Vincenzo Del Fatto

Visual Summaries of Geographic Databases by Chorems

Ph.D. Thesis

April, 07, 2009

University of Salerno, Italy
INSA of Lyon, France

Ph.D.

Jointly submitted at

the Institut National des Sciences Appliquées de Lyon
(Ecole Doctorale Informatique et Information pour la Société)

and the Università degli Studi Di Salerno
(Dipartimento di Matematica e Informatica)

By

Vincenzo Del Fatto

On Avril, 07, 2009

Visual Summaries of Geographic Databases by Chorems

PhD Committee Chair, Prof. Martin Hervé, Université Joseph Fourier, Grenoble

Supervisors

Prof. Robert Laurini, INSA-Lyon

Prof. Giuliana Vitiello, Università di Salerno

Reviewers

Prof. Silvana Castano, Università di Milano

Dr. Anne Ruas, HDR, Director of COGIT, Institut Géographique National, Paris

Examiner

Prof. Maria Francesca Costabile, Università di Bari

To my parents, Michela and Sandro.

*To Marco, Elena and Sylva, my present family,
for the endless patience, continuous support
and encouragements.*

*"Everybody knows that something is impossible to realize till someone inexperienced comes and invents it."
(Albert Einstein)*

This thesis is the result of three years of work. During this period I have been accompanied and supported by many people. In this small section I will attempt to thank all who assisted me in various forms to bring this study to its conclusion. My apologies to anyone inadvertently omitted.

First of all, I would like to thank my supervisors, Prof. Robert Laurini and Prof. Giuliana Vitiello, whose support and guidance made my thesis work possible. I am very grateful to Prof. Robert Laurini and Prof. Giuliana Vitiello for providing me with the opportunity to spend eighteen month at INSA of Lyon, an important experience in my life.

The LabGIS group contributed to the development of this work. Especially the extensive comments, discussions and interactions with Dr. Monica Sebillo had a direct impact on the final form and quality of this thesis. I would like to thank Dr. Monica Sebillo, for our many discussions and providing me advises and tips that helped me a lot, but also for giving me the first opportunity to meet the Research. I am also grateful to the other LabGIS components, Luca Paolino and Davide De Chiara, for sharing ideas, experiences and food. Thanks to all my friends from Dipartimento di Matematica e Informatica of the University of Salerno, for their company, their support and for the soccer matches.

I would also thank Dr. Sandro Bimonte, for sharing his house with me and for being good colleague during the period I stayed in Lyon. Many thanks to my *choremist* colleague at LIRIS, Karla Lopez, for sharing ideas about *chorems*, and my “French” colleagues Ruggero Pensa and Rosario De Giorgi for logistic support in Lyon.

I had the pleasure to supervise and work with students who did their graduation work in our project and have been somehow beneficial for the presented work in this thesis, Mario Mauro, Luca Stabile and Andrea Bruno.

I would thank my parents, Michela and Sandro, for allowing me to study.

Finally, I am forever indebted to my brother Marco, Elena and Sylva for all the support that they gave me during these years, for their understanding, their endless patience and their encouragements.

Vincenzo Del Fatto
Fisciano, March, 02, 2009

Abstract

Traditional cartography is a fundamental tool to visually describe facts and relationships concerning with territory. This is a well-known and well-established approach and decision makers are usually satisfied by its expressiveness when it concerns the cartography of facts. Differently, this kind of cartography may fail when dealing with scenarios referring to heterogeneous issues, such as political, economic and demographic problems, due to the large amount of complex data to represent in a map. Then, more effective solutions in supporting users to locate facts, trends and new patterns should be investigated. In this dissertation the research carried out within an international project is presented, meant to define cartographic solutions able to better represent geographic information extracted from database contents, which refer both to geographic objects and spatio-temporal phenomena. An actual support for human activity to model and analyze the reality of interest may indeed consist of an immediate synthesis of the most relevant data, disregarding details. Such a synthesis may be based on the usage of visual metaphors, which are able to capture and restitute the most salient features of a scenario. Moreover, it may represent the starting point for further processing tasks aimed to derive spatial analysis data, and to support expert users in decision making, thus bridging the gap between the complexity of the adopted applications and the need for rapid and exhaustive responses expected by domain experts.

The major contribution of this work along this line has been to define a methodology to visualize geographic database summaries, expressing them through "schematized representations of territories", known as *chorems*. In particular, two specific contributions have been produced by investigating and implementing the proposed methodology. The former consists of the formal specification of chorems in terms of visual language and structure, in order to both standardize the chorem creation and assembling process and provide a usable framework for computer systems. The latter is represented by the design and the implementation of a system which generates maps containing chorems starting from geographic database content, in a semi-automatic manner.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Contents

This chapter introduces the scope, the purpose and the organization of this thesis. In particular, it identifies motivations which underlie it, presents the objectives of this work, and describes how it is organized.

1.1 Aim and Scope

Visualization of geographic data traditionally belongs to the research area known as geographic visualization, or *geovisualization*, which refers to a set of tools and techniques supporting geospatial data analysis, specifically conceived to handle the complex structure of spatial data, involving space, time and a set of thematic attributes. As in the related fields of scientific visualization and information visualization, geovisualization is focused on knowledge construction and combines geospatial information with human vision and domain expertise, to allow for data exploration and decision-making processes. In particular, geovisualization integrates approaches from scientific visualization, (exploratory) cartography, image analysis, information visualization, exploratory data analysis (EDA) and Geographic Information Systems (GIS) to provide theory, methods and tools for the visual exploration, analysis, synthesis and presentation of geospatial data. Most of the major challenges in the field of the geovisualization are investigated by the International Cartographic Association (ICA). Since then, this discipline has continued to grow as a subject of practice and research and a key-point in its research agenda corresponds to the development of new forms of representations that support the understanding of geospatial phenomena and space-time processes.

Traditional cartography is a fundamental tool to visually describe facts and relationships concerning with territory. It associates geographic concepts with graphic representations based on the use of colours, textures, symbols, and diagrams, which help readers to gain an immediate comprehension of represented data. More advanced computer-enabled techniques, such as map animation and interactive 3D views, can be exploited as well, thus obtaining different perspectives into the data through the use of multiple interactively linked views. This is a well-known and well-established approach and decision makers are usually satisfied by its expressiveness when it concerns the cartography of facts.

A more complex approach is instead required when dealing with scenarios referring to composite issues, such as political, economic and demographic problems, where expert users are provided with a huge amount of data describing geographic objects and spatio-temporal phenomena. As matter of fact, nowadays maps are mainly used as tools both for analysis, problem solving and decision making, and for communication of ideas between people. The concept of cartographic communication has substantially extended in recent years, although the communicative role of maps seems to fully act in accordance with traditional cartography rules. For instance, an analyst may require maps in order to communicate analysis results to a decision maker, or participants of cooperative problem solving may exchange their ideas and arguments using maps. In these cases, the actual goal is not to restitute the reality, i.e. to represent exactly everything with a high level of details, but rather to show the most salient features or characteristics useful to decision-making. In other words, a relevant support for human activity to model and analyze the reality of interest may consist of an immediate synthesis of data of interest, which disregards details. Such a synthesis may then represent the starting point for further processing tasks aimed to derive spatial analysis data, thus bridging the gap between the complexity of the adopted applications and the need for rapid and exhaustive responses expected by domain experts.

A technological support to the realization of this goal may be achieved by integrating results from different research areas, namely geospatial databases, knowledge discovery in databases, and (spatial) data mining. In particular, each discipline may synergically contribute to the definition of a geovisualization methodology for advanced cartographic solutions, by providing techniques for discovering interesting previously unknown, but potentially useful patterns from geospatial databases. As for geospatial database research, it has been an active area for several decades, due to the growing data management and analysis needs of geographic applications. Indeed, the impetus for improving geospatial analysis and functionalities comes from the needs of existing applications, such as GIS, Location Based Services (LBS) and Sensor Networks, in a variety of domains, such as ecology and environmental science, public safety, transportation, Earth science, epidemiology, crime analysis, and climatology. In addition, geospatial database systems have become more and more required because of both the increasing use of geographic data and functionalities over the Web, and the suitability and cheapness of hardware tools, such as Global Positioning System (GPS) technology. Such an explosive growth of geographic data and the widespread use of geospatial databases have then emphasized the need for the automated discovery of geospatial knowledge. Thus, spatial data mining has become an investigated and required process thanks to its capability of discovering patterns and relationships among data. However, the huge volume of geographic information to visualize within a small portion of a map makes it difficult for humans to easily understand and analyze it. In fact, it is known that when used improperly, visual displays may be misleading and not productive. Therefore, it is necessary to find ways for validating patterns and relationships detected in data by means of visualization, in order to generate hypotheses, develop problem solutions and construct knowledge.

1.2 Objectives

Based on the previous motivations, an international project was launched among three research institutions, namely Institut National des Sciences Appliquées (INSA) of Lyon (France), Dipartimento di Matematica e Informatica (DMI), University of Salerno (Italy) and Tecnólogico de Monterrey, Campus de Puebla (Mexico). The project is meant to define cartographic solutions able to synthesize contents of geographic databases and represent them within a map in a readable and intuitive way.

Generally speaking, this thesis, which originates within the framework of this international collaboration, aims to pursue one of main goals proposed in research agenda of the ICA, i.e., to develop new forms of visual representations that support the comprehension of complex scenarios characterized by spatio-temporal properties useful to decision-making tasks. In particular, the purpose of this dissertation is to obtain new modalities of visual representation which provide decision-makers with a means capable to represent an immediate synthesis of data of interest disregarding details. As a matter of fact, the production of such a new visualization means can bridge the gap between the complexity of the adopted applications and the need for rapid and exhaustive responses expected by domain experts, representing a relevant support for human activity to model and analyze the reality of interest.

The expected visualization means can exploit a relevant characteristics of visual languages, namely visual metaphors. Such metaphors can be adopted in order to schematize some territories where facts, trends and patterns are located, thus maximizing the comprehension of the message to deliver. This is in agreement with the shared idea that the message to deliver through a map is more important than the accuracy of details.

The solution proposed as visual metaphor is based on the concept of *chorems*, which are schematic representations of territories potentially suited for visualizing static geographic objects and dynamic spatio-temporal phenomena. Since their appearance, chorems have had a large diffusion in many domains thanks to their capability to provide geographers with a global view of phenomena of interest or an insight into a specific issue.

Starting from this statement, the thesis aims to assign a value added to the conventional role played by chorems, thus accomplishing an innovative function, namely to visually summarize geographic information conveyed by a map of chorems. In order to achieve this goal the following activities have been planned:

- To analyze the concept of *chorem* and study how they have been used in literature;
- To propose a definition of *chorems* which allows to embed them in a software system.
- To propose a system which is meant to reach the following two main goals:
  - to discover and extract interesting, previously unknown, and potentially useful information from geographic databases, and
  - to generate and visualize a schematic visual representation of the extracted information by using the *chorems*,

As for the first two tasks, their relevance derives from the need of a rigorous approach for chorem creation and assembling, which may avoid the failure in conveying the information they are meant to. Part of these activities is being investigated with Karla Lopez's contribution, a PhD student from INSA of Lyon, a partner of the international collaboration. As for the third task, much attention in this dissertation is being devoted to the design and implementation of the visualization system component, aiming to automate the overall process as much as possible. A preliminary extraction system component is being designed in collaboration with Karla Lopez, whereas the final version will be subject of her forthcoming PhD Thesis.

1.3 Main Contributions

This section summarizes the main contributions of this Thesis. Generally speaking, the aim of this thesis was to determine an innovative cartographic solution to visually represent knowledge related to a territory, that manually or automatically could be extracted from a (geographic) dataset. An innovative result of our methodology is represented by the opportunity of using "schematized representations of territories", namely the *chorems*, as a means to represent geographic knowledge. In fact, this approach may confer added value when little-known territories are investigated for different purposes allowing to easily represent significant before unknown aspects through a combination of simple visual metaphors, that convey useful geographic knowledge.

The specific contributions produced by investigating and implementing our methodology are related to different research areas and can be synthesized as follows:

- The formal specification of chorems in terms of visual language and structure is given. This specification is meant to both formalize the structure of chorems and solve ambiguities of terms in the chorems domain, but essentially in order to provide a usable framework for computer systems, which supports the automatic generation of chorematic maps. To satisfy this requirement, a formal definition of an XML-like multi-level language, named *ChorML*, has been introduced to store information about chorems and to support communication among system modules.
- The definition and the implementation of a system which generates chorematic maps starting from geographic database content, in a semi-automatic manner. The system proposed in this dissertation is meant:
  - To extract geographic objects and spatio-temporal phenomena from a geographic database by using SQL queries and (spatial) data mining techniques, to allow expert users to identify interesting patterns and trends, and to associate them with a proper type of chorems.
  - To transform a list of chorems and relationships, stored in a XML-like language, to a visual representation, namely a chorematic map, by implementing a multi-agent system architecture for managing multi-layer simplification and map object placement and labeling. The proposed multi-agent system architecture is based on the chorem classification given in this dissertation. Finally, some work has been done for understanding how different levels of simplification are perceived by users, and how a system can take into account such a difference of perception.

1.4 Overview of the Thesis

The Thesis is organized in seven chapters.

Chapter 2 introduces the context of this dissertation, which is based on several topics, namely Geographic Data and Functionalities, Geographic Information Systems and Science, Cartographic Generalization, Knowledge Discovery in Databases and Data Mining, Visual Languages, and Multi-agent Systems.

Chapter 3 presents the meaning and the history of chorem concept, it discusses aspects that led us to choose chorems in this research work, and illustrates a formal definition and classification of *chorems*.

Chapter 4 proposes the architecture of a system for visually summarizing geographic database content by using chorems, whose aim is to extract geographic objects and spatio-temporal phenomena from geographic databases and to represent them into a map by using chorems.

Chapter 5 describes a prototype of the *Chorem Visualization System* and the technology used in order to implement it. An example of construction of a map containing *chorems* is illustrated in order to demonstrate the effectiveness of the system.

Chapter 6 concludes this dissertation by summarizing our contributions and identifying directions for future works.

Chapter 2

State of the Art Review

Contents

This chapter introduces the context of this dissertation, which is based on several topics, namely Geographic Data and Functionalities, Geographic Information Systems (GIS) and Science, Knowledge Discovery in Databases and Data Mining, Digital Cartography, Visualization of Geographic Information, Visual Languages, and Multi-agent Systems. In particular, some aspects of Data Mining, Visual Languages and Multi-agent Systems related with spatial information and functionalities are thoroughly discussed, such as spatial DBMS (SDBMS), spatial data mining, visual languages as visual summaries and query languages for DBMS and SDBMS.

2.1 Geographic Data

Nowadays, geographic data emerge as central to many applications, such as GIS, Computer-Aided Design (CAD), Robotics, Image Processing [112]. They deal with scenarios of the real world and are organized in themes or layers, which are in turn represented in maps. Typical examples of themes are cities, rivers, counties, etc, which correspond to particular topics and belong to the same typology of data

In the GIS context, several terms, such as *geographic*, *spatial* and *geospatial* are commonly used. Although their meaning is very similar, they slightly differ. The term *geographic* is pertinent to the Earth's properties, the term *spatial* is appropriate to describe any space used in particular to emphasize the generality, while the term *geospatial* is utilized to indicate a subset of spatial data related to the Earth's properties [87].

Geographic data may be characterized according to two different categories, namely discrete data, which describe geographic objects with specific boundaries, and fields, which represent geographic phenomena continuously distributed.

In literature, several works point out the peculiarity of geographic data. In particular, Anselin [4], who argues that 'spatial is special', properly highlights two significant aspects, namely spatial dependence and spatial heterogeneity, which both violate the properties of standard statistical techniques. In particular, the former refers to the Tobler's famous Fist Law of Geography [123] "all things are related but nearby things are more related than distant things". The latter implies that the results of any analysis are always dependent on how the boundaries of the study are drawn [57]. A third characteristic is mentioned in [87], where the authors state that in GIS, geographic data are often managed by means of different layers, according to the data typology. However, such a management incorrectly suggests that each layer captures something unique to it, whereas in practice geographic layers are often highly interrelated [57]. Provided previous properties, from a design point of view geographic data are associated with sets of homogeneous entities, known as geographic object. A geographic object is made up of two components [112]:

- a description, which specifies the object by means of non-spatial attributes. These attributes are also referred to as thematic or descriptive attributes. As an example, the name and the length represent non-spatial attributes of a river;
- a spatial component, also known as spatial or geometric attribute. This component describes both the geometry (location, shape, orientation and size) and the topology (spatial relationships among objects) of a geographic object. For instance, a river may be represented as a line.

As for the geometry of the spatial component, it may be identify through two different primitives [56]. In the former geographic data are conceived as a grid of spatially continuous cell values, which are best suited to represent continuous phenomena. In the latter, the geographic data are expressed by using discrete points, lines and polygons, which are best suited to represent features individually identified. Sometimes, a third primitive, named Surface, may be used to describe geographic data by using 3D planes, which are able to exploit the third dimension of the data. Many areas of GIS, including data models, data quality, analysis, and modeling are based on this component distinction [56].

As for the spatial relationships expressed by the spatial component, different classes are used to represent spatial information, such as direction, distance and topological relationships. In particular, topological relationships are meant to store properties among spatial objects which do not change when their shape undergoes transformations [85], such as scaling, translation and rotation. They are usually classified in three categories, namely adjacency, containment and connectivity [85]. In the following, both the Allen's [2] and the Egenhofer's [45] classification are recalled, due to the role they have for the scope of this work. Allen proposed the following relationships between one-dimensional temporal intervals. Let X and Y be two temporal intervals, the following relationships may hold:

- *before*, represented by symbols $<$ and $>$. For instance X ¡ Y means that X happens strictly before Y;

- *equal*, represented by symbol =. For instance $X = Y$ means that X and Y happen (start and end) simultaneously;
- *meet*, represented by symbols m and mi (m inverse). For instance X m Y means that Y starts when X ends;
- *overlaps*, represented by symbols o and oi (o inverse). For instance X o Y means that Y starts before that Y starts;
- *during*, represented by symbols d and di (d inverse). For instance X d Y means that X starts after that Y starts and X ends before that Y ends;
- *starts*, represented by symbols s and si. (s inverse). For instance X s Y means that X and Y start simultaneously;
- *finishes*, represented by symbols f and *fi*. (f inverse). For instance X f Y means that X and Y end simultaneously.

Table 2.1 summarizes the Allen's Relationships.

Relation	**Symbol**	**Symbol for Inverse**	**Pictorial Example**
X *before* Y	$<$	$>$	XXX YYY
X *equal* Y	=	=	XXX YYY
X *meets* Y	m	mi	XXXYYY
X *overlaps* Y	o	oi	$X\ X\ X\ X$ $Y\ Y\ Y\ Y$
X *during* Y	d	di	$X\ X$ $Y\ Y\ Y\ Y\ Y$
X *starts* Y	s	si	$X\ X$ $Y\ Y\ Y\ Y$
X *finishes* Y	f	fi	$X\ X\ X$ $Y\ Y\ Y\ Y$

Table 2.1: Allen's Relationships. (Source [2]).

In [45], Egenhofer proposed the 9-intersection model, which is a comprehensive model for binary topological spatial relations and applies to objects of type region, line and point. It characterizes a topological relationship between two point sets, A and B, by the set of intersections of A's interior (A°), boundary (δA), and exterior (A^-) with the interior, boundary and exterior of B. With each of these nine intersections (see Figure 2.1) being empty (ø) or non-empty (¬ø), the model has 512 possible topological relations between two point sets, some of which can never occur, depending on the dimensions of the objects, and the dimensions of their embedding space.

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

Figure 2.1: 9-intersection matrix proposed by Egenhofer. (Source [44])

Based on this model, eight topological relationships between regions are identified:

- *disjoint*, the intersection between region A and region B corresponds to an empty set, that is these two regions do not have points in common (Figure 2.2(a)),
- *meet*, the intersection among the boundaries of region A and region B is not empty, that is these two regions share at least a point of the boundaries (Figure 2.2(e)),
- *overlap*, the intersection between region A and region B is not an empty set (Figure 2.2(h)),
- *inside*, a region A is inside a region B if A is completely region in B. The reverse relationship of Inside is called Contains (Figure 2.2(c)),
- *contains*, a region A contains a region B if B is completely contained in A. The reverse relationship of Contains is called Inside (Figure 2.2(b)),
- *covers*, a region A covers a region B if regions A contains region B, and regions A and B share their boundaries (Figure 2.2(f)). The reverse relationship of Covers is called Coveredby,
- *coveredby*, a region A is coveredby a region B if region A is inside region B, and regions A and B share their boundaries (Figure 2.2(g)). The reverse relationship of Coveredby is called Covers,
- *equal*, two regions A and B are equal if A contains B and B contains A, that is these two regions are perfectly coincident(Figure 2.2(d)).

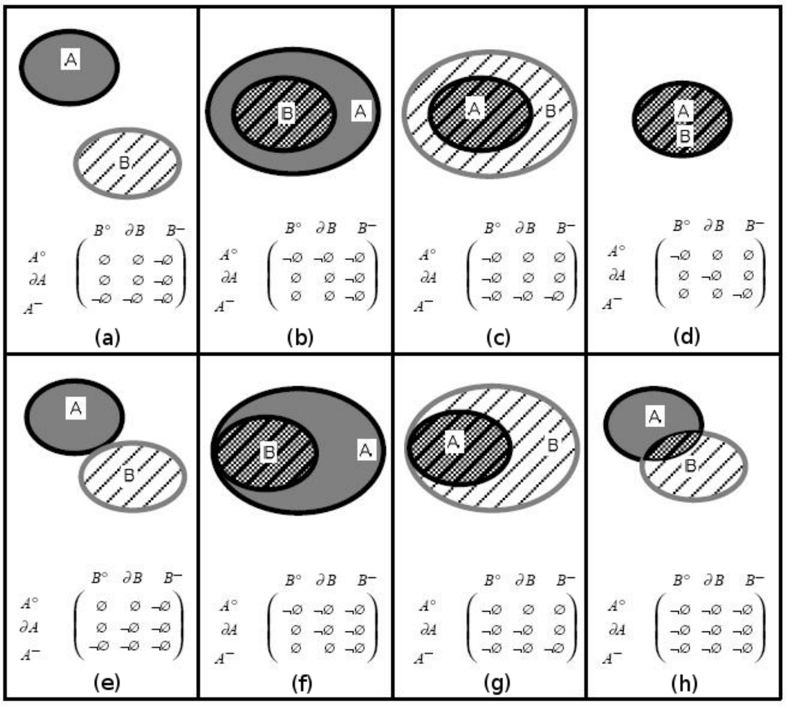

Figure 2.2: Egenhofer's Relationships between two regions. (Source [44])

2.2 Geographic Information Systems and Science

In recent years, geographic information and functionalities have become more and more popular and, as a consequence, integrated in a huge number of (software) applications and systems. Such a growing trend of using spatial information and functionalities has redefined and widened the concept of GIS over the time. Among simple definitions, the United States Geological Survey [133] defines a GIS as "a computer system capable of assembling, storing, manipulating, and displaying geographically referenced information, i.e., data identified according to their locations". In Rhind [110] a GIS is defined as "a system of hardware, software, and procedures designed to support the capture, management, manipulation, analysis, modeling, and display of spatially referenced data for solving complex planning and management problems". In Lo [85] a GIS is defined as a special class of information systems (see Figure 2.3), which contains all their characteristics and is capable to manage data related with real-world features or phenomena described in terms of locations.

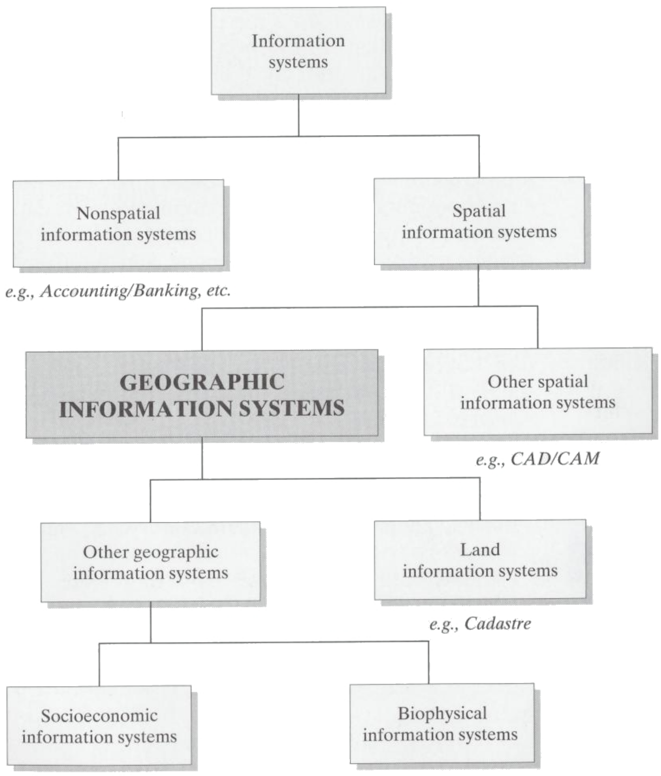

Figure 2.3: The role of GIS in Information Systems hierarchy. (Source [85])

In the last decades, the growing popularity of GIS has also affected the various disciplines which exploit the GIS technology, thus stimulating them to converge into a new field of science [85], which Goodchild [56] names Geographic Information Science or GIScience. It is defined as a set of research issues arising from the handling of geographic information. An alternative definition, given in [56], states that GIScience is the whole of knowledge that GIS implement. The aim of GIScience is providing the theoretical and organizational consistency for the scientific study of geographic information [85]. Once the GIScience context has been well established, a definition of GIS in terms of components has been given. As discussed in [85] a GIS is a system made up of four elements, namely data, technology, application areas and people. Regarding the data component, three basic formats are usually used to represent geographic data in GIS, namely vector, raster and surface. A shown in Figure 2.4, vector data describe geographic information by using points, lines and polygons, which are well-suitable to represent discrete features.

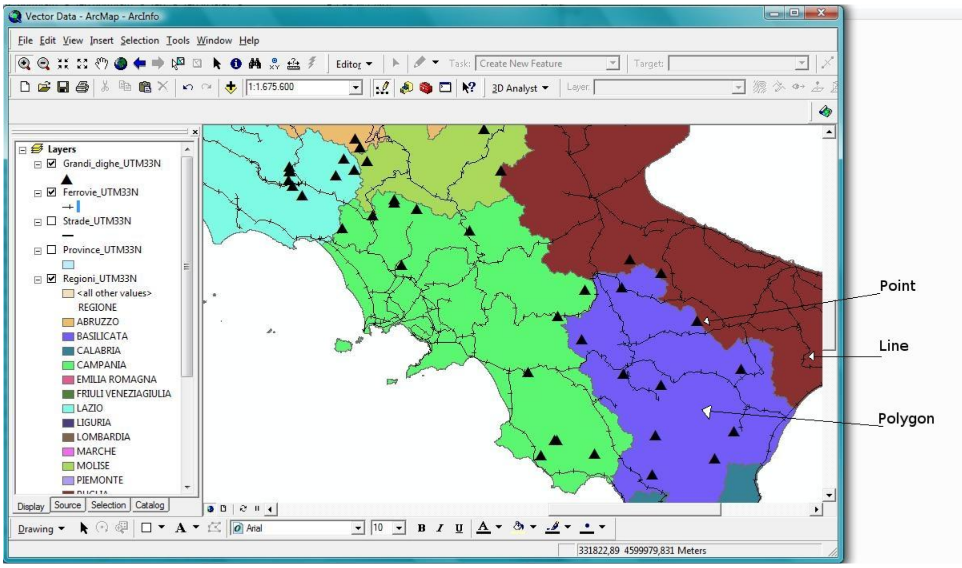

Figure 2.4: Vector data visualized by using ArcGIS ArcMap.

Raster data (see Figure 2.5) describe geographic information by using grids of cell values, which are best suited to represent continuous phenomena.

Finally, surface data, as depicted in Figure 2.6, describe geographic information by using 3D planes composed by points and lines, which are able to exploit the third dimension of the data.

Figure 2.7 shows how a GIS manages the geographic information by means

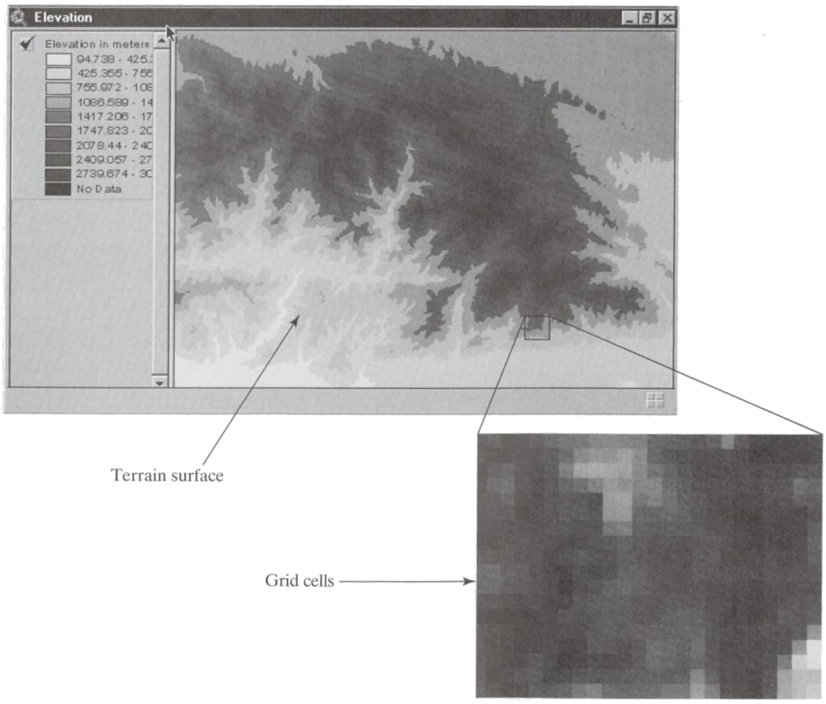

Figure 2.5: Raster data. (Source [85])

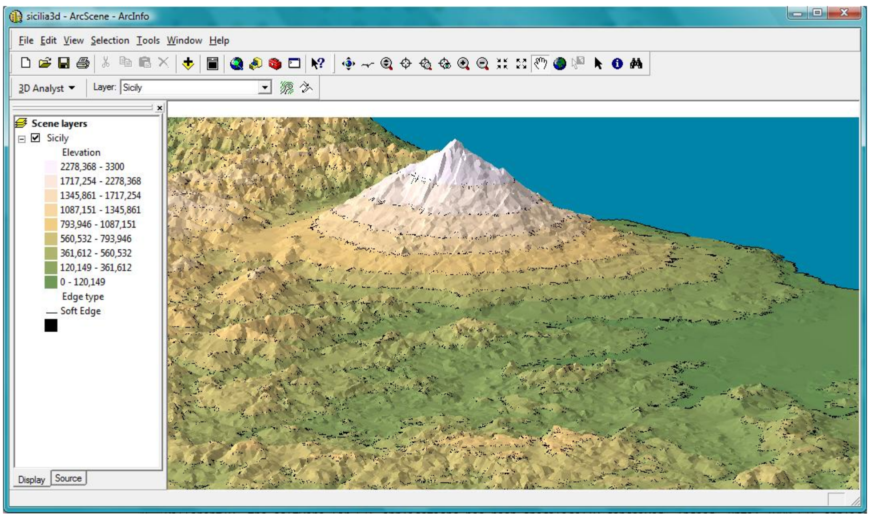

Figure 2.6: Surface data visualized by using ArcGIS ArcScene. (Source [85])

of different layers, according to the typology of data.

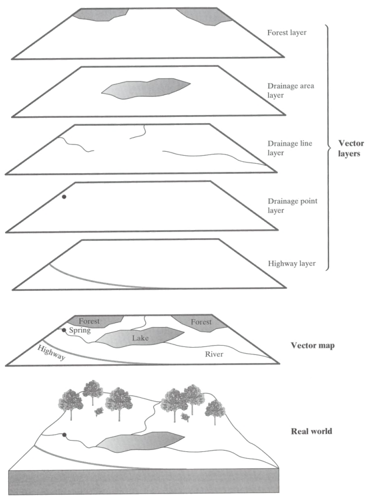

Figure 2.7: Different vector layers. (Source [85])

As for the technology component, the hardware underlying a GIS followed the typical development of the information systems. Differently, the software for GIS applications has been specifically conceived. Indeed, until 1990 GIS applications were conventionally developed through a mixture approach, which handled the descriptive component of data by using a DBMS, and the spatial component by means of a proprietary module. Nowadays, such a model has been substituted by a relational integrated approach, which consists of handling both spatial and descriptive component in a unique database.

As for the application domains, GIS was first developed from the early 1960s until 1980 in particular fields, such as census, surveying and mapping, land and resource management [85]. Nowadays, as shown in Figure 2.8, GIS are widely spread in all sectors of human activities.

Finally, based on the definition given in [85], the people component is classified in three categories, depicted in Figure 2.9: viewers, general users and GIS specialist. Viewers, who represent the largest class of users, sporadically browse a geographic database. General users, who usually affect the success of a GIS in an organization, use GIS for conducting business, performing professional services, and making decisions. GIS specialists, such as GIS managers and programmers, are people who actually make GIS to work. They play the most direct role in the success of GIS implementation.

Sectors	*Application Areas*
Academic	Research in humanities, science and engineering Primary and secondary schools—school district delineation, facilities management, bus routing Spatial digital libraries
Business	Banking and insurance Real estate—development project planning and management, sales and renting services, building management Retail and market analysis Delivery of goods and services
Government	Federal government—national topographic mapping, resource and environmental management, weather services, public land management, population census, election, and voting State/provincial government—surveying and mapping, land and resource management, highway planning and management Local/municipal government—social and community development, land registration and property assessment, water and wastewater services Public safety and law enforcement—crime analysis, deployment of human resources, community policing, emergency planning and management Health care International development and humanitarian relief
Industry	Engineering—surveying and mapping, site and landscape development, pavement management Transportation—route selection for goods delivery, public transit, vehicle tracking Utilities and communications—electricity and gas distribution, pipelines, telecommunications networks Forestry—forest resource inventory, harvest planning, wildlife management and conservation Mining and mineral exploration Systems consulting and integration
Military	Training Command and control Intelligence gathering

Figure 2.8: Major Application Areas of GIS. (source [85])

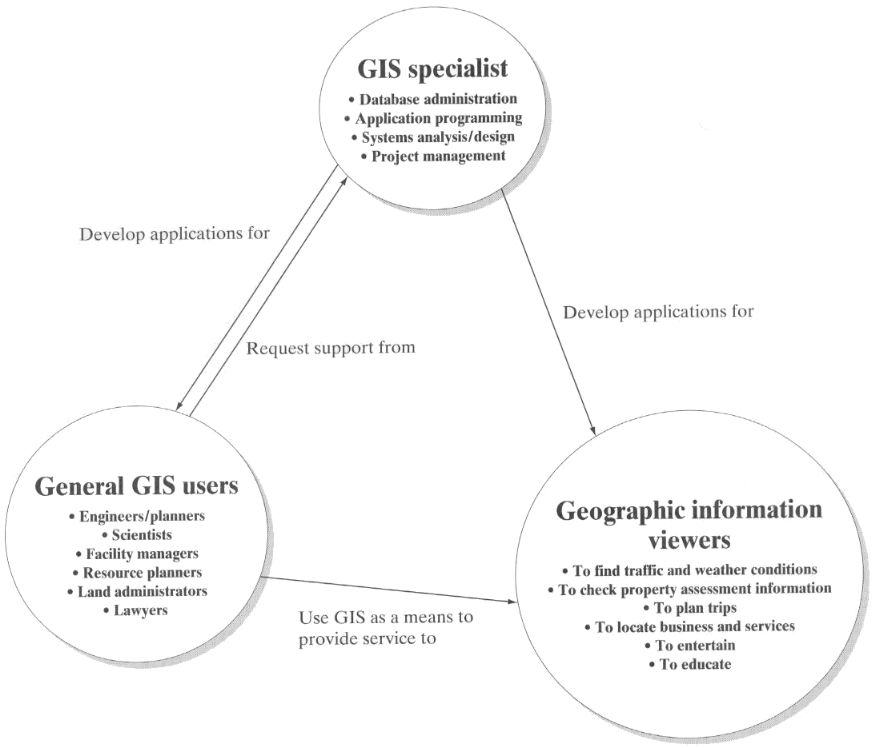

Figure 2.9: GIS users and their relationships. (source [85])

2.2.1 GIS and Spatial DBMS

As mentioned in Rigaux et al. [112], a GIS is able to perform several tasks, such as data input, verification, storage, management, transformation and presentation. Usually, some of these tasks, such as data storage and management, are handled by a database management system (DBMS). Nevertheless, the use of a pure relational DBMS is suitable for managing just the alphanumeric component of a GIS, it is not appropriate for handling the spatial component, mainly because of performance issues in managing spatial types and typical geometric computations [112]. Two methods are used for handling spatial components, a loosely coupled approach and an integrated approach [112]. The former consists of handling the alphanumeric component by a (relational) DBMS and the spatial component by a specific module. Examples of GIS software which use this approach are ArcInfo (ESRI), MGE, TiGRis (Intergraph)[112]. The latter consists of integrating a relational DBMS with a spatial extension.

The method based on the spatial extension is more interesting and suitable to our scope, so in the following it is briefly described. Most of the general-purpose relational DBMS vendors and developer groups, such as Oracle[1] and PostgreSQL[2], propose spatial extensions (Oracle Spatial[3] and PostGIS[4] respectively) in order to integrate them into their data management modules for querying and managing geographic data. The spatial extension allows for integrating into the usual data management modules:

- spatial types, such as points, lines, polygons, arc line, arc polygon,
- spatial indexes, such as R-tree or Quad-tree, meant to limit searches within tables based on spatial criteria,
- spatial functions, such as topological relationships contains, disjoint, equal, overlap, covers, typical spatial operators, such as buffer, and set operations, such as union, difference and intersection, which can be directly integrated into the SQL statements.

[1] Oracle Corporation [online], website, `http://www.oracle.com/index.html`, accessed 01 march 2009.

[2] PostgreSQL [online], webpage, `http://www.postgresql.org/`, accessed 01 march 2009.

[3] Oracle Spatial [online], webpage, `http://www.oracle.com/technology/products/spatial/index.html`, accessed 01 march 2009.

[4] PostGIS [online], webpage, `http://postgis.refractions.net/`, accessed 01 march 2009.

2.3 Knowledge Discovery in Databases and Data Mining

Nowadays, a large amount of digital data can be collected and handled across a huge variety of fields, from business and scientific fields, to satellite images and text reports. Then, a new generation of computational theories and tools are needed to support humans in extracting useful information (knowledge) from these rapidly growing volumes of digital data. Knowledge discovery in databases (KDD) is the resulting research field dealing with these theories and tools [50]. Extracting knowledge from a set of digital data correspond to find a set of meaningful patterns. A pattern can be defined as a statistic summary, such as the mean, median, or standard deviation, or a simple rule such as "If customer bought X, he/she is also likely to buy Y and Z", which may represent an unexpected and interesting finding [118]. In the past, several names has been given to the task of finding useful patterns in data, namely data mining, knowledge extraction, information discovery, information harvesting, data archaeology, and data pattern processing [50]. According to the definition of Fayyad, *et al.* [50] KDD concerns the overall process of discovering useful knowledge from data, while data mining concerns a particular step in this process, that is, the application of specific algorithms for extracting patterns from data. Shekhar [118] defines data mining as "the process of discovering potentially interesting and useful patterns of information embedded in large databases". In this dissertation, we will consider the KDD term equivalent to data mining, thus using both terms without distinction. Data mining promises the capacity to rapidly and automatically search for local and potentially high-utility patterns using computer algorithms [118]. The whole process of data mining, as shown in Figure 2.10, combines many different sub processes, such as data extraction and data cleaning, feature selection, algorithm design and tuning, and the analysis of the output when the algorithm is applied to data [118].

