Potentialities of Chorems as Visual Summaries of Geographic Databases Contents

Vincenzo Del Fatto[1,2], Robert Laurini[1], Karla Lopez[1], Rosalva Loreto[3], Françoise Milleret-Raffort[1], Monica Sebillo[2], David Sol-Martinez[4], and Giuliana Vitiello[2]

[1] LIRIS, INSA-Lyon, France
[2] DMA, Università di Salerno, Italy
[3] Benemérita Universidad Autónoma de Puebla, México
[4] Tecnológico de Monterrey, Puebla, México
Vincenzo.Del-fatto@insa-lyon.fr, Robert.Laurini@insa-lyon.fr, Karla.Lopez@insa-lyon.fr, rloreto13@yahoo.com.mx, Francoise.Raffort@insa-lyon.fr, msebillo@unisa.it, davidr.sol@gmail.com, gvitiello@unisa.it

Abstract. Chorems are schematized representations of territories, and so they can represent a good visual summary 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 based on the own knowledge of the territory. So, an international project was launched in order to automatically discover spatial patterns and layout chorems starting from spatial databases. After examining some manually-made chorems some guidelines were identified. Then the architecture of a prototype system is presented based on a canonical database structure, a subsystem for spatial patterns discovery based on spatial data mining, a subsystem for chorem layout, and a specialized language to represent chorems.

Keywords: Geographic Information Systems, chorem, geographic generalization, semantic generalization, database summarizing.

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. In other words, relevant database summaries can be a good approach [Saint-Paul et al., 2005 also for geographic databases.

A research program was launched between several research institutions in order to test whether cartographic solutions based on chorems can be more satisfying. Invented by [Brunet 1986, 1993], chorems can be defined as a schematized

G. Qiu et al. (Eds.): VISUAL 2007, LNCS 4781, pp. 540–551, 2007.

representation of territories. By schematized, one means that the more important is a sort of short global vision emphasizing salient aspects. 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 [Pech Palacio et al. 2002] 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 of the territory, its history, the climatic constraints and the main sociological and economic problems.

Fig. 1. Conventional administrative map of France

Figure 1 shows a conventional map of France emphasizing administrative divisions. Figure 2 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 axis,
- Major traffic axes and seas are depicted,
- And the French territory is divided in two parts, Eastern part the more developed, and Western part the less developed.

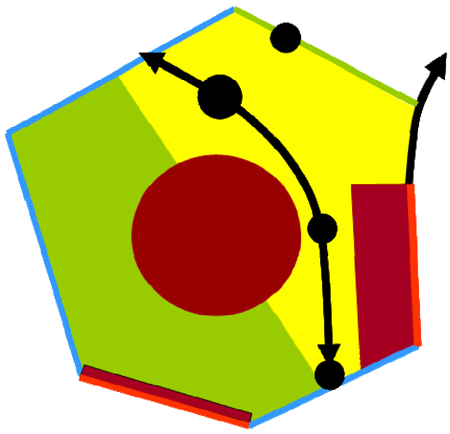

Fig. 2. A chorem map of France

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

Another example is issued from the water problem in Brazil. Indeed, a conventional map only showing main rivers as illustrated in Figure 3, 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.

Lastly, let us examine the chorem given in Figure 4 schematizing the US population in which the country is represented by a rectangle. The exterior arrows show emigration from foreign countries and inner arrows the more important flows coming from North East. The country is split into two zones, the Eastern more populated, and the Western part less populated.

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 bothr as the layout of geographic knowledge, and a kind of summary for geographic databases characterized by:

