Robert Laurini, Monica Sebillo, Giuliana Vitiello, David Sol-Martinez and Françoise Raffort

Computer-generated Visual Summaries of Spatial Databases: Chorems or not Chorems?

revues.org

Revues.org is a platform for journals in the humanites and social sciences run by the CLEO, Centre for open electronic publishing (CNRS, EHESS, UP, UAPV).

Methods

Computer-generated Visual Summaries of Spatial Databases: Chorems or not Chorems?

Robert Laurini[1], Monica Sebillo[2], Giuliana Vitiello[3], David Sol-Martinez[4], Françoise Raffort[5]

[1] LIRIS, Institut National des Sciences Appliquées de Lyon, University of Lyon, France
[2] DMA, Università di Salerno, Italy
[3] DMA, Università di Salerno, Italy
[4] Tecnológico de Monterrey, Puebla, México
[5] LIRIS, Institut National des Sciences Appliquées de Lyon, University of Lyon, France

Correspondence to: Robert.Laurini@insa-lyon.fr

Abstract

Chorems can be defined as representations of elementary structure of a geographic space or as schematized representations of territories, and as such they can represent a good candidate for generating visual summaries of spatial databases. Indeed for spatial decision-makers, it is more important to identify and map problems than facts. Until now, chorems were made manually by geographers who needed an exhaustive knowledge of the territory under study, a clear-cut set of rules to decide what the salient phenomena are, and who had no problems to cartography them. Here we present a methodology based on spatial data mining, that both diminish the requirements in terms of starting knowledge, and provide a more rigorous approach to select the important features.

1. INTRODUCTION

Visual tools and cartography in particular, are often used for decision making. When it comes to fact representations, decision-makers are usually satisfied with the current cartographic tools, 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 especially in other disciplines than geography.

So, a research program was launched between several research institutions in order to test whether cartographic solutions based on chorems can be relevant for summarizing spatial databases. Invented by Brunet (Brunet, 1986, 1993), chorems can be defined as representations of elementary structure of a geographic space or as schematized representations of territories. By schematized, one means that the more important is a short global vision emphasizing salient aspects in order to consider them in a summary (Saint-Paul *et al.* 2005). More, according to Brunet, chorems are a model among others of territories. This definition can be a good starting point to construct maps for spatial decision making. In other words, it is possible to analyze existing databases to extract chorems by spatial data mining (Laurini *et al.* 2006) and visualize them. This paper develop the idea that chorems can serve as new tools for visualizing and summarizing geographic information and the description of the architecture of a prototype system is given to substantiate that view.

2. WHAT ARE CHOREMS?

2.1. FROM CONVENTIONAL CARTOGRAPHY TO CHOREM MAPS

Chorems (For a discussion of wording in english, see Box A) are a schematized representation of a territory. In the past, chorems were drawn manually by geographers, essentially because they had all the required 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. So this knowledge is

a solid background to derive chorems through a rigorous reasoning methodology. A first example is taken from the water problem in Brazil (Figure 1).

IN ENGLISH: CHORÈME, CHOREME OR CHOREM?

Historically speaking, Prof. Brunet from the University of Montpellier, France coined the French word « chorème » from the greek Χώρημα which means place, location. After, the word « chorème » was used in English directly as coming from French, and then sometimes « choreme » without accent. But considering its etymology and linguistic rules for transforming a Greek word into English, those expressions are not acceptable. Look for example at words such as problem, system, etc.

Greek	Spanish	Italian	French	English
Πρόβλημα	Problema	Problema	Problème	Problem
Προβλήματα	Problemas	Problemi	Problèmes	Problems
Σύστημα	Sistema	Sistema	Système	System
Συστήματα	Sistemas	Sistemi	Systèmes	Systems
Χώρημα	Corema	Corema	Chorème	Chorem
Χωρήματα	Coremas	Coremi	Chorèmes	Chorems

Visual languages are a relatively new discipline (Chang, 1990) which tries to use visual icons, symbols and grammars to represent concepts and ideas, especially in information technology. Representative outcomes range from the design of graphic interfaces, visual queries, visual computations and so on. It is extremely popular in cultures which are not based on letters such as the Chinese's one. When visual languages are well designed, they do not need textual explanation and the recourse to legend could even be considered as a sort of failure. Figure 2 presents another example where the reader is left to discover and understand the meaning of the drawing by herself/himself without a legend (Laurini *et al.* 2009).