2.3.1 Spatial Data Mining

The explosive growth of spatial information and functionalities, the widespread use of large spatial databases and the advent of fast computing have inspired new ways of thinking about spatial data and analysis, emphasizing the need for the automated discovery of spatial knowledge in terms of data mining [56]. According to the definition of classical data mining, spatial data mining can be defined as the process of discovering interesting and previously unknown, but potentially useful patterns, from large spatial datasets. In spite of such a similar definition, there is a

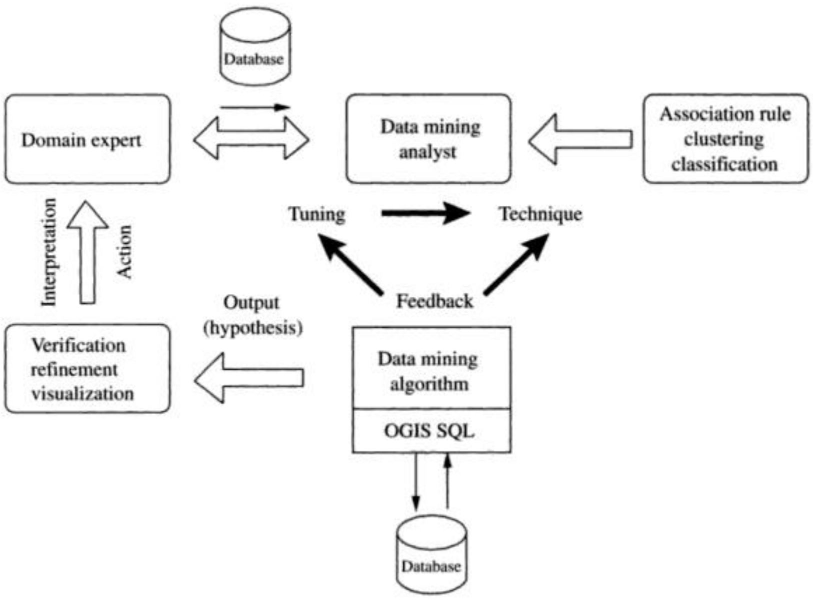

Figure 2.10: Data mining process. (source [118])

significant difference between classical and spatial data mining. In fact, the complexity of spatial data types, spatial relationships and spatial autocorrelations makes more difficult the extraction of interesting and useful patterns from spatial databases than extracting the corresponding patterns from traditional numeric and categorical databases [118].

In literature, some well-known historical examples of spatial data mining findings have been provided in [60]. The authors quote discoveries occurred well before the advent of computers, such as the cause of cholera which infested London in 1855. In the context of spatial databases, a pattern can be defined as a rule which also integrates spatial associations, such as “if a house is close to a beach, then the cost of the house is expensive” or “if a country is adjacent to the North Mediterranean Sea, then it is a wine-exporter” [118].

2.4 Digital Cartography

According to the definition proposed in [67] by Imhof, the cartography is “the art and technique of map reproduction. It is in a narrower sense to be characterized as a refining process between the original source material and the reproduction". Digital Cartography deals with the creation and the analysis of maps through the use of computers. In this dissertation several digital cartography issues are considered. In the following, some well-known activities related to the digital cartography are mentioned, namely Cartographic Generalization, Spatial Conflict Resolution, Map labeling and Flow Displaying.

2.4.1 Cartographic Generalization

In the cartographic visualization the generalization corresponds to the process meant to introduce modifications into a map in order to improve its readability when it is represented at a reduced level of scale. This is one of the most complex processes in cartography because of several reasons. In fact, during this task some operations must be maintained:

- the accuracy, the information content, and readability of the map [19],
- logical and unambiguous relationships between the map objects [140],
- the aesthetic quality of the map [140].

Moreover, also the automation of generalization is a very difficult task, because it is an holistic process due to its subjective nature [19], involving considerable decision making [85]. In order to lower the complexity of the generalization process, it is decomposed in several operation classes, depending on the involved data and the reason for generalizing [85]. In literature, several classifications of these operations have been proposed [96, 113, 117, 140]. Generally, they can be summarized in the following:

- *Simplification*, which eliminates unnecessary details of a spatial object, by selecting a subset of the coordinates that compose it, and preserving its original shape;
- *Smoothing*, which correct small imperfections by shifting the coordinates of the involved spatial object;
- *Aggregation*, which groups adjacent spatial objects and creates a new single one;
- *Amalgamation*, which joins small features into a larger map element;
- *Merging*, which is usually applied to parallel line features;

- *Collapse*, which decomposes area or line features into point features;
- *Refinement*, which adjusts the geometry of an object in order to improve its aesthetic visual aspect;
- *Typification*, which reduces the density of spatial objects, as well as their levels of detail;
- *Exaggeration*, which amplifies the size the geometric representation of the spatial object involved, in order to focus on its importance and to improve its legibility;
- *Displacement*, which shifts objects in a map in order to maintain the separation between them and gain clarity;
- *Classification and symbolization*, which group together objects sharing similar or identical geographic attributes and characteristics, into a new object.
- *Elimination*, which eliminates spatial objects because of their small size or lesser importance in the map;
- *Size reduction*, which reduces the size of the involved spatial object.

In the scope of this work, some of these operations play a more relevant role, namely simplification and displacement. The line simplification is the most frequently used operator in generalization. It consists of a selective reduction of points (pairs of coordinates) that represent the line, preserving as better as possible the original shape of the line involved. Object displacement is another important and well investigated aspect about the map generalization, meant to resolve spatial conflicts among objects which are too close to be well distinguished and visualized. In literature, a variety of algorithms implementing the same operation have been proposed, depending on the data model and its usage. Among the others, it is worth to recall [81, 39, 107, 109, 28] for simplification operation and [91, 86, 114, 8] for the object displacement operation.

2.4.2 Spatial Conflicts Resolution

Strictly related to the generalization process is the topic known as spatial conflicts resolution in map displaying. Also this topic plays a significant role in the context of this dissertation. In literature, the study of spatial conflicts managing and resolution in map displaying is a well-known issue. It deals with the management of spatial conflicts which may occur among map objects during a spatial operation, i.e. a zooming out operation on a

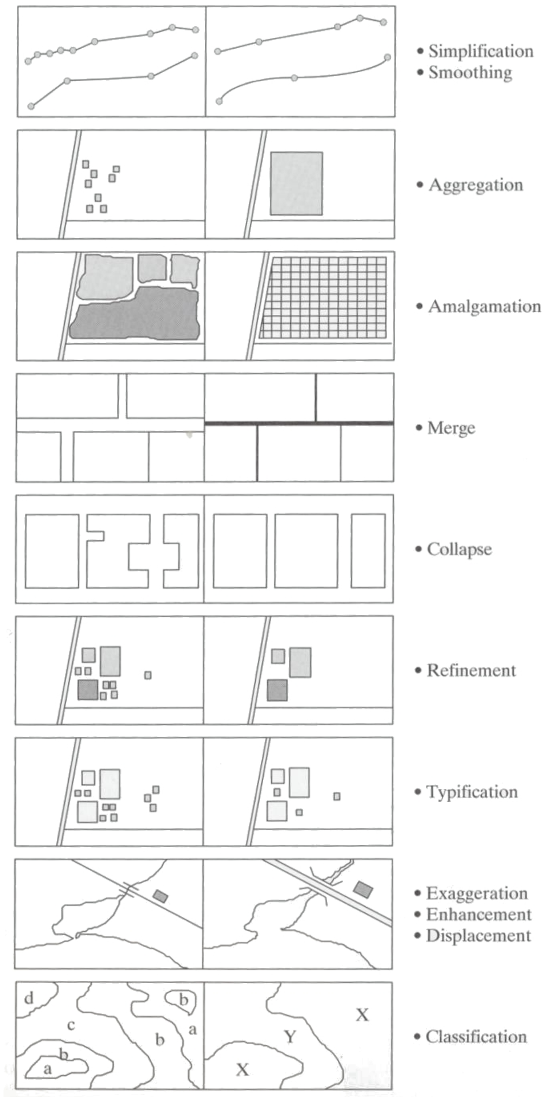

Figure 2.11: A list of typical generalization operators. (source [85])

map, and with the resolution of these conflicts. Many works are essentially based on the resolution with the displacement operator [91, 114, 142, 139, 61, 86], which preserves the original shapes of spatial objects involved, so reducing the loss of spatial information [69]. Other operations have been proposed as resolution of spatial conflicts, such as elimination, amalgamation, and size reduction [40, 139, 115], which improve the onerous computational time of the displacement operator [69].

2.4.3 Map Labeling

The map-labeling problem, also known as *name placement*, represent another important topic connected to digital cartography techniques and in particular to the map layout design process. Dealing with the correct readability of the maps, it is a significant visual problem cartographic design [85], which consists of assigning positions for the labels of graphical features of a map so that these elements can be uniquely identified [34]. A wide work on map labeling exists in literature [34, 92], starting from Imhof [67, 68] and Yoeli [149], which defined the basic rules for labeling maps which have influenced the development of labeling algorithms up to now, namely readability, unambiguity, and avoidance of overlaps. Several solutions to the to the map-labeling problem, are presented in literature, based on expert systems [1, 35, 154], discrete gradient descent methods [25], simulated annealing [24, 25, 155], integer linear programming [152, 153, 120], approximation algorithms [137, 151, 143, 48], tabu search [147] and genetic algorithms [134, 148].

2.4.4 Flow Displaying

Flow displaying deals with representing geographical movement on the maps, such as movement of people, ideas, money, energy, or material [129]. Geographical movement is crucial topic. In [99] affirm that the French engineer Charles Minard was one of the first to use maps to depict social movement phenomena. Nowadays, several work discusses about showing geographical movement in computerized maps [127, 93, 58]. Usually, several ways of rendering geographical movement on maps exist [129]. A *choropleth map* with shading or colors is common technique which shows change of state, but not movement. Moreover a map containing appropriately scaled plus and minus symbols to just show change. Maps which explicitly contain Flows are generally classified in discrete maps and continuous maps. The former consists of bands or arrows whose width is proportional to the volume moved. The latter use vector fields or streamlines to show continuous flow patterns [126, 36, 128]. Finally, Tobler [127] names "migration maps", maps which represent patterns of geographical movement by arrows or bands between places, using information arriving in from-to tables.

2.5 Visualization of Geographic Information

It is a generally shared opinion that"maps are the most effective and intuitive way of presenting geographic information", as natural outcome of geographic data processing [85]. Consequently to the advent of computer technologies, digital maps have gradually replaced printed maps becoming the principal method for visualizing geographic information on screens.

In particular, as suggested in [85], the flexibility and expressive power of a digital map allow for easily abstracting or converting geographic information into other forms of products such as graphs, charts, statistical tables and written reports. Moreover, by integrating sophisticated data analysis tools with digital maps, new application areas may benefit from this means, such as real-time decision support, animation of spatial processes, spatial simulation, and predictive modeling, as well as integration with non spatial statistical analysis [85].

Generally speaking, visualization of geographic information traditionally belongs to the research area known as geographic visualization, or geovisualization [79, 78], which refers to a set of tools and techniques supporting geospatial data analysis through the use of interactive visualization. As in the related fields of scientific visualization [90, 121] and information visualization [17, 62] geovisualization is focused on knowledge construction and combines geospatial information with human vision and domain expertise, to allow for data exploration and decision-making processes [90, 88, 71]. In particular, geovisualization integrates approaches from scientific visualization, (exploratory) cartography, image analysis [20], information visualization [42, 10, 26], exploratory data analysis (EDA) [64] and GIS [85, 87, 57] to provide theory, methods and tools for the visual exploration, analysis, synthesis and presentation of geospatial data [89].

Most of the major challenges in the field of the geovisualization are investigated by the International Cartographic Association (ICA)[5] Commission on GeoVisualization[6].

[5] International Cartographic Association [online], website, http://www.icaci.org

[6] International Cartographic Association, Commission on Geovisualization [online], website, http://geoanalytics.net/ica/

In modern software, geographic data are represented by using both traditional cartographic techniques based on the use of colours, textures, symbols, and diagrams, and by exploiting computer-enabled techniques such as map animation and interactive 3D views. Moreover, maps are used in combination with non-geographic visualization techniques such as scatterplots [21] or parallel coordinates [138], which provide different perspectives into the data through the use of multiple interactively linked views. In particular, mainly in geovisualization environments, maps are used to stimulate visual thinking about geospatial patterns, relationships and trends [79].

Strictly connected to the geovisualization, many different scientific communities studied visual approaches to exploratory data analysis [53], associating them with different terms, *Visual Data Mining* for the database community [74, 17, 111], *Exploratory Visual Analysis* for the statistic community [132, 6, 97, 131], and *Exploratory Spatial Data Analysis* for the geography community. In [52], Gahegan lists several distinct types of visual data mining and exploratory analysis techniques:

- Map-based techniques,
- Chart-based techniques,
- Projection techniques,
- Pixel techniques,
- Iconographic techniques,
- Hierarchical and network techniques

In the scope of this thesis, the map based techniques and the iconographic techniques are particularly interesting. The former allows the user to interactively change the mapped data and its visual appearance [43], also by using the map legends as an interaction means, as shown in [104] and [3]. The latter uses complex symbols, such as stick figures [105] or faces [22, 38] to encode many data dimensions simultaneously. Iconographic techniques are meant to facilitate perception of the 'whole', while still preserving individual variables.

2.6 Visual Languages

A visual language is defined in [65] as a “tightly integrated communication unit” composed of words, images and shapes. Moreover, he asserts that

Visual Languages represent a promising means towards the improvement of human performance in the next years. In [94] a clear definition of Visual Languages is presented as "languages with alphabets consisting of visual representations that are used for human-human or human-computer communication". Currently, several disciplines belonging to different research areas study visual languages. In particular, in computer science, in order to permit communication, interaction and reasoning between humans and computers [94], visual languages adopt different modalities, such as:

- Direct manipulation, allowing the user to directly interact with visual objects in order to perform an action,
- Visualization of information, allowing the user to easily discover hidden characteristics and relationships in textual representation,
- Visualization of software, allowing the user to better understand and debug programs, by visually showing their static and dynamic characteristics,
- Diagrammatic representation and reasoning, allowing the user to better create inferences and prediction by using pictorial displays,
- Graphical simulations, allowing the user to easily understand time-varying processes thanks to the natural dynamic visual representation of the processes being simulated.

Moreover, in [94], the author identifies significant properties of visual languages that outline guidelines, in particular related to HCI:

- "Explicit representation of salient relations of the domain through visually perceivable properties",
- "Spatially localized and visually cued organization of relevant information about domain objects and relations",
- "Facilitation of visual transformation to simulate domain processes",
- "Reduction of complexity through recognition and guidance".

In the scope of this thesis, these aspects of visual languages are of interest. In particular, in the reminder of this section, preliminaries on visual languages related to both (spatial) database quering and visual summaries will be presented.

2.6.1 Visual Queries Languages

Visual query languages represent one of the most effective methods used to reproduce the user's mental model to discover and manipulate data. In an exhaustive survey [18] about visual query languages, significant work has been analyzed as well as relevant issues have been outlined for the design of new visual query systems. A significant aspect of visual query languages is to associate data with visual notations (icons, diagrams, graphs) which recall their meaning. Such visual notations can be spatially manipulated and arranged to compose queries, which are then automatically translated into the underlying textual query language. The expressiveness of visual languages allowed the researcher to explore the applicability of this approach also for spatial databases in several application domains, such as image databases, remote sensing, GIS. In such applications, a visual notation allows to summarize into a homogeneous structure expressed by graphics and labels, several heterogeneous components, ranging from meaning, to geometry, to location and properties. Recently, also the third component of the space, necessary in representing and analyzing particular data [101, 32, 31], was investigated by using 3D icons into a virtual scenario.

In recent literature, many text-based spatial query languages have been proposed. Basically, all those query languages extend the standardized SQL query language, by integrating the SELECT-FROM-WHERE construct with new operators and command sets. Sketch based visual languages represent an important visual approach for GIS querying. Such languages are mainly based on the query-by-example approach where users draw particular configurations representing an example of the result that should be displayed. In [100], the author describes some characteristics of recent visual query languages belonging to the sketch and drawing based approach, such as Sketch![95], Spatial-Query-By-Sketch [46], Cigales [16], LVIS [7], VISCO [141]. A relevant advantage of such an approach is to easily express similarity-based queries, even if it is not suitable for expressing composite queries. A different approach, adopting the visual paradigm for spatial query specification and result visualization, is followed in Spatial Exploration Environment (SEE) [73], and in Phenomena [102]. Between them, Phenomena seems to be more suitable, because it allows users to direct manipulating visual elements also when performing complex spatial computational tasks.

2.6.2 Visual Summaries

Conventional cartography adopts projections in order to minimize the distortion from the reality. However, when dealing with cartography referring to complex issues, the message to deliver thorough a map is more important than the accuracy of details. So, this message can be more understandable when a certain level of distortion is accepted, showing only the more salient features or characteristics. The usage of visual metaphors represents an effective solution in supporting users to locate facts, trends and patterns. An immediate synthesis of data of interest, disregarding details, is a real support for human activity to model and analyze the reality of interest. Such a synthesis may then represent the starting point for further processing tasks aimed to derive spatial analysis data, and to support expert users in decision making. In the remainder of this section, we describe two cartographic representations, namely *cartograms* and *chorems*, which exploit a relevant property of visual languages, the visual metaphors. Such metaphors are adopted in order to schematize some territory where facts, trends and patterns are located so maximizing the understandability of the message to deliver.

Cartograms

Cartograms, invented by Tobler [124, 125], are maps where a particular distortion is chosen and is made explicit. In conventional cartography, territories are represented proportionality to their area, whereas in cartograms, areas are represented proportionally from other data. In an exhaustive survey [130], "thirty five years of cartogram" have been analyzed. According to the definitions in [5], a cartograms is "a small diagram, on the face of a map, showing quantitative information" or "an abstracted and simplified map the base of which is not true to scale". Examples of cartogram, which can be found in the Worldmapper collection[7], are shown in Figure 2.12 and in Figure 2.13, where the world repartition of wealth in 2002 and the world population in Spring 2000 are represented, respectively.

Other types of cartograms can be found in the NCGIA collection[8], namely Contiguous Cartograms, Non contiguous Cartograms and Dorling cartograms. Among them, Dorling cartograms [37] are based on circles selected to be proportional to some data. Figure 2.16 gives an example

[7]Worldmapper. [online], webpage, `http://www.worldmapper.org/index.html`, accessed 01 march 2009.

[8]NCGIA. Cartogram gallery [online], website, `http://www.ncgia.ucsb.edu/projects/Cartogram_Central/index.html`, accessed 01 march 2009.

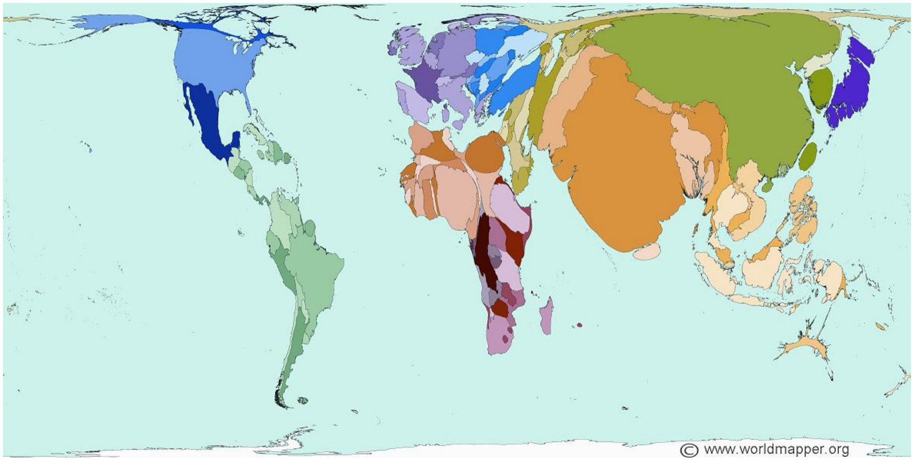

Figure 2.12: Cartogram showing population in Spring 2000. (Source [145])

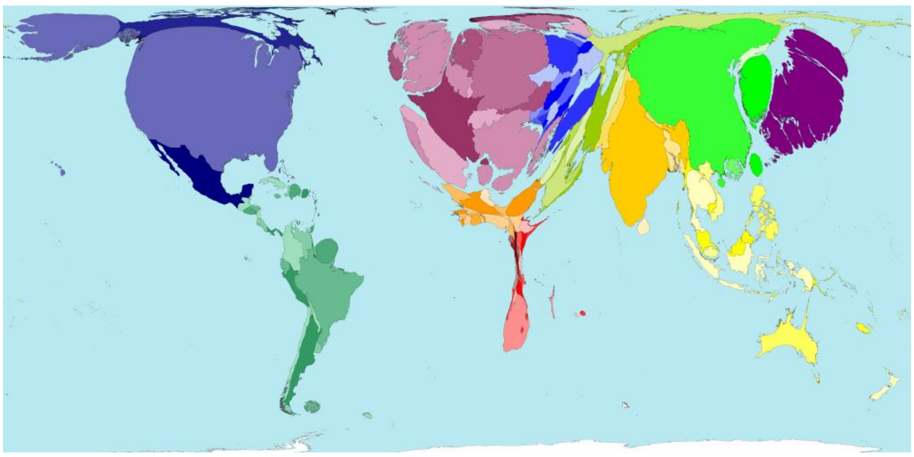

Figure 2.13: Cartogram showing wealth in 2002. (Source [145])

extracted from the 2004 presidential election in the USA. For designing a cartogram, the most important problem is to preserve topology, i.e. vicinity with the neighboring territories. For this purpose, one of the most popular methods is based on the Gastner and Newman's algorithm [54]. According to this method, the construction of a (flat 2D) cartogram involves finding a transformation of a plane to another plane, such that the Jacobian of the transformation is proportional to a specified (population) density. This choice of normalization for the Jacobian ensures that the total area before and after the transformation is the same. Cartograms

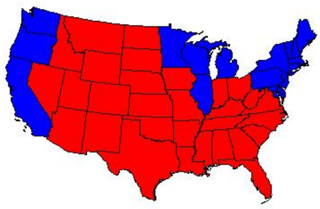

Figure 2.14: Conventional map showing the result of the 2004 presidential election in the USA.

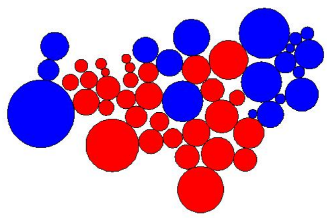

Figure 2.15: Dorling cartogram in which circles are proportional to the population rather than area.

Figure 2.16: Example of Dorling cartogram

represent a very interesting cartographic representation meant to show the relative importance of a variable, and make easy visual comparisons between territories.

Chorems

While cartograms show values of a single variable at a time, chorems allow designers to assemble into a single map more than one thematic layer, thus representing the relative importance of a set of objects and phenomena related each other. The term *chorem* derives from the Greek word $\chi\omega\rho\eta\mu\alpha$, (read chorema) which means space or territory. According to the definition of the French geographer Roger Brunet, who invented chorems [13], a chorem is a schematized territory representation, which eliminates any detail not necessary to the map comprehension. In this definition, the term *schematized* means that the more important characteristic is a sort of synthetic global vision emphasizing salient aspects. Moreover, Brunet defines the chorems as "elementary structures of the space represented by graphic model" used so that "all the spatial configuration rises from the combination, eventually very complex, of simple mechanisms".

Figure 2.17 shows a set of chorems, related to the France country, where some aspects are highlighted, such as the simplified geometric shape, the most important cities, the different development areas, the flows representing displacement of people and equipment.

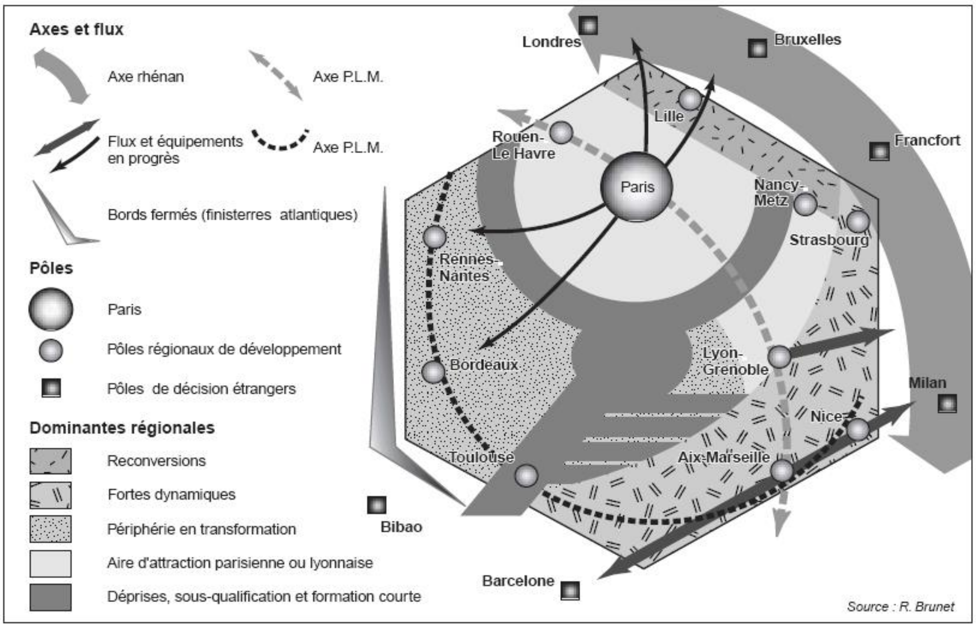

Figure 2.17: Example of map containing chorems about France. (Source [72])

This capability of chorems may be exploited to provide decision makers with a means to acquire syntactic information (what, where and when), as well as semantic aspects (why and what if), useful to support human activity to model, interpret and analyze the reality of interest. As the chorems are central subject of this thesis, more information about them will be given in the reminder of this dissertation.

2.7 Multi-agent Systems

Agent technology is an important area of research in computer science, in particular in disciplines such as artificial intelligence, robotics and psychology. Several definitions have been assigned to the agent concept in literature. An agent is generally defined as an autonomous behavior with a set of peculiar characteristics. In particular, according to [144], agents are software entities holding own will, situated in a given environment, and capable to act autonomously in their environment in order to reach the goals for which they have been conceived. Formally, agents are defined as hardware or software systems with the following properties:

- *autonomy*, they operate without human intervention;
- *proactiveness*, they take initiative to reach a prefixed goal;
- *reactiveness*, they perceive the environment and opportunely change behavior to meet designed objectives;
- *social ability*, they interact and cooperate with other agents to reach their common goals.

A multi-agent system can be described as a distributed system containing several agents which interact in order to accomplish their private and possibly their common goals. During such interactions, the agents communicate directly each other, either through the intermediation of other agents or through the environment.

2.7.1 Multi-agent Systems and Cartography

In cartography, a multi-agent system may offer many advantages, mostly in the domain of cartographic generalization [70]. As discussed in Subsection 2.4.1, the cartographic generalization is a very complex process as well as its automation, due to its holistic nature. A multi-agent systems may offer a proper support to face this complexity through its properties. As a matter of fact, the autonomy provided by a multi-agent system represents a framework for cartographic generalization [108]. Moreover, as discussed in [66] a multi-agent system can follow the holistic nature of the cartographic generalization process. In literature, a large amount of research, based on multi-agent approach, has been carried out in specific topics of cartography, such as automatic cartographic generalization [115, 80, 40, 116], spatial conflict resolution [91, 114, 61, 84, 142, 139] and on-the-fly map generation [55, 70]. In particular, as discussed in [70], in the AGENT project [115, 80, 41] geographic objects contained in databases are translated into software agents aiming to satisfy the existing cartographic constraints as much as possible. In the project proposed in [84], the software agents implement a web-based map information retrieval technique (MAPBOT), whose goal is to search for geographic information. Finally, in [70], a system meant both to generate map on-the-fly and to resolve spatial conflict is presented.

2.8 Conclusion

In this Chapter the context of the research of this dissertation is introduced. Preliminaries on several topics are presented, namely Geographic Data and Functionalities, Geographic Information Systems and Science, Knowledge Discovery in Databases and Data Mining, Digital Cartography, Visualization of Geographic Information, Visual Languages, and Multi-agent Systems. In particular, some aspects of Data Mining, Visual Languages and Multi-agent Systems related with spatial information and functionalities are thoroughly discussed, such as spatial DBMS (SDBMS), spatial data mining, visual languages as visual summaries and query languages for DBMS and SDBMS. In the next chapter the concept of chorem will be presented, in terms of meaning and history, and new definitions and classification will be introduced, meant to support the usage of chorems as visual summary of geographic databases.

Chapter 3

A Formal Specification of Choremes

Contents

This chapter introduces the concept of chorem and discuss aspects that led us to choose chorems in this research work. Section 3.1 describes the meaning and the history of the chorem concept, also showing some examples of map containing chorems. Sections 3.2 and 3.2 illustrate a formal chorems definition and illustrates a formal chorems classification, both resulting from researches and studies carried out. Section 3.3 points out the reasons why the chorems have been chosen as base for this research and how they can be used to visually summarize the contents of geographic databases.

3.1 Chorem History and Meaning

The chorems framework is based on very simple shapes and on a minimal number of composition rules. According to Brunet[13], seven shapes are sufficient to describe the models representing the chorems:

- *point*, meant to represent places, spots, poles;
- *line*, meant to represent contacts, breaking, ties, border, relations;
- *area*, meant to represent extensions, shapes;
- *flow*, meant to represent dynamics, movements, symmetry, intensity;
- *passage*, meant to represent bridges, crossing, bifurcations, tunnels;
- *polarization*, meant to represent focusing, dynamics;
- *gradient*, meant to represent dissymmetry, attraction, repulsion.

Moreover, Brunet proposed 28 elementary chorems, (see Figure 3.1) as a combination of four shapes, namely:

- *point*,
- *line*,
- *area*,
- *network* (a combination of point, line and area),

and seven geospatial dynamics, namely:

- *maillage*: region subdivision patterns,
- *quadrillage*: region's infrastructure patterns,
- *gravitation*: gravity patterns,
- *contact*: front of communication patterns,
- *tropisme*: unilaterally biased movements patterns,
- *dynamique*: conquest diffusion patterns,
- *hirarchie*: hierarchies patterns.

Each chorem represents a basic model of spatial configuration, so by integrating them into a map, a huge variety of spatio-temporal phenomena as well as spatial configurations of geographic objects may be represented at different scales. Generally, a set of chorems can be defined as a visual vocabulary which allows expert users to describe significant aspects of a territory, leaving apart minor characteristics. Starting from Brunet's chorems, some other authors presented modified versions of such table. In particular, Piveteau and Lardon [106] presented a methodology based on a simplified table of seven basic chorems and four *question registers*, as shown in Figure 3.4.

Cheylan et al. [23] presented a table of 21 basic chorems (see Figure 3.5). Fontanabona [51] presented a table of 24 basic chorems. Figure 3.6 shows the 24 chorems presented in [51]. Other authors, starting from the idea of chorems but not considering Brunet's table, formally define and associate a set of chorems with a specific domain. In particular, Klippel et al. presented wayfinding chorems, which specifically refer to route direction domain. They define wayfinding chorems as "a representation vocabulary that characterizes route knowledge and that is based on mental conceptualizations of directions at decision points" [76]. Figure 3.7 shows the basic set of wayfinding chorems.

Thery [122] and Grataloup [59] presented *chrono-chorèmes* and *paléo-chorèmes*, which specifically refer to time domain, in particular related to long-term temporal phenomena in historical topography. Figure 3.8 shows an example of a chrono-chorème, where the development of *Tours* city (France) is described. In particular, the Figure 3.8 shows the *Tours* city, positioned near a river and a crossing, which has been expanded in subsequent phases particularly on a side of the river.

Summarizing, in literature several uses of the concept of chorem can be found, especially from geographers, whose fundamental goal is using

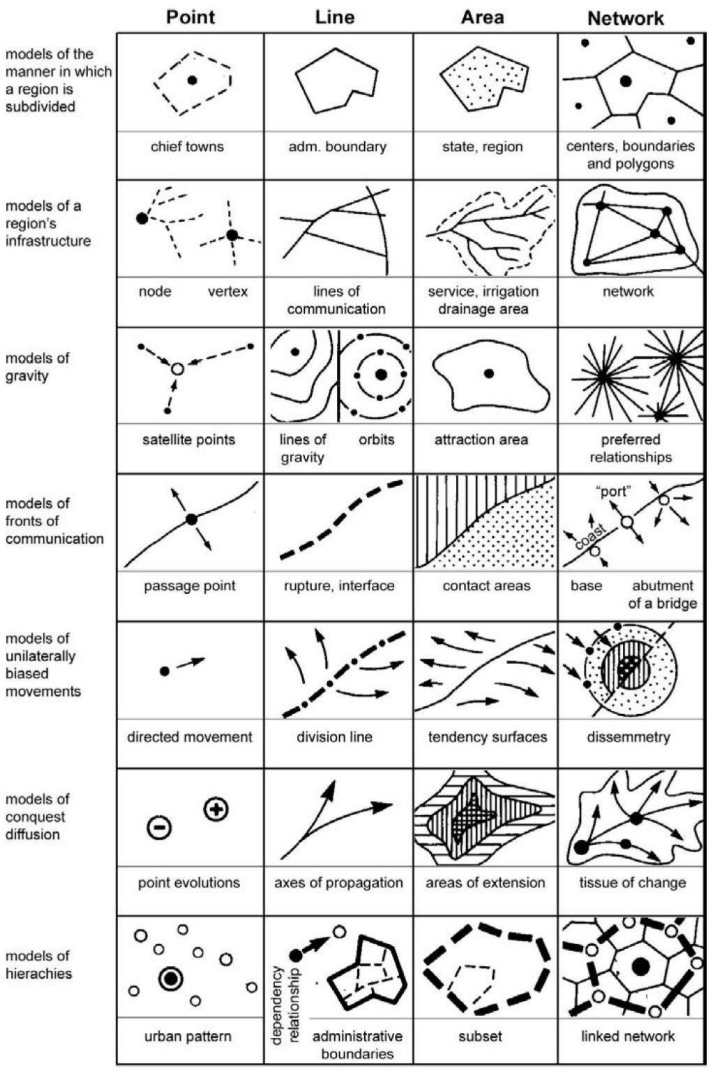

Figure 3.1: The basic set of Brunet. (Source: [13])

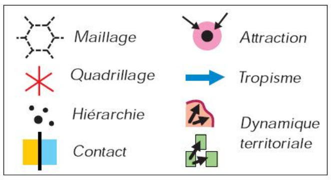

Figure 3.2: The simplified table of seven basic chorems.

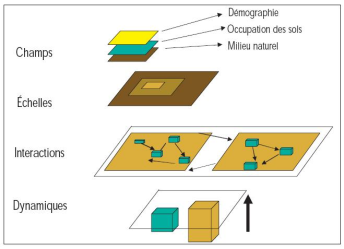

Figure 3.3: The four *question registers*.

Figure 3.4: The methodology presented in Lardon *et al.*.

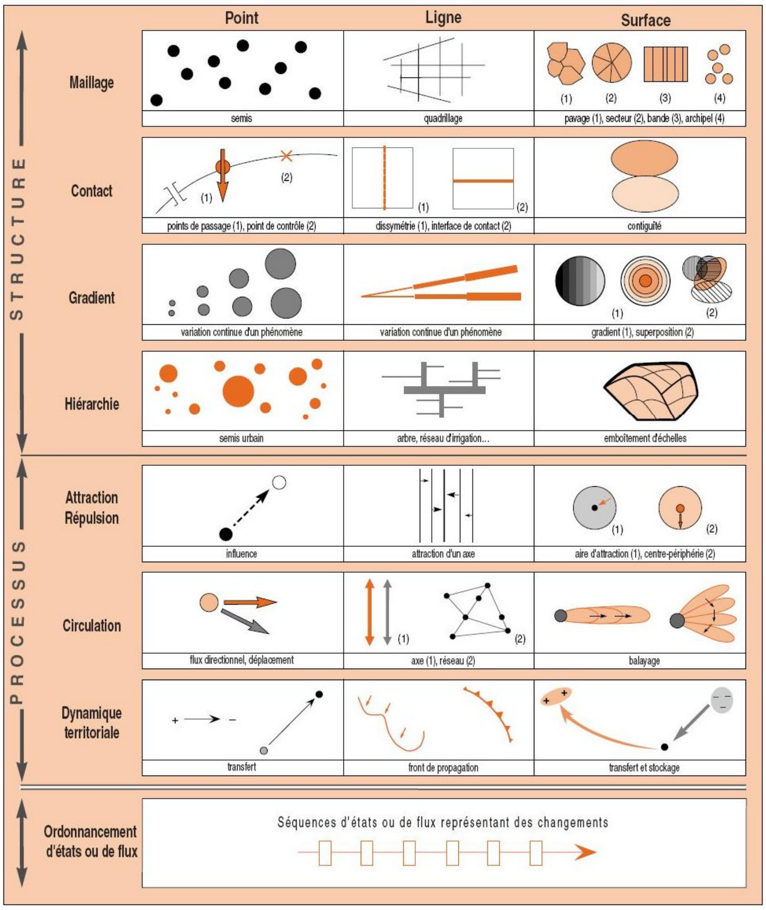

Figure 3.5: A table of 21 basic chorems presented in Cheylan *et al.* (Source: [23])

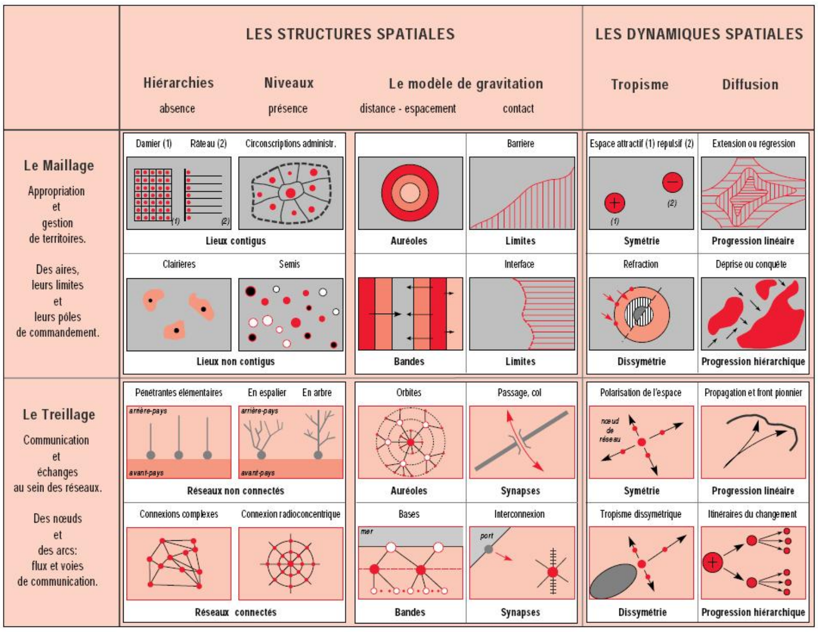

Figure 3.6: A table of 24 basic chorems presented in Fontanabona. (Source: [51])

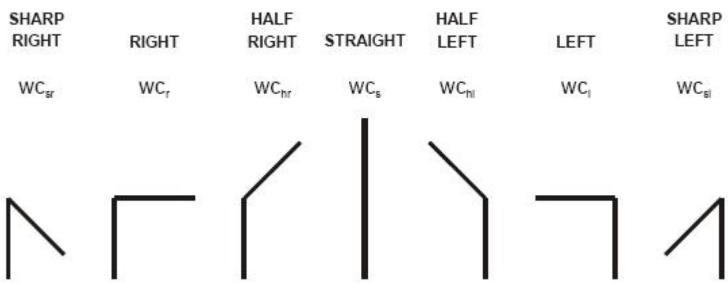

Figure 3.7: Basic set of wayfinding choremes. (Source: [76])

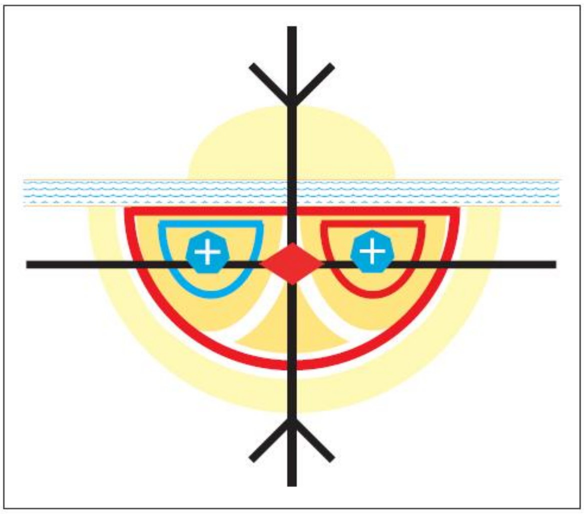

Figure 3.8: Chrono-chorème of *Tours* city. (Source: [59])

chorems as a communication mean, i.e. to represent spatio-temporal phenomena, defining new ad hoc chorems starting from personal knowledge, without formal definition. Recently, a formalization need of the chorem concept arises in literature, in geographers community but also in computer scientist community (visual language experts), in order to allow a general and more diffuse applicability of such concept to several domains.

