Towards Visual Summaries of Geographic Databases Based on Chorems

Robert Laurini

LIRIS, INSA de Lyon, F – 69621 Villeurbanne

Abstract

For many applications, it is very important to get an overview of database contents; and in the case of geographic database, a visual summary can be very helpful for a decision-maker or for any person in charge of a territory. To get such a visual summary, a two-step process is built, first a phase of spatial data mining process extracts geographic knowledge, and a second phase visualizes it by means of chorems – which can be defined as schematized representations of territory. In other words, semantic generalization must be followed by geographic generalization. The scope of this paper is to present the last results of an international project.

1 Introduction

For many decisions, visual tools are necessary, and especially for spatial decision making for which cartography is an essential tool. When it is the cartography of facts, usually decision-makers are satisfied, but when it deals with visualization of problems, conventional cartography is rather delusive: indeed it seems more interesting to locate problems and perhaps to help discover new problems or hidden problems. So the key-idea is to generate a global overview of the database contents, or more to summarize it.

A research program was launched between several research institutions (INSA-Lyon, France, University of Salerno, Italy, Tec de Monterrey,

Puebla, Mexico, etc.) in order to test whether cartographic solutions based on chorems[1] can be more satisfying. Invented by Brunet (1986, 1993), chorems can be defined as a schematized representation of territories. By schematized, one means that the more important is a sort of short global vision emphasizing salient aspects (Saint-Paul et al. 2005). This definition can be a good starting point to construct maps for spatial decision making. In other words, the goal of this research project is starting from existing databases, to analyze them so that to extract chorems by spatial data mining (Laurini et al. 2006) and visualize them.

This paper will be organized as follows. First chorems will be studied essentially as a new tool for visualizing and summarizing geographic information. Then the description of the architecture of a prototype system will be given.

2 What are Chorems?

2.1 From Conventional Cartography to Chorem Maps

As previously said, according to Brunet, chorems are a schematized representation of a territory. In the past, chorems were made manually by geographers, essentially because they had all the knowledge of the territory in their mind. This knowledge was essentially coming from their familiarity with the territory under study, its history, the climatic constraints and the main sociological and economic problems.

Fig. 1a shows a conventional map of France emphasizing administrative divisions whereas Figure 1b gives an example of a chorem map of France, in which the following aspects are stressed:

- the geometric shape is simplified,
- only big cities are mentioned (Paris, Lyon, Marseilles and Lille),
- only important mountains are shown, Alps as a frontier towards Italy, Pyrenees towards Spain, and the Massif Central forcing traffic to follow the Rhone river axis,

---

[1] Some people use the word "choremes" in English; but this is not correct taking into account usual rules for translating ancient Greek to English. Compare system, problem, etc. So, the word "chorem" is better and was selected.
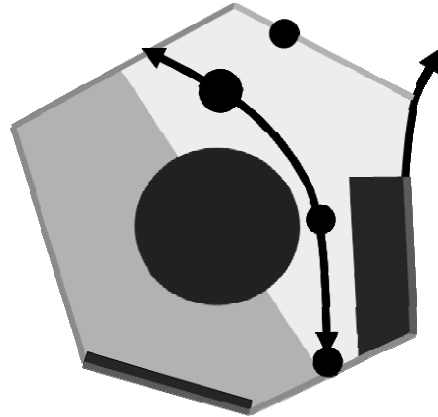
(a) Conventional map of France   (b) A Chorem map of France

Fig. 1. Two maps of France.

- major traffic axes and seas are depicted,
- the blue lines show sea coastlines,
- and the French territory is divided in two parts, Eastern part the more developed, and Western part the less developed.

We claim that such a map is much more informative about the difficulties of France than a flat administrative map.

Another example refers to the water problem in Brazil. Indeed, a conventional map only showing main rivers as illustrated in Figure 2, does not lead to the solution of various problems such as:

- locations of places lacking water
- locations of the places with too much water
- locations of aquatic resources
- locations of humid zones
- locations of the water resources
- locations of the deserts,
- etc.