Since chorems are outcomes of both cartography and visual languages, they face a paradox: as cartographic outcomes they need explanations, and as visual language outcome, they do not;
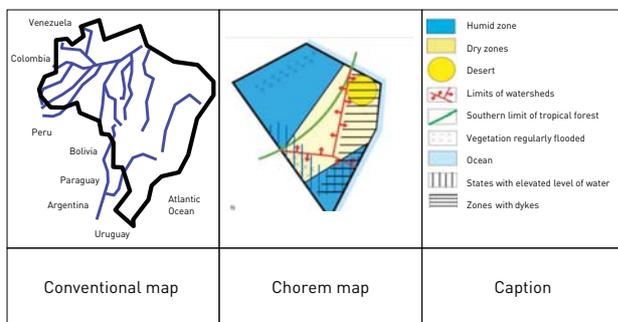

Figure 1. The water problem in Brazil using: (a) a conventional river map and (b) a chorem map. Only the second features locations of (1) places lacking water, (2) places with too much water, (3) aquatic resources, (4) humid zones, (5) the water resources, (6) and deserts. This example is adapted from Lafon *et al.* 2005 with the permission of Baptiste Lafon).
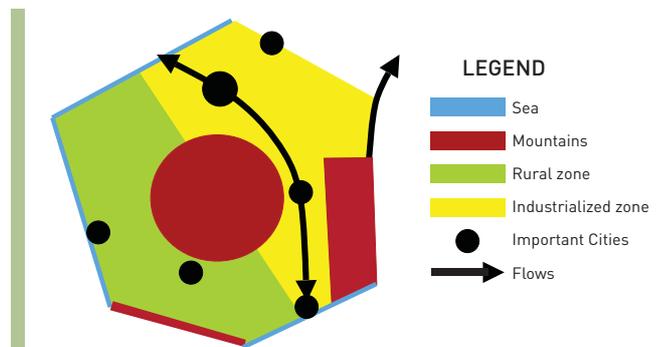

Figure 2. A chorem map of France as an outcome of visual languages.

Quite the opposite, as a visual language, textual legend would be synonym of failure.

These chorem maps can be seen both as the layout of geographic knowledge, and as a kind of summary for geographic databases characterized by: a geographic generalization to simplify the shape of the territory under study, and a semantic generalization to select the more salient aspects of the non-spatial attributes of the geographic database.

Chorems are expected to bring some added value in domains such as:

- geomarketing, to generate a global cartography of the sells and analyse the local variations of market penetration.
- archeology, to uncover the evolution of spatial structures or relationships (sociological structure of a city, history of the dominant commercial flows, etc.)
- sensor-based environmental monitoring and control, to rapidly discover anomalies, inconsistent sensor behavior and places where actions are need
- politics, to analyze the more salient aspects of an election;

2.2. CHOREM REPRESENTATION

2.2.1. ISSUES

The main issues are:

- Chorems can be considered too much simplified and do not restitute the complexity of a territory. In contrast, some chorematic maps can be very sophisticated[6] when representing several phenomena. Such chorems can be very difficult to understand or to explain.
- When some boundaries are laid out, for instance between two zones; the reader must not forget that the lines corresponding to the boundaries are simplified or are approximated.
- Some observers think that a chorem map can have a prescriptive view whereas it has only a descriptive objective.
- and of course, one of the major difficulties is to decide what the salient phenomena are and how to select them.

2.2.2. DEPICTION OF ITEMS ON CHOREM REPRESENTATION

How salient aspects can be depicted? Originally, Brunet established a table to set a completely defined vocabulary (by means of icons) which could be used in any situation. In practice, a study that surveyed 50 manually-made chorem maps gave the following results: (1) even if the chorem concept is used by a lot of geographers, the Brunet's vocabulary is not very used; (2) generally the users define their own chorem vocabulary; (3) usually less than 10 chorems are used in a single chorematic map; (4) the more used patterns can be lumped into main categories such as main cities, main regions and main flows, which can be retrieved by SQL SELECTs, clustering, and by both

clustering and SELECTs respectively. Users seem to prefer to define their own vocabulary by providing an ad-hoc caption (Karla Lopez, PhD thesis, in preparation).

2.3. TOWARDS NEW CONCEPTS FOR GEOGRAPHIC DATABASES

In addition to the initial definition *(schematized representation of territories)*, chorems can also be used to give: (1) a visual summary of spatial database contents, (2) a global vision of a spatial database (Shneiderman, 1997), (Del Fatto *et al.* 2007), (3) a representation of visual geographic knowledge, (4) or a new strategy to access spatial database.