In the remainder of this section some of maps discovered in literature containing chorems, which we used as a basis of our research, are shown. It is worth to note two relevant aspects about such maps:

- all maps discovered in literature are manually-made and are based on the personal knowledge of the map maker.
- To the best of our knowledge, no computer-based system that builds maps containing chorems exists in literature.

In figure B.3 a map containing chorems about France is presented. In particular, in this map several dynamics are highlighted, circles and boxes represent major cities, arrows represent flows and axes, different areas represent regional development trends. In figure 3.10 a map containing chorems about U.S.A. is shown. In this map several dynamics are highlighted, circles and represent major cities, arrows represent dynamics, stained areas represent different regions.

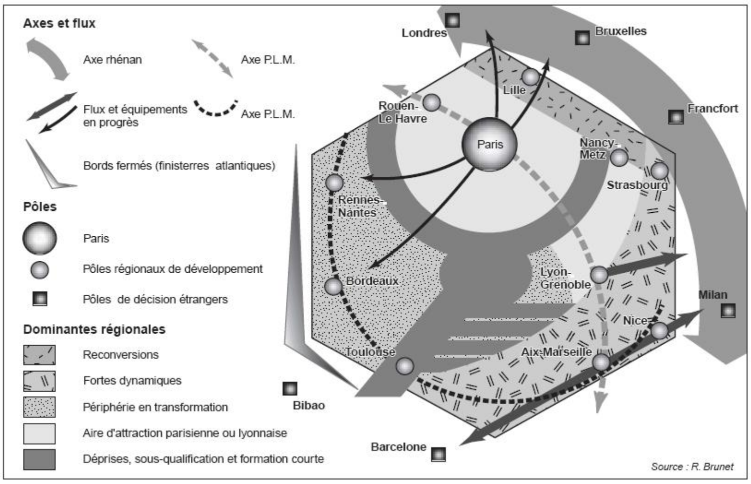

Figure 3.9: Map containing chorems about France. (Source [72])

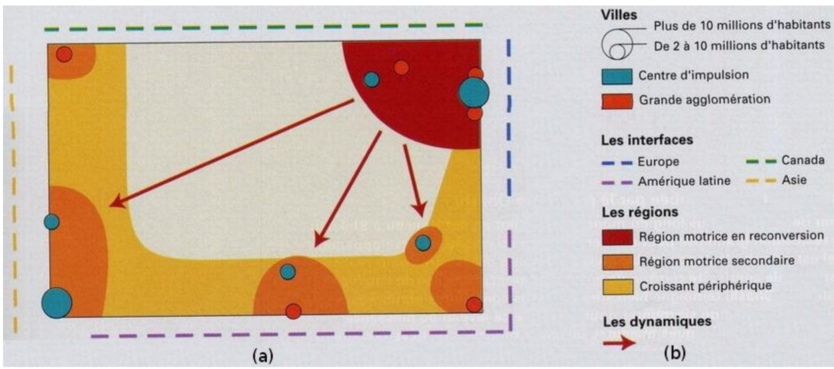

Figure 3.10: Map containing chorems about USA.

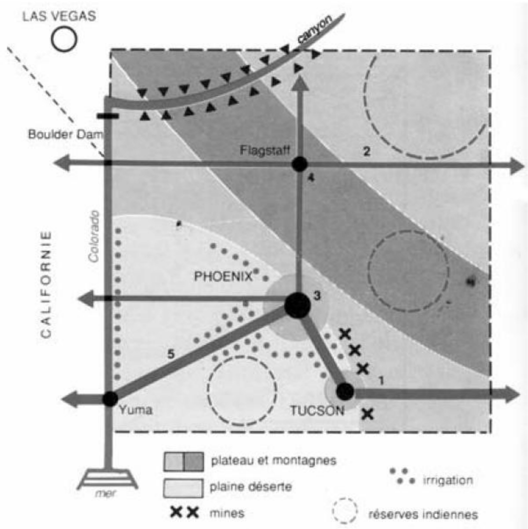

Figure 3.11: Map containing chorems about Arizona.

In Figure 3.11 a map containing chorems about Arizona is presented. In particular, in this map territorial characteristics are highlighted, black circles represent major cities, dotted circles represent Indian reservations, arrows represent principal routes, different areas represent territorial characteristics. In figure B.4 a map containing chorems about Spain is shown. In this map several dynamics are highlighted, circles and stars represent major cities, red arrows represent national routes, stained areas represent territorial characteristics, such as expanding attractive and repulsive areas or urbanization areas.

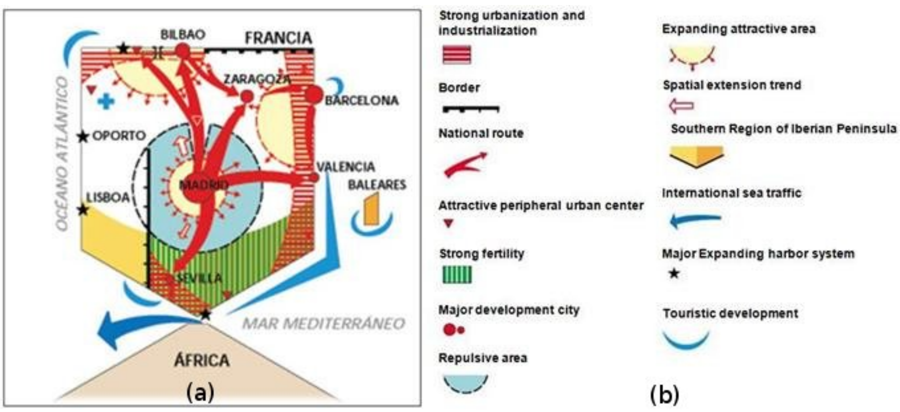

Figure 3.12: Map containing chorems about Spain.

3.2 Basic Definitions and Classification

The aim of this section is to introduce a formal specification of chorems in terms of visual language and. In order to formalize such a specification, we first carried out a study meant to analyze the extent to which the set of 28 chorems proposed by Brunet is used. Starting from approximately hundred manually-made maps containing chorems, we observed that expert users typically prefer to adopt a minimal set of Brunet's chorems or to generate an ad hoc chorem vocabulary instead. Moreover, we realized that the most used chorems can be grouped into two classes, namely chorems with inner geographic properties, able to represent typical geographic elements such as cities, rivers and regions, and chorems which describe phenomena among geographic elements, such as flow and regression. Based on those observations, starting from the seven basic shape which describe chorems and from the analysis of the other chorems tables presented in literature, we propose a formal specification of chorems in terms of components and structure, and we describe a proper classification which takes the chorem meaning into account.

Basic Definitions. In this section basic definitions about chorems are presented, meant both to formalize the terms that are used in the remainder of the dissertation and to solve the ambiguity of terms characterizing the chorems domain.

Definition 3.1: a *Geographic Pattern* or *Pattern* is an interesting regularity discovered in a geographic database by using functions for extracting relevant unknown information. Patterns can be used as a starting point to identify spatio-temporal phenomena and relationships among them. An example of typical geographic pattern is "In regions of Mexico, where the agave is cultivated, tequila is produced"

Definition 3.2: *Proto-chorems* can be defined as data items, properly cleaned and organized, on which SQL queries, (spatial) data mining functions or other analysis methods can be applied, in order to discover patterns. As an example, a Proto-choreme can be a database table with data referring to population movements and spatial data related to the regions containing such population. From this table several chorems can be extracted.

Definition 3.3: a *Chorem element* is a basic graphic element that may represent either a single geographic object or a single phenomenon. For instance, considering the map in Figure 3.10, a single point representing a city, or a single arrow representing a flow, as shown in Figure 3.13.

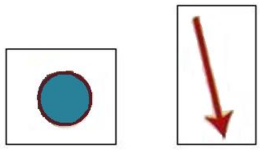

Figure 3.13: Point and arrow.

Definition 3.4: a *Chorem* is a set of chorem elements of the same typology. For instance, considering the map in Figure 3.10, the most important cities of the USA, the main flows between such cities, as shown in Figure 3.14.

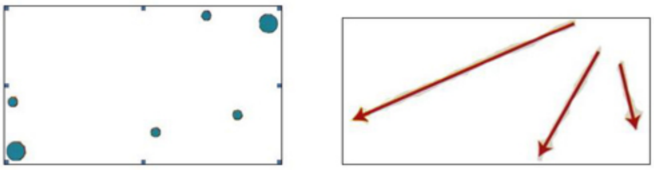

Figure 3.14: Cities and flows.

Definition 3.5: a *Chorematic Map* is a set of chorems which schematizes data of interest related to a specific place or region. A legend is associated with chorematic map, which explains the meaning of each chorem, as highlighted in Figure 3.12(b).

Chorem Classification. Chorems can be classified into three main categories, namely:

- *geographic chorems*,
- *phenomenal chorems* and
- *annotation chorems.*

Geographic chorems represent objects with simple spatial features, such as points, lines, polygons, and objects made up of their combinations, such as networks. As an example, stained areas with peculiar characteristics and cities represented by circles and, depicted in Figure 3.12, represent instances of geographic chorems, respectively polygons and points.

Phenomenal chorems describe spatio-temporal phenomena involving one or several geographic chorems. The initial set of phenomenal chorems we have identified consists of three types, namely *Flow*, *Tropism*, and *Spatial Diffusion*. The Flow represents population or object movement between points and/or polygons. Red arrows shown in Figure 3.12 represent an example of phenomenal chorems corresponding to Flow. The Tropism represents an homogeneous attractive or repulsive space, around a geographic chorem. Following the definition in [51] the Tropism considers the modification of one or more attributes of a place, in function of the action of an external phenomenon. For instance, it can be used for describing a city or a region which attracts the population. In Figure 3.12(b) the areas symbolizing with plus and minus sign represents an example of a phenomenal chorem corresponding to Tropism. Finally, the Spatial Diffusion represents a spatial progression or regression, from a geographic chorem towards a specific direction. According to the definition in [51] the spatial diffusion points out the modification of a place, step by step, under effect of a innovation or strategic decision in a territorial logic. As an example, it can be used for describing the development trend of a city in a particular region. In Figure 3.12 a Spatial Diffusion phenomenal chorem is represented by the yellow arrows around the city of Madrid.

Annotation chorems represent map labels or remarks. For instance, as shown in Figures 3.11 and 3.12, names near points symbolizing cities represent instances of annotation chorems. Further annotation chorems can be added by designers in order to complete a chorematic map. As an example, *Californie* or *Las Vegas* labels in Figure 3.11 , or *Africa* label in Figure 3.12 are elements which have been added to provide users with additional information that probably is not in the database which collects information about the chorematic map.

The previous definition and classification are meant to support expert users to both discover patterns by using available chorems, and possibly to generate new ones, through a uniform approach which takes into account both their structure and meaning. To this aim, it is valuable to use a chorem dictionary where chorems available so far are stored in agreement with the given definition and classification. In particular, the dictionary may be represented as a relation table where each entry is associated with a chorem, expressed in terms of category, type, meaning and visual representation. When new chorems are specified, the dictionary may be enriched by adding a new entry which characterizes it in terms of definition and classification. In Table 3.1 the schema of a chorem dictionary is depicted, where chorems we have specified are defined and classified.

3.3 Summarizing the Content of a Geographic Database

This section points out the reasons why the chorems have been chosen as base of this research project and how they can be used to visually summarize the contents of geographic databases. As van Elzakker [135] states in his PhD Thesis, traditional maps usually preserve as much as possible the geometric aspects of reality (bearings, distances, area relationships). Many research proposals in the contest of map projections and cartographic generalization are based on this characteristic. But, for many applications and activities, a correct geometric representation of the real scale of a map is not always needed, and sometimes not even desirable. A famous example of such a situation is provided in [135], i.e. the graphic representation of the London Underground (shown in Figure 3.15), which is originally conceived by Henry C. Beck.

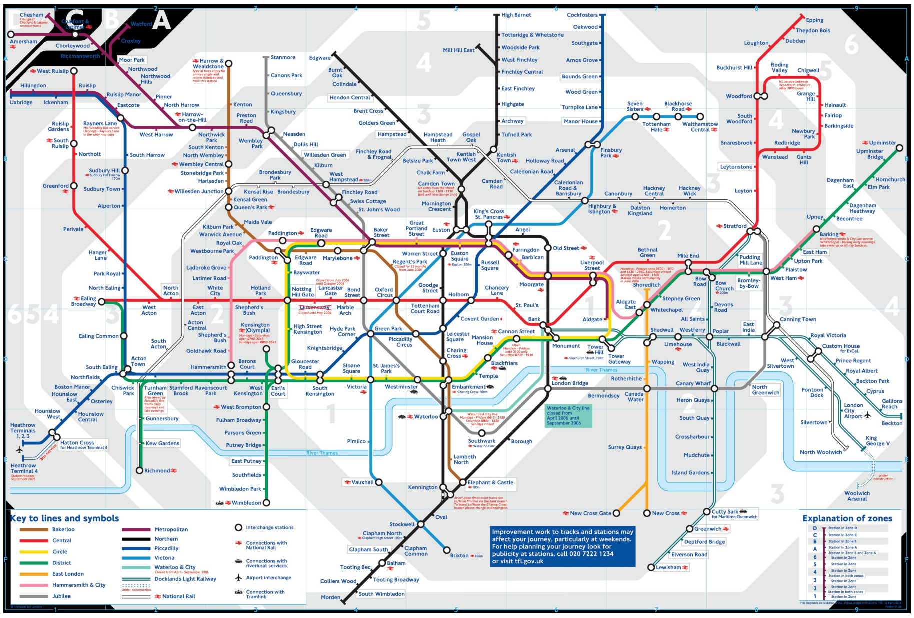

Figure 3.15: London Underground Map (source: [135])

In the graphic representation of the London Underground, geographic reality is intentionally distorted, but the result is a very simple, effective and intuitive map. Two reasons justify such lack of accuracy:

1. it is unnecessary to exactly reproduce geometric details of the underground that cannot yet be known,

2. the aim of this map is to convey rough ideas about the possible spatial organization of a large area.

The same is also true for chorematic maps. Indeed, in the London Underground map as well as in all the maps containing chorems in this chapter, we notice low accuracy in the geometry of geographic objects but the topological relationships are preserved. This is the most important characteristic when the structure, spatial relationships and dynamics of a geographic region have to be advanced.[135]

In our research project, we needed to display maps in conceptual manner, by adopting a graphical representation able to highlights salient aspects leaving out topographic details. Along this direction, chorems seemed to be a good approach that attempts to associate conceptual spatial representations to traditional maps. Lee at al. [83] assert that chorems are in practice a good convergence of two general approaches on the process of map construction, namely:

- the cognitive conceptual approach (CCA) which "is characterized by taking conceptual spatial representations (CSRs) as a starting point, and produces richer (more detailed, more precise) representations by concretizing, combining, and contextualizing them, for example, in a representational medium" and
- the data-driven approach (DDA), more generally used, which "starts with rich representations of spatial environments and derives representations that are more schematic by systematic abstraction, for example, by cartographic generalization".

In [23] Cheylan *et al.* present the diagram shown in Figure 3.16 which point out that chorems can be used as means to describe phenomena as well as to better understand phenomena.

Since their appearance, chorems have had a large diffusion in many domains ranging from politics, economy and demography, to climatology, evolution in epidemiology, agricultural activity and natural risks, thanks to their capability to provide decision makers with a key tool to get either a global view of a phenomenon of interest or an insight into a specific issue.

In order to exploit the chorem ability of visually summarize the geographic information which composes a map, the idea of our research project was to test whether cartographic solutions based on chorems could be used to represent visual summaries of spatial datasets. In the presented approach, the geographic information conveyed by a chorematic map, does not derive

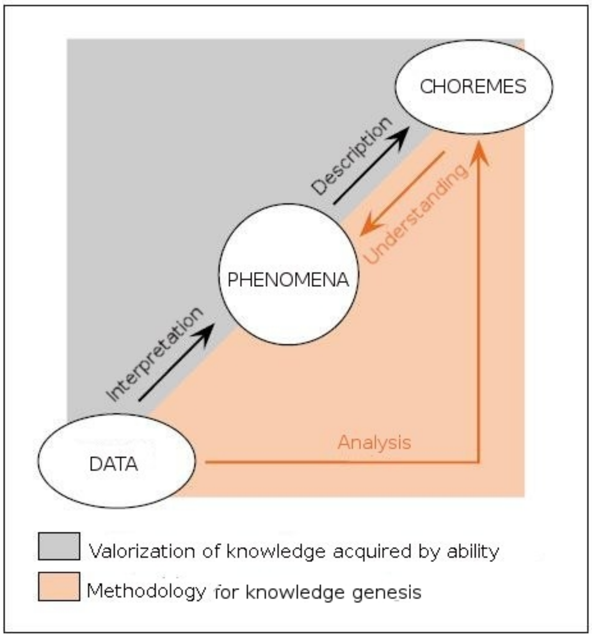

Figure 3.16: A methodological approach to chorems. (Source: [23])

from the personal experience of an expert geographer, but it is rather extracted from the contents of a spatial database. Indeed, as previously discussed in chapter 2, a spatial database contains many (unknown) information which can be first discovered, but also proposed in a suitable manner, in order to be effective. The aim of this thesis is to propose a method to visually summarize information extracted from geographic databases by using chorems. This idea can be schematized as the pyramid shown in Figure 3.17, where the basement is represented by the geographic database contents, the intermediate part is composed of several level of geographic and semantic generalization, and the apex corresponds to the chorematic map. Such a capability of chorems to visually describe the thematic content of a spatial dataset may be the starting point to construct maps for spatial analysis as well as spatial decision making. Finally, chorems seem present some interesting aspects other than the characteristic of visually summarize spatial information. Such aspects will be presented in the section dedicated to suggestions for further research.

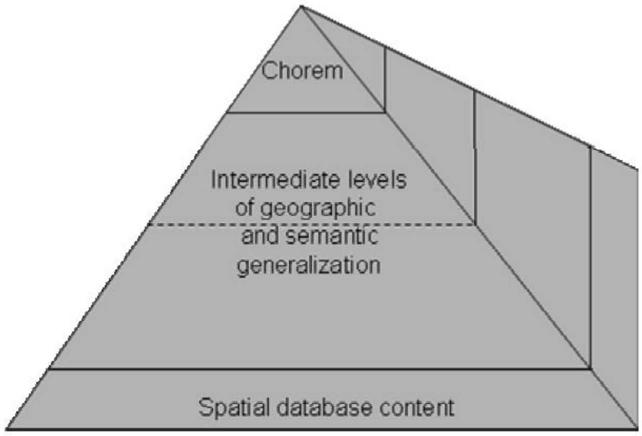

Figure 3.17: A pyramid of contents. (Source: [29])

3.4 Conclusion

In this chapter a presentation of the chorem concept is provided as well as the aspects that led as us to choose chorems in this research project. In particular, in Section 3.1 the meaning and the history of the chorem concept are illustrated, also showing some examples of maps containing chorems. In Sections 3.2 and 3.2, a formal chorems definition and a formal chorems classification are presented, both resulting from researches and studies carried out in this research project. These studies, made in collaboration with Karla Lopez, started from the analysis of approximately 100 manually-made maps containing chorems. Finally, in Section 3.3 the reasons why the chorems have been chosen as basis for this research project are provided and the idea of using the chorems as visual summary of geographic databases is proposed. In the next chapter an architecture meant to extract chorems from a geographic databases and visualize them in a chorematic maps will be presented.

Category	**Type**	**Meaning**	**Visual representation**
Geographic	Polygon	Administrative limits, thematic areas	Colored filled or dashed area
Geographic	Point	Cities, point of interest	Circles, stars, points
Geographic	Line	Geographic separations (rivers, mountains)	Lines
Geographic	Network	Combination of point and lines emphasizing the main characteristics of a network	Combination of circles, stars, points and lines
Phenomenal	Flow	Movements among point and areas	Colored arrows
Phenomenal	Tropism	Attractive and repulsive space around a geographic chorem	Colored filed area with arrows
Phenomenal	Spatial Diffusion	Spatial extension and regression toward specific directions	Colored arrows
Annotation	Label	Map labels	Text

Table 3.1: Chorem definitions stored into the Chorem Dictionary.

Chapter 4

ChEViS: Chorem Extraction and Visualization System

Contents

This chapter presents the architecture of ChEViS (Chorem Extraction and Visualization System), a system for visually summarizing geographic database content by using chorems. The aim of this system is to extract geographic objects and spatio-temporal phenomena from geographic databases and to represent them into a map by using chorems. Section 4.1 introduces some preliminaries about existing methods of constructing and using chorems. Section 4.2 proposes the general architecture of the system, showing the two major components which make up this framework, namely the extraction and the visualization subsystems. Section 4.3 gives a brief presentation of the extraction subsystem, namely *Chorem Extraction System (CES)*, which is meant to extract spatial objects and phenomena from a spatial database and associate them with different chorems types. A preliminary structure is proposed in collaboration with other authors [29], but relevant results, which are not in the scope of this dissertation, are awaited from forthcoming PhD Thesis of Karla Lopez. Section 4.4 presents a significant part of the result of this thesis, the design of the visualization subsystem, namely *Chorem Visualization System (CVS)*, whose aim is to transform a list of chorems extracted by the extraction subsystem and stored in an XML-like language, to a visual representation, that is a chorematic map. Section 4.5 presents the ChorML, an XML-like language, meant to store information about chorems, support communication between the extraction and the visualization subsystems as well as among the different subsystem modules.

4.1 Preliminaries

In literature, some methods and approaches suggest how to construct and use maps containing chorems. As discussed in the previous chapter, in Cheylan et al. [23] two different approaches to chorems generation and use are presented, as shown in Figure 4.1, namely enhancement of knowledge acquired by ability and methodology for knowledge genesis. In the former approach a domain expert user uses her/his personal knowledge and experience in order to understand (interpretation) the phenomena observed, then s/he formalizes her/his interpretation (description) by using a combination of chorems. In the latter, a domain expert user directly applies chorems chosen from a table to the circumstances observed. In other words, s/he recognizes as similar specific spatial structures thus assigning them the same phenomena (analysis). Finally, s/he verifies the validity of her/his suppositions (understanding). As reported in [135], Ormeling [98] states that two ways exist of constructing maps by using chorems as “building-stones”, namely producing and combining basic chorems from scratch (as shown in Figure 4.2) or by starting from a given

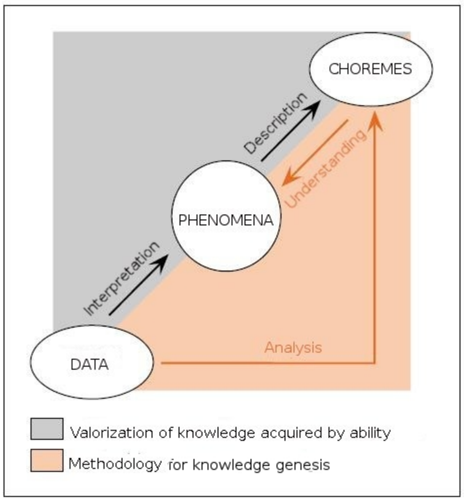

Figure 4.1: A methodological approach to chorems presented in [23]

map and reaching a schematic representation where all irrelevant aspects, including geometric accuracy, are removed [135]. Generally, in literature start from the assumption that all information composing chorematic maps originate from the expert user's knowledge about the domain of the map. In our approach, based on the definitions presented in the previous chapter, the starting point of the creation of a chorematic map corresponds to the information stored in a geographic database, and not derived from the scientist' s knowledge or geographer's experience. Moreover, the aim of the proposed system in this dissertation is also to guide the expert user to the semi-automatic generation of chorematic maps.

4.2 The General Architecture of ChEViS

In Figure 4.3 the architecture of ChEViS system is given, which consists of two major components, namely the Chorem Extraction System and the Chorem Visualization System. The former is meant to derive and manipulate the information from available geographic datasets, the latter handles such information by assigning them a visual representation in terms of chorems and chorematic maps. It is worth to note that both subsystems handle the same data structure which stores information about chorems. However, they manage, use and/or modify different subparts of the whole chorem structure depending on the running step. In particular, the Chorem Extraction System handles the conceptual properties of

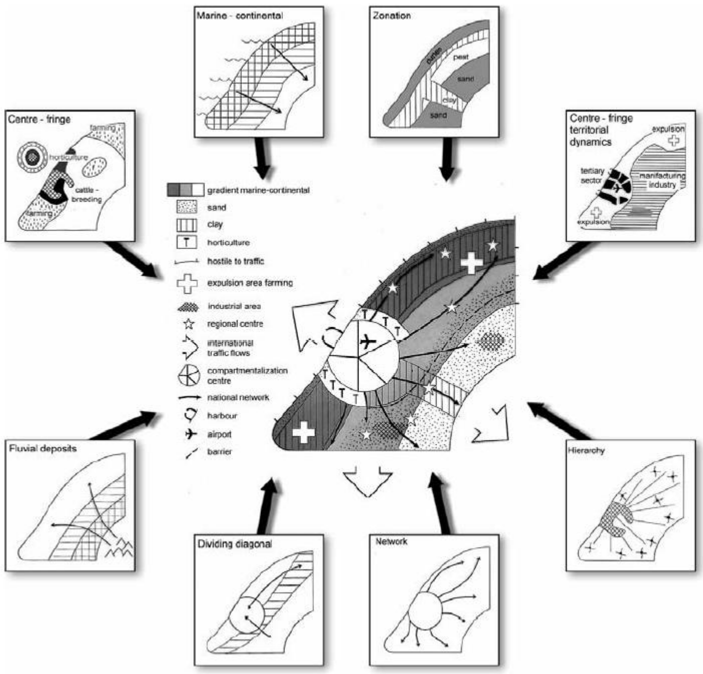

Figure 4.2: Construction of a Map of the Netherlands starting from basic chorems. (Source [98], legends translated by [135])

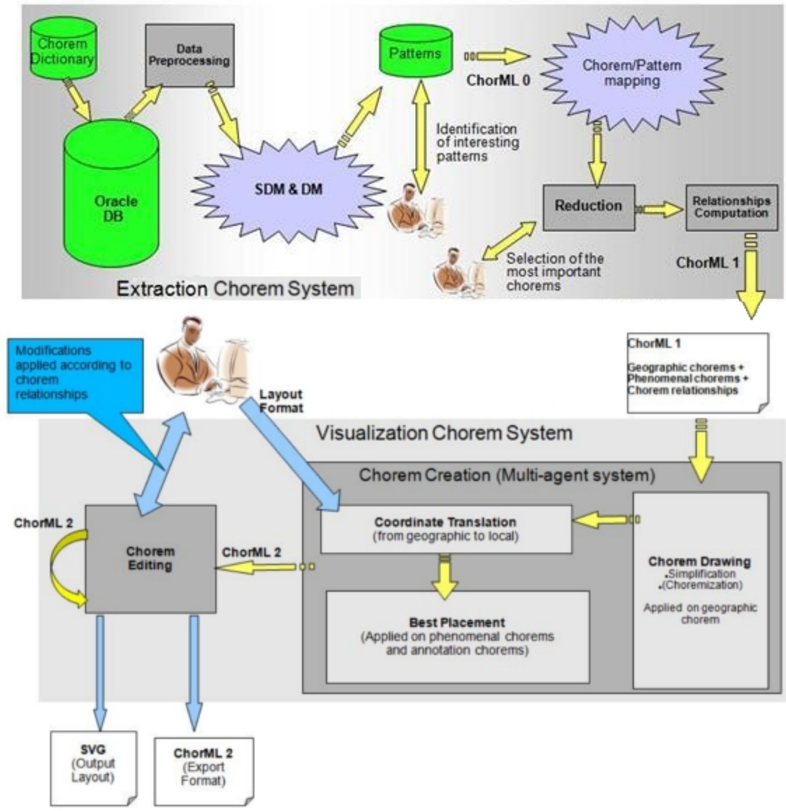

Figure 4.3: The architecture of the proposed system

several chorems types, such as name, type, size and coordinates, by expressing them through alphanumeric attributes. As for the Chorem Visualization System, it produces, elaborates and modifies the chorem geometric shapes, starting from the conceptual properties generated by the previous phase, in order to assign them a proper visual representation.

The storage of chorem characteristics and the communication among system components are based on a multi-level language, named ChorML. In particular, based on the involved components, a proper ChorML level is used in order to map different specifications and derive formats useful to the running process. In the following of this chapter a description of subsystems and underlying procedures is given, and the specification of each ChorML level used to import/export information among modules is detailed. A preliminary design of the architecture of both subsystems is introduced in [82, 29, 30].

4.3 The Chorem Extraction System

This section briefly presents the architecture of the Chorem Extraction System, shown in figure 4.4, whose purpose is to extract geographic objects and spatio-temporal phenomena from a geographic database. The preliminary structure here proposed results from a international project discussed in [82]. Compared with the initial idea, the architecture of the system has been refined in terms of subsystems, modules and languages. In particular, the modules of subsystem devoted to the extraction of significant patterns from spatial databases has been enhanced by integrating (spatial) data mining techniques [77, 49], procedures for pattern reduction, and SQL-based data filtering. The whole process starts

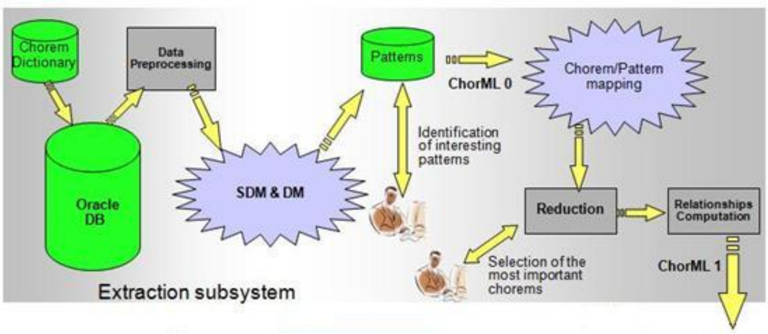

Figure 4.4: The architecture of the system which extracts chorems.

by taking into account a geographic database or a dataset containing data of interest, and a dictionary which contains the description of available chorems in terms of name, type and structure. In particular, the chorem dictionary is imported into the database and is exploited to support pattern extraction. As for the underlying technologies devoted to data preprocessing and (spatial) data mining, the extraction subsystem embeds the Oracle[1] DBMS and its spatial extension[2], the Subdue System [75], enhanced by the PAM (Partitioning Around Medoids) algorithm [103]. Moreover, it provides users with relevant feedback about partial results, that is to say, at each step users may interact with it in order to refine queries or functions, thus enhancing the quality of results. Afterwards, a Pattern/Chorem mapping process and a reduction process are applied, meant to derive a significant set of chorems. In particular, the former associates each pattern with a chorem, the latter determines a reduced list of the most important chorems. In fact, a lot of patterns can be found by the previous steps, thus overcrowding the potential chorematic map under construction. Both processes allow user intervention in order to select the most interesting patterns and chorems. Finally, the relationship computation process is invoked in order to obtain a list of relationships between the recognized chorems. To this aim Spatial Oracle functions are used and results are coded in a specific ChorML level, which is then sent to the visualization subsystem.

4.4 The Chorem Visualization System

This section presents a significant part of this thesis, the design of the Chorem Visualization System. The aim of this subsystem is to transform a list of chorems and relationships, extracted by the extraction subsystem and stored in a XML-like language, to a visual representation, namely a chorematic map. Starting from the previous versions presented in [82]and in [30], the architecture of the visualization subsystem has been refined and improved. In particular, modules have been properly restructured to take into account the evolving chorem specifications. Moreover, recent techniques for shape simplification have been investigated in order to provide solutions for a multilayered simplification, which preserves the underlying relationships. As shown in Figure 4.5, two distinct tasks are performed by this subsystem, namely the chorem creation and the chorem editing. The former is meant to create the visual component of chorems and arrange them in a map through three subsequent steps, namely the

[1] Oracle Corporation [online], website, `http://www.oracle.com/index.html`
[2] Oracle Spatial [online], webpage, `http://www.oracle.com/technology/products/spatial/index.html`

chorem drawing, the coordinate translation and the best-placement of chorems. The latter allows users to add further details and/or refinements to the chorematic map resulting from the chorem creation phase. Details on a prototype which implements such modules will be illustrated in the next chapter.

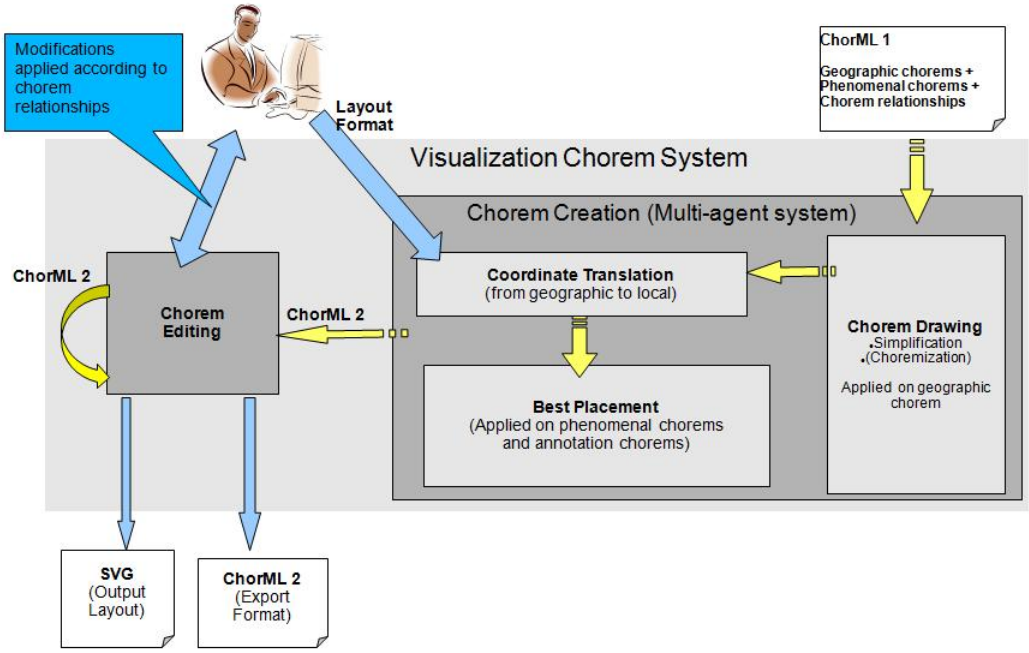

Figure 4.5: The architecture of the visualization subsystem

4.4.1 The Chorem Creation task

This subsection describes the Chorem Creation task, whose aim is to generate the visual component of chorems which will be then spatially arranged in a chorematic map. This task is accomplished by using a multi-agent system, whose aim is threefold, namely:

- to generate the visual component of chorems, according to the specifications arising from the extraction subsystem. The Chorem Drawing phase is meant to this aim,
- to adapt the chorem components according to the visualization layout. A Coordinate Translation phase is meant to perform this task.

- to determine the chorems best placement, preserving structural and topological constraints. A Best Placement phase is meant to execute this phase.

In the reminder of this subsection, a description of the architecture of the multi-agent system is given. Then, the Chorem Drawing, the Coordinate Translation and the Best Placement phases are discussed.

The architecture of the multi-agent system Several motivations, discussed in Section 2.7.1, have led to choose a multi-agent system to accomplish the operations necessary to this task. In particular, in order to structure this multi-agent system, several hints have been considered found in [70], which describes a system meant both to generate a map on-the-fly and to resolve spatial conflicts. In figure B.11 the architecture of the multi-agent system is shown. This architecture is composed by three hierarchic levels containing the *Chorematic Map Agent* (level 1), a set of *Chorem Agents* (level 2), and a combination of a set of *Chorem Element Agents* and a set of *Relationship Agents* (level 3), respectively. The

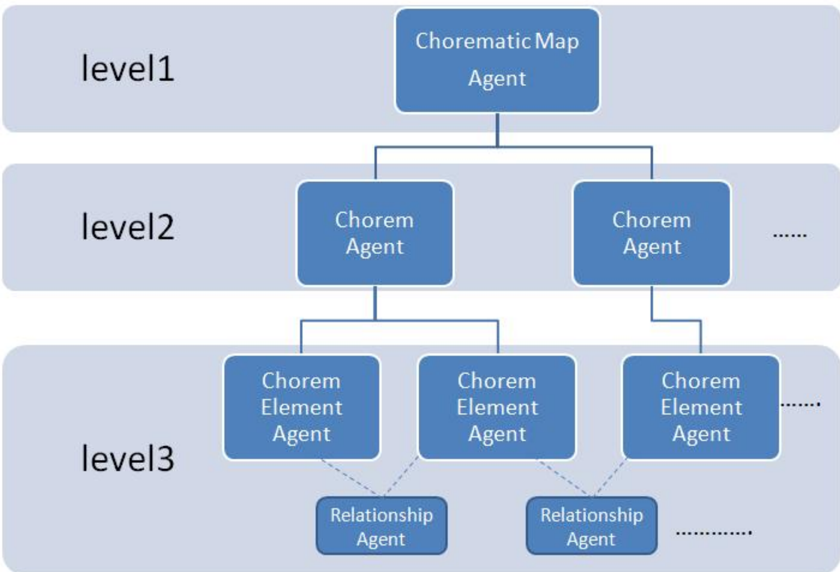

Figure 4.6: The architecture of the multi-agent system.

Chorematic Map Agent is responsible of creating all Chorem Agents belonging to level 2 and all *Relationship Agents* belonging to level 3. In particular, by parsing the ChorML code this agent:

- generates a Chorem Agent for each chorem parsed and assigns it the portion of ChorML code that it has to manage.
- generates a Relationship Agent for each relationship parsed and coordinates the created chorem agents in order to calculate further topological relationships, which are subsequently used to create other Relationship Agents.