Bearing in mind all those examples, we claim that those chorem maps are much more informative and helpful to decision-makers. Those chorem maps can be seen both as the layout of geographic knowledge, and a kind of summary for geographic databases characterized by:

- a geographic generalization in order to simplify the shape of the territory under study,
- and a semantic generalization in order to select the more salient aspects of the non-spatial attributes of the geographic database.
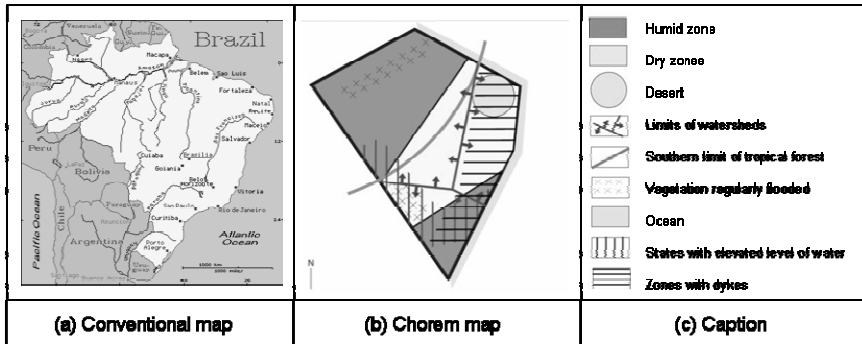
Fig. 2. The water problem in Brazil using a conventional river map (a) and a chorem map (b) issued from (Lafon et al. 2005).

However, some problems exist, especially due to their definitions:

1. For some people, chorems are too much simplified and do not restitute the complexity of a territory. By contrast, some chorematic maps can be very sophisticated (see for instance Peru's chorem in http://flodemon.club.fr/choreme.htm) which tries to represent several phenomena, but it is very difficult to understand or to explain.
2. When some boundaries are laid out, for instance between two zones; the reader must not forget that the lines corresponding to the boundaries are over-simplified or are approximated.
3. One of the major difficulties is to decide what the salient phenomena are and how to select them.

In order to overcome those limitations, our research program was based on the following assumptions:

1. The starting point will be an existing geographic database, not a so-called exhaustive knowledge of a territory under study;
2. The selection of important features will be based on spatial data mining;
3. Only a small subset of chorems will be used, not the entire Brunet's table.

Based on those assumptions, we do not want to re-make chorems of well-known territories, but rather explore some little-known database contents: in other words, we want to use our methodology only in domains which can bring some added value, such as:

• in geomarketing, when the CEO of an enterprise wants to have a global cartography of the places where his products are sold; and eventually to detect the anomalies in order to adapt a marketing strategy to sell more;

- in archeology, especially when spatial and spatio-temporal relationships must be exhibited and discovered;
- in sensor-based environmental monitoring and control, to rapidly discover anomalies, inconsistent sensor behavior and places where actions are need;
- in politics, especially after some elections to study the more salient aspects.
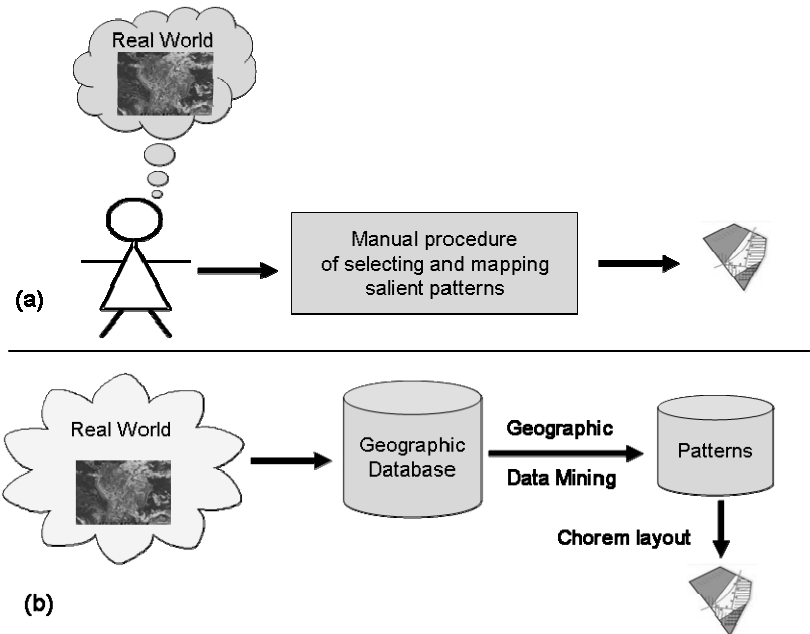
Fig. 3. Comparison of the conventional way of making chorems and our methodology. (a) the conventional way based on a partial vision of a territory; (b) based on a geographic database and geographic data mining for selecting the salient aspects.