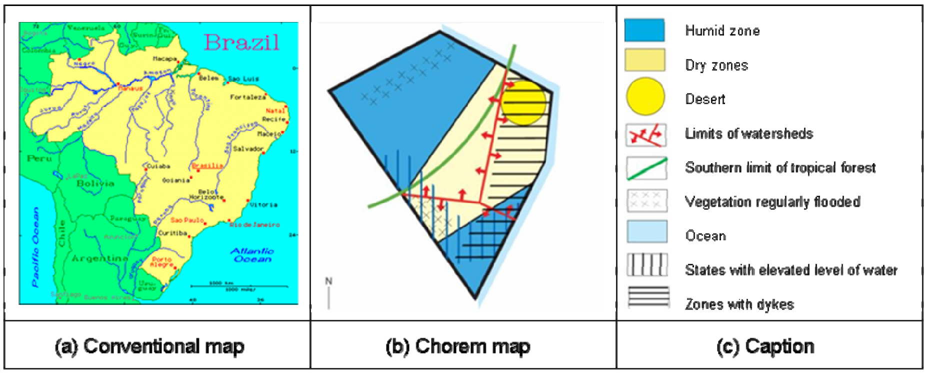

Fig. 3. The water problem in Brazil using a conventional river map (a) and a chorem map (b) issued from [Lafon et al. 2005]

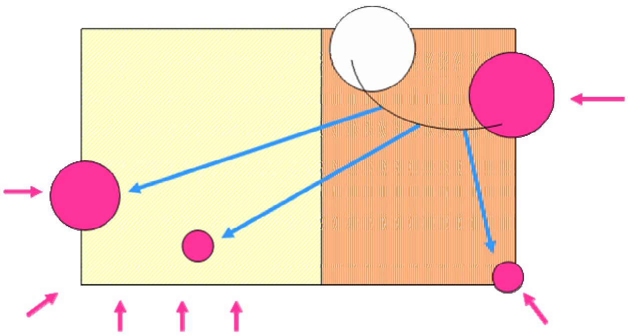

Fig. 4. Chorem for the US population [USA]

- 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.

Now the question is how to represent visually those information chunks? Two possibilities exist:

- either to define a complete vocabulary (by means of icons) which can be used in any situation (this was the Brunet's attitude when defining his chorems by means of a table),
- or to let the user define himself his own vocabulary by providing a 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 very used;
- generally the users define their own chorem vocabulary,
- usually less than 10 chorems are used in a 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 Summarizing Geographic Databases

To conclude this paragraph, 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, Laurini et al. 2006]
- and representing visual geographic knowledge.

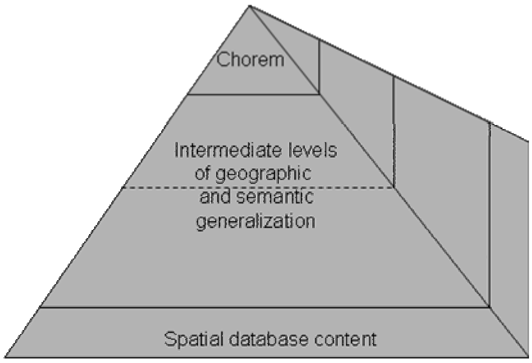

Fig. 5. A pyramid of contents

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 basement the database contents. At intermediate levels, several levels of geographic and semantic generalization can be defined. See Figure 5 for such a pyramid.

To explore those new possibilities, some prototypes must be designed, implemented and tested. Let us examine a proposed architecture.

3 Architecture of the System

An explorative system has been designed according to the main following specifications (Figure 6):

1 – chorem discovery based on spatial data mining, the result being a set of geographic patterns or geographic knowledge (upper part of Figure 6),
2 – chorem layout including geometric generalization, selection, algorithms for visualization (lower part of Figure 6).

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.

Let us examine all of those issues.