Indeed, for geographic database access, it can be interesting to follow Ben Shneiderman's mantra for designing human interfaces *"Overview, zoom and filter, details on demand"* (Shneiderman1997), i.e. macroscopic versus microscopic approach. So, we can state that chorems can be an excellent candidate at an "overview" level when studying a territory.

As a chorem can be seen as a visual summary, other layers of visual schematization can be defined from the database contents defining a sort of pyramid in which the apex is the chorem map, and the basement the database contents. At intermediate levels, several levels of geographic and semantic generalization can be defined. See Figure 3 for such a pyramid.

For conventional databases, approaches such as starfield or space filling treemaps were created for relational or object-oriented databases. The starfield system is targeted to layout instances of a database object or a relation into a screen: a procedure is given for selecting the two axes from attributes, and then a third axis is selected for colours; the result is called a starfield. The best known example is the starfield system made for Hollywood movies (Ahlberg-Shneiderman 1994). For databases with different objects, another metaphor is used based on so-called space filling treemaps; personally, we would prefer to name this approach the "bookshelf" metaphor.
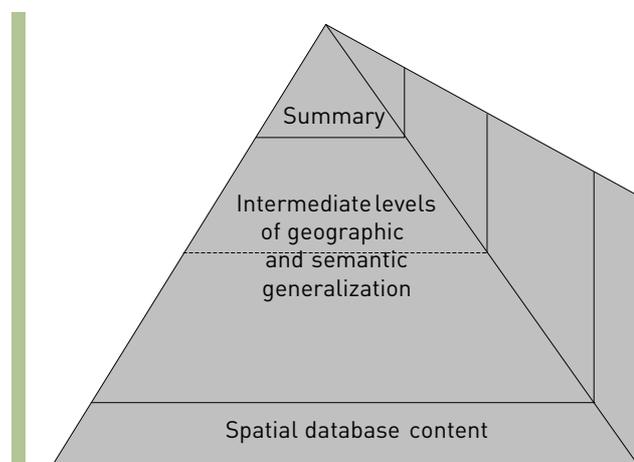

Figure 3. A pyramid of contents.

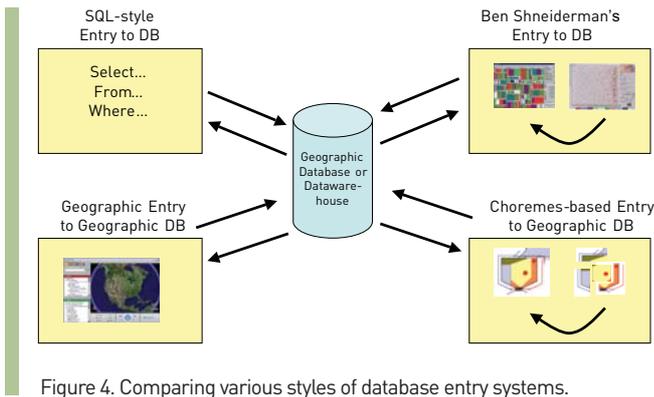[6] See for instance Peru's chorem in http://flodemon.club.fr/choreme.htm

Figure 4. Comparing various styles of database entry systems.

Back on geographic databases and datawarehouses, the chorems approach can have a similar target. In this case the chorem gives an overview of the situation of the territory, whereas the *"details on demand"* step can be represented by a detailed mapping. And by *"zooming and filtering"*, we can gracefully and gradually reduce the search space. Here zooming will mean using different geographic scales or thematic disaggregation, whereas filtering reflects conditions and criteria (geographic and semantic zooming). By zooming and filtering, a sub-chorem can be defined. By sub-chorem, we mean a chorem made for a smaller territory. For instance, one can generate a chorem for a whole country, then chorems for regions and so on.

In other words, chorems can be seen as a new way to enter geographic databases. Table 1 gives a comparison between conventional databases, geographic databases and datawarehouses. Figure 4 schematized the comparison of various types of database entry systems.