Fig. 3 illustrates the difference between the conventional manual way of making chorems, and our methodology based on geographic databases and data mining.

Now the question is how to visually represent those salient aspects? Two possibilities exist:

- either to define a complete vocabulary (by means of icons) which can be used in any situation (this was Brunet's attitude when defining his chorems by means of a table),
- or to let the user define his own vocabulary by providing an ad-hoc caption.

2.2 Results of a Study of Existing Manually-Made Chorem Maps

A study was conducted about the chorems as they were used in several maps. Approximately 50 manually-made chorem maps were studied. The results are:

- even if the chorem concept is used by a lot of geographers, the Brunet's vocabulary is not often used;
- generally the users define their own chorem vocabulary,
- usually less than 10 chorems are used in a single chorematic map,
- the more used patterns can be regrouped into main categories such as (1) main cities (which can be retrieved by SQL SELECTs), (2) main regions which can be retrieved by clustering and (3) main flows which can be retrieved by both clustering and SELECTs.

2.3 Towards New Concepts for Geographic Databases

To conclude this section, it appears that chorems in addition to the initial definition (schematized representation of territories) can be potentially used for other goals such as:

- visually summarizing spatial database contents,
- global vision of a spatial database (Shneiderman 1997; Del Fatto et al. 2007),
- representing visual geographic knowledge,
- and new strategy for accessing spatial database.

As a chorem can be seen as a visual summary, some other layers of visual schematization can be defined from the database contents. So a sort of pyramid can be defined in which the apex is the chorem map, and the base is the database contents. At intermediate levels, several levels of geographic and semantic generalization can be defined. See Fig. 4 for such a pyramid.

To explore those new possibilities, some prototypes must be designed, implemented and tested. Let us examine a proposed architecture.
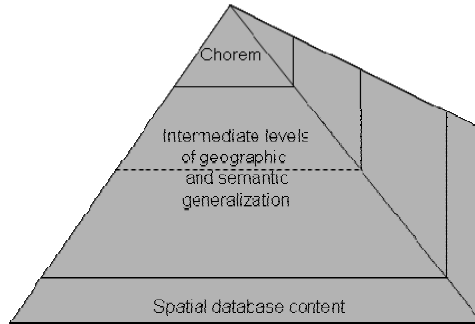
Fig. 4. A pyramid of contents.

3 Architecture of the System

An explorative system has been designed according to the main following specifications (Fig. 5):

1. chorem discovery based on spatial data mining, the result being a set of geographic patterns or geographic knowledge (upper part of Fig. 5),
2. chorem layout including geometric generalization, selection, algorithms for visualization (lower part of Fig. 5).

To facilitate spatial data mining and extract relevant semantics, a canonical database structure has to be defined. As an intermediary between chorem discovery and chorem layout, a language has be defined, named ChorML.