The *Chorem Agent* is responsible of creating a *Chorem Element Agent* for each Chorem Element parsed in the portion of the code which the *Chorematic Map Agent* has assigned to it. The *Chorem Element Agents* handle the geometry associated to the corresponding Chorem Element, and execute all operations required on its geometry in the various processes composing the Chorem Creation Task. Finally, the *Relationship Agents* are responsible of managing (topological) relationships among chorems which are stored in the ChorML code. In particular, the task of these agents is to store the Chorem Elements involved in the relationships and to verify whether relationships hold during all the processes involved during the Chorem Creation Task.

Comparing the system illustrated in figure 4.7 [70], the *Chorematic Map Agent*, the *Chorem Agent* and the *Chorem Element Agent* behave in a similar way as to the *Coordinator Agent*, the *Type Agent*, and the *Chorem Element Agent*, respectively. The *Relationship Agent* slightly differs from the *Container Agent*. In fact, while the *Relationship Agents* store (topological) relationships among chorems, the *Container Agent* is responsible of the generalization of adjacent map object, encompassing all the objects that should be affected by modifications, when changes of the map scale are required by the user. The multi-agent system proposed in this work manages different types of relationships in order to execute the processes of the Chorem Creation task:

- topological and non-topological relationships among Chorem Element belonging to the same Chorem,
- topological and non-topological relationships among Chorem Element belonging to the different Chorems.

In figure 4.9, an example of the structure of the multi-agent system is depicted, which manages the map of Figure 4.8, containing two chorems, namely a chorem representing one area and a chorem representing four points. Let the blue agent represent the Chorematic Map Agent, the green agents represent the Chorem Agents, the pink agents represent the Chorem Element Agents and the orange agents represent the Relationship

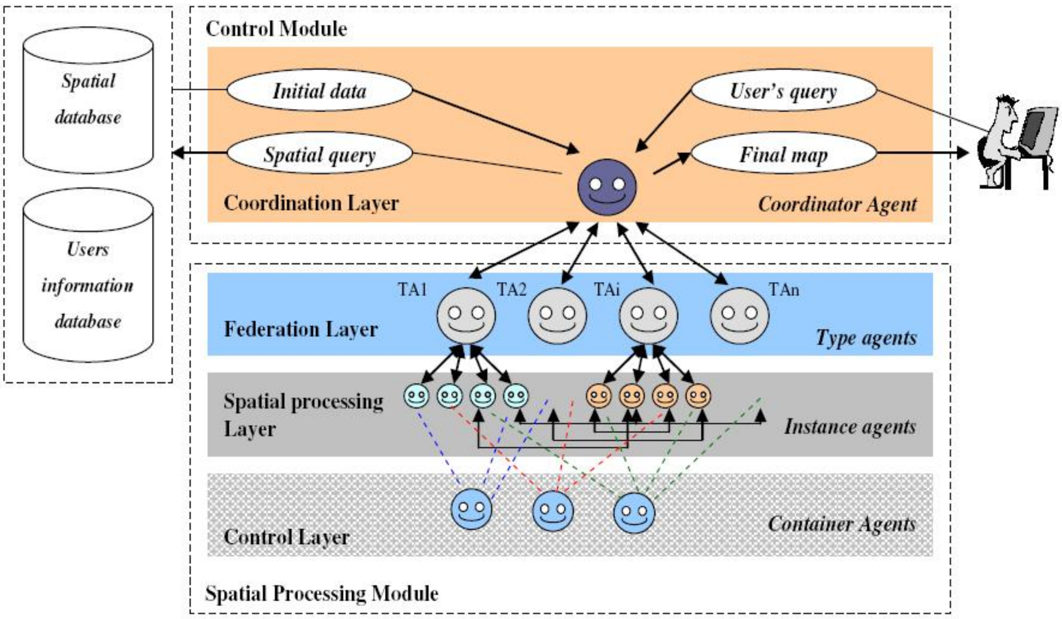

Figure 4.7: The multi-agent system proposed in [70].

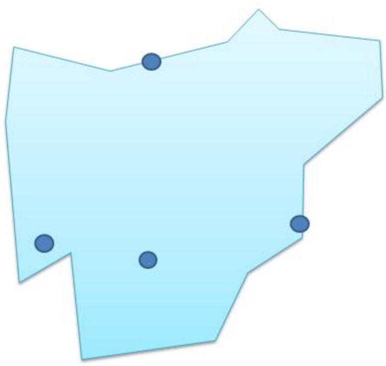

Figure 4.8: A map containing two simple chorems (area and points).

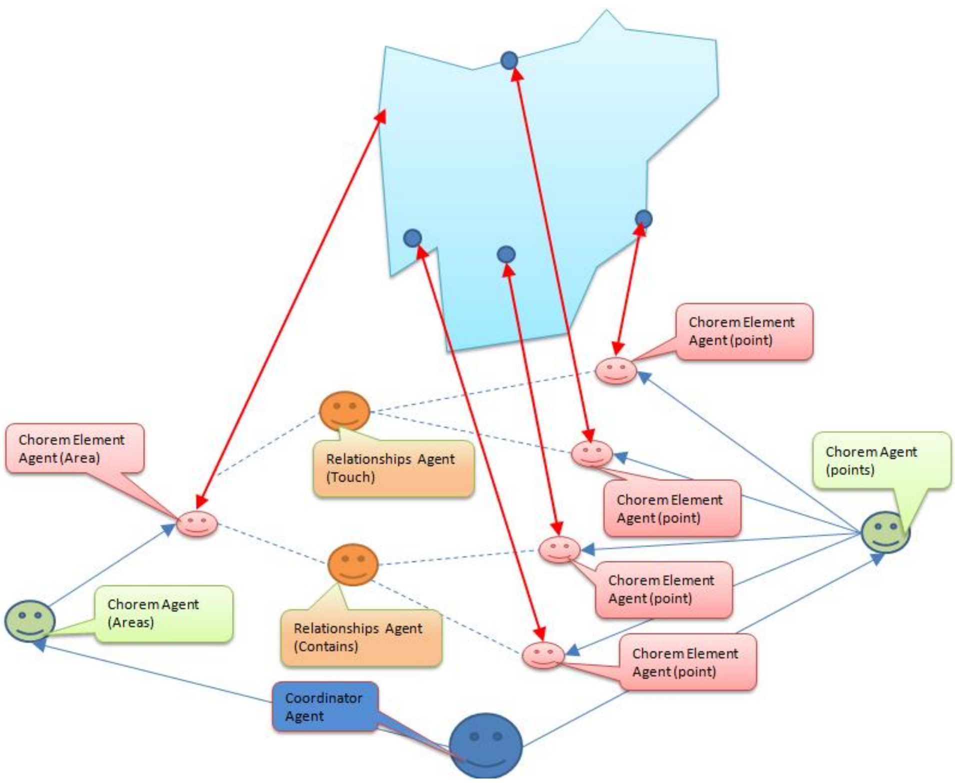

Figure 4.9: An example of the multi-agent system structure

Agents. Then, two Chorem Agents are created by the Chorematic Map Agent and in turn two Chorem Agents create one Chorem Element Agent representing the area, and four Chorem Element Agents representing the points, respectively. Moreover, two Relationships Agents are created by the Chorematic Map Agent, namely, a Relationship Agent representing the contains relationship and a Relationship Agent representing the touch relationship. The former stores the relationship among the Chorem Element Agent representing the area and two Chorem Element Agents representing the two points contained into the area. The latter stores the relationship among the Chorem Element Agent representing the area and two Chorem Element Agents representing the two points positioned on the boundary of the area.

The Chorem Drawing phase During the Chorem Drawing phase, the multi-agent system performs two, not necessary interconnected steps, named simplification and choremization. Such transformations invoke some procedures and spatial operators, as illustrated in figures 4.10 and 4.12. In particular, the simplification step determines a simplified version (see Figure 4.10(b)) of the data geometry, by reducing the number of vertices of the original shape (see Figure 4.10(a)). As for the choremization step, it associates a regular shape (see Figure 4.11(b)) with the possible simplified geometry of data (see Figure 4.11(a)). In this phase, the potential chorematic map contains only Geographic Chorems, whose geometry is simplified and in case *choremization*. That is, only the chorems which have explicit geometric objects, such as points, lines and polygons, are considered.

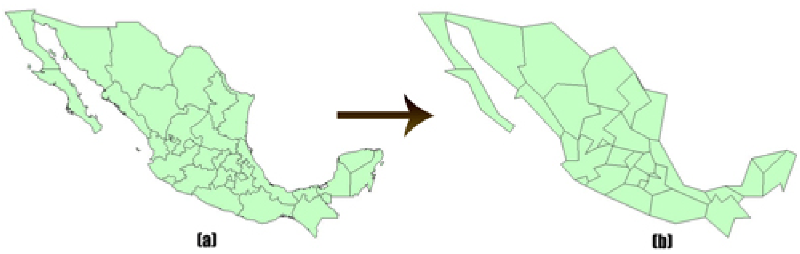

Figure 4.10: An example of the Simplification process.

As for the simplification step, it is the crucial task of the Chorem Drawing phase, due to the peculiarity of the nature of chorems, namely the extremely simplified shape. During this process, spatial conflicts may occur, which have to be solved. This step is based on spatial data

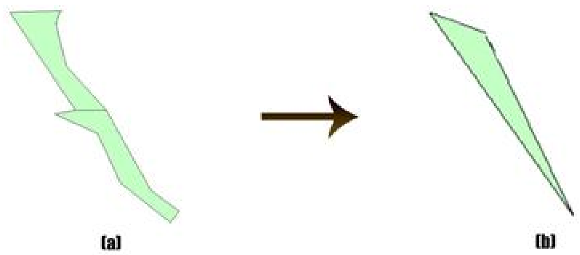

Figure 4.11: An example of the Choremization process.

functions, embedding the Ramer-Douglas-Peucker (RDP) [107, 39] algorithm and its improvements [63, 146]. Briefly, the RDP algorithm is a well known algorithm in literature, which uses a threshold in order to iteratively remove vertices from the input line, preserving the original shape.

In this Thesis, modifications to the RDP algorithm are investigated and added in order to manage topological violations which happen during the simplification step. Such modifications are motivated on the remark that the simplification method must respect the fact that the lines and /or polygons to be simplified have a context [12]. In fact, our approach deals with a particular generalization processes, i.e., simplification, whereas it is meant to resolve cartographic displaying problems, such as map labeling and flow displaying. A significant feature of the proposed multi-agent system consists of its capability to perform a multi-layer simplification task. In literature, the simplification problem is a relevant topic of investigation, although little research has been carried out on simplification executed in a context. In fact, the simplification applied on a single geographic object, such as a line or a polygon, corresponds to a trivial task because it affects just its shape. On the contrary, a simplification task may violate (topological) relationships when it is autonomously applied on different adjacent objects belonging either to the same layer or to different overlapped layers. This situation is depicted in Figure 4.12, where the point representing the Dutch city of Maastricht, after the simplification process falls in Belgium. Not much approaches exist in literature about this topic. In [150], Zhan *et al.* presented a cognitive study on conflict resolution due to simplification, which suggests as solution first to identify the conflicts and then to try to patch up them. Differently, the approach proposed in [12] tries to avoid conflicts from the start by using geometric algorithms. In our simplification step, this latter approach has been chosen and investigated, because the aim of this specific step is to preserve as much as possible the shape and the proportions of the map.

Finally, some work has been done about the user perception of the simplification [33], for understanding how different levels of simplification are perceived by the user. In fact, the system can take advantage from this knowledge, considering that such a difference of perception can influence the execution of the simplification process.

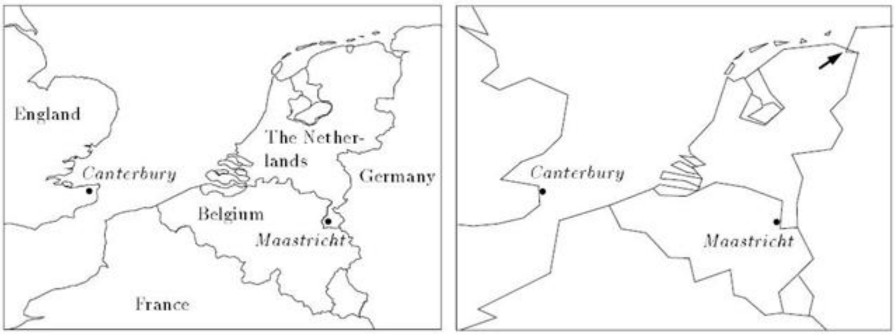

Figure 4.12: An example of the multi layer simplification process which introduces topological errors. (source [12])

As for the choremization step, it is based on spatial data functions, embedding algorithms which overstress the simplification of a shape in order to obtain a very simple polygon, such as triangle, rectangle and square. It is worth to note that the choremization is an optional step which can be shifted by the user, because contrary to algorithms implementing the simplification step, the choremization algorithms lose the original shape, so potentially making the identification of the shape very difficult. Finally, the Chorem Drawing phase guarantees the maintenance of spatial constraints among elements of the original map, also by adopting an underlying geographic reference system. That is to say, the applicability of any transformation is checked with respect to the spatial relationships it affects.

The Coordinate Translation phase Once the Chorem Drawing task is concluded and the drawing of the expected chorems is obtained, the Coordinate Translation phase is carried out. In particular, users are asked to specify details about the output map, such as the number of colors and the visualization layout (for instance A4). This phase is essentially accomplished by converting coordinates of the geographic object belonging to the chorems from a geographic reference system to the local reference system. The choice of the visualization layout affects the number and the size of chorems that can be introduced onto a map, since it is necessary to

Category	**Type**	**Priority**
Geographic Chorem	Polygon	Highest
Geographic Chorem	Line	...
Geographic Chorem	Point	...
Phenomenal Chorem	Tropism	...
Phenomenal Chorem	Spatial Diffusion	:.
Phenomenal Chorem	Flow	...
Annotation Chorem	Label	Lowest

Table 4.1: Chorem Element Agent's priorities

guarantee the readability requirement.

The Chorem Best Placement phase The next phase is the Chorem Best Placement. It is meant to integrate phenomenal and annotation chorems in the chorematic map, assigning them a proper position. Like the previous one, also this phase involves the resolution of spatial conflicts resulting from the interaction of the geographic objects belonging to the Chorem Element Agents. Such spatial conflicts are checked by the Relationship Agents, and managed by the Chorematic Map Agent.

In order to resolve spatial conflicts, unlike the simplification step, in this phase only the displacement operator is considered by the multi-agent system. It is a typical operator used in the spatial conflicts resolution, introduced in Section 2.4. Moreover, several different techniques have been investigated in order to:

- insert the annotation chorems into the map. In particular, techniques for label placement discussed in Section and algorithms for the placement of diagrams on Maps [136].
- insert the phenomenal flow chorems into the map. In particular, techniques for Flow Displaying discussed in Section 2.4.4,

As for the spatial interaction among the geographic object of the map, in the multi-agent system proposed in this thesis we adopt the method suggested in [70, 69], where the objects is based on different priorities assigned to the different agents belonging to the system. In other words, when a conflict rises between two elements of the map, the element with lower priority is forced to shift. The Table 4.1 the priority hierarchy used in the multi-agent system proposed in this thesis are summarized.

In particular, Geographic Chorems have higher priority than Phenomenal Chorems have and Phenomenal Chorems have higher priority than Annotation Chorems have. Moreover, also among Geographic Chorems different priorities are assigned. Geographic Chorems which manage polygons have higher priority than Geographic Chorems which manage lines have and Geographic Chorems which manage lines have higher priority than Geographic Chorems which manage points have. It is worth to note that during the resolution of spatial conflicts the Chorem Element Agents with lower priorities are more affected by displacement than Chorem Element Agents with higher priorities. Thanks to this strategy, the agents with the higher priorities are benefited respect to agents with the lower priorities, and the geographic objects appertaining to these agents will be visualized in more legible ways in the final map .

Finally, it is worth to point out that in order to satisfy 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 chorematic maps. As an example, from a chorematic map project which contains 10 chorems, two different chorematic maps may be derived, each containing five chorems.

4.4.2 The Chorem Editing

After the semi-automatic process for obtaining a chorematic map, some difficulties may occur regarding chorem placement and layout. Moreover, further refinements affecting semantic and graphic properties may be required by users. The aim of the Chorem Editing phase is to support users in taking part in the construction of chorematic map, by editing, moving or deleting some elements, e.g., refining the expected output map. In particular, the Chorem Editing phase allows users to perform the following tasks:

- to import a ChorML representation of a chorematic map; i.e. chorematic maps coming from other systems can be used to carry out modifications,
- to display chorems; indeed, the main purpose of this phase is to create the graphical representation of chorems under construction,
- to modify graphical representation of chorems; in order to solve problems regarding placement and layout, i.e. resizing or displacing a circle shape representing a city in a map,

- to modify semantics representation of chorems; in order to solve problems regarding placement and layout, i.e. changing the meaning which the system assigned to some chorems.
- to export both a graphical (Scalar Vector Graphics, SVG[3]) and a ChorML representation of chorematic maps. Indeed, the former can be utilized only for visualizing purposes, because it is able to store only graphic characteristics of the map. The latter can be utilized also to store information about relationships among chorems, allowing to modify the choremamatic maps without losing topological characteristics.

A Chorem Editor has been created in order to implement previous tasks. The Chorem Editor has been built as an extension of the Magelan[4] Graphics Editor, an open source 2D vector graphics editor, based on Java programming language. More details about such an implementation are given in the next chapter.

4.5 The ChorML Language

A multi-level XML-like language, named ChorML is defined in order to store information about chorems as well as to support communication among the different system modules. Compared with the preliminary idea presented in [29], the initial four levels of ChorML have been restructured and enriched with new tags meant to handle data in terms of attributes and values, and information about implicit properties, such as the sequence of operations applied on a dataset in order to derive a set of patterns. The ChorML is generally composed of

- a not standardized set of XML tags which store information about the structure of the chorems and implicit properties,
- a standardized sets of XML tags which store geographic information and visualization information

Three different levels of ChorML are defined as following:

Level 0 of ChorML consists of eXtensible Markup Language (XML)[5] and Geography Markup Language (GML)[6] tags. This is the level of ChorML

[3]Scalable Vector Graphics [online], website, `http://www.w3.org/Graphics/SVG/`
[4]The Magelan 2D Java Graphics Editor [online], website, `http://magelan.hit.bg/`
[5]eXtensible Markup Language [online], website, `http://www.w3.org/XML/`
[6]Geography Markup Language [online], website, `http://www.opengis.net/gml/`

utilized in the extraction subsystem. As depicted in Figure 4.13, it is meant to store information about proto-chorems, lineage of data, that is information about data source, and functions applied to them in order to obtain chorems.

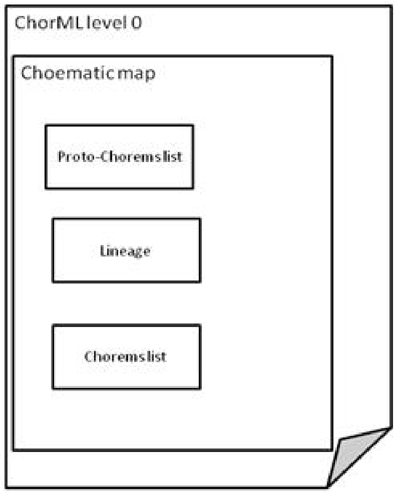

Figure 4.13: . ChorML level 0 structure.

level 1 of ChorML, shown in Figure 4.14, is still a combination of XML and GML tags. It specifies the results of the application of (spatial) data mining algorithms. This is the level of ChorML utilized for communication between the extraction and the visualization subsystem. In particular, such a level collects:

- general information containing id, project name, author name, creation date, layout, reference system, original database name, last update;
- the chorem list where geographic data are coded in GML;
- pre-legend which contains a description of each chorem;
- topological and structural relationships between chorems.

Presently, this level is totally specified with an XML grammar presented in [27].

Level 2 of ChorML corresponds to the level of ChorML utilized in the visualization subsystem. Depending on the running step, different types of

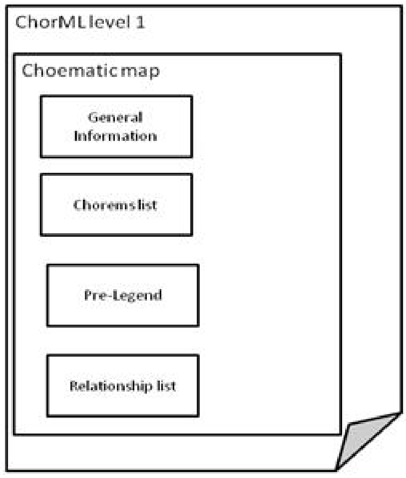

Figure 4.14: . ChorML level 1 structure.

tags are involved, namely of XML and SVG tags and one of their possible combinations. As shown in Figure 4.15, the elements of this level are:

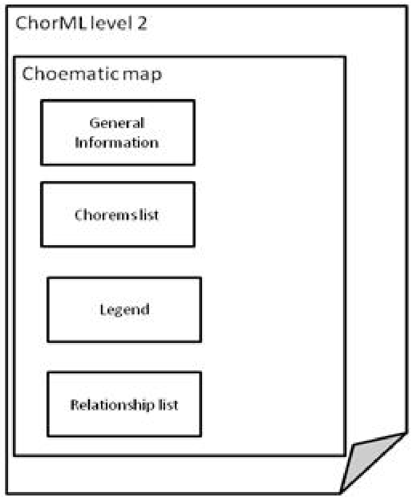

Figure 4.15: . ChorML level 2 structure.

- general information containing id, project name, author name, creation date, layout, reference system, original database name, last update;
- the list of simplified chorems resulting from the best-placement task;

- the list of modified chorems resulting from the chorem editing task;
- a legend describing the meaning of each chorem;
- topological and structural relationships between chorems.

Finally, the chorem editor, described in the next chapter, is able to export the chorematic map expressed in ChorML2 code to a version expressed in SVG code for mere visualization purposes, losing information about topological properties of chorems.

4.6 Conclusion

In this chapter an architecture of a system for visually summarizing geographic databases by chorems is presented. The aim of this system is to extract geographical object and spatio-temporal phenomena from a geographic databases and represent them into a map by using chorems. In literature, some methods and approaches suggest how to construct and use chorems. All approaches founded in literature start from the assumption that all information composing map containing chorems originates from the knowledge of a expert user of the domain tied to the map. Instead, in our approach, the starting-point of the creation of a map containing chorems is the information stored in a geographic database and not derived from the knowledge and the experience of a scientist or geographer. Moreover, the aim of the system proposed in this dissertation is also to guide the expert user, in order in the semi-automatically generation of chorematic maps. In Section 4.2 the general architecture of such a system has been proposed, showing the two major component which compose this framework, namely the extraction and the visualization subsystems. In Section 4.3 a brief presentation of the Chorem Extraction System is given. Such a subsystem is meant to extract spatial objects and phenomena form a spatial database and associate them with different chorems types. A preliminary structure, conceived in collaboration with other authors, has been proposed [29], but relevant results are awaited from forthcoming PhD Thesis of Karla Lopez, and are not in the scope of this dissertation. In Section 4.4 a relevant part of the purpose of this thesis is presented, namely the conceiving of the visual subsystem, whose aim is to transform a list of chorems extracted by the extraction subsystem and stored in a XML-like language, to a visual representation, that is a map containing such chorems, by using a multi-agent system. Finally, in section 4.5 the ChorML is presented, a XML-like language, meant to store information about chorems, which support communication between the extraction and the visualization subsystems as well as among the different subsystem modules. In the next chapter, details on the prototyping implementation of the visualization system will be presented and discussed.

Chapter 5

The Prototype of the Chorem Visualization System

Contents

This chapter presents a prototyping implementation of the Visualization Chorem System, describe the underlying technologies in terms of software components and specifies how it works. In particular, Section 5.1 illustrates the software components underlying the two most relevant modules have been illustrated, namely Chorem Creation and Chorem Editing, and software technologies embedded into the *ChorML* language. Section 5.2 shows an example of a chorematic map construction in order to demonstrate effectiveness of the system functionality.

5.1 Technologies adopted for the Implementation

As discussed in the previous chapter, a multi-level XML-like language, named ChorML, has been defined in order to both store information about chorems, and support communication among different modules of the system. To reach this aim, in ChorML proper tags have been defined, which can be classified into three groups in agreement with their purpose, namely Extensible Markup Language (XML), Geography Markup Language (GML) and Scalar Vector Graphics (SVG) tags. As for the first group, in order to store alphanumerical information and relationships about chorems, proper XML tags have been specified, thanks to the XML capability of defining ad hoc vocabularies by exploiting the Document Type Definition (DTD) and the XML Schema. As for the geographic component of chorems, it is managed in ChorML by using GML, which represents the specification by the Open Geospatial Consortium[1] to model, store and exchange geographic content. It is worth to note that several related XML technologies can be embedded within GML, such as XML Schema, XSLT, XLink and XPointer, thus proving a seamless processing also in heterogeneous environments. GML is used to describe geographic information, but it does not specify how this information is visualized. To this aim, the visualization of the geographic component in ChorML is handled by using SVG, a vector standard model where geometric objects are described by their geometry, topology and thematic attributes. The choice of using XML-based languages, such as SVG and GML, has been affected by the advantage that they are open standards, thus enabling an easy integration with other technologies based on XML.

As for the implementation of the prototype of the Visualization Chorem System, several open source Java frameworks and libraries have been used and the resulting components have been integrated trough the Eclipse

[1]The Open Geospatial Consortium [online], website, `http://www.opengeospatial.org/`

platform. Eclipse[2] is an open development platform comprised of extensible frameworks, tools and runtimes for building, deploying and managing software. In its standard form it is an Integrated Development Environment (IDE) for Java developers, consisting of the Java Development Tools (JDT) and the Eclipse Compiler for Java (ECJ).

5.1.1 Implementing the Chorem Creation task

The Chorem Creation task is generally meant both to create the visual component of chorems and to arrange them into a chorematic map. Such a phase is accomplished through a multi-agent system, whose aim is to produce the visual components, to spatially arrange them according to the visualization layout and to determine their best placement, preserving topological and non-topological relationships among them.

In order to implement this multi-agent system, the Java Agent DEvelopment Framework (JADE)[3] [11] has been used.In Figure 5.1 the JADE graphical user interface is depicted. It is an open source software framework to develop agent-based applications in compliance with the Foundation for Intelligent Physical Agents (FIPA)[4] specifications for interoperable intelligent multi-agent systems. The choice of JADE among several existing platforms for multi-agent system development, such as AgentBuilder[5], Jack[6], MadKit[7], Zeus[8], is essentially due to the high level facilities of implementation with which JADE supplies users. For instance, it provides developers with predefined agent's behaviors such as *simple*, *cyclic*, *sequential*, *parallel* and *finite state machine*. Moreover, developers may benefit from both the open source nature of this framework and the portability of the system, being the agents developed by using Java. In order to simplify the software development of JADE agents within Eclipse, the EJADE[9] plug-in has been used. It is a (non-profit and free) Eclipse plug-in that helps JADE agent developers to launch agent platform and to deploy agents. This plug-in takes care of Java classpaths and it bridges the JADE Java agents specifications to the Eclipse debugging framework. Such a functionality allows for debugging agents similarly to any Java

[2] Eclipse [online], website, `http://www.eclipse.org`

[3] Java Agent DEvelopment Framework, [online], website, `http://jade.tilab.com/`

[4] Foundation for Intelligent Physical Agents specification [online], website, `http://www.fipa.org/repository/standardspecs.html`

[5] AgentBuilder [online], website, `http://www.agentbuilder.com/`

[6] Jack [online], website, `http://www.agent-software.com`

[7] Madkit [online], website, `http://www.labs.bt.com/projects/agents/zeus`

[8] Zeus [online], website, `http://www.madkit.org`

[9] Eclipse IDE Plug-in for Java Agent Development Environment [online], website, `http://disi.unitn.it/~dnguyen/ejade/`

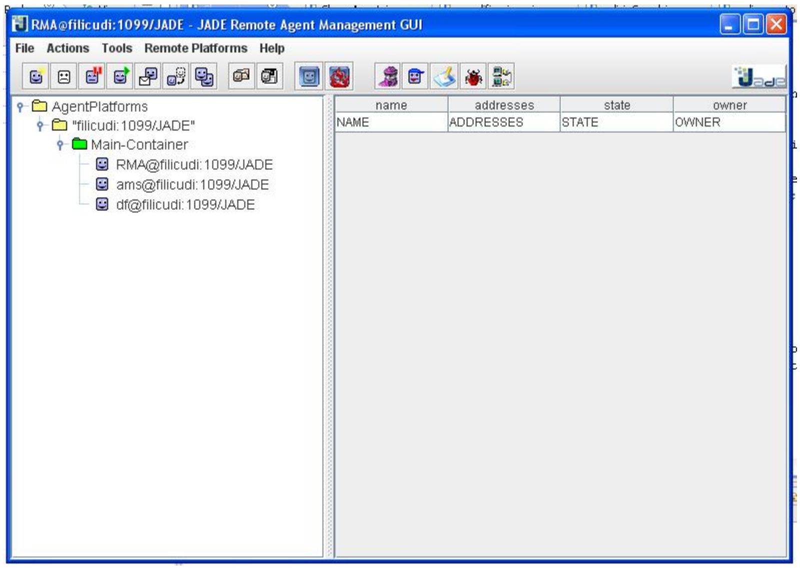

Figure 5.1: The JADE graphical user interface.

class, i.e., by inspecting run-time values of variables, stepping java code, and so on. The EJADE graphical user interface is shown in Figure 5.2.

Finally, in order to provide the multi-agent system with spatial managing and analysis capabilities, two open source Java libraries have been used, namely Geo Tools[10], which provides standards compliant methods for geospatial data manipulation, and Java Topology Suite (JTS)[11], which is an implementation of the Simple Features Specification for SQL, by the OGC[12], which provides both a geometry model and a complete, consistent and robust implementation of fundamental 2D spatial algorithms.

5.1.2 Implementing the Chorem Editing task

The Chorem Editing phase aims to support users in refining the chorematic map obtained from the Chorem Creation step, by editing, moving or deleting the elements composing the map. In particular it performs the following tasks:

[10] The Open Source Java GIS Toolkit [online], website, `http://geotools.codehaus.org`

[11] Java Topology Suite [online], website, `http://www.vividsolutions.com/jts/jtshome.htm`

[12] The Open Geospatial Consortium [online], website, `http://www.opengeospatial.org/`

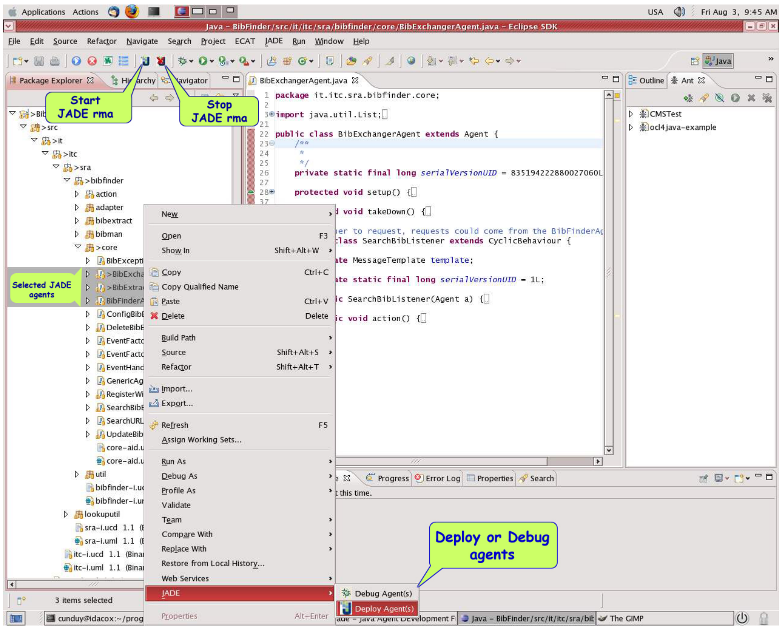

Figure 5.2: The EJADE plug-in embedded in JADE graphical user interface.

- to import a ChorML representation of a chorematic map,
- to display chorems,
- to modify both graphical representation and semantics of chorems, in order both to solve problems regarding placement and layout,
- to export both a SVG and a ChorML representation of chorematic maps.

A Chorem Editor has been implemented in order to carry out previous tasks. Figure 5.3 shows the visual interface of the Chorem Editor, where a simple chorematic map is shown, containing two points with different sizes representing cities, a flow between them, representing a migration, and a region which spatially contains these chorems. 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, (depicted in Figure 5.4), allows users to interact with and modify chorem properties, also affecting their visual representation.

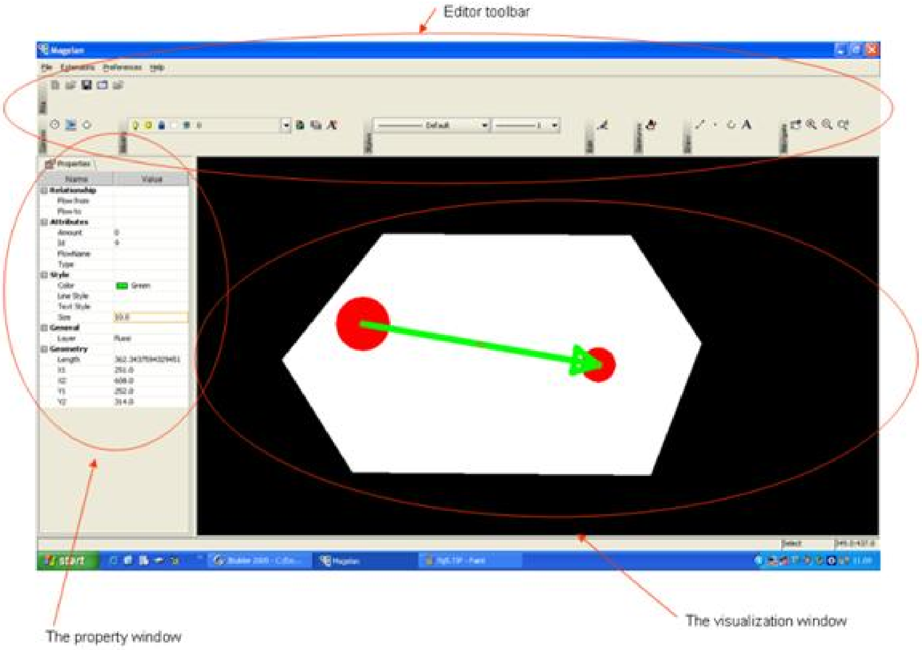

Figure 5.3: The visual interface of the Chorem Editor.

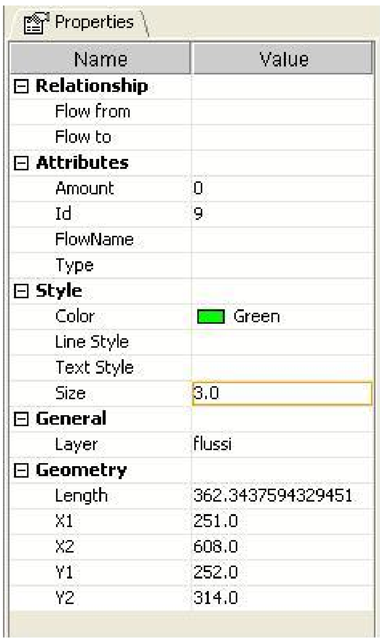

Figure 5.4: The property window of the Chorem Editor.

Analogously, the visualization window, which is meant to display the chorematic map under construction, allows users to manipulate its graphic components, also affecting properties displayed into the property window. This characteristics is shown in Figure 5.5, where the city colour is changed by interacting with the colour palette. As a consequence, the value of the corresponding attribute is modified, in turn. Finally, Figure 5.6 illustrates

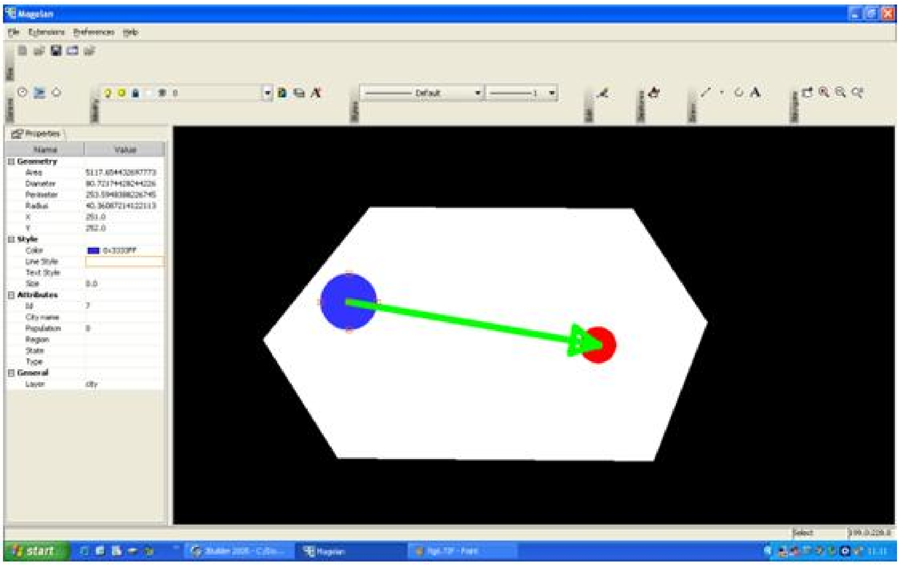

Figure 5.5: User interaction for chorem manipulation.

the Layer Manager, which allows users to handle each graphic object independently of the others by distributing chorems on different layers.

5.2 An example of Chorematic Map Generation

In order to illustrate the functioning and the effectiveness of the system, this section exemplifies the analysis of demographic scenarios related to the Italian population in the last decade. Starting from a dataset containing demographic data about Italy, the example describes the whole process the system accomplishes to construct one of the chorematic maps necessary to represent the migration of the Italian population. In particular, details about the different steps and the modules involved are described. For safe of completeness also the steps performed by the Chorem Extraction System are described in this example, even if much attention is dedicated to the Chorem Visualization System.

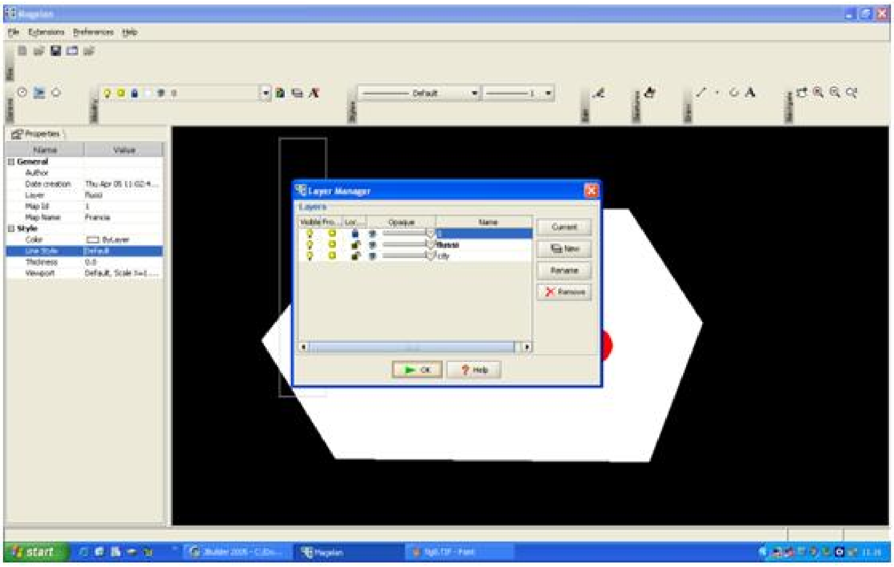

Figure 5.6: The Layer Manager.