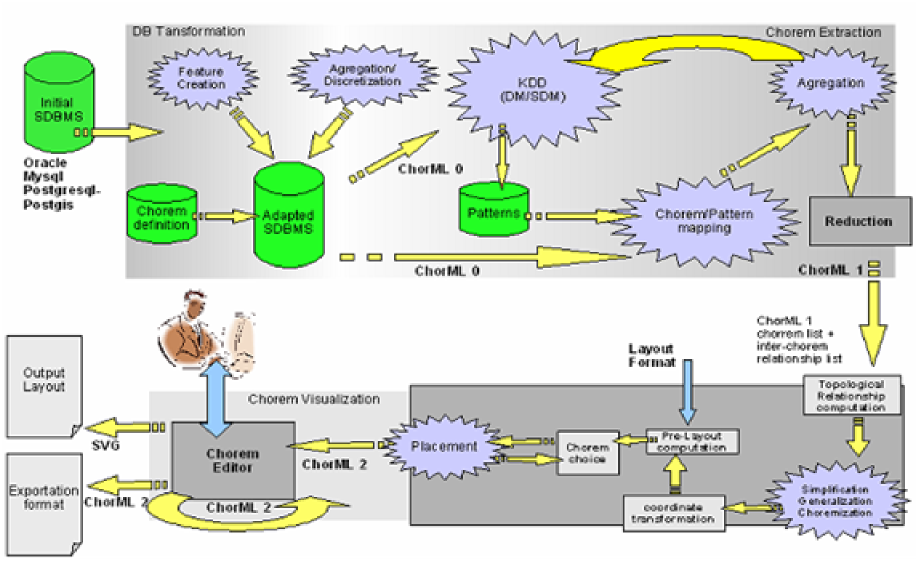

Fig. 6. Architecture of an explorative system for chorem discovery and layout

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 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 can 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.

3.2 Spatial Pattern Discovery

As previously said, spatial patterns will be extracted from spatial data techniques. See [Ester et al, 1997] or [Holder-Cook, 2005] 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.

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.

As shown in Figure 6, 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 Figures 7-9, such transformations are illustrated. In particular, the *simplification* step determines a simplified version (see fig. 7(b)) of the data geometry, by reducing the number of vertices of the original shape (see fig. 7(a)).

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 11(a) and 11(b) depict such a transformation.

The *choremization* phase associates a regular shape (see fig. 10(b)) with the possible simplified geometry of data (see fig. 10(a)).

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 harbours which are located along the boundary must move with the boundary; otherwise, harbours 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.