Another problem is the vicinity of the territory. Indeed, in several encountered manual chorem maps, external information must be added, such as the names of sea, of adjacent countries and so one. In order to provide this information, which is currently not in the initial database, a special table of the canonical database was defined. For instance, a canonical database (spatial and non-spatial) at country level will include:

- basic information such as cities, regions, main hydrology, main roads, mountains, etc.
- more elaborated information such as networks, flows, barriers
- external information such as boundary types, names of seas and of adjacent countries
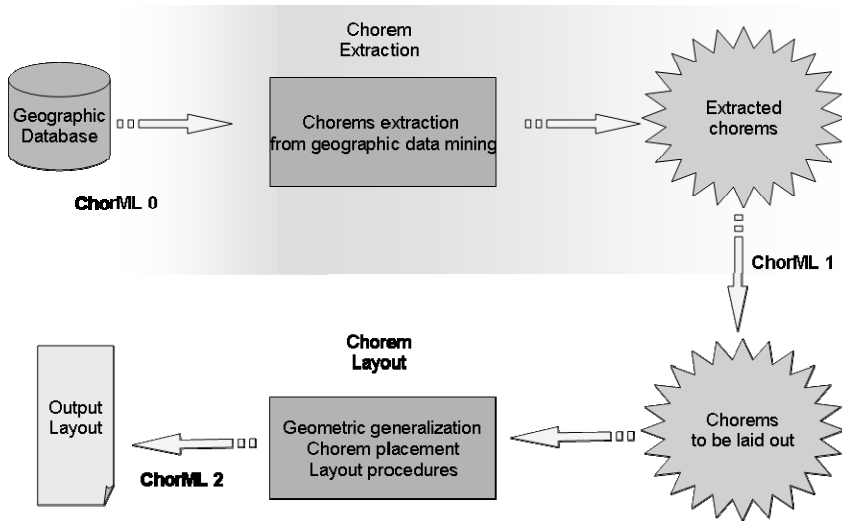- etc.

Fig. 5. Architecture of the system in two parts, the upper part corresponds to chorem extraction, and the lower part, chorem layout.

3.1 Spatial Pattern Discovery

As previously said, spatial patterns will be extracted from spatial data techniques. See (Ester et al. 1997) or (Pech et al. 2002) for details. However, in data mining it is well known that a lot of patterns can be retrieved. Two problems exist:

- setting of list of techniques to be used taking our context into account,
- selecting chorems from patterns.

So, among the relevant techniques, we have chosen to use first clustering and aggregation procedures together with SELECTs.

The next phase is how to identify chorems from spatial patterns, taking into consideration that a maximum of 10 chorems must be chosen. Those ten chorems must correspond to the more important spatial patterns. At this point, we have no clear-cut solution to reduce the number of patterns. In our first prototype we have decided not to implement an automatic solution: for that a visual interface will help the user choose the more important patterns (chorems) for the layout phase.

For the moment being, two kinds of spatial data mining techniques are used, those existing in ORACLE Spatial, and by using SUBDUE (Pech 2005).

3.2 Chorem Layout

Once the list of chorems and the set of constraints among them are obtained from the Chorem Extraction Subsystem, they are sent to the Visualization Subsystem in order to derive a visual representation of chorems and chorem maps, both in terms of layout and semantic content.

(a)   (b)
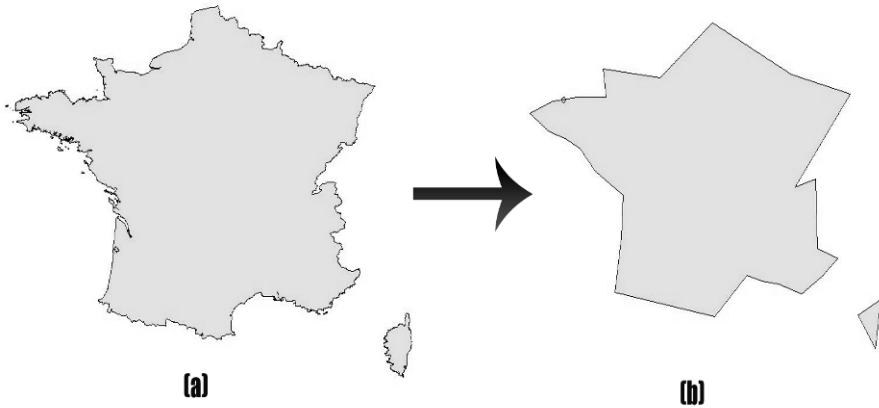
**Fig. 6.** An example the shape simplification process (generalization).

