered: such a thing exists in the real world, but it should not exist according to a precise rule; i.e. do not forget exceptions, exemptions and transgressions;
- the so-called best practices can be considered both as recommendations and as local rules;
- spatial data mining considers items that are frequent linked or things that appear frequently together; do not forget the support and confidence in those association rules;
- a rule is valid only on a clearly defined territory; this space can be called the jurisdiction of the rule (for example: a country);
- some geographic local rules can supersede global rules; in this case, the places where the local rules applied must be inside the jurisdiction of the global rule;
- the language in which the rule is defined, is an important characteristic of a rule; the language defines place names and ontology; generally there is only one language in one place, but there are places where many languages must be considered; example, the rules of the European Union.

As a consequence, the following model can be proposed (see Figure 29).

Figure 29. Relationships between places and rules

Requirements for formal modeling of the geographic rules

The purpose of this section is not to give detailed formalism for modeling spatial rules, but simply to present some requirements. In fact, the design of the modeling language must be designed with the key-requirements of a future inference engine.

The context must integrate:

- the jurisdiction and language for the knowledge system
- a dictionary of place names (gazetteer)
- an ontology that defines the concepts and, above all, the organization of types of geographical features (rivers, mountains, cities, etc.).

In addition:

- a knowledge base containing geographic objects, their attributes, and geographic relations between them;
- and finally the list of rules; these rules should refer to the ontology and the gazetteer of the context.

However, before submitting a new rule, you must specify whether this rule derives from a metarule or supersedes another rule, otherwise because of possible contradictions, there can be failures in the inference engine.

Location should be defined through place names and with the list of coordinates of the polygons. In some cases, one can imagine a geometric formula with unions, intersections and differences. In the case of paths, you must specify two locations (departure and arrival).

Also there will be conditions (IF) written as predicates, the consequences (THEN). In some cases, a further consequence of ELSE type might be interesting for example for sanctions.

In some case, the user can be interested to get information about the automatic reasoning by giving traces. So, some explanatory text can be interesting to store along rules.

To finish this description, statistics or rules from data mining, you must add measures as confidence about the association rules.

COMPONENTS OF GEOGRAPHIC KNOWLEDGE BASES

Any geographic knowledge base will consist of (Figure 30):

- an ontology describing the geographic objects, their relationships between them and their attributes;
- a gazetteer giving the list of place names;
- a set of geographic objects with their spatial type and attributes in accordance with the ontology and sometimes a toponym as given in the gazetteer (remember that all features do not bear a name, f. i. electricity poles);
- a set of relations between geographic objects which are not mentioned in the ontology;
- some physico-mathematical models especially to describe the geographic processes;
- a set of geographic rules which will use the vocabulary of the ontology and place names described in the gazetteer and sometimes some mathematical models.

Figure 30 Contents of a geographic knowledge base.

In addition, Figure 31 illustrates the links between a geographic object and the gazetteer and the ontology.

In addition, a geographic inference engine is a system which will be able to make reasoning about geographic knowledge. Whereas a conventional inference engine is only based on logic, such a system must integrate topology, computational geometry, statistics and spatial analysis because geographic rules necessitate those aspects to be modeled.

Figure 31. A geographic object with its links with gazetteer and ontology

The general structure is illustrated in Figure 32. The core consists in a geographic inference engine working with the geographic knowledge base together with an input and an output. In input, there is the description of a geographic project such as:

- Where to put a new airport, a new hospital, a new stadium, etc.?
- Is this new construction project compliant with planning rules?

- What is the best mode or the best way to get from A to B?
- Where and what is the priority?
- How to organize a plan for green spaces in a city?
- How to reorganize common transportation?
- What are the best places to visit and how to set a touristic tour?

Figure 32. General architecture of a geographic reasoning system.

As output, the more common seems to be a feasibility report consisting in textual or cartographic issues. Among textual issues, let us mention, outside error or inconsistency reports, essentially explanations regarding the possible achievement of the project, comparison and ranking of alternatives. In the cartographic issues, maps can be good candidates and sometimes, chorems (De Chiara et al. 2011) can be an elegant way to summarize the result.

FUTURE RESEARCH DIRECTION AND CONCLUSION

The goal of this chapter was to examine some semantic aspects of geographic knowledge. But, more investigation must be made to have a set of complete semantics in order to present all requirements for a formal model to represent geographic knowledge.

Presently, from an IT point of view, the description of any geographic feature must be characterized by the following aspects:

- its names derive from a gazetteer,
- its types and attributes derive from an ontology,
- its relations with other geographic features also are governed by an ontology,
- its geometric description is based on crisp or fuzzy geometry,
- it can be present in rules both as in antecedents and in consequents.
- and it can be deduced from geographic reasoning and rules.

The first step was to give a starting model for geographic objects and rules. Then the design of such an engine will likely imply to revisit those models.

Concerning geographic rules, different types have been identified, but all of them appear to have different semantics. More works must be done to get a unified model for representing them.