As first, proto-chorems are produced which contain data about cities and regions, as schematized in Table 5.1 and Table 5.2, respectively. In particular, in Table 5.1 data related to cities in terms of city name (*Capoluogo*), position (*Lat,Lon*), population (*Pop2000*), city identifier (*Codistat*) and shape (*Shape*), are shown.

In Table B.2 a matrix is shown, where the x value associated with the *(a, b)* pair corresponds to the x number of citizens who migrated from the a region towards the b region. As an example, the *(Calabria, Lombardia) = 5308* implies that 5308 citizens migrated from the *Calabria* region towards the *Lombardia* region. Figure 5.7 depicts spatial components associated with both tables, in terms of geometry and position. By applying spatial data mining techniques and SQL queries on data about regions, five macro-regions are obtained. In particular, the SDO_SAM.SPATIAL_CLUSTERS spatial function determines five groups of regions, clustered according to the closeness property. Based on this result, demographic data depicted in Table 5.2 are then summarized through a proper aggregate function. A threshold is used by the system to select the most important displacements of people. Table 5.3 synthesizes the number of displacements among the derived macro-regions, by highlighting the displacements exceeding 25000 people. Demographic data depicted in Table 5.1 are instead used to determine a set of important cities in term of population. A proper threshold is used by the system to determine the 6 most populated cities through a SQL query, by filtering

Capoluogo	**Lat**	**Lon**	**Pop2000**	**CodIstat**	**Shape**
Agrigento	13,5896	37,3219	54619	084001	Point
Alessandria	8,6200	44,9134	85438	006003	Point
Ancona	13,5110	45,5991	100507	042002	Point
Aosta	7,3177	45,7383	34062	007003	Point
Arezzo	11,8700	43,4733	91589	051002	Point
Ascoli Piceno	13,6313	42,8481	51375	044007	Point
Asti	8,2079	44,8989	71276	005005	Point
Avellino	14,7896	40,9154	52073	064008	Point
Bari	16,8520	41,1187	316532	072006	Point
Belluno	12,2171	46,1503	35050	025006	Point
...	...	...	...	...	...

Table 5.1: Demographic data related to the Italian cities.

Shape	**Regione**	**Piemonte**	**VdA**	**Lombardia**	**Veneto**	**...**
Polygon	Abruzzo	235	4	792	337	...
Polygon	Basilicata	377	6	769	166	...
Polygon	Calabria	2244	167	5308	1083	...
Polygon	Campania	2648	69	8591	3183	...
Polygon	Emilia-Romagna	852	55	4168	1953	...
Polygon	Lazio	1230	55	3701	1968	...
Polygon	Liguria	3184	77	2274	379	...
Polygon	Lombardia	5595	128	0	3579	...
...	...	...	...	...	...	...

Table 5.2: Demographic data related to the Italian regions.

Figure 5.7: Spatial components of Italian regions and cities.

	Nord	Centro	Sud	Sicilia	Sardegna
Nord	0	18112	**31130**	13150	5248
Centro	20903	0	10594	2084	1428
Sud	**66668**	22132	0	2521	847
Sicilia	**26489**	4172	2483	0	319
Sardegna	6323	1420	588	269	0

Table 5.3: Number of displacements among the 5 macro-regions in Italy

the cities exceeding 500000 people. It is worth to note that the extraction system produces such results in terms of tables and alphanumeric data. In order to illustrate results obtained during this phase, a fictitious graphic representation of them is shown in B.15. Here, 5 macro-regions clustered from 20 regions, and 6 cities filtered from 96 cities are depicted.

Figure 5.8: Clustered Regions obtained from the Oracle Spatial function SDO_SAM.SPATIAL_CLUSTERS, and filtered cities.

According to the previous definition, three different chorems are now available, namely:

- macro-regions chorem which contains five chorem elements corresponding to five groups of regions,
- migrating people chorem which contains three chorem elements

corresponding to the three most important flows, and

- most important cities chorem which contains six chorem elements corresponding to the six most populated cities.

Finally, the Chorem Extraction System computes both structural and topological relationships between chorem elements, useful during the visualization phase. In particular, structural relationships represent non-spatial relationships between chorem elements. As an example, a migration between the North macro-region and the South macro-region is represented as a (from, to) relationship between them. Topological relationships represent spatial relationships between chorem elements, expressed through to the Egenhofer's relationships, discussed in Chapter 2. As an example, macro-regions and cities are related each other by a spatial *contains* relationship.

Chorems and relationships obtained by the Chorem Extraction System are then sent to the Chorem Visualization System in terms of *ChorML1* code, in order to derive a visual representation. In Figure 5.9 a *ChorML1* code fragment is listed, which refers to the Italy chorematic map, where the highlighted code portion describes the GML component associated with the chorem element associated with Rome city.

```
<?xml version="1.0" encoding="utf-8"?>
<chorml>
  <ChorematicMap>
    <MapId>map1</MapId>
    <MapName>Italy</MapName>
    <GeneralInformation>
      <Date/>
      <Author/>
      <Format/>
    </GeneralInformation>
    <ChoremList>
      <Chorem>
        <ChoremElementList>
          <ChoremElement>
            <ElementId>elem1</ElementId>
            <ElementName>Roma</ElementName>
            <ElementSize>4</ElementSize>
            <_SHAPE_>
              <gml:MultiPoint srsName="epsg:8307">
                <gml:pointMember>
                  <gml:Point>
                    <gml:coordinates>12.52,
                                     42.88
                    </gml:coordinates>
                   </gml:Point>
                </gml:pointMember>
              </gml:MultiPoint>
            </_SHAPE_>
          </ChoremElement>
.   .   .
.   .   .
.   .   .
```

Figure 5.9: An example of *ChorML1* code, resulting from the Chorem Extraction System.

As discussed in chapter 4, two tasks are performed by the Chorem

Visualization System, namely Chorem Creation and Chorem Editing. During the Chorem Creation task, the multi-agent system creates the agents representing the geographic chorems on the basis of the procedure described in Section 4.3. In particular, Figure 5.10 illustrates the agent structure composed of a Chorematic Map Agent (CMA), three Chorem Agents (CA) and fourteen Chorem Element Agents (CEA). Is worth to note that in this figure the relationship agents which also belong to the structure are not shown in order to preserve the readability.

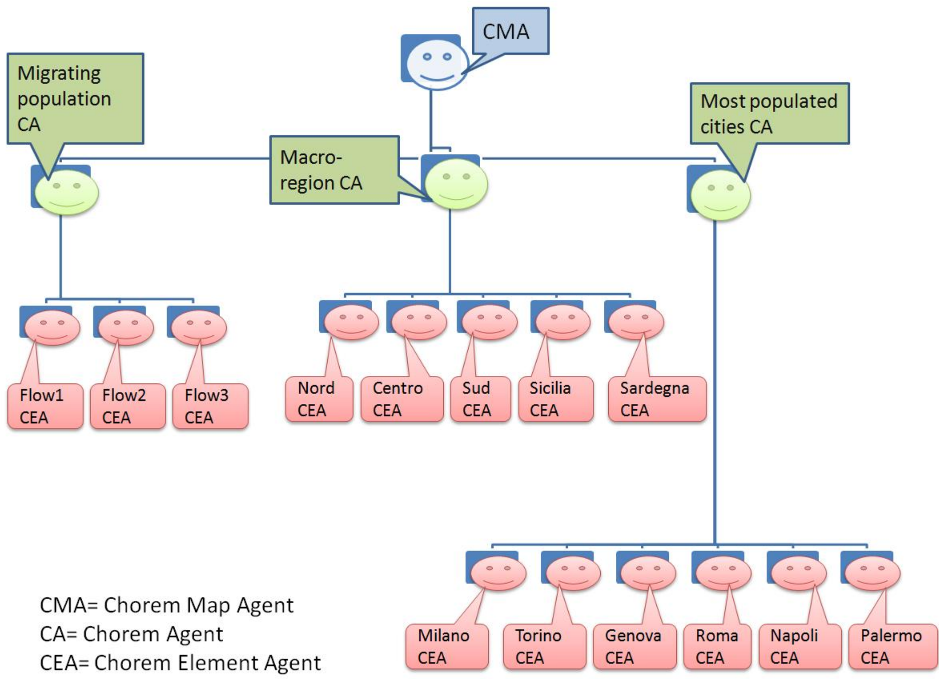

Figure 5.10: The structure of the multi-agent system (without relationship agents) of the running example.

Moreover, also the relationship agents are generated by the system, i.e., a relationship agent for each topological relationship listed in table 5.4 is created. In this table, topological relationships among the five macro-regions (rows) and the six cities (columns) are shown. In particular, the *Touch* relationship between a region and a city means that the city is positioned near the boundary of the region, the *Contains* relationship between a region and a city means that the city is within the region.

Once the structure of the system is generated, the Chorematic Map Agent coordinates the computation of additional topological relationships which are devoted to support the multi-layer simplification task, as discussed in

	Milano	Torino	Genova	Roma	Napoli	Palermo
Nord	Contains	Contains	Touch			
Centro				Contains		
Sud					Touch	
Sicilia						Touch
Sardegna						

Table 5.4: Topological relationships among the five macro-regions and the six cities.

Section 4.4.1. For instance, in this phase the system examines the macro-region geometries in order to verify whether they share some boundaries. In fact, topological errors may occur if two regions sharing a boundary are simplified separately by the two Chorem Element Agents which handle them. Such a situation, may result in gaps or overlapping between regions, as depicted in Figure 5.11. After the computation of additional topological relationships, both macro-region and most populated cities chorems undergo a simplification and a choremization step.

Figure 5.11 illustrates chorems derived from a successful simplification, where topological relationships have been preserved. Figure 5.12 illustrates

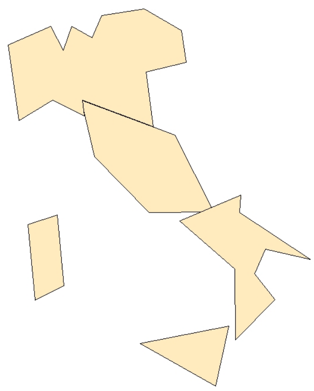

Figure 5.11: Macro-regions chorem simplified not considering sharing boundaries.

chorems derived from the correct simplification step. Once simplified, geometries handled by the Chorem Element Agents may be *choremized*. Such a step is based on an algorithm which stresses simplification by associating geometries with simpler shapes approaching convex polygons with a few number of edges, as discussed in Section 4.4.1. Figure 5.13

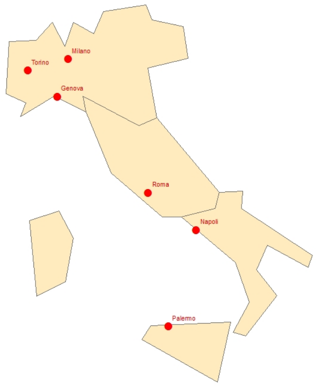

Figure 5.12: Chorems obtained from the simplification phase.

shows the final result of the Choremization step.

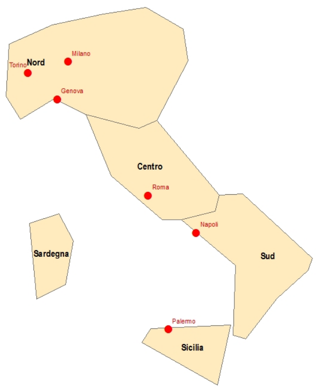

Figure 5.13: Chorems obtained from the *choremization* phase.

In the subsequent Coordinate Translation task, coordinates of the *choremized* geometries are translated into a reference system, local to the visualization layout chosen by the user. In the running example, we have adopted a well-established cartographic procedure. It modifies the *ChorML1* code replacing all the GML tags through a simple tag containing the x and y coordinates in agreement with the chosen layout. Figure 5.14 shows such a modification applied to the code depicted in Figure 5.9.

In order to complete the chorematic map construction, migrating people chorems and annotations chorems have to be managed by the multi-agent system, whose aim is to both spatially arrange them onto the chosen

```
<?xml version="1.0" encoding="utf-8"?>
<chorml>
  <ChorematicMap>
    <MapId>map1</MapId>
    <MapName>Italy</MapName>
    <GeneralInformation>
      <Date/>
      <Author/>
      <Format>
        <Height>600</Height>
        <Width>800</Width>
      </Format>
    </GeneralInformation>
    <ChoremList>
      <Chorem>
        <ChoremElementList>
          <ChoremElement>
            <ElementId>elem1</ElementId>
            <ElementName>Rome</ElementName>
            <ElementSize>4</ElementSize>
            <ElementShape>x=23,y=56</ElementShape>
          </ChoremElement>
.    .    .
.    .    .
.    .    .
```

Figure 5.14: An example of *ChorML1* code, resulting from the coordinate translation phase.

visualization layout and determine their best placement, preserving structural and topological constraints among them.

Once the chorematic map is completed, the multi-agent system transforms the *ChorML1* in *ChorML2*. In particular, the Chorem Element Agents convert the geographic component of chorems into SVG tags. In Figure 5.15 the final ChorML2 code is illustrated, which may be then imported in and edited by the Chorem Editor. Figure 5.16 shows the resulting chorematic map displayed by the Chorem Editor, where the green arrows represent migrating people flows, and labels represent city and macro-region names.

By using the Chorem Editor, the user can modify the chorematic map leaving the editor the management of the relationships existing among chorems. In fact, the Chorem Editor advises the user when a relationship is being violated, letting her/him to choose the behavior to adopt. Figure 5.17 shows a situation, when the user tries to displace the red circle representing the *Genova* city. The Chorem Editor alerts her/him that this displacement violates the *Touch* topological relationship and, if confirmed, the topological relationship will be set to *Disjoint*.

Finally, the user can export the chorematic map into an interoperable visualization format, usable in other systems. In particular, the Figure 5.18 shows the resulting chorematic map displayed through a common

```
<?xml version="1.0" encoding="utf-8"?>
<chorml>
  <ChorematicMap>
    <MapId>map1</MapId>
    <MapName>Italy</MapName>
    <GeneralInformation>
      <Date/>
      <Author/>
      <Format>
        <Height>600</Height>
        <Width>800</Width>
      </Format>
    </GeneralInformation>
    <ChoremList>
      <Chorem>
        <ChoremElementList>
          <ChoremElement>
            <ElementId>elem1</ElementId>
            <ElementName>Rome</ElementName>
            <ElementSize>4</ElementSize>
            <svg>
                <ellipse style="fill: #ff0000" cx="10.7406"
                      cy="4.15625" rx="0.55625" ry="0.55625"/>
            <ellipse style="fill: none; fill-opacity:0;
                     stroke-width: 0.1; stroke: #ff0000" cx="10.7406"
                     cy="4.15625" rx="0.55625" ry="0.55625"/>
            <ellipse style="fill: none; fill-opacity:0;
                     stroke-width: 0.01; stroke: #000000"
                     cx="10.7406" cy="4.15625" rx="0.55625"
                     ry="0.55625"/>
            <text style="fill: #000000; text-anchor:start;
                  font-size: 0.7; font-family: sans; font-style: normal;
                  font-weight: normal" x="8.5" y="4.375">Paris</text>
           </svg>
          </ChoremElement>
.    .    .
.    .    .
.    .    .
```

Figure 5.15: An example of *ChorML2* code.

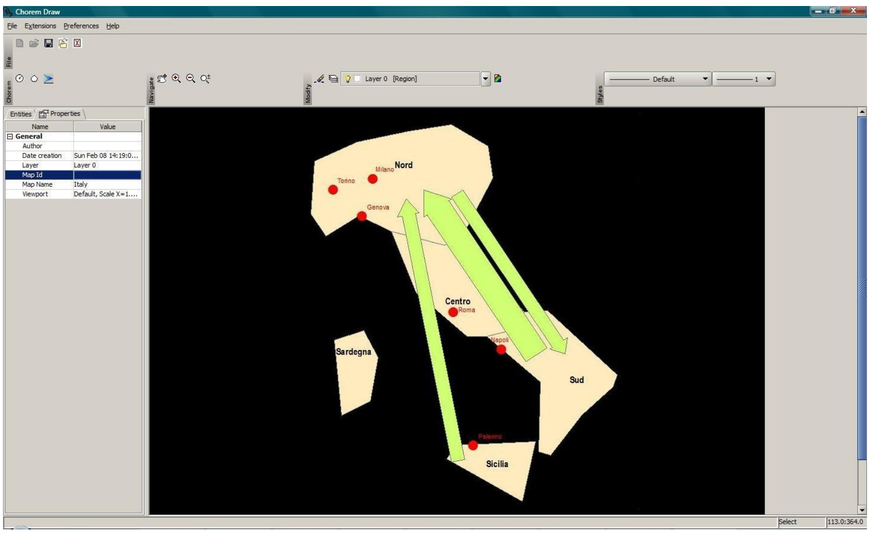

Figure 5.16: The completed chorematic map.

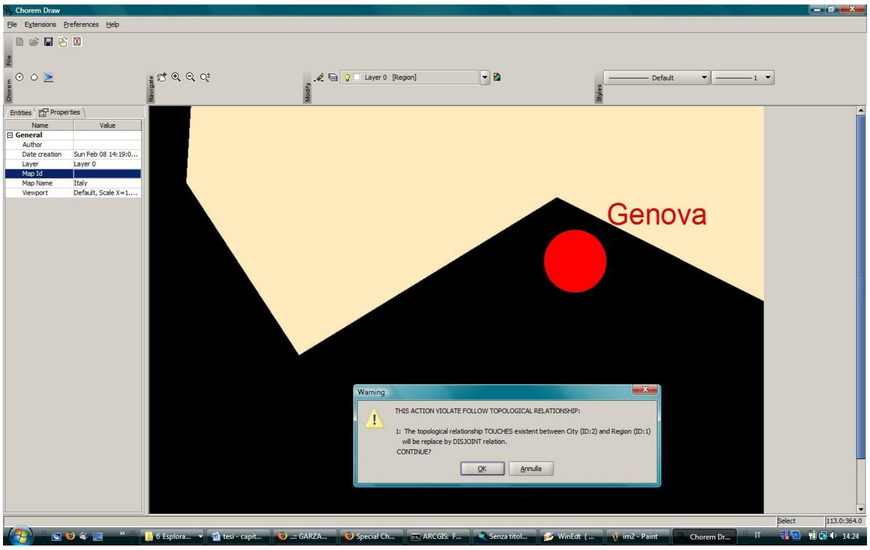

Figure 5.17: Violation of a topological relationship reported by the Chorem Editor.

SVG viewer. It is worth to note that the export task adds a legend the chorematic map in order to support users in interpreting the chorem meaning.

5.3 Conclusion

In this chapter a prototype of the Visualization Chorem System has been introduced, the adopted technology in terms of software components have been described and the system functionalities implementation have been explained. In particular, in Section 5.1 the software components underlying the two most relevant modules have been illustrated, namely Chorem Creation and Chorem Editing, and software technologies embedded into the *ChorML* language have been listed. Finally, in Section 5.2 the construction of a chorematic map has been exemplified in order to demonstrate the effectiveness of the system. Conclusions, contributions and directions for future work of this dissertation will be presented in the next chapter.

Figure 5.18: The Chorematic map showing the flows of migrating population among macro-regions in Italy.

Chapter 6

Conclusion

Contents

After an initial general discussion, this chapter summarizes the content of this Thesis, highlighting the contribution to the research, and discusses possible scenarios for the future work. In particular, the General Discussion Section presents the characteristics of the chorem concept and the main features of our system, whose aim is to extract geographic objects and spatio-temporal phenomena from geographic databases and to represent them into a chorematic map. The subsequent section presents the main contribution of this thesis, analyzing its significant aspects and the impact of them on the related disciplines. Finally, the Future Work Section discusses some final remarks about the research that we aim to carry out in order to extend and improve functionalities of our system, and to investigate other possible applications of chorems in the geographic information field.

6.1 General discussion

Traditional cartography is an essential tool to describe facts and relationships concerning with territory. It associates geographic concepts with graphic symbols which help readers to gain an immediate comprehension of represented data. This approach is well-established and decision makers are usually satisfied by its expressiveness when dealing with the cartography of facts. A similar solution to adopt when decision makers have to face issues related to problem visualization has investigated in this dissertation. In this case, expert users are provided with a large amount of data which describe both objects and spatio-temporal phenomena. Moreover, correlations among data are not explicitly stated, then a big effort is required to detect and then represent dynamics, movements and changes which underlie possible problems.

In order to determine an innovative cartographic solution meant to overcome acknowledged limits of the traditional cartography, an international research activity has started involving experts from different domains. From the ambit of this international collaboration this thesis has originated, where, first of all, we have proposed the usage of the chorems as visual summaries of geographic database contents. An initial study was carried out in order to analyze different solutions. Among them, chorems seemed to be the most suitable, thanks to their expressiveness and synthesis capabilities. Moreover, the results of this study motivated us to initially provide a formal definition and classification, meant to homogenize chorem construction and usage.

Then, a system to semi-automatically generate chorematic maps has been designed and implemented. This system is composed of two main subsystems, namely the Chorem Extraction System and the Chorem Visualization System. The former is meant to extract spatial objects and phenomena form a spatial database and associate them with different chorems types. The latter is meant to transform a list of chorems extracted by the extraction subsystem and stored in a XML-like language, to a visual representation by using a multi-agent system. Different technologies have been adopted, ranging from SQL queries and (spatial) data mining, to simplification and best-placement algorithms. Moreover, when necessary, open source platforms and Java programming language have been employed both to realize new ad hoc modules and to integrate existing ones into a homogeneous framework. Finally, an XML-like multi-level language, named ChorML, has been introduced to store information about chorems and to support communication among system modules.

6.2 Contribution

This section summarizes the main contributions of our work. Generally speaking, the aim of this thesis was to determine an innovative cartographic solution to visually represent knowledge related to a territory, that manually or automatically could be extracted from a (geographic) dataset. Several solutions have been proposed in literature and their benefits have been investigated. Our major contribution along this line has been to define a methodology to visualize geographic database summaries, expressing them through "schematized representations of territories", known as chorems, which looked suitable candidates for this purpose. These summaries are characterized both by a geographic generalization, which simplifies the shapes of the territory under investigation, and by a semantic generalization, which selects the relevant aspects of spatial and non-spatial attributes of the geographic database. A further innovative result of our methodology is also represented by the opportunity of using the chorems as representation for geographic knowledge. In fact, this approach may confer added value when little-known territories are investigated for different purposes, such as for geo-marketing, environmental studies and archaeology, allowing to easily represent significant before unknown aspects through a combination of simple visual metaphors, that convey useful geographic knowledge. It is worth to point out that also conventional geography may benefit from our methodology because it can be used to support construction of maps containing chorems, whose purpose still remain to schematize relevant factors of a territory. The specific contributions produced by investigating and implementing our methodology are related to different research areas and can be synthesized as follows:

- The formal specification of chorems in terms of visual language and structure is given. This specification is meant to both formalize the structure of chorems and solve ambiguities of terms in the chorems domain. In fact, in literature several definitions of chorems have been presented, most of them starting from Brunet's definition. However, we have observed that expert users typically prefer to either adopt a minimal set of Brunet's chorems or generate an *ad hoc* chorem vocabulary, in spite of the existing definitions in literature. Therefore, a formal specification of chorems has been necessary, in order to provide a usable framework for computer systems, which supports the automatic generation of chorematic maps. In order to satisfy this requirement, a formal definition of an XML-like multi-level language, named ChorML, has been introduced to store information about chorems and to support communication among system modules.
- The definition and the implementation of a system which generates chorematic maps starting from geographic database content, in a semi-automatic manner. The system proposed in this dissertation is meant:
  - To extract geographic objects and spatio-temporal phenomena from a geographic database by using SQL queries and (spatial) data mining techniques, to allow expert users to identify interesting patterns and trends, and to associate them with a proper type of chorems. In literature, the extraction of information from geographic databases has been a relevant topic of investigation. However, many features makes this issue innovative, such as the attempt of automating the overall process, the method of associating the extracted patterns and trends with pre-defined visual categories, and the integration of the extraction component with a system able to display the visual representation of such information.
  - To transform a list of chorems and relationships, stored in a XML-like language, to a visual representation, namely a chorematic map, by implementing a multi-agent system architecture for managing multi-layer simplification and map object placement and labeling. The proposed multi-agent system architecture is based on the chorem classification given in this dissertation. It is similar to some approaches proposed in literature meant to execute cartographic generalization, while it differs for its specific purposes. In fact, our approach not only deals with a particular generalization process, i.e.,

simplification, but it is also meant to resolve cartographic displaying problems, such as map labeling and flow displaying. A significant feature of the proposed multi-agent system consists of its capability to perform a multi-layer simplification task. In literature, the simplification problem is a relevant topic of investigation, although little research has been carried out on simplification executed in a context. In fact, the simplification applied on a single geographic object, such as a line or a polygon, corresponds to a trivial task because it affects just its shape. On the contrary, a simplification task may violate (topological) relationships when it is autonomously applied on different adjacent objects belonging either to the same layer or to different overlapped layers. Finally, some work has been done for understanding how different levels of simplification are perceived by users, and how a system can take into account such a difference of perception.

6.3 Future work

As future works we intend:

- To execute comparative experimental studies in order to evaluate the effectiveness of our approach. In particular, such experiments are meant to measure expert users' capability in using chorematic maps produced by the system, in order to locate factors generating anomalies, difficult to detect when supported by traditional cartography.
- To enhance techniques embedded in the system in order to reduce the human intervention required, at level of both extraction and visualization of chorems.
- To investigate the integration of the system proposed in this dissertation within a decision support system.
- To further investigate Egenhofer's topological model (which is used to characterize all relationships holding among areas, lines and points) in order to improve the accuracy of the map simplification task.
- To verify whether chorems can be used as means to navigate geographic database contents. In fact, an interesting approach may consist of applying Ben Shneiderman's mantra for designing human

interfaces “Overview, zoom and filter, details on demand” for geographic database access [119]. According to this approach, a chorematic map may represent an initial ”overview” level when studying a territory. Then, a geographic phenomenon represented within the map by a set of chorems, could be thoroughly analyzed by zooming, thus obtaining more details about it. For this purpose, cartographic and semantic generalization functionalities and Map Multiple Representation techniques could be added to the system.

- To verify whether chorems can be used as query language for geographic databases. In fact, according to the query by sketch approach, a user could draw a set of chorems which represent the geographic object or the phenomenon which s/he is looking for within the database.

- To execute studies in order both to investigate cognitive and semiotic aspects of chorems and to support the interoperability of the proposed system.

Part I

Annexes

Appendices

Appendix A

Riassunto in italiano

A.1 Introduzione

La visualizzazione di dati geografici appartiene tradizionalmente all'area di ricerca nota come *Geographic Visualization* o *Geovisualization*. In particolare, la Geovisualization integra metodi da diversi campi di ricerca, quali la *Scientific Visualization*, la Cartografia, l'Analisi di Immagini, l'*Information Visualization*, e i Sistemi Informativi Geografici (GIS), per fornire teorie, metodi e strumenti per l'esplorazione visuale, l'analisi, la sintesi e la presentazione di dati geospaziali.

La cartografia tradizionale è un mezzo essenziale per descrivere i fatti e le relazioni che riguardano un territorio. Essa associa dei simboli grafici, che aiutano i lettori ad avere una comprensione immediata dei dati rappresentati, a concetti geografici. Questo è un metodo affermato da tempo, e gli utenti esperti che prendono decisioni dall'analisi di mappe, sono solitamente soddisfatti dalla loro capacità espressiva, quando ci si occupa della cartografia che riguarda esclusivamente fatti. Un approccio più complesso è richiesto invece quando bisogna affrontare scenari che fanno riferimento a questioni complesse, come ad esempio la politica, i problemi economici e demografici, nei quali gli utenti esperti hanno a disposizione un gran numero di dati che descrivono sia elementi geografici che fenomeni spazio-temporali.

In questi casi, l'obiettivo principale non è restituire esattamente tutti i particolari ad un alto livello di dettaglio, ma piuttosto mostrare gli aspetti piu salienti. In altre parole, un supporto significativo alle attività di modellazione e analisi, può consistere in un'immediata sintesi dei dati di interesse, che trascura dettagli poco significativi. Un supporto tecnologico alla realizzazione di questo obiettivo può essere ottenuto integrando risultati da diverse aree di ricerca, quella dei Database Geografici, del Knowledge Discovery in Databases e dello Spatial Data Mining. In particolare, ognuna di queste discipline può contribuire alla definizione di una metodologia di Geovisualization per soluzioni cartografiche avanzate, fornendo tecniche per scoprire informazioni non conosciute e potenzialmente significative dal contenuto di Database Geografici..

Per determinare una soluzione cartografica innovativa, mirata ad oltrepassare i limiti riconosciuti della cartografia tradizionale, è stato promosso un progetto di ricerca internazionale fra i seguenti tre istituti di ricerca: (i) Institut National des Sciences Appliquées (INSA) de Lyon (Francia), (ii) Universitá degli Studi di Salerno (Italia) e (iii) Tecnólogico de Monterrey, Campus de Puebla (Mexico). Tale progetto intende definire soluzioni cartografiche in grado di rappresentare in maniera appropriata informazioni, estratte da database geografici, che si riferiscono sia a elementi geografici che a fenomeni spazio-temporali.

L'obiettivo di questa Tesi, che nasce nell'ambito di questo progetto internazionale, è quello di ottenere nuove modalità di rappresentazione visuale che possano fornire ai decisori uno strumento in grado di restituire una sintesi immediata dei dati di interesse, che trascuri particolari non utili.

Tale sintesi può rappresentare il punto di partenza per ulteriori attività, volte ad ottenere l'analisi di dati territoriali e per coadiuvare gli utenti esperti nei processi decisionali, in modo da colmare il divario tra la complessità delle applicazioni adottate e la necessità di una risposta rapida ed esauriente attesa dagli esperti di dominio.

La soluzione proposta in questa direzione, si basa sul concetto di *Corema* e sulla sua capacità di sintetizzare scenari che comprendono elementi geografici e fenomeni spazio-temporali, associando a questi notazioni visuali schematizzate.

Il riassunto in italiano della tesi è organizzato come segue. La definizione e la classificazione dei coremi sono riportate nel paragrafo A.2. Nel Paragrafo A.3 sono descritti l'architettura del sistema ChEViS (Chorem Extraction and Visualization System) e il linguaggio ChorML. In particolare, vengono dettagliati gli algoritmi e le tecniche su cui sono basati i due sotto-sistemi che compongono il sistema ChEViS, ovvero il sistema di estrazione dei coremi, detto *Chorem Extraction System*, e il sistema di visualizzazione dei coremi, detto *Chorem Visualization System*. Un esempio di costruzione di mappa corematica è illustrato nel Paragrafo A.4. Le conclusioni, i contributi alla ricerca e le idee per futuri lavori sono riportate nel Paragrafo A.5.

A.2 Coremi

Il termine *corema* proviene dalla parola greca $\chi\omega\rho\eta\mu\alpha$, che significa spazio o territorio. Secondo la definizione del geografo francese Roger Brunet [13], i coremi sono una rappresentazione schematica di un territorio, che elimina ogni dettaglio non necessario alla comprensione di una mappa. Brunet ha proposto 28 coremi elementari, ciascuno dei quali rappresenta una semplice configurazione spaziale. Integrando tali configurazioni spaziali in una mappa, vari fenomeni spazio-temporali possono essere rappresentati a diverse scale.

La Figura A.3 mostra un confronto tra una tradizionale mappa della

Francia e una mappa della Francia che contiene coremi, nella quale diversi aspetti sono messi in evidenza:

- la forma semplificata dei confini,
- le città più importanti,
- le aree con differente livello di sviluppo,
- i flussi che rappresentano le principali rotte nazionali.

Figure A.1: Una tradizionale mappa della Francia.

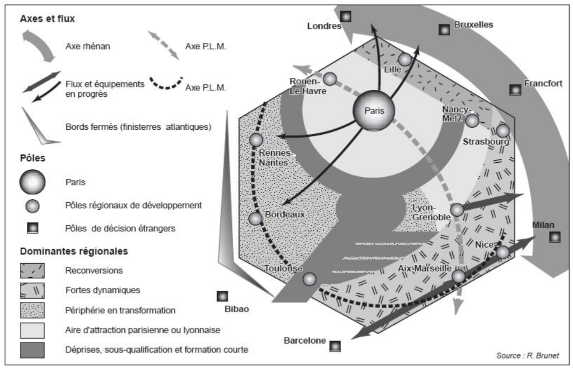

Figure A.2: Mappa della Francia che contiene coremi. (Fonte [72])

Figure A.3: Confronto tra una mappa tradizionale e una mappa contenente coremi

Fin dalla loro comparsa, i coremi hanno avuto una grande diffusione nella comunità dei geografi, dove sono stati ampiamente discussi e utilizzati [13, 14, 15, 23, 9]. Diverse definizioni e classificazioni dei coremi sono state enunciate a partire dalla proposta di Brunet, così come una grande quantità di mappe contenti coremi sono state prodotte manualmente, con lo scopo sia di dare una visione globale di una realtà di interesse sia di ottenere un'analisi di uno specifico problema. Ad esempio, in Figura A.4 viene mostrata una mappa che contiene coremi sulla Spagna, nella quale vengono evidenziate differenti dinamiche sullo sviluppo del territorio.

Tuttavia, la mancanza di un approccio rigoroso per la creazione e la costruzione di coremi ha determinato la proliferazione di soluzioni *ad hoc*, che in molti casi non riescono a trasmettere le informazioni che dovrebbero trasmettere. I primi risultati della ricerca su cui si basa questo lavoro si possono trovare in [29], in cui è definita una versione preliminare dell'architettura del sistema proposto in questa Tesi, e quella un linguaggio

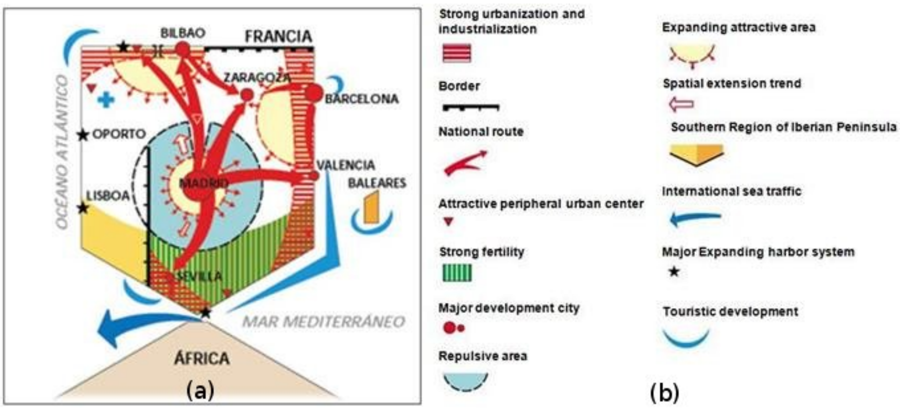

Figure A.4: Mappa che contiene coremi che riguardano la Spagna.

XML-like, chiamato *ChorML*, ideato per consentire la comunicazione tra i moduli del sistema. Sulla base di tali risultati preliminari e delle varie definizioni dei Coremi, in questa Tesi viene fornita una definizione formale e una classificazione dei coremi in termini di struttura e significato, e viene proposto un sistema in grado supportare gli utenti esperti nella generazione di mappe contenenti coremi.

A.2.1 Definizioni di base

In questo paragrafo sono presentate alcune definizioni di base che riguardano i coremi. Tali definizioni mirano sia a formalizzare i termini che vengono utilizzati nella Tesi sia per risolvere alcune ambiguità dei termini che caratterizzano il campo dei coremi.

Definizione 1: un *Pattern Geografico* o *Pattern* è una interessante regolarità di qualche fenomeno geografico scoperta in database geografico utilizzando funzioni di Data Mining (spaziale). I pattern possono essere utilizzati come punto di partenza per individuare fenomeni spazio-temporali e relazioni tra di essi. Un esempio di pattern geografico: "in Messico, nei luoghi in cui si coltiva l'agave, viene prodotta la tequila".

Definizione 2: I *Proto-coremi* possono essere definiti come dati, ripuliti e organizzati, in modo tale che su di essi possano essere applicate query SQL, funzioni di Data Mining (spaziale) o di altri metodi di analisi, al fine di scoprire interessanti pattern.

Definizione 3: Un *Elemento di Corema* è ogni elemento grafico di base che può rappresentare un unico oggetto geografico o un singolo fenomeno. Ad

Esempio, un punto che rappresenta una città, o una freccia che rappresenta un flusso, come mostrato in Figura A.5

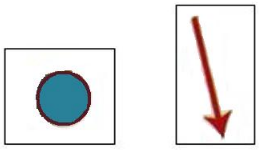

Figure A.5: Punto e freccia che rappresentano una città e un flusso.

Definizione 4: Un *Corema* è un insieme di *elementi di corema* della stessa tipologia. Ad esempio, le città più importanti di un paese o i principali flussi tra le città, come mostrato in Figura A.6.

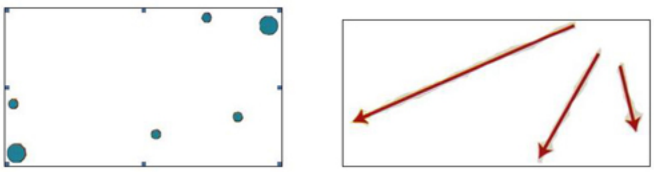

Figure A.6: Città e flussi.

Definizione 5: Una *Mappa Corematica* è un insieme di coremi che schematizzano dati di interesse legati a un territorio. Una leggenda è generalmente associata ad una mappa corematica, utile a chiarire il significato di ogni corema contenuto nella mappa, come illustrato Figura A.7.

A.2.2 Classificazione dei coremi

I coremi possono essere classificati in tre categorie principali, *coremi geografici*, *coremi di fenomeno* e *coremi di annotazione*.

I coremi geografici rappresentano elementi con semplici caratteristiche territoriali come punti, linee e poligoni e gli elementi costituiti da loro combinazioni, come ad esempio le reti. A titolo di esempio, le zone tratteggiate con peculiari caratteristiche e le città rappresentate da cerchi e stelle, raffigurate in Figura A.8, rappresentano istanze di coremi geografici, contenenti rispettivamente poligoni e punti.

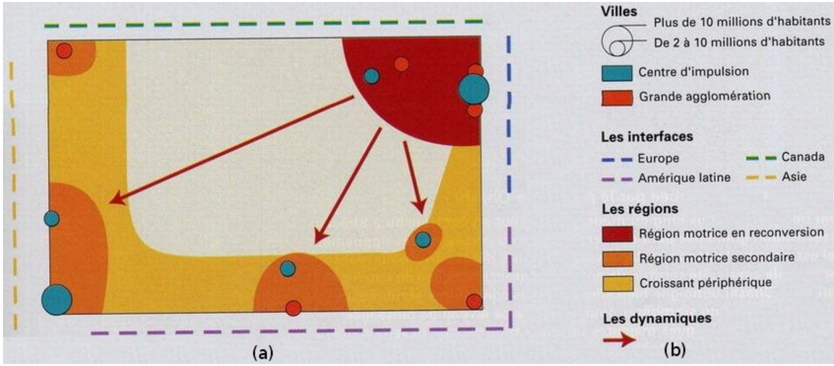

Figure A.7: Mappa che contiene coremi che riguardano gli USA.