Five different tasks are performed by this subsystem, namely chorem drawing, coordinate translation, best-placement of chosen chorems, pre-layout computation and chorem editing. As for the chorem drawing, it is performed through three, not necessary interconnected, steps, named simplification, choremization and generalization, where some procedures and spatial operators are invoked. In Figure 6, such transformations are illustrated.

As for the generalization step, which is a well known set of techniques in cartography (Buttenfield and McMaster 1991), it may be invoked to group features that share some common properties, both geometric and descriptive, and generate a unique geometric representation of the involved elements. Fig. 9a and 9b depict such a transformation.

The choremization phase associates a regular shape with the possible simplified geometry of data.

One of the problems which may arise when simplifying and generalizing chorems, is related to the possible loss of crucial spatial constraints among elements of the original map. Thus, when the boundary is simplified, cities such as harbors which are located along the boundary must move with the boundary; otherwise, harbors would be positioned in the middle of the sea, or in the middle of the land. In order to preserve the spa-

tial consistency among geographic elements, topological constraints are checked and, if a violation occurs, the Visualization Subsystem modifies the city location, accordingly.
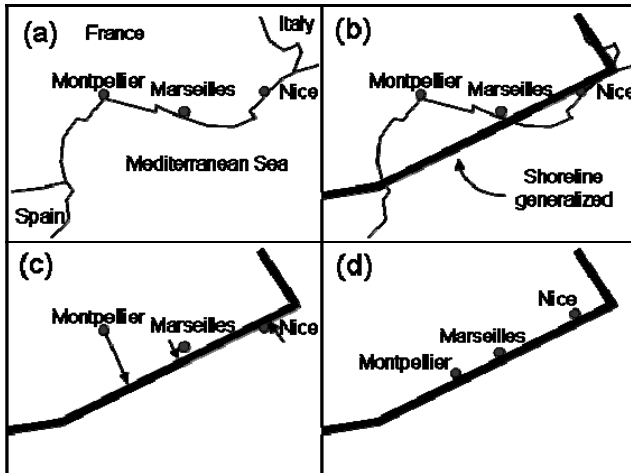
Fig. 7. Projecting harbors onto generalized shoreline. (a) situation before generalization. (b) generalized shoreline. (c) harbors must be moved. (d) final layout.

See Fig. 7 for an example along the French Mediterranean shoreline.

It is interesting to mention that as harbors must follow the topological relation "meet inside", some places must follow "meet outside"; for instance consider the city of Geneva regarding France and generalized Eastern boundary.

It is worth to notice that in order to both preserve topological constraints and properly apply spatial operators, an underlying geographic reference system is maintained during the chorem drawing phase.

Once the drawing of the expected chorem is obtained, users are asked to specify details about the output map, such as the number of colours and the final layout format (for instance A4). The latter affects the number of chorems that can be introduced onto a map, since it is necessary to guarantee the readability requirement.

Based on the information provided by users, the next phase translates the chorem coordinates, acquired with respect to the original geographic reference system, into new coordinates defined with respect to a reference system local to the chosen visualization format.

At this stage, chorems extracted by the Chorem Extraction Subsystem are associated with a locally georeferenced visual representation. The goal of next step consists of aggregating chorems onto the output map. This is

accomplished by a multi-agent system whose aim is to spatially arrange chorems onto the chosen visualization format and determine their best placement (Jones 1989), preserving structural and topological constraints among them. It is worth to point out that in order to guarantee the best placement requirement, independent sets of interrelated chorems may be aggregated onto different maps, in order to provide users with more intuitive and readable chorem maps.