Ben Shneiderman's mantra	Conventional databases	Chorem-based approach
Starting point	Relational or object-oriented database of an organization	Any kind of data which can be useful
1 – Overview	Generally the "overview" is visually presented by means of starfield or space filling treemaps; they are both structure- and content-oriented.	The territory-level chorem can give an overview, perhaps more linked to problems than to data contents.
2 – Zoom and filter	Criteria can be used to reduce the search space.	The territory can perhaps be split in different zones, each of them with a sub-chorem (geographic zoom). A second way can be to reduce the number of topics (semantic zoom)
3 – Details on demand	The final step delivers what could be necessary for the user, usually as a table.	Here both tables and maps can be the final steps, depending on the user's needs.

Table 1. Comparing accesses to conventional and geographic databases.

3. ARCHITECTURE OF THE SYSTEM

To overcome the limitations of the manual generated chorem (see 2.2.1), we designed a research program based on the following assumptions:

- The starting point should be an existing geographic database, not a so-called exhaustive knowledge of a territory under study;
- The selection of important features should be based on spatial data mining;
- Only a small subset of chorems should be used, not the entire

A new way of entering a geographic database can be sketched. At the opening, a global chorem map can be displayed, then by semantic and geographic filtering some sub-chorem maps can be visualized and finally, the final query answer (map or table) can be displayed. To explore those new possibilities, an explorative system has been designed (see figure 5).

The chorem discovery is based on spatial data mining, the result being a set of geographic patterns or geographic knowledge (upper part). The chorem layout includes geometric generalization, selection, algorithms for visualization (lower part).

To facilitate spatial data mining and extract relevant semantics, a canonical database structure is defined. ChorML is a language that acts as an intermediate between chorem discovery and chorem layout.

3.1. CANONICAL DATABASE

The system begins by a database to be mined in order to extract spatial patterns. However, the data mining algorithms are not flexible enough to deal with any kind of spatial databases. In order to solve this problem, or in other words to avoid the problem of interoperability between our system and any kind of geographic databases, a structure has been designed, named canonical database. A canonical database is defined as a fixed structure of a geographic database so that any data mining algorithm must be
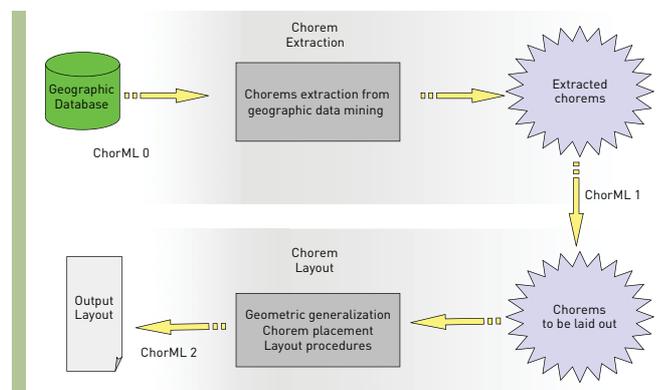

Figure 5. Architecture of the system.

LAURINI *ET AL* COMPUTER-GENERATED VISUAL SUMMARIES OF SPATIAL DATABASES: CHOREMS OR NOT CHOREMS?
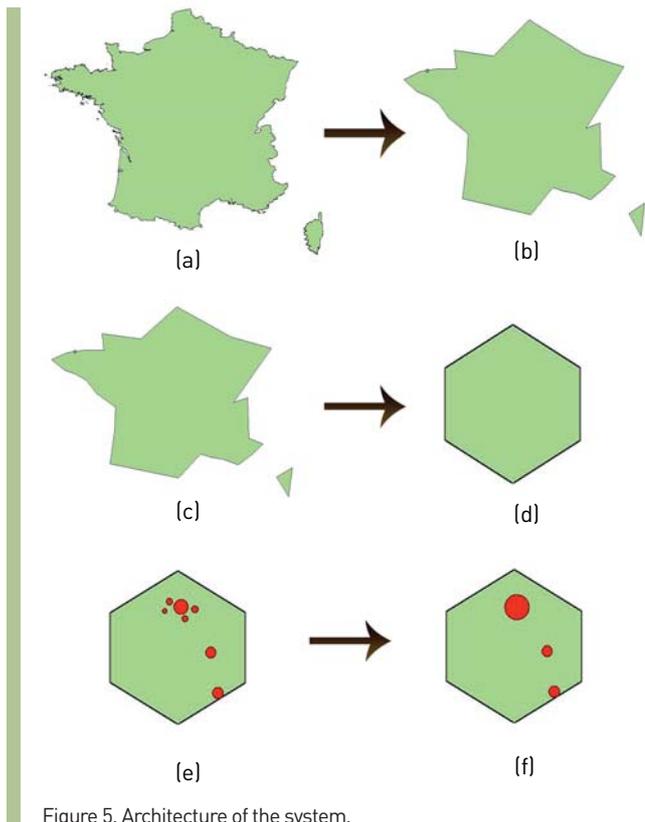
S·A·P·I·E·N·S

Figure 5. Architecture of the system.

applied without modification. Thus, the users must transform their initial database into this structure, either by a list of views, or by creating new tables with this structure.