Concerning the design of a geographic inference engine, it is too early to give the main characteristics and functionalities. But it will be certain that in consequences, some modifications must be made in geographic rule representation.

The scope of the chapter was to explore the nature and the semantics of geographic knowledge bases. Whereas the models for geographic objects and ontologies can be directly used, much work must be done to get efficient models for gazetteers and geographic rules.

In order to design efficient tools for territorial intelligence by using geographic knowledge and especially geographic rules, there is still a long way to go.

I want to thanks all the people who have helped me examine several aspects regarding semantics of geographic knowledge.

REFERENCES

Allen J.F. (1983) Maintaining knowledge about temporal intervals. In: *Communications of the Association for Computing Machinery,* 26 November 1983. ACM Press. pp. 832-843, ISSN 0001-0782.

Bertacchini Y., Rodríguez-Salvador M., Souari W. (2007) From territorial intelligence to competitive & sustainable system, Case studies in Mexico & in Gafsa University. *Second International Annual Conference of Territorial Intelligence*, Oct 2007, Spain, pp. 37-54.

Clementini E., Di Felice P. & van Oosterom P. (1993) A small set of formal topological relationships suitable for end-user interaction. In Abel, D. & Ooi, B.-C. *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.

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.

De Chiara D., DelFatto V., Laurini R., Sebillo M., & Vitiello G. (2011) A chorem-based approach for visually analyzing spatial data, *Journal of Visual Languages and Computing*, Vol. 22 (3), 173-193.

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.

Egenhofer M. & R.D. Franzosa R.D. (1991) Point-set topological spatial relations, *International Journal of Geographical Information Science*, vol.5, no.2, pp. 161-174.

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

Faiz S., (1999) *Systèmes d'Information Géographique : Information qualité et Datamining*, Editions CLE, Tunisie, 362 pages, 1999.

Girardot J.-J., Brunau E. (2010) Territorial Intelligence and Innovation for the Socio-Ecological Transition. *9th International Conference of territorial intelligence*, ENTI, Nov 2010,Strasbourg, France. 2010, INTI-International Network of Territorial Intelligence.

Golledge, R.G. (2002) The Nature of Geographic Knowledge. *Annals of the Association of American Geographers*, 92(1), 1-14.

Goodchild M. F. & Hill L. L. (2008). Introduction to Digital Gazetteer Research. *International Journal of Geographical Information Science*, 22(10), pp. 1039-1044.

Graham I. (2006) *Business Rules Management and Service Oriented Architecture: A Pattern Language.* London, John Wiley

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

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

Hećimović Ž. & 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..

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.

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

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.

Kim, T.J., Wiggins, Lyna L. & Wright, J.R. (Eds.) *Expert Systems: Applications to Urban Planning*, Springer-Verlag 1980.

Laurini R. (2012) Importance of Spatial Relationships for Geographic Ontologies. In *Planning Support Tools: Policy Analysis, Implementation and Evaluation Proceedings* of the Seventh International 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.

Laurini R. (2014) A Conceptual Framework for Geographic Knowledge Engineering, *Journal of Visual Languages and Computing* 25, pp. 2-19.

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.

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.

Lejdel B., Kazar O., & Laurini R. (2015) Mathematical framework for topological relationships between ribbons and regions, *Journal of Visual Languages & Computing*. 2015, Volume 26, pp. 66-81. DOI 10.1016/j.jvlc.2014.11.004.

Lévy, P. (1994) *L'intelligence collective.* Pour une anthropologie du cyberspace. La Découverte, París.

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.

Miedes Ugarte B. (2008) Territorial intelligence and the three components of territorial governance. In *International Conference of Territorial Intelligence*, Besançon 2008., Oct 2008, Besançon, France. pp.10,

Morgan T. (2008) *Business Rules and Information Systems: Aligning IT with Business Goals*. Addison-Wesley.

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.

Servigne S, Ubeda T, Puricelli A, & Laurini R (2000) A Methodology for Spatial Consistency Improvement of Geographic Databases. In: *GeoInformatica*. 2000, Vol.4, Issue 1 pp. 7-34.

Ross R. G. (2011) More on the If-Then Format for Expressing Business Rules: Questions and Answers, *Business Rules Journal*, Vol. 12, No. 4 (Apr. 2011), URL: http://www.BRCommunity.com/a2011/b588.html

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

KEY TERMS AND DEFINITIONS

Gazetteer: initially a list of place names; now a database for place names.
Geographic objects: whereas features are in the real word (mountain, river, church, etc.), geographic objects are computer representation of features.
Geographic ontology: an ontology of geographic objects with spatial relations.
Geographic relations: relations between geographic objects.
Geographic rules: a rule dealing with geographic objects.
Ribbon: a line with a width.
Spatial relations: mathematic relations between objects.
Territorial intelligence: business intelligence applied to countries, regions, cities, etc.
Toponyms: another word for place names.