I coremi di fenomeno descrivono fenomeni spazio-temporali e coinvolgono uno o più coremi geografici. L'insieme iniziale di coremi di fenomeno che è stato individuato si compone di tre tipi, rispettivamente *Flusso*, *Trofismo* e *Diffusione Spaziale*. Il corema di tipo flusso rappresenta lo spostamento di popolazione o di oggetti tra punti e/o poligoni, ovvero tra punti di interesse e aree. Le frecce rosse e blu mostrate in Figura A.8 rappresentano un esempio corema di fenomeno flusso. Il corema di tipo trofismo rappresenta uno spazio omogeneo attrattivo o ripulsivo, intorno a un corema geografico. Ad esempio, può essere utilizzato per descrivere una città o una regione che attira popolazione da altri luoghi. In Figura A.8, la zona semi-circolare intorno alla città di Barcellona rappresenta un esempio di un corema di fenomeno trofismo. Infine, la diffusione spaziale rappresenta una progressione o regressione spaziale, che parte da un corema geografico verso una direzione specifica. Per fare un esempio, può essere utilizzato per descrivere la tendenza allo sviluppo di una città in una determinata direzione. In Figura A.8, un corema di fenomeno diffusione spaziale è rappresentato dalla frecce gialle intorno la città di Madrid.

I Coremi di annotazione rappresentano etichette o annotazioni di una mappa. Ad esempio, come illustrato in Figura A.8, i nomi delle città vicino ai punti che le simbolizzano, rappresentano le istanze di coremi di annotazione. Ulteriori coremi di annotazione possono essere aggiunti in seguito, al fine di completare una mappa corematica. A titolo di esempio, le etichette *Mar Mediterraneo* e *Oceano Atlantico* sono elementi che sono stati aggiunti per fornire agli utenti informazioni aggiuntive che non sono contenute nel database da cui la mappa è stata estratta.

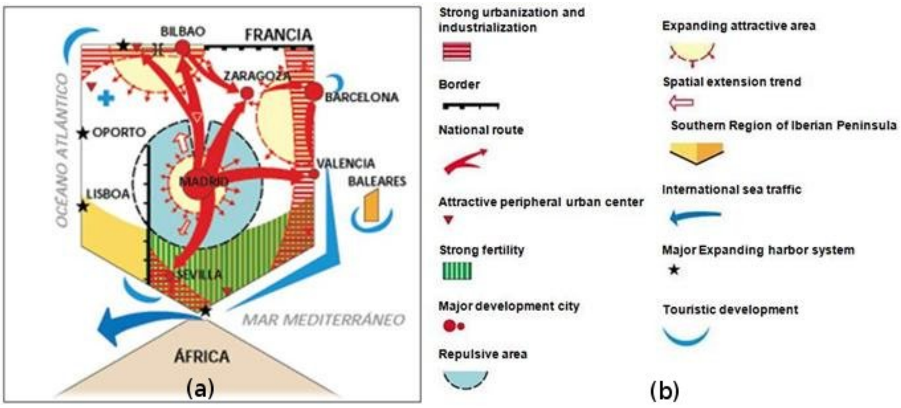

Figure A.8: Mappa che contiene coremi che riguardano la Spagna.

A.3 ChEsViS: Chorem Extraction and Visualization System

In Figura A.9, è mostrata l'architettura del sistema ChEViS (Chorem Extraction and Visualization System), che consiste di due componenti principali, ovvero i sottosistemi per l'estrazione e la visualizzazione dei coremi. Il primo sottosistema, o *Chorem Extraction System*, è destinato a gestire le informazioni disponibili nei dataset geografici per estrarre informazioni rilevanti, mentre il secondo sottosistema, o *Chorem Visualization System*, si occupa di tali informazioni assegnando ad esse una rappresentazione visuale in termini di coremi e mappe corematiche. È importante notare che la struttura dei coremi è gestita da entrambi i sottosistemi. Tuttavia, essi usano e/o modificano differenti sottoparti della struttura del corema a seconda della fase di esecuzione. In particolare, il sistema di estrazione si occupa delle proprietà concettuali dei coremi, come il nome, il tipo, la dimensione e le coordinate, esprimendoli attraverso attributi alfanumerici. Per quanto riguarda il sistema di visualizzazione, esso crea e modifica le componenti geometriche dei coremi, a partire dalle proprietà concettuali generate dalla fase precedente, al fine di assegnare loro una corretta rappresentazione visuale.

La comunicazione tra le componenti del sistema e la memorizzazione della struttura dei coremi è basata su un linguaggio multi-livello *XML-like*, che prende il nome di *ChorML*. In particolare, un opportuno livello di ChorML viene utilizzato in base alle componenti coinvolte, in modo da poter essere compatibile con le diverse specifiche della struttura che memorizza i coremi e ricavare differenti formati utili al processo in esecuzione.

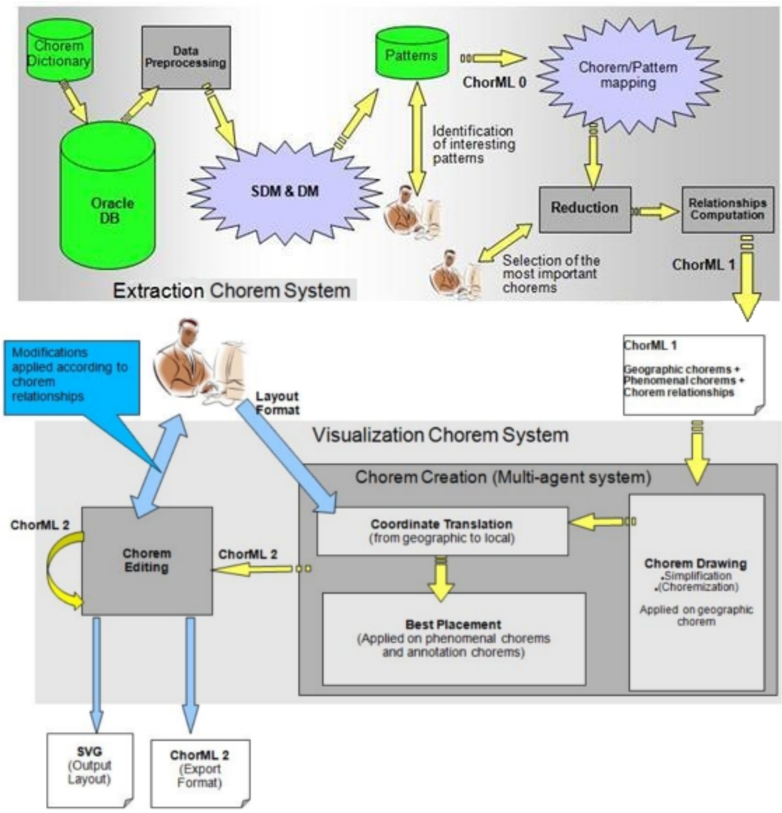

Figure A.9: Architettura del sistema ChEViS.

A.3.1 Chorem Extraction System

Questa sezione presenta brevemente il *Chorem Extraction System*, la cui architettura mostrata in Figura A.10, il cui scopo è di estrarre oggetti geografici e fenomeni spazio temporali da database geografici. La struttura preliminare presentata in questa sezione deriva dal progetto internazionale discusso in [82]. Confrontata con l'idea iniziale, l'architettura del sistema è stata perfezionata nei diversi moduli che lo compongono. L'intero processo

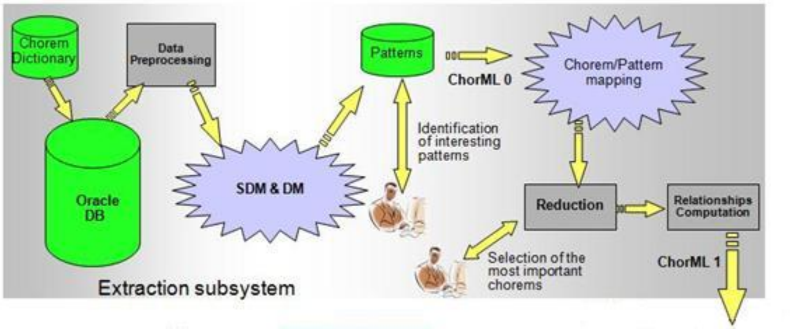

Figure A.10: Architettura del sistema che estrae i coremi.

parte considerando un database geografico, contenente dati di interesse e di un dizionario che contiene la descrizione dei coremi in termini di nome, tipo e struttura. In particolare, il dizionario dei coremi viene importato nel database e viene impiegato nelle successive fasi del processo di estrazione dei coremi. Per quanto riguarda le tecnologie utilizzate per i processi di pre-elaborazione dei dati e di data mining (spaziale), il sistema di estrazione include il DBMS Oracle[1] e la sua estensione spaziale[2] e il sistema SubDue [75], migliorato con l'uso dell'algoritmo PAM (Partitioning Around Medoids) [103]. Inoltre, il sistema fornisce agli utenti feedback sui risultati parziali. Ovvero, ad ogni passo gli utenti possono interagire con il sistema, al fine di perfezionare le query e le funzioni, migliorando così la qualità dei risultati. Successivamente, sono eseguiti un processo di mappatura tra pattern e coremi e un processo di riduzione, teso ad ottenere un significativo, ma ridotto, insieme di coremi. In particolare, il primo processo associa ciascun patter estratto con un corema, mentre il secondo processo determina una riduzione del numero di coremi in una lista dei coremi più significativi. Infatti, una gran quantità di pattern può essere estratta dalla procedura precedente, rischiando in tal modo di sovraccaricare la mappa corematica in costruzione con troppi

[1] Oracle Corporation [online], website, `http://www.oracle.com/index.html`

[2] Oracle Spatial [online], website, `http://www.oracle.com/technology/products/spatial/index.html`

coremi. Entrambi i processi consentono l'intervento degli utenti, al fine di selezionare i pattern e i coremi più significativi. Infine, viene eseguito il calcolo delle relazioni per ottenere un elenco delle relazioni spaziali tra ogni coppia di elementi di corema. A questo scopo, vengono utilizzate funzioni di Oracle Spatial e i risultati sono codificati in codice ChorML, che viene in seguito inviato al sistema di visualizzazione.

A.3.2 Chorem Visualization System

Questo paragrafo presenta una parte significativa del lavoro di Tesi, il progetto e l'implementazione del sistema di visualizzazione dei coremi, detto *Chorem Visualization System*. Lo scopo di questo sistema è di trasformare una lista di coremi e di relazioni tra coremi, proveniente dal sistema di estrazione in termini di testo, in una rappresentazione visuale, ovvero in una mappa corematica. I risultati ottenuti in [82] e in [30] sono stati raffinati e migliorati, considerando sia l'evoluzione delle definizioni della classificazione dei coremi, sia di alcune tecniche di semplificazione cartografica, che mirano a garantire la semplificazione multi-layer della mappa. Come mostrato in Figura A.11, due diverse fasi vengono svolte da questo sistema, ovvero la creazione de coremi, o *Chorem Creation*, e l'editing dei coremi, o *Chorem Editing*.

Chorem Creation

La fase di creazione dei coremi viene svolta attraverso un sistema multi-agenti che esegue tre passi, *Chorem Drawing*, *Coordinate Translation* e *Best Placement*. L'architettura del sistema multi-agente progettato per questa fase è mostrata in Figura A.12.

Per quanto riguarda il passo di Chorem Drawing, esso è applicato ai coremi geografici, che esplicitamente contengono componenti geometriche. Tale passo viene effettuato attraverso due operazioni, la semplificazione e la *coremizzazione*, mostrate in Figura A.13 e Figura A.14.

In particolare, la semplificazione determina una versione semplificata dei dati geometrici, riducendo il numero di vertici della forma originale. Questa tecnica tecnica è basata su funzioni spaziali, che incorporano l'algoritmo di *Ramer-Douglas-Peucker* (RDP) [107, 39] e i suoi miglioramenti [63, 146], e su metodi che tengono conto che la semplificazione viene effettuata in un contesto [12], e non deve violare le relazioni topologiche che intercorrono tra gli elementi dello stesso corema o tra elementi di coremi diversi. L'operazione di coremizzazione stressa

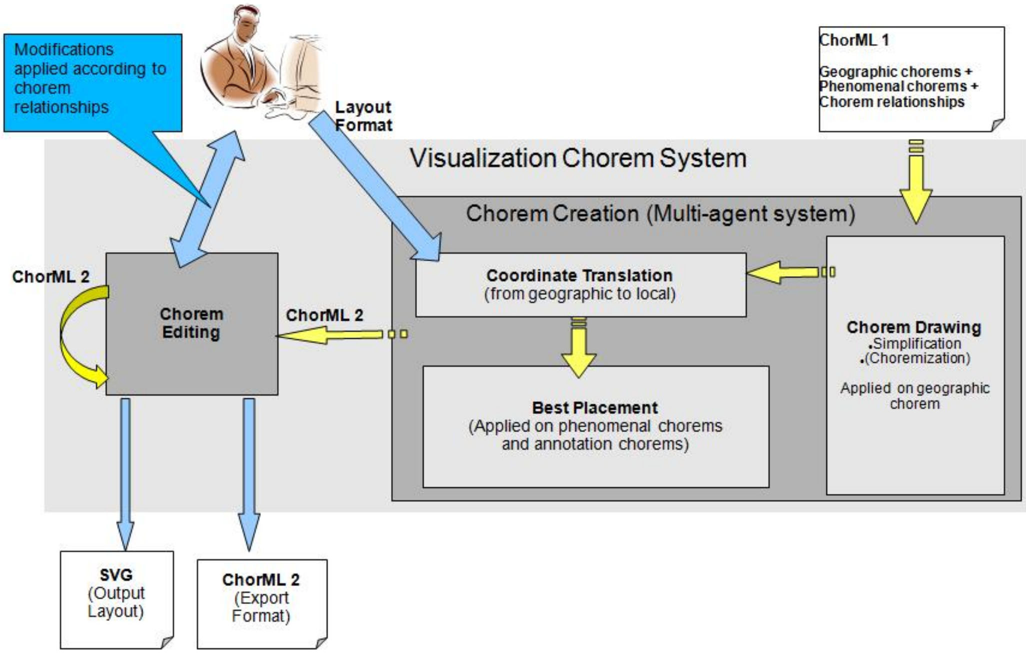

Figure A.11: Architettura del sistema di visualizzazioni dei coremi.

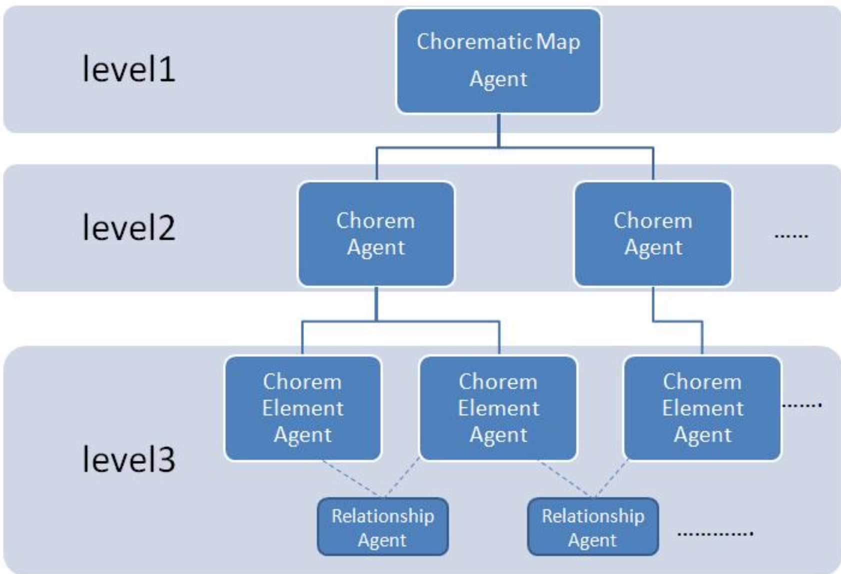

Figure A.12: Architettura del Sistema Multi-agente.

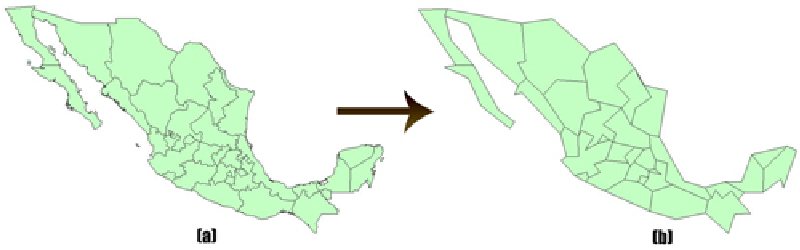

Figure A.13: Esempio dell'operazione di semplificazione.

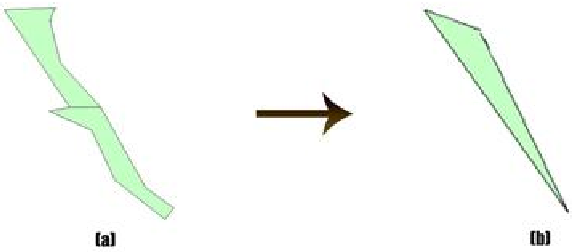

Figure A.14: Esempio dell'operazione di *coremizzazione*.

ulteriormente la semplificazione, associando alla componente spaziale dei coremi gia semplificata una forma di un poligono regolare con pochi lati. Il passo di *Chrorem Drawing* garantisce il mantenimento delle relazioni topologiche tra gli elementi della mappa, sia durante la semplificazione che durante la coremizzazione.

Il passo successivo è il *Coordinate Translation*, che essenzialmente converte le coordinate, appartenenti alla componente spaziale dei coremi, da un sistema di riferimento geografico a un sistema di riferimento locale che fa riferimento al layout di visualizzazione scelto dall'utente in questa fase.

Infine, il passo di *Best Placement*, destinato a integrare coremi di fenomeno e coremi di annotazioni nella mappa corematica, assegnando loro la posizione più corretta. Il sistema multi-agente compie questa operazione usando essenzialmente il noto operatore di *displacement* della generalizzazione cartografica, e considerando che gli elementi di corema presenti nella mappa corematica possiedono diverse priorità. Ovvero, se nasce un conflitto tra due elementi, quello con priorità minore è costretto a spostarsi.

Infine, per soddisfare la condizione di leggibilità della mappa, insiemi indipendenti di coremi possono essere aggregati in mappe corematiche differenti. Ad esempio, da una mappa corematica contente dieci coremi potrebbero derivare due mappe corematiche contenenti cinque coremi ciascuna.

Chorem Editing

Nella fase di creazione dei coremi potrebbero verificarsi alcune difficoltà nel posizionamento dei coremi, così come potrebbero essere richiesti dagli utenti ulteriori raffinamenti che riguardano le proprietà semantiche e grafiche dei coremi. L'obiettivo della fase di editing dei coremi è quello di supportare gli utenti nella modifica della mappa corematica, aggiornando, spostando o eliminando alcuni elementi. Per supportare questa fase, il Chorem Editor è stato implementato come estensione di Magelan[22], un editor di grafica vettoriale 2D open source, basato sul linguaggio di programmazione Java. In particolare, il Chorem Editor può svolgere i seguenti compiti:

- importare una rappresentazione in ChorML di una mappa corematica,
- visualizzare la mappa corematica,
- modificare sia la rappresentazione grafica che la semantica dei coremi
- esportare sia una rappresentazione grafica (SVG)[3] che una rappresentazione in ChorML della mappa corematica.

A.3.3 Il linguaggio ChorML

Un linguaggio *XML-like* multi-livello, detto *ChorML* è stato definito al fine di memorizzare le informazioni riguardanti i coremi e per consentire la comunicazione di tali informazioni tra i diversi moduli del sistema. I suoi livelli sono descritti come segue:

- il livello 0 è composto da tag XML[4] e GML[5]. E' stato pensato per memorizzare le informazioni sui proto-coremi, l'origine dei dati (lineage) e le funzioni ad essi applicate al fine di ottenere i coremi.

[3] Scalable Vector Graphics [online], website, `http://www.w3.org/Graphics/SVG/`
[4] eXtensible Markup Language [online], website, `http://www.w3.org/XML/`
[5] Geography Markup Language [online], website, `http://www.opengis.net/gml/`

- il livello 1 del ChorML è ancora una combinazione di tag XML e GML. Esso specifica i risultati degli algoritmi di Data Mining (spaziale). In particolare, gli elementi del linguaggio sono:
  - informazioni di carattere generale contenenti id, il nome del progetto, il nome dell'autore, la data di creazione, il layout, il sistema di riferimento, il nome del database originale, ultimo aggiornamento;
  - la lista dei coremi in cui i dati geografici sono codificati in GML;
  - pre-legenda, che contiene una descrizione testuale di ogni corema;
  - la lista di relazioni topologiche e non topologiche tra coremi.
- Il livello 2 corrisponde al codice importazione/esportazione del Chorem Editor, ed è una combinazione di tag XML e SVG. Gli elementi di questo livello comprendono:
  - informazioni di carattere generale contenenti id, il nome del progetto, nome dell'autore, la data di creazione, il layout, sistema di riferimento, il nome del database originale, l'ultimo aggiornamento;
  - l'elenco dei coremi semplificati risultanti dall'operazione di Chorem Creation;
  - l'elenco dei coremi modificati risultanti dall'operazione di editing del corema;
  - una legenda che descrive il significato di ciascun corema;
  - la lista di relazioni topologiche e non topologiche tra coremi.

  Infine, a partire dal ChorML2, il Chorem Editor può esportare la mappa in formato SVG, esclusivamente per scopi di visualizzazione.

A.4 Un esempio di generazione di Mappa Corematica

Per illustrare il funzionamento e l'efficacia del sistema, in questo paragrafo verrà esemplificata la generazione di una mappa corematica. A partire da un dataset che contiene i dati demografici riguardanti l'Italia, l'esempio descrive l'intero processo che il sistema compie per costruire una mappa corematica che rappresenta la migrazione della popolazione italiana nel 2000.

Capoluogo	**Lat**	**Lon**	**Pop2000**	**CodIstat**	**Shape**
Agrigento	13,5896	37,3219	54619	084001	Point
Alessandria	8,6200	44,9134	85438	006003	Point
Ancona	13,5110	45,5991	100507	042002	Point
Aosta	7,3177	45,7383	34062	007003	Point
Arezzo	11,8700	43,4733	91589	051002	Point
Ascoli Piceno	13,6313	42,8481	51375	044007	Point
Asti	8,2079	44,8989	71276	005005	Point
Avellino	14,7896	40,9154	52073	064008	Point
Bari	16,8520	41,1187	316532	072006	Point
Belluno	12,2171	46,1503	35050	025006	Point
...	...	...	...	...	...

Table A.1: Dati demografici correlati alle città italiane.

Shape	**Regione**	**Piemonte**	**VdA**	**Lombardia**	**Veneto**	**...**
Polygon	Abruzzo	235	4	792	337	...
Polygon	Basilicata	377	6	769	166	...
Polygon	Calabria	2244	167	5308	1083	...
Polygon	Campania	2648	69	8591	3183	...
Polygon	Emilia-Romagna	852	55	4168	1953	...
Polygon	Lazio	1230	55	3701	1968	...
Polygon	Liguria	3184	77	2274	379	...
Polygon	Lombardia	5595	128	0	3579	...
...	...	...	...	...	...	...

Table A.2: Dati demografici correlati alle regioni italiane.

In una prima fase vengono prodotti dei proto-coremi che contengono i dati di città e regioni, schematizzati in Tabella A.1 e Tabella A.2, rispettivamente. In particolare, la Tabella A.1 contiene dati relativi alle città italiane in termini di nome (Capoluogo), posizione (Lat, Lon), popolazione (Pop2000) e codice identificativo della città (Codistat).

La Tabella A.2 rappresenta una matrice nella quale il valore x, associato alla coppia *(a, b)* corrisponde al numero x dei cittadini che si spostano dalla regione a verso la regione b. Ad esempio, *(Calabria, Lombardia) = 5308* implica che 5308 cittadini sono emigrati dalla regione *Calabria* verso la regione *Lombardia*. La Figura A.15 descrive le componenti spaziali connesse ai dati di entrambe le tabelle, in termini di geometria e posizione.

Applicando tecniche di Data Mining (spaziale) e queries SQL sui dati delle

Figure A.15: Componenti geografiche dei dati su regioni e città italiane.

	Nord	Centro	Sud	Sicilia	Sardegna
Nord	0	18112	**31130**	13150	5248
Centro	20903	0	10594	2084	1428
Sud	**66668**	22132	0	2521	847
Sicilia	**26489**	4172	2483	0	319
Sardegna	6323	1420	588	269	0

Table A.3: Migrazioni tra le 5 macro-regioni

regioni, si ottengono cinque macro-regioni. In particolare, la funzione spaziale SDO_SAM.SPATIAL_CLUSTERS determina cinque gruppi di regioni, raggruppati secondo la proprietà di prossimità. Sulla base di questo risultato, i dati demografici sono quindi ricalcolati con una funzione di aggregazione. Una soglia definita dall'utente è usata dal sistema per selezionare le migrazioni di popolazione più significative. La Tabella A.3 sintetizza il numero delle migrazioni fra le macro-regioni ottenute, evidenziando quelle che superano le 25000 unità. I dati demografici in Tabella A.1 sono invece usati per determinare un insieme delle città importanti in termini di popolazione. Una soglia adeguata definita dall'utente è usata per determinare le sei città più popolate con una query SQL, in questo caso scegliendo le città con più di 500000 abitenti. Effettivamente, il sistema di estrazione fornisce tali risultati in termini di tabelle e dati alfanumerici. Per illustrare i risultati ottenuti durante questa fase, una rappresentazione grafica fittizia è mostrata in Figura A.16, nella quale si possono vedere 5 macro-regioni, ottenute dall'aggregazione di venti regioni, e le sei città più popolose, ottenute filtrando 96 città.

Figure A.16: Regioni raggruppate con la funzione SDO_SAM.SPATIAL_CLUSTERS di Oracle Spatial, e città filtrate con una query SQL.

Basandosi sulle precedenti definizioni, tre coremi sono identificati:

- un corema di macro-regioni (Macro-region Chorem), che contiene cinque elementi di corema che corrispondono a cinque gruppi di regioni.
- un corema di migrazione della popolazione (Migrating population Chorem), che contiene tre elementi di corema che corrispondono ai tre flussi più importanti.
- un corema delle città più popolate (Most populated cities Chorem), che contiene sei elementi di corema che corrispondono alle sei città più popolate.

Per concludere, il sottosistema di estrazione calcola le relazioni spaziali fra gli elementi del corema, utili durante la fase di visualizzazione. Per esempio, alcune macro-regioni e alcune città sono in relazione tra loro attraverso la relazione spaziale *contains*, che esprime contenimento. I coremi ed e le relazioni ottenuti dal sottosistema di estrazione sono quindi trasmessi al sottosistema di visualizzazione, per generare una rappresentazione visiva. Tali informazioni sono espresse in termini di *ChorML1*. In Figura A.17 è mostrato un frammento di codice della mappa corematica che si sta generando, nel quale la parte evidenziata descrive la componente spaziale dell'elemento di corema della città di Roma, espresso in GML.

Come discusso nel paragrafo precedente, due differenti fasi sono eseguite

```
<?xml version="1.0" encoding="utf-8"?>
<chorml>
  <ChorematicMap>
    <MapId>map1</MapId>
    <MapName>Italy</MapName>
    <GeneralInformation>
      <Date/>
      <Author/>
      <Format/>
    </GeneralInformation>
    <ChoremList>
      <Chorem>
        <ChoremElementList>
          <ChoremElement>
            <ElementId>elem1</ElementId>
            <ElementName>Roma</ElementName>
            <ElementSize>4</ElementSize>
            <_SHAPE_>
              <gml:MultiPoint srsName="epsg:8307">
                <gml:pointMember>
                  <gml:Point>
                    <gml:coordinates>12.52,
                                    42.88
                    </gml:coordinates>
                   </gml:Point>
                </gml:pointMember>
              </gml:MultiPoint>
            </_SHAPE_>
          </ChoremElement>
.   .   .
.   .   .
.   .   .
```

Figure A.17: Codice *ChorML1*, ottenuto dal sistema di estrazione dei coremi.

	Milano	Torino	Genova	Roma	Napoli	Palermo
Nord	Contains	Contains	Touch			
Centro				Contains		
Sud					Touch	
Sicilia						Touch
Sardegna						

Table A.4: Relazioni topologiche tra le marco-regioni e le città.

dal sistema di visualizzazione, vale a dire la fase di creazione dei coremi, detta Chorem Creation, e la fase di editing dei coremi, detta Chorem Editing. Durante la fase di Chorem Creation, viene creato il sistema multi-agente che contiene gli agenti che rappresentano i coremi geografici. In particolare, la Figura A.18 illustra la struttura ad agenti composta da un Chorematic Map Agent, tre Chorem Agents e quattordici Chorem Element Agents.

Inoltre, il sistema genera anche i relationship agents, e in particolare, un agente per ogni relazione presente in Tabella A.4, nella quale sono elencate le relazioni topologiche tra le cinque macro-regioni e le sei città.

Una volta che la struttura del sistema multi-agente è completa, il

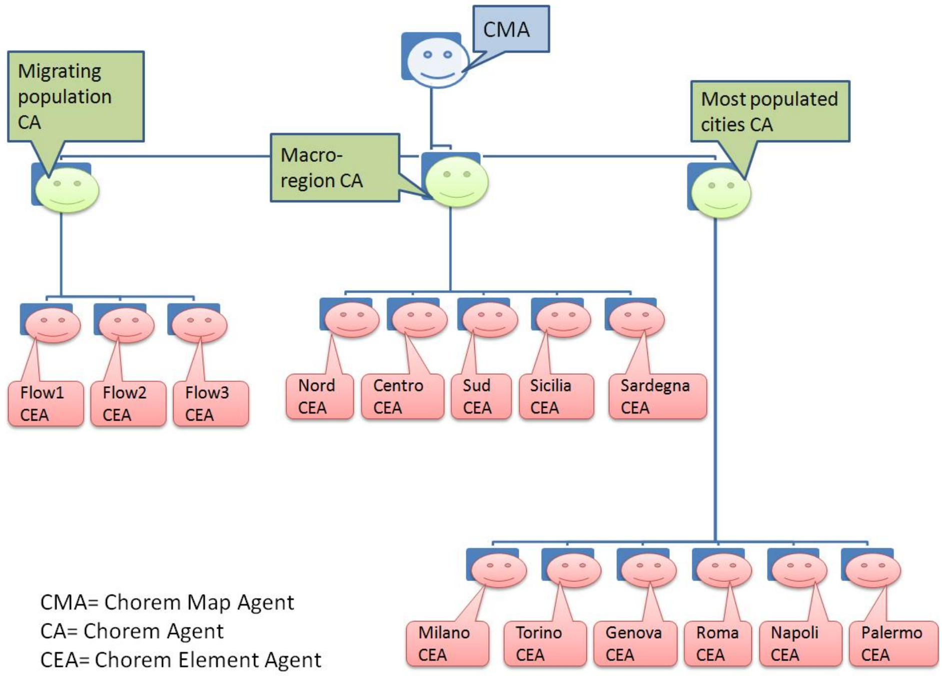

Figure A.18: La struttura del sistema ad agenti creato in questo esempio.

Chorematic Map Agent coordina il calcolo di ulteriori relazioni topologiche utili alla fase di semplificazione multi-layer. Ad esempio, in questa fase il sistema analizza le macro-regioni per verificare se esse condividono qualche confine. Questa situazione, se non gestita, può far nascere errori topologici in fase di semplificazione, come mostrato in Figura A.19.

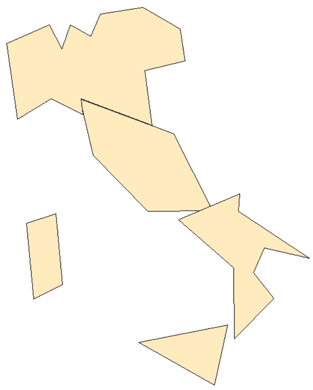

Figure A.19: Macro-regioni semplificate non considerando i confini in comune.

Ultimata questa fase, sia i coremi delle macro-regioni sia quelli delle città vengono sottoposti alla semplificazione e alla coremizzazione. La Figura A.20 mostra i coremi derivati della fase di semplificazione.

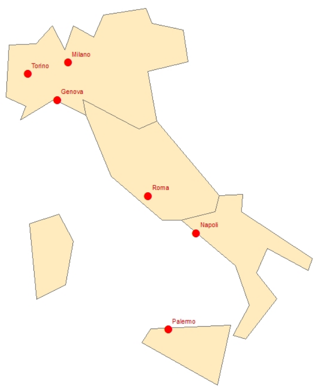

Figure A.20: Coremi ottenuti dalla fase di semplificazione.

Una volta semplificate, le geometrie gestite dai Chorem Element Agents possono subire una fase di coremizzazione. Questa fase è basata su un algoritmo che stressa ulteriormente la semplificazione per ottenere delle forme molto semplici, simili a poligoni convessi con pochi lati. La Figura A.21 mostra il risultato della fase di coremizzazione.

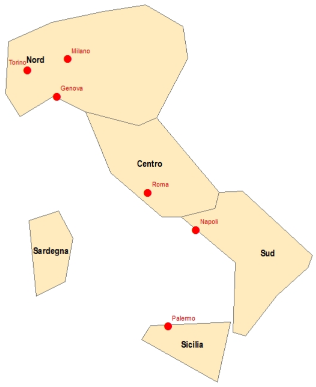

Figure A.21: Coremi ottenuti dalla fase di coremizzazione.

Nella successiva fase di *Coordinate Translation*, le coordinate delle componenti geometriche dei coremi vengono convertite in un sistema di riferimento compatibile con il layout di visualizzazione scelto dall'utente. In particolare, le parti di codice in GML nel *ChorML1* vengono sostituite da semplici tag contenenti le coordinate nel nuovo sistema di riferimento. La Figura A.22 mostra una tale modifica applicata al codice in Figura A.17.

Per ultimare la costruzione della mappa corematica, il sistema multi agente deve gestire i coremi che rappresentano i flussi e le annotazioni, cercando il miglior posizionamento possibile che preserva le relazioni topologiche tra i coremi. Una volta completata la mappa, il sistema trasforma il *ChorML1* in *ChorML2*, convertendo le componenti spaziali dei coremi da GML in SVG. La Figura A.23 illustra il codice ChorML2, che può essere importato e modificato nel Chorem Editor.

Grazie all'uso del *Chorem Editor*, l'utente può modificare la mappa corematica, lasciando all'editor la gestione delle relazioni topologiche. Infatti, il Chorem Editor avvisa l'utente quando una relazione topologica viene violata da una modifica della mappa, permettendogli di scegliere quale comportamento seguire. La Figura A.25 mostra una tale situazione, in cui l'utente cerca di spostare il cerchio rosso che rappresenta la città di Genova.

Infine, l'utente può esportare la mappa corematica ottenuta, eventualmente modificata, in un formato di visualizzazione utilizzabile

```
<?xml version="1.0" encoding="utf-8"?>
<chorml>
  <ChorematicMap>
    <MapId>map1</MapId>
    <MapName>Italy</MapName>
    <GeneralInformation>
      <Date/>
      <Author/>
      <Format>
        <Height>600</Height>
        <Width>800</Width>
      </Format>
    </GeneralInformation>
    <ChoremList>
      <Chorem>
        <ChoremElementList>
          <ChoremElement>
            <ElementId>elem1</ElementId>
            <ElementName>Rome</ElementName>
            <ElementSize>4</ElementSize>
            <ElementShape>x=23,y=56</ElementShape>
          </ChoremElement>
.   .   .
.   .   .
.   .   .
```

Figure A.22: Codice *ChorML1*, risultante dalla fase conversione del sistema di riferimento.

```
<?xml version="1.0" encoding="utf-8"?>
<chorml>
  <ChorematicMap>
    <MapId>map1</MapId>
    <MapName>Italy</MapName>
    <GeneralInformation>
      <Date/>
      <Author/>
      <Format>
        <Height>600</Height>
        <Width>800</Width>
      </Format>
    </GeneralInformation>
    <ChoremList>
      <Chorem>
        <ChoremElementList>
          <ChoremElement>
            <ElementId>elem1</ElementId>
            <ElementName>Rome</ElementName>
            <ElementSize>4</ElementSize>
            <svg>
                <ellipse style="fill: #ff0000" cx="10.7406"
                      cy="4.15625" rx="0.55625" ry="0.55625"/>
            <ellipse style="fill: none; fill-opacity:0;
                     stroke-width: 0.1; stroke: #ff0000" cx="10.7406"
                     cy="4.15625" rx="0.55625" ry="0.55625"/>
            <ellipse style="fill: none; fill-opacity:0;
                     stroke-width: 0.01; stroke: #000000"
                     cx="10.7406" cy="4.15625" rx="0.55625"
                     ry="0.55625"/>
            <text style="fill: #000000; text-anchor:start;
                  font-size: 0.7; font-family: sans; font-style: normal;
                  font-weight: normal" x="8.5" y="4.375">Paris</text>
           </svg>
          </ChoremElement>
.   .   .
.   .   .
.   .   .
```

Figure A.23: Il codice *ChorML2*.

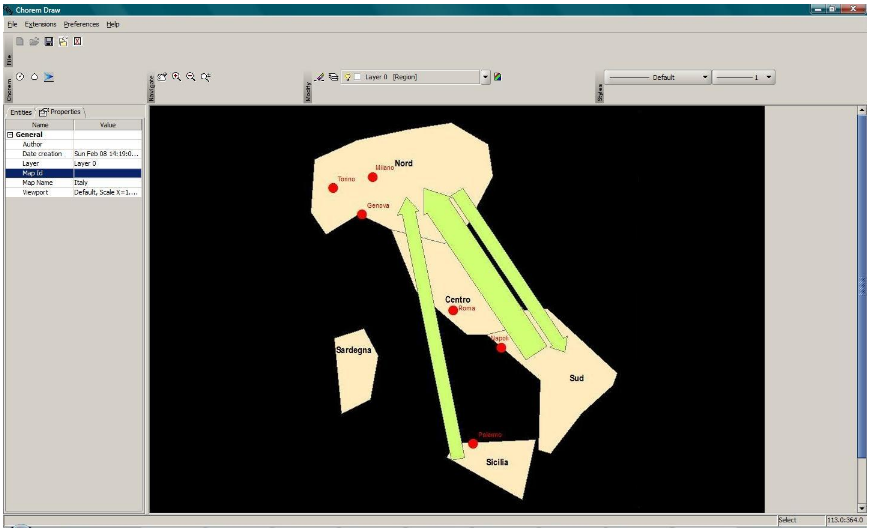

Figure A.24: La mappa corematica completa.

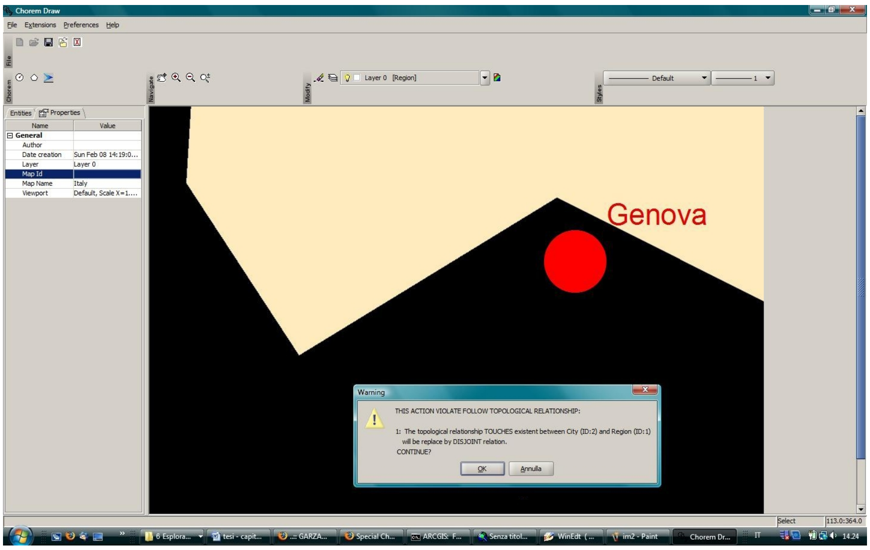

Figure A.25: Una violazione di una relazione topologica riportata dal Chorem Editor.

anche in altri sistemi. La Figura A.26 mostra la mappa corematica visualizzata in un comune visualizzatore di SVG. Nella fase di esportazione il sistema aggiunge una legenda al layout di visualizzazione, per supportare l'utente nella comprensione del significato dei coremi.