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 seas, adjacent countries and so on. 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: (1) basic information such as cities, regions, main hydrology, main roads, mountains, etc. (2) more elaborated information such as networks, flows, barriers, (3) external information such as boundary types, names of seas and of adjacent countries, etc.

3.2. SPATIAL PATTERN DISCOVERY

Spatial patterns are extracted using 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, and 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, there is 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 to choose the more important patterns (chorems) for the layout phase.

3.3. 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.

The *simplification* step determines a simplified version (see Fig. 6b) of the data geometry, by reducing the number of vertices of the original shape (see Fig. 6a). As for the *generalization* step, which is a well known set of techniques in cartography (Buttenfield-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. Figures 6c and 6d depict such a transformation. The *choremization* phase associates a regular shape (see Fig. 6e) with the possible simplified geometry of data (see Fig. 6f).

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 (see figure 6).

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,
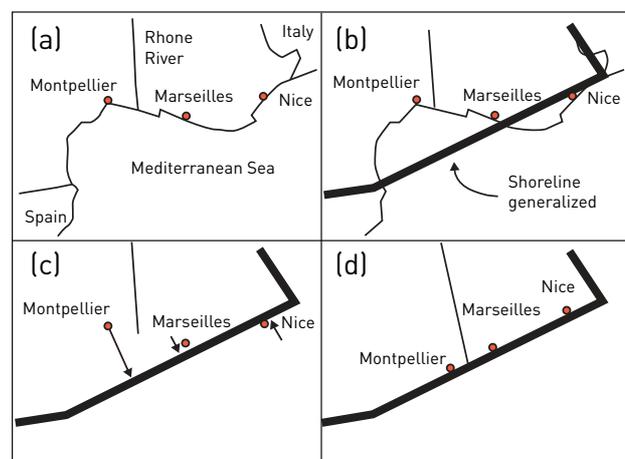

Figure 7. Projecting harbors onto generalized shoreline. (a) situation before generalization. (b) generalized shoreline. (c) harbors must be moved. (d) final layout.

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 spatial consistency among geographic elements, topological constraints are checked and, if a violation occurs, the Visualization Subsystem modifies the city location, accordingly. Figure 7 gives 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 its generalized Eastern boundary. It is worth noticing 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 that spatially arranges chorems onto the chosen visualization format and determines their best placement (Jones, 1989), preserving structural and topological constraints among them. To guarantee the best placement requirement and provide users with more intuitive and readable chorem maps, independent sets of interrelated chorems may be aggregated onto different maps.

Difficulties can occur regarding chorem placement and layout, and 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 may perform the following tasks:

1. import of a list of chorems positioned onto a chorem map;
2. chorem display starting from the information derived from the previous steps;
3. 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;

4. generation of a graphical representation based on SVG (Scalable Vector Graphics)[7]
5. export of both a graphical representation (SVG) and a proper ChorML-based representation of chorems.

A visual interface of the Chorem Editor has been built as an extension of the Magelan Graphics Editor, an open source 2D vector graphics editor, based on Java programming language. The Chorem Editor consists of two working areas, namely a property window and a visualization window, and a toolbar containing both a set of buttons and a tabbed list by which functionality may be invoked. In particular, the property window allows users to interact with and modify chorem properties, also affecting the visual representation. Analogously, the visualization window, which is meant at displaying the chorem map under construction, allows users to manipulate its graphic components, also affecting properties displayed into the property window.

3.4. CHORML

Based on XML, ChorML is a language used to store chorems. It is structured in three levels.

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.

At level 0, the structure is as follows: heading (database name, custodian, lineage, etc.), and database contents in GML.

At level 1, the heading and complimentary 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 four kinds of patterns are of particular interest for 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, 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.

Finally, 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.

---

[7] http://www.w3.org/Graphics/SVG/

Laurini et al Computer-generated Visual Summaries of Spatial Databases: Chorems or not Chorems?

S·A·P·I·EN·S

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.