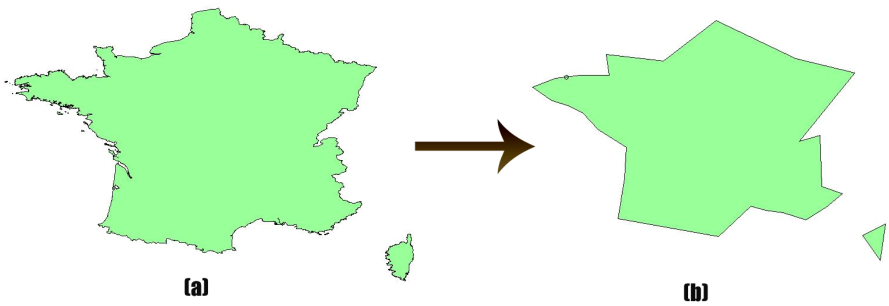

Fig. 7. An example the simplification process

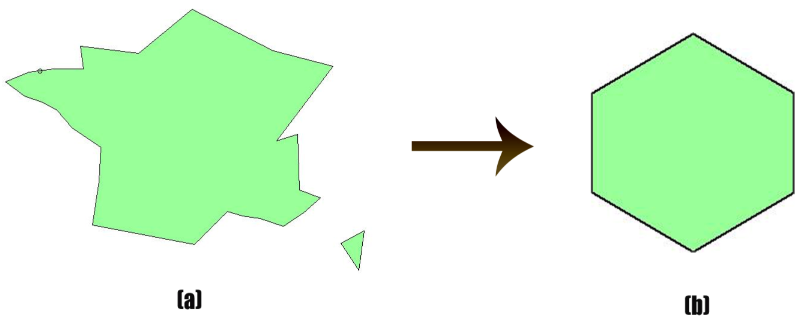

Fig. 8. An example the choremization process

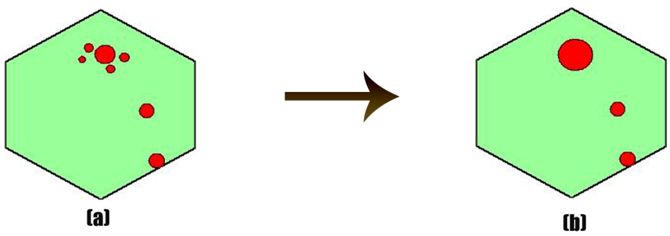

Fig. 9. An example the generalization process

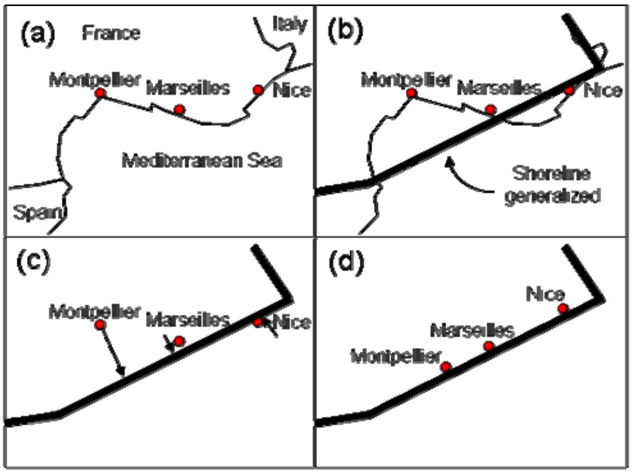

Fig. 10. Projecting harbors onto generalized shoreline. (a) situation before generalization. (b) generalized shoreline. (c) harbors must be moved. (d) final layout.

See Figure 12 for an example along the French Mediterranean shoreline.

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 may perform 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 of both a graphical representation (SVG) and a proper ChorML-based representation of chorems.

Figure 11 shows the visual interface of the Chorem Editor, which 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.

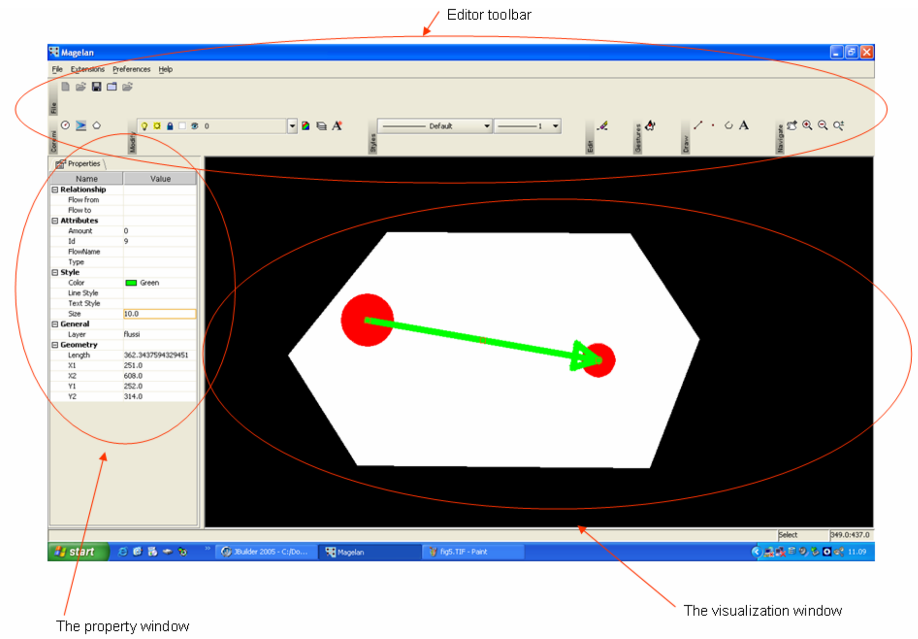

Fig. 11. The visual interface of the Chorem Editor

3.4 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 corresponds to positioned chorems
- level 3 is a subset of SVG.

For instance, in level 1, the feature coordinates can be longitude/latitude and feature attributes, whereas in level 2, we deal with pixel coordinates, line styles, colors and textures.

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.

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.

Present experimentations are processed based on ORACLE 10g 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.

References

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

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

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

Ester, M., Kriegel, H.-P., Sander, J.: Spatial Data Mining: A Database Approach. In: Scholl, M.O., Voisard, A. (eds.) SSD 1997. LNCS, vol. 1262, pp. 47–66. Springer, Heidelberg (1997)

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

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

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

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

Laurini, R., Milleret-Raffort, F., Lopez, K.: A Primer of Geographic Databases Based on Chorems. In: Meersman, R., Tari, Z., Herrero, P. (eds.) On the Move to Meaningful Internet Systems 2006: OTM 2006 Workshops. LNCS, vol. 4278, pp. 1693–1702. Springer, Heidelberg (2006)

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

Shneiderman, B.: Designing the User Interface, 3rd edn. Addison-Wesley, Reading (1997)

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

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

USA, http://lettres-histoire.info/lhg/geo/geo_usa/cartes_usa_geo.htm