Anyway, some difficulties can occur regarding chorem placement and layout, as well as further refinements affecting semantic and graphic properties may be required by users. To this aim, users are provided with a tool for chorem editing which allows them to refine the expected output map.

In particular, the Chorem Editor performs the following tasks:

- import of a list of chorems positioned onto a chorem map;
- chorem display starting from the information derived from the previous steps;
- modification of both visual representation and semantic structure of chorems, without loss of consistency between them; in order to solve problems regarding chorem placement and layout the Chorem Editor can change chorem positions, colours and shape;
- generation of a graphical representation based on SVG (Scalable Vector Graphics) (SVG);
- export in graphical representation (SVG).

See Fig. 8 for an example of a visual summary after the study of an Italian database for population, and Fig. 9 for a ChorML excerpt.

3.3 ChorML

A language was designed to store chorems. Based on XML, ChorML presents several levels:

- level 0 corresponds to the initial database in GML (Geographic Markup Language) (GML);
- level 1 corresponds to the list of extracted patterns;
- level 2 is a subset of SVG (SVG).

For instance, at level 0, the feature coordinates can be longitude/latitude and feature attributes, whereas at level 1 the feature remains only if it belongs to a selected pattern, and finally at level 2, we deal with pixel coordinates, radius, line styles, colors and textures.
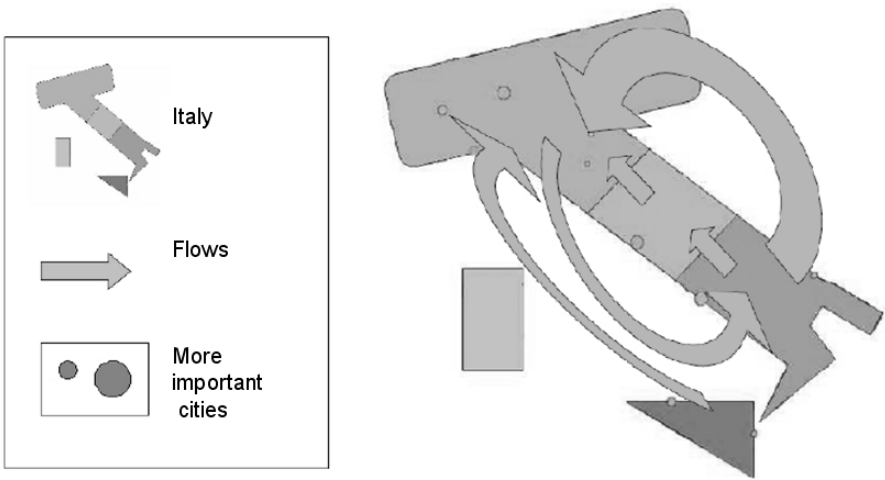
Fig. 8. Example of a visual summary from an Italian population database.