4. FINAL REMARKS: CHOREMS OR NOT CHOREMS?

This paper gives some elements for the visual summarizing of spatial databases based on automatic discovery and layout of chorems. A rapid analysis of existing manually-made chorems provided some guidelines to design a prototypic architecture consisting in a semantic simplification (chorem discovery) and a geometric simplification (chorem layout).

In the conventional way of designing chorems, the user—the so-called "choremist"—was supposed to have an exhaustive knowledge of the territory under study, a clear-cut set of rules to decide what the salient phenomena are, and not to have problems to cartography them. Our hypothesis is that the proposed methodology based on spatial data mining both restricts the starting knowledge, and provides a more rigorous approach to select the important features: with this method, the absence of an important issue on a chorematic map reflects on a deficit in the database and can not be attributed to an arbitrary choice of the user. In doing so, we are aware that the definition of chorems has gradually evolved from "representations of elementary structure of a geographic space" or "schematized visual representation of a territory" to "schematized representation of a geographic database" or even to "visual summary of geographic databases".

Chorems are interesting candidates to visualize geographic database summaries and have the potential to be used as representations of geographic knowledge. Even though our methodology could be applied to re-do well known chorems for example in conventional geography, we claim that our methodology would confer more added value when applied to little-known territories such as geo-marketing, environmental studies (such as sensor-based systems for environmental monitoring), archaeology, etc.

We recognize that the architecture of our system is not yet stabilized. More applications are needed to validate the overall structure and when the structure of our system will be sufficiently robust, real applications will be developed. In other words, only when fundamental problems in computing will be solved, fundamental problems in geography will be faced, such as the validity of chorematic approach in geography.

We do not want to enter into the so-called chorem controversy. According to some colleagues our approach is not consistent with the chorem methodology. Indeed, our chorems are very different from Brunet's one but our goal is to simplify a spatial database both at semantic and geometric points of view: We need a visual language for representing geographic knowledge and geographic database summaries and the word chorem seems to be the more adequate for this purpose. In practice, it is well accepted by the community of information technology.

Finally, based on our methodology, a lot of research and practical experimentations are needed to prove that Brunet's list of chorem is a relevant and exhaustive set of primitives to model territories. In other words, it can constitute a fresh research field per se in geography, but no more in information technology. Results not before a decade.

REFERENCES

Ahlberg C. & B Shneiderman (1994). Visual Information Seeking: Tight Coupling of Dynamic Query Filters with Starfield Displays", Proc. of ACM CHI94 Conference, 313-317.

Brunet R. (1986). La carte-modèle et les chorèmes, Mappemonde 86/4 pp. 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. & R. McMaster (1991). Map Generalization: Making Rules for Knowledge Representation, Longman, London

Chang S. K. (1990). (Ed). Visual Languages and Visual Programming, Plenum Publishing Corporation, New York.

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

Del Fatto V. *et al.* (2007). Potentialities of Chorems as Visual Summaries of Spatial Databases Contents, VISUAL 2007, 9th International Conference on Visual Information Systems, Shanghai, China, 28-29 June 2007, Edited by Qiu G., C Leung, X Xue & R Laurini., Springer Verlag LNCS, Volume 4781 "Advances in Visual Information Systems", pp. 537-548.

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

Holder L.B. & D. Cook. (2005). Graph-based Data Mining, J. Wang (ed.), Encyclopedia of Data Warehousing and Mining, Idea Group Publishing.

Jones C.B. (1989). Cartographic Name Placement with Prolog. IEEE Computer Graphics and Applications.Volume 9, Issue 5, pp. 36 – 47.

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

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

Laurini R. *et al*. (2009). Chorem Maps: towards a Legendless Cartography? Proceedings of DMS 2009, 15th International Conference on Distributed Multimedia Systems, September 2009, Organized by Knowledge Systems Institute. In press.

Pech Palacio M., D. Sol Martinez & J González (2002). Adaptation and Use of Spatial and Non-Spatial Data Mining. Proceedings of International Workshop Semantic Processing of Spatial Data (GEOPRO 2002), Centre for Computing Research, Instituto Politécnico Nacional, México.

Shneiderman B. (1997). Designing the User Interface, Third edition. Addison-Wesley Publishing Company, 600 pp.

Saint-Paul R., G. Raschia & N Mouaddib (2005). General Purpose Database Summarization. In Int. Conf. on Very Large Databases (VLDB 2005), Trondheim, Norway, Morgan Kaufmann Publishers, p. 733–7