Figure A.26: La mappa corematica che mostra i maggiori flussi di migrazione in Italia nel 2000.

A.5 Conclusione

La cartografia tradizionale è un mezzo essenziale per descrivere i fatti ed le relazioni che interessano un territorio. Essa associa a concetti geografici dei simboli grafici che aiutano i lettori ad avere una comprensione immediata dei dati rappresentati. Questo è un metodo affermato, e gli utenti esperti prendono decisioni dall'analisi di mappe, sono solitamente soddisfatti dalla loro espressività quando ci si occupa della cartografia dei fatti. In questa Tesi stata studiata una soluzione da adottare quando i decisori devono invece affrontare questioni legate a complessi problemi di visualizzazione. In questo casi, infatti, gli utenti esperti hanno a disposizione un gran numero di dati che descrivono sia elementi geografici che fenomeni spazio-temporali. Inoltre, le correlazioni fra i dati complessi non sono immediatamente chiare, quindi è richiesto un grande sforzo per individuare e quindi rappresentare le dinamiche, i movimenti ed i cambiamenti che sono alla base di possibili problemi da analizzare. Per determinare una soluzione cartografica innovativa, mirata ad oltrepassare i limiti riconosciuti della cartografia tradizionale, è stato concepito un progetto di ricerca al quale partecipano esperti dai domini differenti. Uno studio iniziale è stato effettuato per analizzare le soluzioni differenti. Fra loro, i coremi sono sembrati essere il mezzo più adatto, grazie alla loro espressività e capacità di sintesi. In questa Tesi, nata nel contesto di questo progetto internazionale, sono stati proposti una definizione e una classificazione dei coremi, destinate a standardizzarne la costruzione e l'uso. Inoltre, è stato progettato e realizzato un sistema per generare in maniera semi-automatica le mappe corematiche. Tecnologie differenti sono state adottato, dalle query SQL al Data Mining (spaziale), dalla semplificazione cartografica alla risoluzione di conflitti tra elementi di una mappa. inoltre, un linguaggio XML-like multi-livello è stato introdotto per consentire la memorizzazione dei dati riguardanti i coremi e garantire la comunicazione fra i moduli di sistema.

I lavori futuri possono essere sintetizzati nelle seguenti attività:

- Eseguire degli studi sperimentali comparativi per verificare l'efficacia dell' approccio proposto. In particolare, tali studi mirano a misurare l'abilità degli utenti esperti nell'usare le mappe corematiche prodotte dal sistema.
- Migliorare le tecniche utilizzate nel sistema per ridurre l'intervento richiesto dall'utente.
- Investigare l'integrazione del sistema proposto con un sistema di supporto alle decisioni.
- Investigare il modello topologico di Egenhofer per migliorare l'accuratezza della fase di semplificazione.
- Verificare se i coremi possono essere utilizzati come strumento per navigare il contenuto di DB geografici.
- Verificare se i coremi possono essere utilizzati come linguaggio di interrogazione per DB geografici.
- Effettuare studi sia sull'aspetto cognitivo sia sull'aspetto semiotico dei coremi e migliorare l'interoperabilità dell'approccio proposto.

Appendix B

Resumé en français

B.1 Introduction

La visualisation des données géographiques appartient traditionnellement au domaine de la recherche connu sous le nom de visualisation géographique ou de géovisualisation. En particulier, elle intègre des approches de champs différents comme la cartographie, la visualisation scientifique, l'analyse d'images, la visualisation des informations, et les système d'information géographique (SIG), pour fournir des théories, des méthodes et des outils pour l'exploration visuelle, l'analyse, la synthèse et la présentation des données géospatiales. La cartographie traditionnelle est un moyen essentiel pour décrire les faits et les relations qui concernent un territoire. Elle associe à des concepts géographiques des symboles graphiques qui aident les lecteurs à comprendre immédiatement les données représentées. Cette méthode est connue depuis longtemps et les utilisateurs experts qui doivent prendre des décisions à partir de l'analyse des cartes, sont habituellement satisfaits de son pouvoir d'expression lorsqu'il s'agit de la cartographie qui concerne les faits. Une approche plus complexe est souhaitée quand il faut affronter des domaines complexes, comme par exemple la politique, l'économie et les problèmes démographiques dans lesquels les utilisateurs experts disposent d'un grand nombre de données qui décrivent les objets géographiques et les phénomènes spatio-temporels. Dans ces cas, l'objectif principal n'est tant pas de restituer exactement avec un haut niveau de détail, mais plutôt de montrer les aspects les plus saillants. En d'autres termes, une aide significative aux activités de modélisation et d'analyse peut consister en une synthèse immédiate des données d'intérêt, qui écarte les détails les moins significatifs.

Un support technologique à la réalisation peut être obtenu en complétant les résultats de différents domaines de recherche comme les bases de données géographiques, la découverte de connaissances dans les bases des données, et la fouille de données spatiales. En particulier, chaque discipline peut contribuer à la définition d'une méthodologie de géovisualisation pour obtenir des solutions cartographiques avancées, en fournissant des techniques pour découvrir des informations connues et potentiellement significatives.

Pour déterminer une solution cartographique innovante, en ayant l'intention d'aller au delà des limites reconnues de la cartographie traditionnelle, un projet de recherche est né entre trois instituts de recherche: (i) Institut National des Sciences Appliquées (INSA) de Lyon (France), (ii) Universitá degli Studi di Salerno (Italie) et (iii) Tecnólogico de Monterrey, Campus de Puebla (Mexique). Un tel projet vise à définir des solutions cartographiques capables de représenter convenablement les informations extraites d'une base de données géographiques, c'est-à dire des informations qui se réfèrent aux objets géographiques et aux phénomènes spatio-temporels.

L'objectif de cette thèse, née de ce projet international, est de fournir des nouvelles modalités de représentation visuelle à l'aide d'un outil capable de rendre une synthèse immédiate des données d'intérêt tout en éliminant les détails les moins significatifs. Une telle synthèse peut représenter le point de départ pour des activités complémentaires, pour obtenir l'analyse de données territoriales, et pour aider les utilisateurs experts dans l'évaluation des décisions, afin de combler la différences que l'on trouve dans la complexité des applications adoptées et la nécessité d'une réponse rapide et exhaustive.

La solution proposé dans cette direction, se fonde sur le concept de *chorèmes* et sur leur capacité à synthétiser des scènes qui comprennent des objets géographiques et des phénomènes spatio-temporels, en associant à ceux-ci des notations visuelles schématisées.

Le résumé en français de la thèse est organisé comme suit. Le paragraphe B.2 introduit la définition et la classification des chorèmes. Le paragraphe B.3 décrit l'architecture du système proposé dans la thèse et le langage ChorML. En particulier, il détaille les algorithmes et les techniques sur lesquels se basent les deux sous-systèmes dont il est composé, à savoir le système d'extraction des chorèmes, et le système de visualisation des chorèmes. Un exemple de construction d'une carte chorématique est illustré dans le paragraphe B.4. Les conclusions et les idées pour les futurs travaux sont décrites dans le paragraphe B.5.

B.2 Chorèmes

Le terme *chorème* provient du mot grec $\chi\omega\rho\eta\mu\alpha$, qui signifie espace ou territoire. Selon la définition du géographe français Roger Brunet [13], les chorèmes sont une représentation schématique d'un territoire qui élimine chaque détail non nécessaire pour la compréhension d'une carte. Brunet a proposé 28 chorèmes élémentaires, chacun représentant une configuration spatiale simple. En combinant les chorèmes élémentaires dans une carte, plusieurs phénomènes spatio-temporels peuvent être représentés à des échelles différentes.

La Figure B.3 montre une comparaison entre une carte traditionnelle de la France et une carte de France contenant des chorèmes, dans lequel des aspects différents ont mis en évidence:

- la forme simplifiée des limites,
- les villes les plus importantes,
- les zones avec niveau différent de développement,
- les flux qui représentent les routes nationales principales.

Figure B.1: Carte traditionnelle de France.

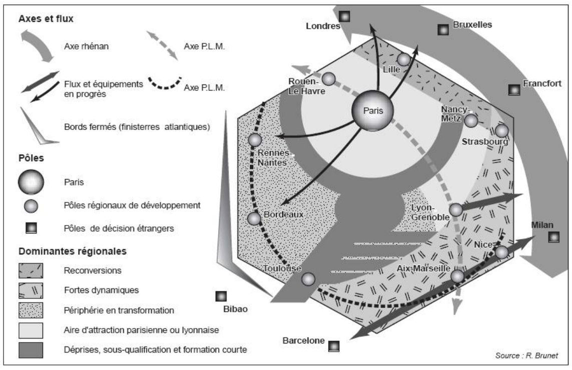

Figure B.2: Carte de France avec des chorèmes. (Source [72])

Figure B.3: Comparaison entre une carte traditionnelle de la France et une carte de France contenant des chorèmes

Depuis leur apparition, les chorèmes ont eu une grande diffusion dans la communauté des géographes, où ils ont été largement développés et été utilisés [13, 14, 15, 23, 9]. En partant de la définition de Brunet,différentes définitions et classifications des chorèmes ont été proposées, et une grande quantité de cartes ont été produites manuellement, pour donner une vision globale d'une réalité d'intérêt comme pour obtenir une analyse d'un problème spécifique. Par exemple, la Figure B.4 montre une carte contenant des chorèmes de l'Espagne, ou différentes dynamiques du développement du territoire sont mises en évidence.

Toutefois, l'absence d'une approche rigoureuse pour la création et la construction des chorèmes a provoqué la prolifération de solutions *ad hoc* qui, dans beaucoup de cas, ne réussissent pas à transmettre l'information qu'ils devraient transmettre. Les premiers résultats de la recherche peuvent être trouvés dans [29], article dans lequel une version provisoire de l'architecture de notre système est définie. Sur la base de tels résultats provisoires et en partant de la définition des chorèmes de Brunet, cette thèse présente une définition et une classification des chorèmes quant à structure et leur signification, et elle propose un système dont l'objectif est

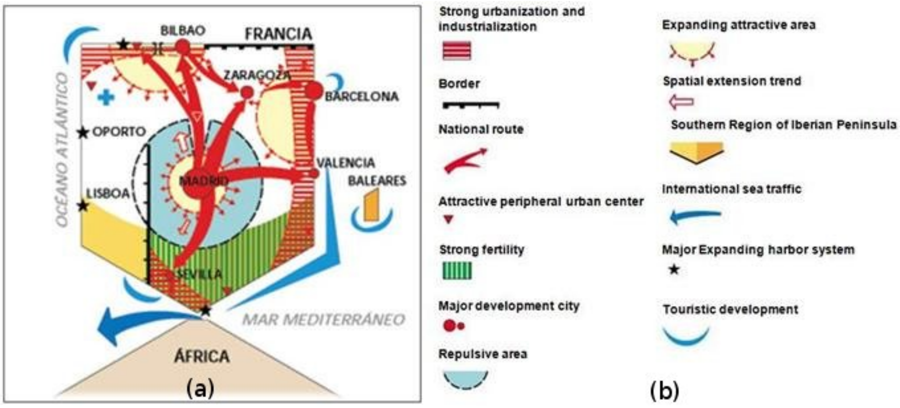

Figure B.4: Une carte contentant des chorèmes d'Espagne.

de permettre la création semi-automatique de cartes contenant des chorèmes.

B.2.1 Définitions de base

Ce paragraphe propose les définitions de base sur les chorèmes. De telles définitions visent à formaliser les termes que seront utilisés dans la thèse comme pour résoudre quelques ambiguïtés des termes qui caractérisent le domaine des chorèmes.

Définition 1: un *Pattern Géographique* ou *Pattern* est une régularité intéressante d'un certain phénomène découvert dans une base de données géographique en utilisant des fonctions de fouille de données spatiales. Les patterns peuvent être utilisés comme point de départ pour la localisation des phénomènes spatio-temporels et des relations entre eux. Un exemple de pattern géographique: *"au Mexique, dans les régions dans lesquelles l'agave est cultivé, la tequila est produite"*.

Définition 2: Les *Proto-chorèmes* peuvent être définis comme données, nettoyées et organisées, afin de pouvoir lancer des requêtes SQL, des fonctions de fouille de données spatiales ou d'autres méthodes d'analyse.

Définition 3: Un *Élément de Chorème* correspond à chaque élément graphique de base qui représente un unique object géographique ou un unique phénomène. Par exemple, un point qui représente une ville, ou une flèche qui représente un flux, comme montré à la Figure B.5.

Définition 4: Un *Chorème* est un ensemble de *éléments de chorème* de

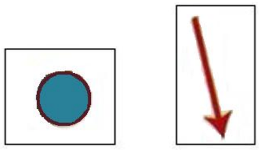

Figure B.5: Un point et une flèche que représentent une ville et un flux.

même type. Par exemple, les villes les plus importantes, ou les flux principaux entre les villes, comme montré à la Figure B.6.

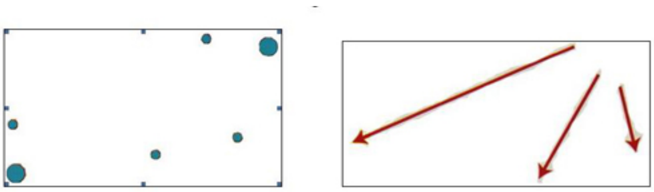

Figure B.6: Villes et flux.

Définition 5: Une *Carte Chorématique* est un ensemble de chorèmes qui schématisent des données d'intérêt liées aux dynamiques d'un territoire. Une légende est généralement associé à la carte chorématique, afin de éclairer la signification de chaque chorème dans la carte, comme montré à la Figure B.7.

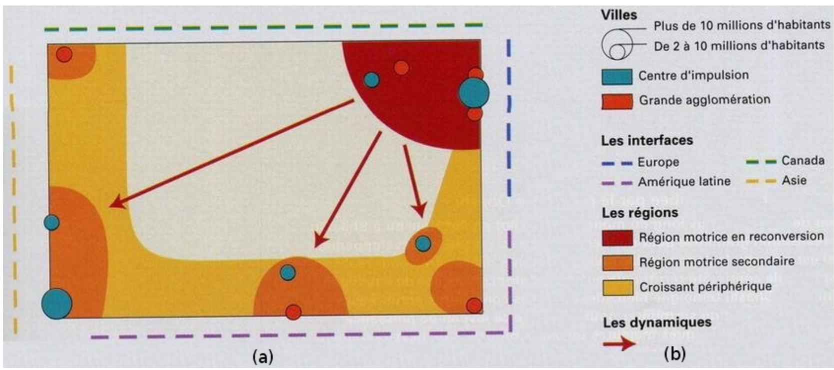

Figure B.7: Carte chorématique des Etats-Unis d'Amérique.

B.2.2 Classification des chorèmes

Les chorèmes peuvent être classifiés selon trois catégories principales, *chorèmes géographiques*, *chorèmes phénoménologiques* et *chorèmes d'annotation.*

Les chorèmes géographiques représentent des éléments avec des caractéristiques territoriales simples comme des points, lignes et polygones et les objets constitués par leurs combinaisons, par exemple les réseaux. Les chorèmes phénoménologiques décrivent des phénomènes spatio-temporels et concernent un ou plusieurs chorèmes géographiques. Un ensemble initial de chorèmes phénoménologiques se compose de trois types, respectivement *flux*, *tropisme* et *diffusion spatiale.* Le chorème de type flux représente le déplacement de populations ou d'objets entre les points et polygones, ou entre des points d'intérêt et zones. Le chorème de type tropisme représente un espace attractif ou répulsif, autour d'un chorème géographique. Finalement, le chorème de type diffusion spatiale représente une progression ou régression spatiale, qui part d'un chorème géographique vers une direction spécifique. Les chorèmes d'annotation représentent des étiquettes ou annotations d'une carte.

B.3 ChEViS: un système pour résumer visuellement le contenu de BD géographiques en utilisant les chorèmes

La Figure B.8 montre l'architecture du système ChEViS (Chorem Extraction and Visualization System), qui est composé de deux sous-systèmes, le système d'extraction des chorèmes, appelé *Chorem Extraction System*, et le système de visualisation des chorèmes, appelé *Chorem Visualization System.* Le *Chorem Extraction System* génère les données disponibles des BD géographiques afin d'en extraire des informations significatives. Le *Chorem Visualization System* gère ces informations en leur assignant une représentation visuelle en termes de chorèmes et cartes chorèmatiques.

La communication entre les composantes du système et la mémorisation de la structure des chorèmes se fonde sur un langage de type XML à plusieurs niveaux, appelé *ChorML.*

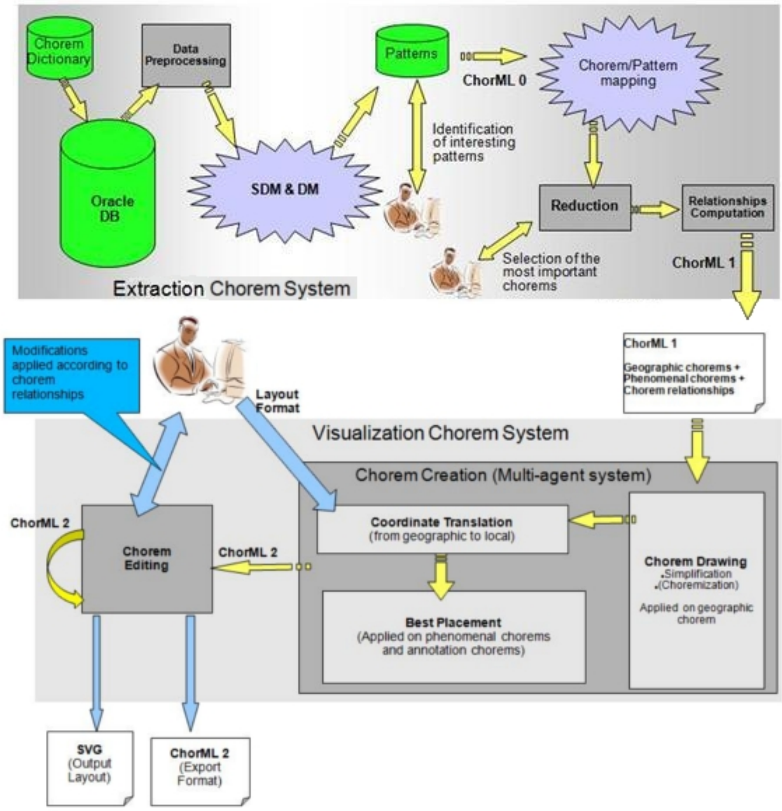

Figure B.8: Architecture du système proposé.

B.3.1 Chorem Extraction System

Ce paragraphe présente brièvement le *Chorem Extraction System*, montré à la Figure B.9, dont l'objectif est l'extraction des objets géographiques et des phénomènes spatio-temporels issus de BD géographiques. La structure initiale introduite dans ce paragraphe dérive du projet international discuté dans [82].

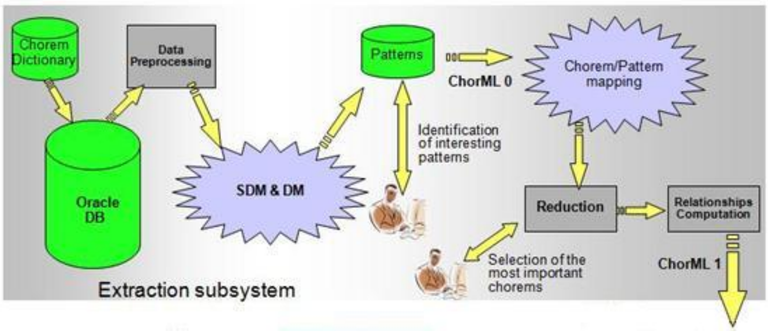

Figure B.9: Architecture du système d'extraction des chorèmes.

Dans une première phase, un dictionnaire des chorèmes est importé dans la base de données pour être employé dans les phases suivantes du processus d'extraction des chorèmes. Par la suite, le système exécute un processus d'association entre les patterns et les chorèmes, et un processus de réduction, afin d'obtenir un ensemble restreint de chorèmes. En effet, du processus d'extraction, une grande quantité de patterns peuvent être obtenus, en risquant ainsi de surcharger d'un nombre excessif de chorèmes la carte chorémematique en cours de construction. Le deux phases permettent la participation des usilisateurs, afin de sélectionner les patterns et les chorèmes les plus significatifs. Finalement, le système exécute le calcul des relations spatiales existantes parmi les éléments de chorème. Dans ce but, des fonctions d'Oracle Spatial[1] sont utilisées par le système et les résultats sont codés en ChorML, lesquels résultats sont ensuite envoyés au système de visualisation.

B.3.2 Chorem Visualization System

Ce paragraphe introduit une partie significative du travail de thèse, le projet et la mise en œuvre du système de visualisation des chorèmes, ou

[1] Oracle Spatial webpage [online], `http://www.oracle.com/technology/products/spatial/index.html`

Chorem Visualization System. La fonction de ce système est de transformer une liste de chorèmes et relations entre chorèmes, en venant du système d'extraction en termes de texte, dans une représentation visuelle, ou bien dans une carte chorématique. Les résultats obtenus dans [82] et [30] ont été raffinés et améliorés, en considérant l'évolution des définitions et de la classification des chorèmes, comme l'application de techniques de simplification cartographique dont l'objectif est de garantir la simplification multicouches de la carte. Comme montré à la Figure B.10, deux phases différentes sont développées par ce système, la création des chorèmes et la modification des chorèmes.

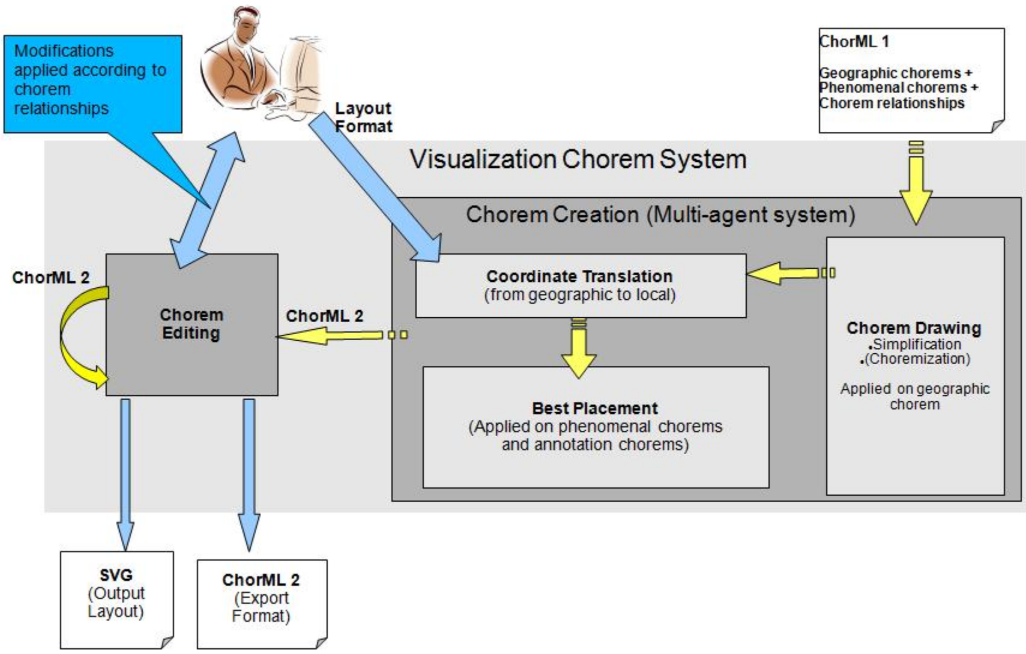

Figure B.10: Architecture du système de visualisation des chorèmes.

Création des chorèmes

La phase de création des chorèmes est développée à travers un système multi-agents qui exécute trois phases, Chorem Drawing, Coordinate Translation et Best Placement. L'architecture du système multi-agents projetée afin d'exécuter cette phase est montré à la Figure B.11.

La phase de création des chorèmes est appliquée aux chorèmes géographiques qui formellement contiennent des composantes

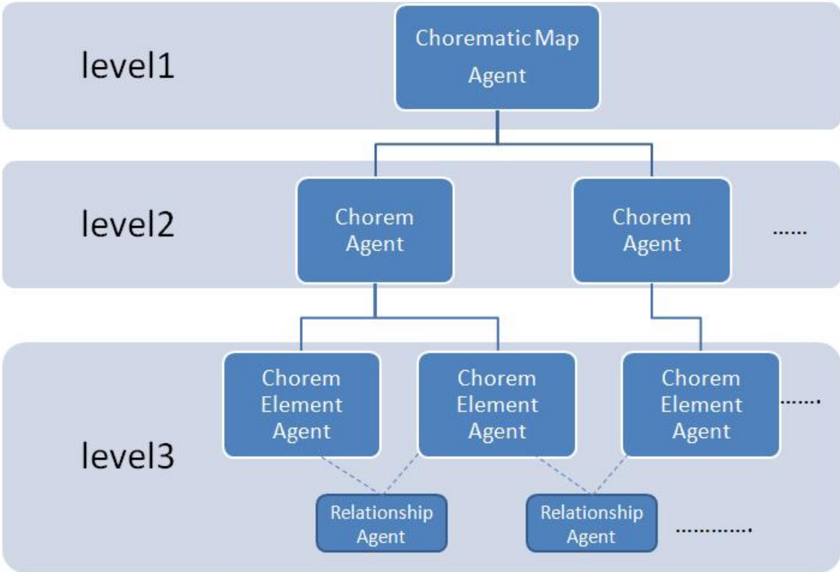

Figure B.11: Architecture du système multi-agents.

géométriques. Elle est effectué à travers deux operations, la simplification e la *chorémisation*, montrées à la Figure B.12 et la Figure B.13.

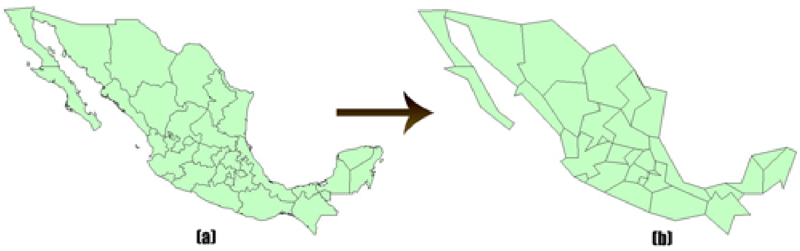

Figure B.12: Exemple de l'operation de simplification.

En particulier, la simplification détermine une version simplifiée des données géométriques, en réduisant le nombre de sommets de la forme originale. Cette technique est basée sur des fonctions spatiales, qui incorporent l'algorithme de *Ramer-Douglas-Peucker* (RDP) [107, 39] et ses variantes [63, 146], et sur méthodes qui considèrent que la simplification est effectuée dans un contexte [12]; de plus elle ne doit pas violer les relations topologiques qui interviennent parmi les éléments du même chorème ou parmi les éléments de chorèmes différents. L'opération de

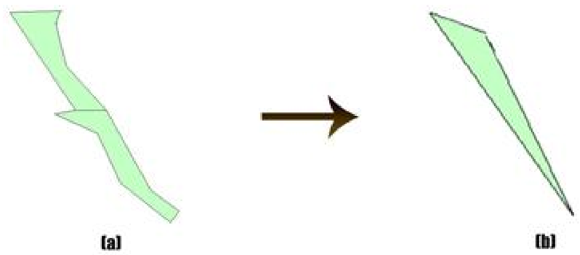

Figure B.13: Exemple de l'operation de *chorémisation*.

chorémisation conditionne ultérieurement la simplification, en associant à la composante spatiale du chorème déjà simplifié, la forme d'un polygone régulier avec un faible nombre de côtés. La phase de *Chrorem Drawing* garantit le maintien des relations topologiques parmi les éléments de la carte, pendant la simplification et pendant la chorémisation.

La phase suivante est le Coordinate Translation, qui essentiellement convertit les coordonnées appartenant à la composante spatiale des chorèmes, d'un système de référence géographique vers un système de référence locale qui se réfère au format de visualisation choisi dans cette phase par l'utilisateur.

Finalement, la phase de Best Placement, dont l'objectif est d'intégrer les chorèmes phénoménologiques et les chorèmes d'annotations dans la carte chorématique, en leur assignant la position la plus correcte. Le système multi-agent complète cette opération, en utilisant essentiellement le bien connu opérateur de déplacement de la généralisation cartographique, et en considérant que les éléments des chorèmes présents dans la carte chorématique possédent des priorités différentes. Ou bien, si un conflit naît parmi deux éléments dans la carte, celui avec plus petite priorité est forcé à se déplacer. Pour satisfaire la condition de lisibilité de la carte, des ensembles indépendants des chorèmes peuvent être unis dans des cartes chorématiques différentes. Par exemple, d'une carte chorématique contenant dix chorèmes pourrait engendrer deux cartes contenants cinq chorèmes chacune.

Chorem Editing

Dans la phase de création des chorèmes, pourraient advenir des difficultés dans le positionnement des chorèmes, car les utilisateurs peuvent demander des raffinements supplémentaires qui concernent les propriétés sémantiques et graphiques des chorèmes. L'objectif de la phase de modifications des chorèmes est de fournir aux utilisateurs la possibilité de modifier la carte chorématique, en mettant à jour, en bougeant ou éliminant des éléments de la carte. Le Chorem Editor a été mis en oeuvre, en étendant, un éditeur de graphiques vectoriels 2D code open source, basé sur le langage de programmation Java. En particulier, Le Chorem Editor peut dérouler les activités suivantes:

- importer une représentation ChorML des cartes chorématiques,
- visualiser les cartes chorématiques,
- modifier la représentation graphique et sémantique des chorèmes
- exporter une représentation graphique (SVG)[2] et une représentation ChorML des cartes chorématiques.

B.3.3 Le langage ChorML

Un langage de type XML multi-niveaux, appellé *ChorML* a été défini, avec la fonctionnalité de mémoriser l'information des chorèmes et de permettre la communication de telles informations entre les différents modules du système. Ses niveaux sont décrits comme suit:

- le niveau 0 est composé de XML[3] et GML[4]. Il a été pensé pour stocker les informations sur les proto-chorèmes, l'origine des données (lineage) et les fonctions appliquées afin d'obtenir les chorèmes.
- le niveau 1 de *ChorML* est également une combinaison de XML et de GML. Il spécifie les résultats des algorithmes de fouille de données spatiales. Les éléments du langage sont en particulier:
  - les informations de caractère général, contenant l'identificateur de la carte, le nom de la carte, le nom de l'auteur, la date de création, le système de référence, le nom de la base de données originale, la dernière date de mise à jour, etc.;
  - la liste des chorèmes dans lequel les données géographiques sont codées en GML;
  - une pré-légende, qui contient une description en format texte de chaque chorème;

[2]Scalable Vector Graphics [online], website, `http://www.w3.org/Graphics/SVG/`
[3]eXtensible Markup Language [online], website, `http://www.w3.org/XML/`
[4]Geography Markup Language [online], website, `http://www.opengis.net/gml/`

  - la liste des relations topologiques et non-topologiques parmi les chorèmes.

- le niveau 2 de *ChorML* est le code d'importation et d'exportation du Chorem Editor, une combinaison de XML et de SVG. Les éléments du langage sont en particulier:

  - les informations de caractère générale, contenant l'identificateur de la carte, le nom de la carte, le nom de l'auteur, la date de création, le système de référence, le nom de la base de données originale, la dernière date de mise à jour, etc.;
  - la liste des chorèmes simplifié résultant de la phase de Chorem Creation;
  - la liste des chorèmes résultant de l'operation de Chorem Editing;
  - une légende, qui contient une description en format texte et visuelle de chaque chorème;
  - la liste des relations topologiques et non-topologiques parmi les chorèmes.

  Finalement, en utilisant *ChorML2*, le Chorem Editor peut exporter la carte en format SVG dans le seul but de visualisation.

B.4 Un exemple de création d'une carte chorématique

Pour illustrer l'exploitation et l'efficacité du système, la génération d'une carte chorèmatique sera donnée en exemple dans ce paragraphe. En partant d'un lot de données qui contient les données démographiques de l'Italie, l'exemple décrit la totalité du processus que le système accomplit pour construire une carte chorématique qui représente les migrations de la population italienne pendant l'année 2000.

Dans une première phase le système produit des proto-chorèmes qui contiennent les données des villes et des régions italiennes, comme schématisés au Tableau B.1 et au Tableau B.2. En particulier, le Tableau B.1 contient des données sur les villes en termes de nom *(Capoluogo)*, position *(Lat, Lon)*, population *(Pop2000)* et code de la ville *(Codistat)*.

Le Tableau B.2 représente une matrice dans laquelle la valeur x, associée au couple *(a, b)* correspond au nombre x de citoyens qui ont migré de la région a vers la région b. Par exemple, *(Calabre, Lombardie) = 5308* implique que 5308 citoyens ont émigré de la *Calabre* vers la *Lombardie*. La

Capoluogo	Lat	Lon	Pop2000	CodIstat	Shape
Agrigento	13,5896	37,3219	54619	084001	Point
Alessandria	8,6200	44,9134	85438	006003	Point
Ancona	13,5110	45,5991	100507	042002	Point
Aosta	7,3177	45,7383	34062	007003	Point
Arezzo	11,8700	43,4733	91589	051002	Point
Ascoli Piceno	13,6313	42,8481	51375	044007	Point
Asti	8,2079	44,8989	71276	005005	Point
Avellino	14,7896	40,9154	52073	064008	Point
Bari	16,8520	41,1187	316532	072006	Point
Belluno	12,2171	46,1503	35050	025006	Point
...	...	...	...	...	...

Table B.1: Données démographiques des villes italiennes.

Shape	Regione	Piemonte	VdA	Lombardia	Veneto	...
Polygon	Abruzzo	235	4	792	337	...
Polygon	Basilicata	377	6	769	166	...
Polygon	Calabria	2244	167	5308	1083	...
Polygon	Campania	2648	69	8591	3183	...
Polygon	Emilia-Romagna	852	55	4168	1953	...
Polygon	Lazio	1230	55	3701	1968	...
Polygon	Liguria	3184	77	2274	379	...
Polygon	Lombardia	5595	128	0	3579	...
...	...	...	...	...	...	...

Table B.2: Données démographiques des régions italiennes.

Figure B.14 décrit les composantes spatiales des deux Tableaux, quant à la géométrie et la position.

En appliquant des techniques de fouille de données spatiales et des requêtes SQL sur les données des régions, le système produit cinq macro-régions. Sur la base de ces résultats, le système réunit les données démographiques et les utilisateurs peuvent établir un seuil afin de sélectionner les déplacements les plus significatifs. Le Tableau B.3 synthétise cette opération, en mettant en évidence les migrations qui dépassent les 25.000 unités. Les données démographiques du Tableau B.1 sont utilisées pour déterminer les villes les plus peuplées, en utilisant une requête SQL. L'utilisateur choisit les villes avec plus de 500.000 habitants. Pour mieux illustrer, la Figure B.15 montre les résultats de cette

Figure B.14: Composantes spatiales des données démographiques.

	Nord	Centro	Sud	Sicilia	Sardegna
Nord	0	18112	**31130**	13150	5248
Centro	20903	0	10594	2084	1428
Sud	**66668**	22132	0	2521	847
Sicilia	**26489**	4172	2483	0	319
Sardegna	6323	1420	588	269	0

Table B.3: Migrations entre les cinq macro-régions

opération, même si à ce niveau le système fournit ces résultats en termes de tableaux et données alphanumériques.

En se fondant sur les définitions précédentes, trois chorèmes sont identifiés:

- un chorème de macro-régions,
- un chorème de migration de la population,
- un chorème des villes les plus peuplées.

Pour conclure, le système d'extraction calcule les relations spatiales entre les éléments d'un chorème, utiles pendant la phase de visualisation. Les chorèmes et les relations obtenues par le système d'extraction sont transmis au système de visualisation en termes de *ChorML1*. La Figure B.16 montre un fragment de code de la carte chorématique, où la partie marquée en jaune décrit la composante spatiale du élément de chorème de la ville de Rome, en termes de GML.

Figure B.15: Les cinq macro-régions et les six villes obtenues.

```
<?xml version="1.0" encoding="utf-8"?>
<chorml>
  <ChorematicMap>
    <MapId>map1</MapId>
    <MapName>Italy</MapName>
    <GeneralInformation>
      <Date/>
      <Author/>
      <Format/>
    </GeneralInformation>
    <ChoremList>
      <Chorem>
        <ChoremElementList>
          <ChoremElement>
            <ElementId>elem1</ElementId>
            <ElementName>Roma</ElementName>
            <ElementSize>4</ElementSize>
            <_SHAPE_>
              <gml:MultiPoint srsName="epsg:8307">
                <gml:pointMember>
                  <gml:Point>
                    <gml:coordinates>12.52,
                                    42.88
                    </gml:coordinates>
                   </gml:Point>
                </gml:pointMember>
              </gml:MultiPoint>
            </_SHAPE_>
          </ChoremElement>
.   .   .
.   .   .
.   .   .
```

Figure B.16: Code *ChorML1*, obtenu par le système d'extraction.

Comme mentionné dans les paragraphes précédents, deux phases différentes sont exécutées par le système de visualisation, la phase de création des chorèmes et la phase de rédaction des chorèmes. Pendant la phase de création des chorèmes, un système multi-agent est créé. En particulier, la Figure B.17 montre la structure de ce système multi-agent créé dans cet exemple, composé d'un Chorematic Map Agent, trois Chorem Agents et quatorze Chorem Element Agents.

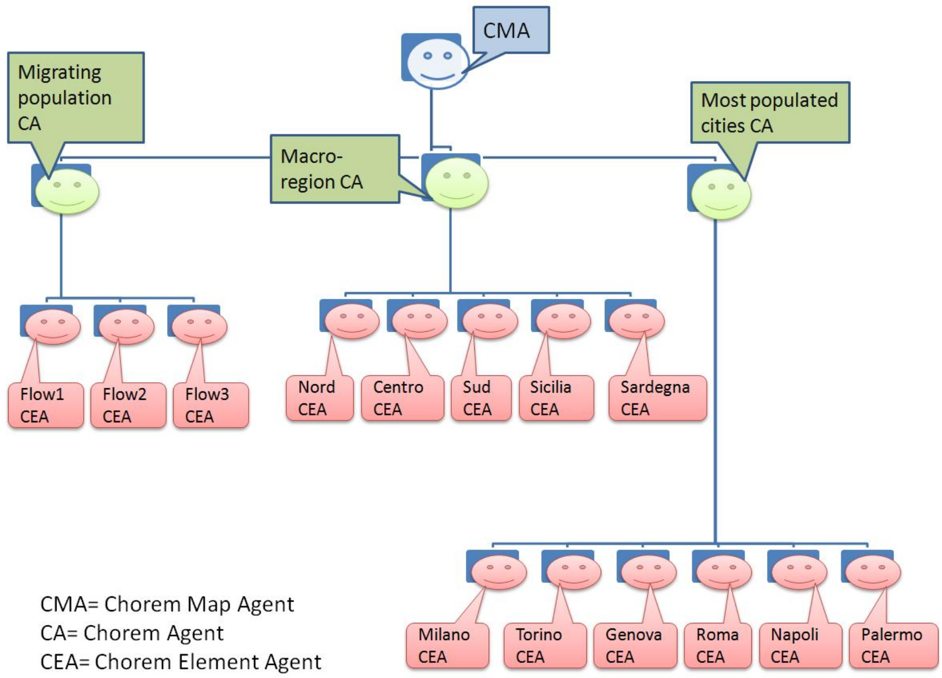

Figure B.17: La structure du système multi-agent.