Example of Flows of the Chorem of Italy

```
<Flow>
        <Name>Flow South to North</Name>
        <Id>Flow1</Id>
        <SessionId>ItalyDatabase.15-04-2008.18:35.Session1</SessionId>
        <SourceTable>ItalyDatabase.ItalyPopFlow</SourceTable>
        <ImportanceDegree>1</ImportanceDegree>
        <Type>
                <PathFlow>
                        <Origin>ClusteringModel1Cluster1</Origin>
                        <Destination>ClusteringModel1Cluster3</Destination>
                        <_SHAPE_>
                                <gml:LineString gml:id="p1">
                                        <gml:posList dimension="2">
                                                45.67 18.56 35.56 15.44
                                        </gml:posList>
                                </gml:LineString >
                        </_SHAPE_>
                </PathFlow>
        </Type>
</Flow>
<Flow>
        <Name>Flow North to South</Name>
        <Id>Flow2</Id>
        <SessionId>ItalyDatabase.15-04-2008.18:35.Session1</SessionId>
        <SourceTable>ItalyDatabase.ItalyPopFlow</SourceTable>
        <ImportanceDegree>2</ImportanceDegree>
        <Type>
                <PathFlow>
                        <Origin>ClusteringModel1Cluster3</Origin>
```
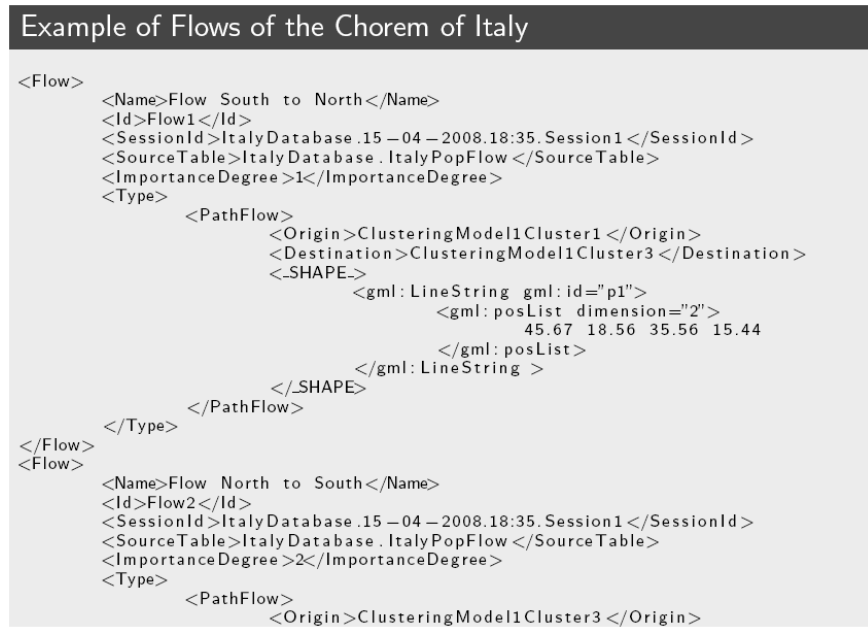
Fig. 9. Example of ChorML describing flows of Fig. 8.

3.3.1 Level 0

At level 0, the structure will include heading (database name, custodian, lineage, etc.) and database contents in GML.

3.3.2 Level 1

At level 1, the heading and additional information are practically not modified, but in place of the GML database contents, we have the list of patterns together with the way to obtain them (lineage). (Coimbra 2008) has shown that that there are four kinds of patterns that result from data mining that appear to be the most interesting in chorem discovery:

- **facts**, for instance the name of a country capital,
- **clusters**, for instance any spatial regrouping of adjacent sub-territories,
- **flows** (one way or both ways)
- **co-location patterns**, especially to describe geographic knowledge; for instance "*when there is a lake and a road leading to that lake, there is a restaurant*".

    In addition to that, we need to include

- **topological constraints**, for instance that a harbor must be inside a territory, not in the middle of the sea,
- and **boundary description**, especially because outside information are usually not included in database, such as sea or neighboring country names.

Presently, this level is totally specified with an XML grammar.

3.3.3 Level 2

Once the patterns are selected, and the output format is known (for instance A4), we need to lay them out. At level 2, the selected patterns are now transformed into drawings encoded in SVG. This information is then sent to the chorem editor to finalize the result.

4 Final Remarks

The objective of this paper was to give some elements for the visual summarizing of spatial databases based on automatic discovery and layout of chorems. After a rapid analysis of existing manually-made chorems, some guidelines were exhibited, so that a prototype architecture can be proposed. This architecture can be split into two parts, the first one corres-

ponding to semantic simplification (chorem discovery) and the second to geometric simplification (chorem layout).

In the conventional way of designing chorems, the so-called "choremist" was supposed to have an exhaustive knowledge of the territory under study, to have clear-cut rules to decide what the salient phenomena are, and not to have problems to cartography them. Our assumption is that based on spatial data mining, the proposed methodology will both restrict the starting knowledge, and will provide a more rigorous approach to select the important features: if an important issue is missing in one chorematic map, it is because it is not included directly or indirectly within the database; not because the "choremist" wants to hide something important. Finally, gradually, we have changed the definition of chorems from "schematized visual representation of a territory" to "schematized representation of a geographic database" and to "visual summary of geographic databases".

Regarding architecture, some modules have already been written and tested (for instance the chorem editor) whereas the specifications of the ChorML language and of the canonical database structure must be finalized. As shown in this paper, chorems look an interesting candidate to visualize geographic database summaries. Another interesting direction of research can be to use chorem as representation for geographic knowledge.

Even so our methodology can be applied to re-make very well known chorems, we claim that our methodology is much more interesting when studying little-known territories such as for geo-marketing, sensor-based GIS for environmental monitoring, archaeology, etc.

Present experimentations are processed based on ORACLE 11g using data from Italy. Next study will extract chorems from an historical database of the Mexican city of Puebla during the XVIIth and the XVIIIth centuries.

Among the difficult problems to solve, there are issues concerning spatial data mining, especially for getting the more important patterns. Presently, some difficulties appear to define mathematically "what is important" and to determine the good methodology to apply spatial data mining tools: so a new avenue for research must be outlined.

References

Brunet R (1986) La carte-modèle et les chorems. Mappemonde 86(4): 4–6

Brunet R (1993) Les fondements scientifiques de la chorématique. In: La démarche chorématique, Centre d'Études Géographiques de l'Université de Picardie Jules Verne

Buttenfield B, McMaster R, (1991) Map Generalization: Making Rules for Knowledge Representation, Longman, London

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

Del Fatto V (2009) Visual Summaries of Geographic Databases by Chorems. Ph Dissertation, April 2009, jointly-awarded by University of Salerno, Italy, and INSA-Lyon, France

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. In: Qiu G, Leung C, Xue X-Y, Laurini R (eds) VISUAL 2007, 9th Int'l Conference on Visual Information Systems, Shanghai, China, 28-29 June 2007, Springer LNCS, Volume 4781 Advances in Visual Information Systems, pp 537–548

Ester M, Kriegel H-P, Sander J (1997) Spatial Data Mining: A Database Approach. In: Proc. of the Fifth Int. Symposium on Large Spatial Databases (SSD '97), Berlin, Germany, Lecture Notes in Computer Science, Vol. 1262, Springer, pp 47–66

GML: http://www.opengis.net/gml/

Holder LB, Cook DJ, (2005) Graph-based Data Mining. In: Wang J (ed) Encyc. of Data Warehousing and Mining, Idea Group Publishing

Jones CB (1989) Cartographic Name Placement with Prolog. IEEE Computer Graphics and Applications 9(5): 36–47

Lafon B, Codemard C, Lafon F (2005) Essai de chorème sur la thématique de l'eau au Brésil, http://histoire-geographie.ac-bordeaux.fr/espaceeleve/bresil/eau/eau.htm

Laurini R, Milleret-Raffort F, Lopez K (2006) A Primer of Geographic Databases Based on Chorems. In: Proceedings of the SebGIS Conference, Montpellier, Springer, LNCS 4278, pp 1693–1702

Pech Palacio M (2005) Spatial Data Modeling and Mining using a Graph-based Representation. Ph Dissertation, December 2005, jointly-awarded by Universidad de las Américas, Mexico, and INSA-Lyon, France

Pech Palacio M, Sol Martinez D, González J (2002) Adaptation and Use of Spatial and Non-Spatial Data Mining. In: Proceeding of International Workshop Semantic Processing of Spatial Data (GEOPRO 2002), Instituto Politécnico Nacional, México, December

Shneiderman B (1997) Designing the User Interface, Third edition, Addison-Wesley Publishing Company

SVG: http://www.w3.org/Graphics/SVG.

Saint-Paul R, Raschia G, Mouaddib N (2005) General Purpose Database Summarization. In: Int. Conf. on Very Large Databases (VLDB 2005), Trondheim, Norway, Morgan Kaufmann Publishers, pp 733–744.