En outre, le système multi-agent produit même les Relationship Agents, et en particulier, un agent pour chaque relation présente dans le Tableau B.4.

Une fois la structure du système multi-agent complétée, le Chorematic Map Agent coordonne le calcul de relations topologiques supplémentaires, utiles à la phase de la simplification multicouche. Par exemple, dans cette phase le système analyse les macro-régions pour vérifier s'ils partagent quelque limite. Cette situation, mal gérée, peut conduire à des erreurs topologiques en phase de simplification, comme montré à la Figure B.18.

Ensuite, les chorèmes des macro-régions comme les chorèmes des villes sont soumis aux phases de simplification et de *chorémisation*. La Figure B.19 montre les chorèmes dérivant de la phase de simplification.

	Milano	Torino	Genova	Roma	Napoli	Palermo
Nord	Contains	Contains	Touch			
Centro				Contains		
Sud					Touch	
Sicilia						Touch
Sardegna						

Table B.4: Relations topologiques parmi les marco-régions et les villes.

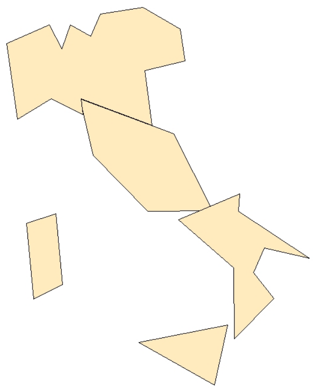

Figure B.18: Macro-régions simplifiées sans considérer les limites en commun.

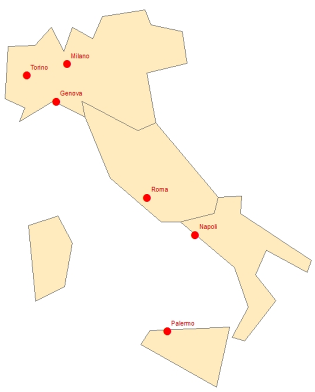

Figure B.19: Chorèmes dérivant de la phase de simplification.

La Figure B.20 montre les chorèmes dérivant de la phase de *chorémisation*.

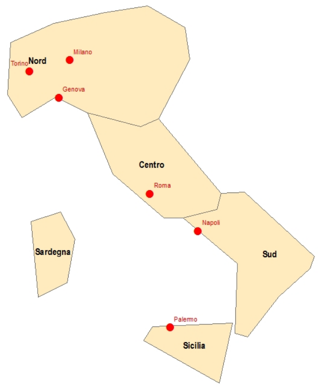

Figure B.20: Chorèmes dérivant de la phase de *chorémisation*.

Pour compléter la construction de la carte chorématique, le système multi-agent doit gérer les chorèmes qui représentent les migrations et les étiquettes, en cherchant le meilleur positionnement possible qui conserve les relations topologiques entre les chorèmes. Une fois la carte complétée, le système transforme le code *ChorML1* en code *ChorML2*, en convertissant les composantes spatiales des chorèmes de GML en SVG. La Figure B.21 montre le code ChorML2.

En utilisant le Chorem Editor, l'utilisateur peut modifier la carte chorématique en laissant la gestion des relations topologiques au Chorem Editor. En effet, le Chorem Editor avertit l'utilisateur quand il viole une relation topologique. La Figure B.23 montre le système qui avertit l'utilisateur qui essaie de déplacer la ville de *Genova*.

Finalement, l'utilisateur peut exporter la carte chorématique obtenue, éventuellement modifiée, dans un format de visualisation utilisable par d'autres systèmes. La Figure B.24 montre la carte chorématique visualisée avec un afficheur graphique commun.

B.5 Conclusion

La cartographie traditionnelle est un moyen essentiel pour décrire les faits et les relations qui concernent un territoire. Elle associe à des concepts géographiques des symboles graphiques qui aident les lecteurs à comprendre immédiatement les données représentées.

```xml
<?xml version="1.0" encoding="utf-8"?>
<chorml>
  <ChorematicMap>
    <MapId>map1</MapId>
    <MapName>Italy</MapName>
    <GeneralInformation>
      <Date/>
      <Author/>
      <Format>
        <Height>600</Height>
        <Width>800</Width>
      </Format>
    </GeneralInformation>
    <ChoremList>
      <Chorem>
        <ChoremElementList>
          <ChoremElement>
            <ElementId>elem1</ElementId>
            <ElementName>Rome</ElementName>
            <ElementSize>4</ElementSize>
            <svg>
                <ellipse style="fill: #ff0000" cx="10.7406"
                      cy="4.15625" rx="0.55625" ry="0.55625"/>
            <ellipse style="fill: none; fill-opacity:0;
                     stroke-width: 0.1; stroke: #ff0000" cx="10.7406"
                     cy="4.15625" rx="0.55625" ry="0.55625"/>
            <ellipse style="fill: none; fill-opacity:0;
                     stroke-width: 0.01; stroke: #000000"
                     cx="10.7406" cy="4.15625" rx="0.55625"
                     ry="0.55625"/>
            <text style="fill: #000000; text-anchor:start;
                  font-size: 0.7; font-family: sans; font-style: normal;
                  font-weight: normal" x="8.5" y="4.375">Paris</text>
           </svg>
          </ChoremElement>
.   .   .
.   .   .
.   .   .
```

Figure B.21: Code *ChorML2*.

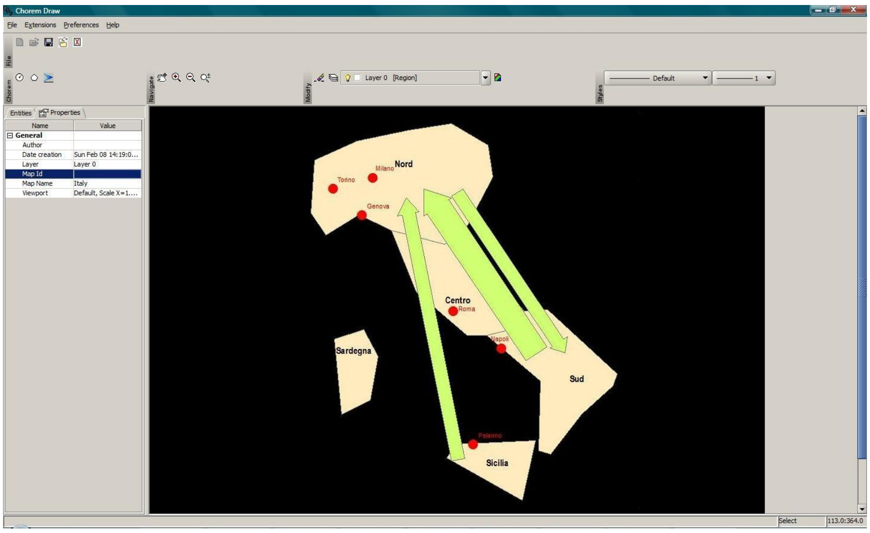

Figure B.22: La carte chorèmatique achevée.

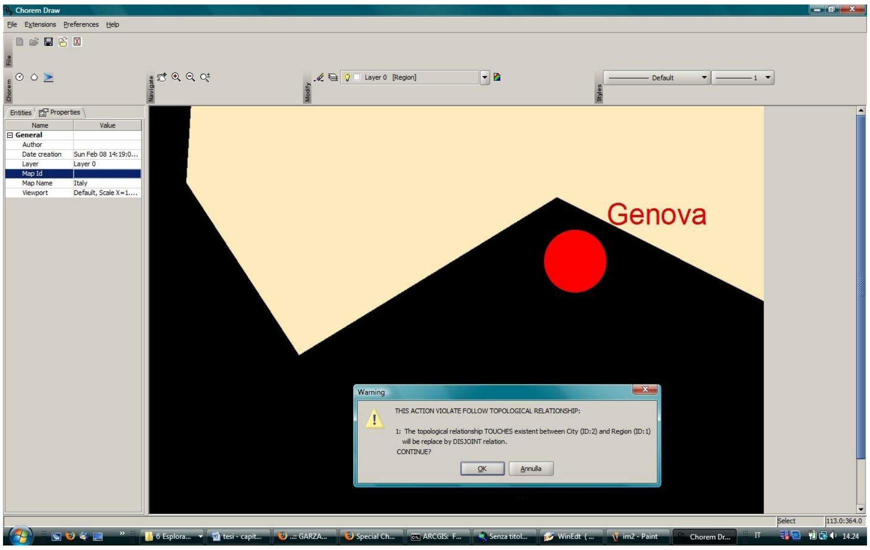

Figure B.23: Une violation des relations topologiques signalée par le Chorem Editor.

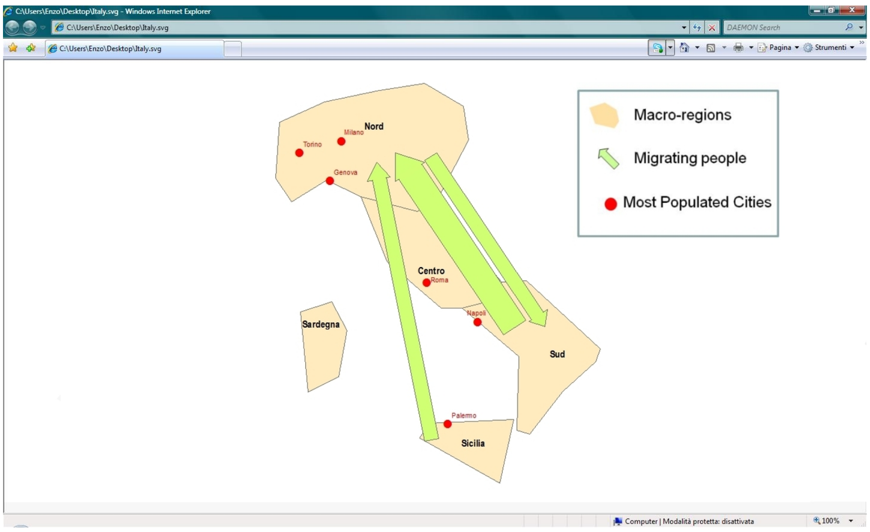

Figure B.24: La carte chorèmatique qui montre les migrations les plus importantes en Italie pendant l'année 2000.

Les utilisateurs experts sont habituellement satisfaits du pouvoir d'expression de la cartographie traditionnelle, quand il s'agit d'affaires simples. Mais, dans le cas d'affaires complexes qui comprennent un grand nombre de données, les utilisateurs experts ont besoin d'une carte qui souligne les aspects les plus saillants plutôt que d'avoir une carte avec un haut niveau de détails.

L'objectif d'un projet de recherche, conçu parmi trois instituts de recherche, est définir des solutions cartographiques capables de représenter convenablement les informations extraites d'une base de données géographiques La solution proposée se fonde sur le concept de *chorèmes* et sur sa capacité de synthétiser des scènes qui comprennent des objets géographiques et des phénomènes spatio-temporels, en leur associant des notations visuelles schématisées.

Cette thèse, née de ce projet international, a proposé une définition et une classification des chorèmes afin de standardiser la construction et l'usage. En outre, un système pour résumer visuellement le contenu des base de données géographiques en utilisant les chorèmes a été conçu et mis en œuvre. Finalement, l'objectif de un langage de type XML multi-niveaux a été de stocker les données des chorèmes et de permettre la communication entre les modules du système proposée.

Les futurs travaux peuvent être synthétisés dans les activités suivantes:

- Effectuer des études expérimentales comparatives pour vérifier l'efficacité de la méthode proposée. En particulier, de telles études viseront à mesurer la capacité des utilisateurs experts à utiliser les cartes chorématiques produites par le système.
- améliorer les techniques usagées dans le système pour réduire l'intervention des utilisateurs.
- étudier l'intégration du système proposé avec un système d'aide à la décision.
- étudier le modèle topologique d'Egenhofer pour améliorer la précision de la phase de simplification.
- vérifier si les chorèmes peuvent être utilisés comme outil de navigation sur le contenu de BD géographiques.
- vérifier si les chorèmes peuvent être utilisés comme un langage d'interrogation et d'accès aux BD géographiques.
- effectuer des études sur l'aspect cognitif comme sur l'aspect sémiotique des chorèmes,

- et enfin améliorer l'interopérabilité du méthode proposé.

Appendix C

Publications

R1 **V. Del Fatto**, L. Paolino, F. Pittarello. A usability-driven approach to the development of a 3D web-GIS environment. In *Journal of Visual Languages and Computing*, vol. 18, n. 3, June 2007, pp. 280-314, Publisher: Elsevier

R2 **V. Del Fatto**, R. Laurini, K. Lopez, M. Sebillo, G. Vitiello. A Chorem-based Method for Visually synthesizing complex phenomena. In *Special Issue of the Information Visualization Journal*, Vol.7, No.3, pp. 253-264, August 2008. Editors: G. Andrienko, N. Andrienko, J. Dykes, S. Fabrikant and M. Wachowicz.

R3 **V. Del Fatto**, V. Deufemia, L. Paolino. Map Integrity Constraint Verification by using Visual Language Parsing. In *Journal of Digital Information Management*, vol. 6 num. 4, 2008, pp. 332-341, Publisher: Digital Information Research Foundation.

C1 **V. Del Fatto**, L. Paolino, F. Pittarello, M. Sebillo, G. Tortora and G. Vitiello. WebMGISQL 3D - Iterating the Design Process passing through a Usability Study. In *Proceedings of the "20th Human-Computer Interaction - HCI2006*, Londra, September 11-15, 2006, pp. 69-73.

C2 **V. Del Fatto**, L. Paolino, F. Pittarello. Representing Topological Relationships by Using 3D objects: An Empirical Survey. In *Proceeding of "the International Workshop on Visual Languages and Computing*, August, 30- September 01, 2006, Grand Canyon (USA), pp. 199-204.

C3 **V. Del Fatto**, L. Paolino, M. Sebillo, G. Vitiello, G. Tortora. Spatial Factors Affecting User's Perception in Map Simplification: An Empirical Analysis. in *Lectures Notes in Computer Sciences - Web and Wireless Geographical Information Systems - (Proceedings of the 8th International Symposium on Web and Wireless GIS W2GIS'08)*, Shangai (China), 11-12 December 2008.

C4 R. Laurini, K. Lopez, **V. Del Fatto**. First Findings in Chorem Discovery and Layout. In *Proceedings of the ISGI 2007, 2nd International Symposium on Generalization of Information*, Geneva, Switzerland, October 1-3, 2007, Proceedings edited by Horst Kemers, CODATA Germany, ISBN 978-3-00-022382-2, pp. 85-104.

C5 **V. Del Fatto**, R. Laurini, K. Lopez, R. Loreto, F. Milleret-Raffort, M Sebillo, D. Sol-Martinez, G Vitiello. Potentialities of Chorems as Visual Summaries of Spatial Databases Contents. *Visual 2007. 9th Conference on Visual Information Systems*, 2007. pp. 540-550

C6 **V. Del Fatto**, R. Laurini, K. Lopez, M. Sebillo, G. Tortora, M. Tucci, G. Vitiello. Chorem Editing - Visual Summary of Spatial Database Content. In *Proceedings of International Workshop on Visual Languages and Computing*, 6-8 September 2007, San Francisco USA

C7 **V. Del Fatto**, M. Giordano, G. Polese, M. Sebillo, G. Tortora. Early Performance Analysis in the Design of Spatial Databases. In P*roceedings of 2nd International Conference on Software and Data Technologies.* 22-25 July 2007 in Barcelona/Spain, volume 3, pp. 172-177

C8 **V. Del Fatto**, V. Deufemia, L. Paolino. Guaranteeing Integrity Constraints During Interactive Map Editing with Visual Language Parsing. In *Proceedings Of The Second International Conference On Digital Information Management (ICDIM 07).* October 28 - 31 2007, pp. 19-24.

C9 **V. Del Fatto**, L. Paolino, M. Sebillo, G. Tortora, G. Vitiello. A Comparison between 2D and 3D Iconic Visual Arrangements for Representing Topological Relationships. In *Proceedings of the Visual Languages and Computing 2007. VLC 2007*, 6-8 September 2007, pp. 209-214

C10 D. De Chiara, **V. Del Fatto**, L. Paolino, M. Sebillo, G. Tortora, G. Vitiello. Human-Computer Interaction in Emergency Management. In *Proceedings of the Visual Languages and Computing 2008, VLC 2008.* Boston, USA, September 4 - 6, 2008, pp. 255-259

CN1 **V. Del Fatto**, L. Paolino, M. Sebillo, G. Tortora, G. Vitiello. Condividere il patrimonio cartografico tramite un sistema basato su un core di metadati geografici. In *Atti della Conferenza Tematica AMFM2006: Informazione Geografica: Data Harmonisation, Interoperabilit e Standard, Reti e Facilities Management*, Ubiquitous GIS, Roma, 21-22 Settembre 2006, pp. 157-164.

Appendix D

Agreement INSA of Lyon - University of Salerno

In this Thesis the research carried out within an international project has been described. In this project three research institutions are involved:

- Laboratoire d'InfoRmatique en Image et Systèmes d'information, Institut National des Sciences Appliquées (INSA) of Lyon (France),
- Dipartimento di Matematica e Informatica (DMI), University of Salerno (Italy) and
- Tecnólogico de Monterrey, Campus de Puebla (Mexico).

This dissertation is a PhD Thesis jointly submitted at the INSA of Lyon, in particular at *Ecole Doctorale Informatique et Information pour la Société*, under supervision of Prof. Robert Laurini and at the University of Salerno (Italy), in particular at *Dipartimento di Matematica e Informatica (DMI)* under supervision of Prof. Giuliana Vitiello.

In the following the Agreement between the University of Salerno (Italy) and The INSA of Lyon (France) is shown, where more details can be found about the jointly PhD Thesis .

CONVENTION DE CO-TUTELLE DE THÈSE

CONVENZIONE DI CO-TUTELA DI TESI

Dans le cadre de l'Université franco-italienne, créée suite au protocole signé à Florence le 6 octobre 1998 par les Ministères des Affaires Etrangères, les Ministères des Universités et de la Recherche, français et italiens, et également, pour la partie italienne, en vertu de la loi du 26/5/2000, n. 161. et pour la partie française de l'arrêté du 6 janvier 2005
In applicazione del protocollo firmato a Firenze il 6 ottobre 1998 dai Ministeri degli Affari Esteri e dell'Università di Francia e Italia, nonché, per la parte italiana, in virtù della Legge 26/5/2000, n.161, istitutivi dell'Università italo-francese, e per la parte francese in virtù del decreto ministeriale del 6 gennaio 2005.

il est établi :
si stipula :

entre / *tra*

Institut National des Sciences Appliquées de Lyon – 20, rue Albert Einstein 69621 Villeurbanne Cedex, France - ci-dessous désigné par "INSA", représenté par le directeur adjoint de la recherche, Prof. Jean-Michel JOLION agissant par délégation du Directeur de l'INSA, le Professeur Alain STORCK,

et / *e*

Università degli Studi di Salerno– Via Ponte don Melillo 84084 Fisciano (SA), Italia - rappresentata dal Rettore Prof. Raimondo Pasquino

la convention de co-tutelle de thèse concernant Del Fatto Vincenzo, de nationalité italienne
la convenzione di co-tutela di tesi a favore di Del Fatto Vincenzo, di nazionalità italiana

TITRE 1 - MODALITÉS ADMINISTRATIVES / *CAPITOLO 1 – PROCEDURE AMMINISTRATIVE*

1.1. Inscription / *Iscrizione*

Del Fatto Vincenzo remplit toutes les conditions nécessaires à son inscription en thèse dans les deux Établissements.
Il dottorando Del Fatto Vincenzo è in possesso di tutti i requisiti richiesti per la sua iscrizione al dottorato di ricerca presso le due università.

L'inscription administrative À INSA sera effectuée dès la signature de cette convention. Elle se fera chaque année dans les deux Établissements.
L'iscrizione all'INSA sarà effettuata successivamente alla firma di questa convenzione, e verrà ripetuta ogni anno presso le due università.

À INSA, l'inscription sera effectuée dans la spécialité de l'Ecole Doctorale Informatique, Information pour la Société, n°335
Presso l'INSA l'iscrizione sarà effettuata nel dottorato di ricerca nella Scuola Dottorale Informatica, ed informazione per la società, n° 335

À Università degli Studi di Salerno l'inscription était effectuée dans la spécialité de la « Formation Doctorale » en Informatica
*All'*Università degli Studi di Salerno, *l'iscrizione è stata effettuata al dottorato di ricerca in Informatica*

1.2. Droits d'inscription / *Tasse d'iscrizione*

Les droits d'inscription seront acquittés entièrement au moins une fois à l'INSA ; l'étudiant sera exonéré à l'Université de Salerno.
Le tasse di iscrizione saranno versate una volta almeno interamente all'INSA di Lione; le tasse di iscrizione all'Università di Salerno non saranno versate perché lo studente Del Fatto è esonerato.

En aucun cas le doctorant ne peut être contraint à acquitter les droits dans plusieurs établissements
Il dottorando non può essere costretto in nessun caso a pagare le tassi di iscrizione presso parecchie università.

1.3. Couverture sociale / *Assistenza sanitaria*

Le candidat bénéficiera de la couverture sociale inhérente à l'acquittement des droits d'inscription dans un des établissements. Il justifiera le cas échéant auprès de l'autre établissement de sa couverture sociale ainsi que d'une assurance civile valable dans chacun des deux pays. Les justificatifs correspondants seront produits au moment de l'inscription.
Il candidato, all'inizio della sua attività presso l'Università partner dovrà dimostrare di essere un possesso di un'assicurazione sanitaria (modello E111/E128 per i cittadini membri della UE). Potrà essergli altresì richiesta la stipula di una polizza di responsabilità civile .

1.4. Conditions d'hébergement et d'aides financières offertes à l'étudiant / *Condizioni di alloggio e sostegno finanziario offerte allo studente*

L'étudiant Vincenzo Del Fatto recevra de l'Université de Salerno une bourse d'ètude triennale pour la durée de la thèse de l'école doctorale d'informatique de Salerno, et l'exonération du paiement des taxes d'inscription. En outre, l'étudiant pouvra utiliser les structures (bibliothèques, laboratoires, restaurant des étudiants, centre sportif, etc.) et les services (email, internet, etc) disponibles près de de l'Université de Salerno.

A l'INSA de Lyon, le doctorant Vincenzo Del FATTO pourra jouir de toutes les prérogatives des doctorants, à savoir l'accès aux bibliothèques, aux laboratoires, aux restaurants universitaires, aux équipements sportifs, etc.

Lo studente Vincenzo Del FATTO riceverà dall'Università di Salerno una borsa di studio triennale per la frequenza al dottorato di ricerca in Informatica e l'esonero dal pagamento delle tasse di iscrizione. Inoltre, lo studente potrà usufruire delle strutture (biblioteche, laboratori, mense, impianti sportivi) e dei servizi (email, internet, etc) a disposizione presso l'Università di Salerno.

All'INSA, lo studente Vincenzo Del FATTO potrà usufruire tutte le strutture per gli studenti (biblioteche, laboratori, mense, impianti sportivi) e dei servizi (email, internet, etc) a disposizione presso l'INSA di Lione.

<u>TITRE 2 - MODALITÉS PÉDAGOGIQUES /*CAPITOLO 2 - MODALITÀ PEDAGOGICHE*</u>

2.1. Travaux de recherche / *Lavori di ricerca*

Les travaux porteront sur : Systèmes d'information géographique et langages visuels
La ricerca verterà su : Sistemi Informativi Geografici e Linguaggi visuali

Les travaux seront dirigés :
La ricerca sarà coordinata :

- à INSA, par le Prof. Robert Laurini du Laboratoire d'Informatique en Images et Systèmes d'Information LIRIS,
- *all'INSA, dal Prof Robert Laurini, del Laboratorio d'Informatique en Images et Systèmes d'Information LIRIS,*

- à Università di Salerno, par le Prof. Giuliana Vitiello du Laboratoire Matematica e Informatica DMI
- *all'Università di Salerno dalla Prof.ssa Vitiello Giuliana delDipartimento di Matematica e Informatica*

Les travaux seront réalisés (répartition du temps de travail entre les deux Etablissements) : approximativement 18 mois auprès de l'INSA et 18 mois auprès de l'Université de Salerno.

La ricerca sarà realizzata (ripartizione dei tempi di ricerca tra i due atenei) :trascorrendo circa 18 mesi presso l'INSA e circa 18 mesi presso l'Università di Salerno.

2.2. Rédaction de la thèse / *Redazione della tesi*

- La thèse sera rédigée en anglais.
- *La tesi sarà redatta in lingua inglese.*

- Deux résumés substantiels (20 pages) seront rédigés en français et italien.
- *Una sintesi sostanziale sarà redatta in francese e in italiano (20 pagine per ciascuna lingua).*

2.3. Soutenance / *Discussione*

La thèse donnera lieu à une soutenance unique à Università di Salerno

Questa tesi si concluderà con un unico esame finale presso l'Università di Salerno

La soutenance sera reconnue par les deux Etablissements. Soit sous forme de diplôme délivré conjointement soit sous forme de diplôme de chaque établissement.

L'esame finale sarà riconosciuto dalle due istituzioni, sia sotto forma di un diploma rilasciato congiuntamente da entrambe le università, sia sotto forma di un diploma rilasciato da ogni università.

Langue utilisée pour la soutenance : anglais ; la présentation orale comportera également au minimum un résumé en français et en italien.

Lingua utilizzata per l'esame finale: inglese ; l'esposizione orale prevederà inoltre una sintesi in italiano e in francese.

Le jury de soutenance sera composé à parité de scientifiques des deux pays concernés. Les deux Directeurs de thèse participeront à la soutenance. Le jury sera constitué conformément aux règles des deux pays concernés et sera approuvé par le directeur de l'INSA, et le Recteur de l'Université de Salerno
Le jury comprend des membres de chaque établissement et des personnalités extérieures. Le nombre des membres ne peut excéder huit.
La commissione giudicatrice sarà costituita da un egual numero di studiosi italiani e francesi. I due Direttori di tesi parteciperanno alla discussione. La commissione sarà composta conformemente alle regole dei due paesi e sarà approvata dal direttore dell'INSA e dal Rettore dell'Università di Salerno
La commissione viene composta da membri di ogni università e da studiosi esterni. Il numero totale dei membri non può essere superiore a otto.

2.4. Délivrance des diplômes/*Rilascio del titolo*

Le grade de docteur sera délivré soit conjointement par les deux établissements soit simultanément dans chacun d'entre eux. Dans tous les cas, sur le diplôme de docteur figureront l'indication de spécialité ou de discipline, le titre de la thèse ou l'intitulé des principaux travaux, la mention de la co-tutelle, les noms et titres des membres du jury et la date de soutenance.
Il titolo di dottore sarà rilasciato congiuntamente dalla due Università oppure contemporaneamente in ciascuna di esse. In ogni caso saranno iscritti sul diploma di dottore la specializzazione o la disciplina, il titolo della tesi o l'intitolato delle ricerche principali, la votazione della co-tutela, i nomi e le qualifiche dei membri della commissione e la data dell'esame finale.

L'INSA délivrera à Del Fatto Vincenzo le titre de: "Docteur de l'INSA de Lyon en Informatique et Information pour la Société" conformément à la réglementation française en vigueur
L'INSA rilascerà a Del Fatto Vincenzo il titolo di "Docteur de l'INSA de Lyon en Informatique et Information pour la Société", in conformità alla normativa francese in vigore

L'Université de Salerno délivrera à Del Fatto Vincenzo le titre de "Dottore di ricerca in Informatica" conformément à la réglementation italienne en vigueur
L'Università di Salerno rilascerà a Del Fatto Vincenzo il titolo di "Dottore di ricerca in Informatica" conformemente alla legislazione italiana in vigore

<u>TITRE 3 – PROPRIÉTÉ/*CAPITOLO 3 - PROPRIETA'*</u>

Les connaissances nouvelles sont la propriété du contractant qui les a obtenues et restent sous son autorité pour leur exploitation et leur diffusion. Elles doivent être protégées conformément aux procédures spécifiques à chacun des pays.
I risultati della ricerca sono proprietà di chi li ha conseguiti e restano sotto la sua disponibilità per il loro sfruttamento e diffusione, e devono essere tutelati conformemente alla normativa vigenti in ciascun paese.

TITRE 4 - MODIFICATION – RÉSILIATION/CAPITOLO 4 - MODIFICHE-RISOLUZIONE

La présente convention est modifiable et résiliable par voie d'avenant établi d'un commun accord et signé par les représentants légaux des deux Établissements.
La presente convenzione è modificabile e rescindibile per comune accordo tra le parti, sottoscritto dai rappresentanti legali delle due Università.

Le ... *(date de la dernière signature)*
(data dell'ultima firma)
21/02/2006

Pour le Directeur de l'INSA de Lyon
Par délégation
Professeur Jean-Michel JOLION
Adjoint au Directeur de la Recherche
Délégué aux Etudes Doctorales

Le Directeur de la Recherche adjoint de l'INSA
Prof. Jean-Michel Jolion

Il Rettore dell'Università degli Studi di Salerno
Prof. Raimondo Pasquino

Le directeur de l'Ecole Doctorale Informatique, Information pour la Société
Pr. Lionel Brunie

Il servizio presso l'Ufficio ... dell'Università degli Studi di Salerno, con qualifica di ... attesta, ai sensi degli art.18 e 19 del D.P.R. N. 445 del 28/12/2000 che la presente copia, costituita da n. 4 pagine è conforme all'originale esibito all'Ufficio ☐ in possesso dell'Ufficio
Fisciano, 22.2.2006

UNIVERSITA' DEGLI STUDI SALERNO

Le directeur de la composante INSA du LIRIS
Pr. Robert Laurini

INSTITUT NATIONAL des SCIENCES APPLIQUÉES de LYON
L.I.R.I.S.

Le Directeur de thèse à l'INSA
Pr. Robert Laurini

Il Direttore della tesi presso Università di Salerno
Drssa Filomena Ferrucci

GIULIANA VITIELLO

Le doctorant / Il Dottorando

Avenant relatif à la convention de these en co-tutelle de Vincenzo Del Fatto

Clausola relativa alla convenzione di testi in co-tutela di Vincenzo Del Fatto

En complément à la convention de thèse en co-tutelle signée le 21 février 2006, entre l'Università degli Studi di Salerno et l'Institut National des Sciences Appliquées de Lyon (INSA), l'article 2.3 est ainsi modifié:
A completamento della convenzione di tesi in co-tutela firmata il 21 febbraio 2006 , tra l'Università degli Studi di Salerno e l'Institut National des Sciences Appliquèes (INSA), l'articolo 2.3 viene modificato come segue:

2.3.1: Soutenance / *Discussione*

Une unique soutenance aura lieu à l'Università degli Studi di Salerno, sera effectuée en langue anglaise et durera 45 minutes.
Una unica discussione avrà luogo presso l'Università degli Studi di Salerno, sarà effettuata in inglese e durerà circa 45 minuti.

2.3.2: Composition du jury / *Composizione della commissione giudicatrice*

Le jury sera composé de six personnes dont trois nommés par l'Università degli Studi di Salerno et trois par l'INSA de Lyon :
La commissione giudicatrice sarà composta da sei membri, di cui tre nominati dall'Università degli Studi di Salerno e tre dall'INSA di Lione:

1. les deux directeurs de thèse seront automatiquement membres du jury, pour l'Università degli Studi di Salerno, Prof.ssa Giuliana Vitiello, et par l'INSA, Prof. Robert Laurini;
1. due direttori di tesi saranno automaticamente membri della commissione, per Università degli Studi di Salerno la Prof.ssa Giuliana Vitiello, per l'INSA prof. Robert Laurini;
2. les deux rapporteurs, l'un proposé par l'Università degli Studi di Salerno et l'autre par l'INSA; celui choisi par l'INSA devra être au moins habilité à diriger des recherches ; celui choisi par l'Università degli Studi di Salerno doit être un expert appartenant à structures (italienne ou étrangère) de recherche publique ou privée;
2. due relatori, uno proposto dall'Università degli Studi di Salerno, esperto appartenente a struttura di ricerca pubblica o privata anche straniera e uno dall'INSA, abilitato a dirigere la ricerca;
3. les deux examinateurs, l'un proposé par l'Università degli Studi di Salerno et l'autre par l'INSA; celui choisi par l'INSA devra être professeur d'université ; celui choisi par l'Università degli Studi di Salerno doit être professeur d'université ;
3. due esaminatori, uno proposto dall'Università degli Studi di Salerno, professore universitario di ruolo e uno dall'INSA, professore universitario.

2.3.3: Président du jury / *Presidente della commissione giudicatrice*

Le président du jury sera choisi parmi les membres français du jury avant la soutenance, puisque la soutenance aura lieu à l'Università degli Studi di Salerno.
Considerato che la discussione della tesi avrà luogo presso l'Università degli Studi di Salerno, il presidente della commissione giudicatrice sarà scelto tra i membri francesi eletti dall'INSA.

2.3.4: Absence d'un des membres du jury / *Assenza di un membro della commissione giudicatrice*

De manière à garantir la parité du jury, si, au dernier moment, un membre du jury était dans l'obligation de ne pas pouvoir participer à la soutenance, la soutenance aura lieu malgré tout et il devra d'urgence communiquer au président du jury un rapport en anglais donnant son jugement sur la thèse.
Per garantire un egual numero di membri della commissione, se, all'ultimo momento, un membro della commissione non può partecipare alla discussione, la discussione avrà luogo ugualmente e il commissario assente dovrà trasmettere al presidente della commissione un rapporto in inglese contenente il suo giudizio sulla tesi.

2.3.5: Procès-verbal / V*erbale di seduta*

Un seul rapport sera rédigé par le président et signé par tous les membres du jury. Ce rapport, long d'une page environ, sera rédigé en anglais.
Un unico verbale sarà redatto dal presidente e firmato da tutti i membri della commissione. Tale verbale, di circa una pagina, sarà redatto in inglese.

Le .... date de la dernière signature *(data dell'ultima firma).*

27 NOV. 2008

Per l'Università degli Studi di Salerno

IL RETTORE
Raimondo Pasquino

Pour l'INSA de Lyon
Professeur Daniel BARBIER, chargé des études doctorales avec délégation de signature du directeur de l'INSA de Lyon, le Professeur Alain ~~STORCK~~

Professeur Attila BASKURT, directeur du LIRIS/INSA

Professeur Robert LAURINI, directeur de thèse

INSTITUT NATIONAL des SCIENCES APPLIQUEES de LYON
L.I.R.I.S.

Bibliography

[1] John Ahn and Herbert Freeman. AUTONAP - an expert system for automatic map name placement. In *Proc. Internat. Sympos. Spatial Data Handling (SDH'84)*, pages 544–569, 1984.

[2] James F. Allen. Maintaining knowledge about temporal intervals. *Commun. ACM*, 26(11):832–843, 1983.

[3] Gennady L. Andrienko and Natalia V. Andrienko. Interactive maps for visual data exploration. *International Journal of Geographical Information Science*, 13(4):355–374, 1999.

[4] Luc Anselin. What is special about spatial data? alternative perspectives on spatial data analysis. In D.A. Griffith, editor, *Spatial Statistics, Past, Present and Future*, pages 63–77, 1990.

[5] ASCE. *The Glossary of the Mapping Sciences.* 1994. New York, ASCE.

[6] Daniel Asimov. The grand tour: a tool for viewing multidimensional data. *SIAM J. Sci. Stat. Comput.*, 6(1):128–143, January 1985.

[7] Marie-Aude Aufaure-Portier and Christine Bonhomme. A high level visual language for spatial data management. In *VISUAL '99: Proceedings of the Third International Conference on Visual Information and Information Systems*, pages 325–332, London, UK, 1999. Springer-Verlag.

[8] M. Bader and R Weibel. Detecting and resolving size and proximity conflicts in the generalization of polygonal maps. In *Proceedings 18th International Cartographic Conference.* Stockholm, pp. 1525-1532, 1997.

[9] Frédéric Barbe. État-nation, le modèle français. *Mappemonde*, 2:44–45, 1994.

[10] Benjamin B. Bederson and Ben Shneiderman. *The Craft of Information Visualization: Readings and Reflections.* San Francisco: Morgan Kaufmann, 2003.

[11] Fabio Bellifemine, Agostino Poggi, and Giovanni Rimassa. Jade: a fipa2000 compliant agent development environment. In *AGENTS '01: Proceedings of the fifth international conference on Autonomous agents*, pages 216–217, New York, NY, USA, 2001. ACM.

[12] M.T. de Berg and M.J. van Kreveld. A new approach to subdivision simplification. 1995.

[13] Roger Brunet. La carte-modèle et les chorèmes. *Mappemonde*, 4:2–6, 1986.

[14] Roger Brunet. *La carte, mode d'emploi.* Fayard, 1987.

[15] Roger Brunet. *La France dans l'espace européen.* 1989.

[16] D. Calcinelli and M. Mainguenaud. Cigales, a visual query language for a geographical information system: the user interface. *Journal of Visual Languages and Computing*, 5:113–132, 1994.

[17] Stuart K. Card, Jock D. Mackinlay, and Ben Shneiderman, editors. *Readings in information visualization: using vision to think.* Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1999.

[18] Tiziana Catarci, Maria F. Costabile, Stefano Levialdi, and Carlo Batini. Visual query systems for databases: A survey. *Journal of Visual Languages & Computing*, 8(2):215 – 260, 1997.

[19] Alessandro Cecconi. *Integration of Cartographic Generalization and Multi-scale Databases for Enhanced Web Mapping.* PhD thesis, University of Zurich, 2003.

[20] Tony Chan and Jianhong Shen. *Image Processing And Analysis: Variational, Pde, Wavelet, And Stochastic Methods.* Society for Industrial and Applied Mathematic, 2005.

[21] Chaomei Chen. Top 10 unsolved information visualization problems. *IEEE Comput. Graph. Appl.*, 25(4):12–16, 2005.

[22] Herman Chernoff. The use of faces to represent points in k-dimensional space graphically. *Journal of the American Statistical Association*, 68(342):361–368, 1973.

[23] Jean-Paul Cheylan, Jean-Pierre Deffontaines, Sylvie Lardon, and Hervè Thery. Les chorèmes: un outil pour l'étude de l'activité agricole dans l'espace rural? *Mappemonde*, 4:2–4, 1990.

[24] Jon Christensen, Joe Marks, and Stuart Shieber. Placing text labels on maps and diagrams. In Paul Heckbert, editor, *Graphics Gems IV*, pages 497–504. Academic Press, Boston, MA, 1994.

[25] Jon Christensen, Joe Marks, and Stuart Shieber. An empirical study of algorithms for point-feature label placement. *ACM Trans. Graphics*, 14(3):203–232, 1995.

[26] William S. Cleveland. *Visualizing Data.* Hobart Press, 1993.

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

[28] R. G. Cromley and G. M. Campbell. Integrating quantitative and qualitative aspects of digital line simplification. *The Cartographic Journal*, 29:25–30, 1992.

[29] Vincenzo Del Fatto, Robert Laurini, Karla Lopez, Rosallva Loreto, Franoise Milleret-Raffort, Monica Sebillo, David Sol-Martinez, and Giuliana Vitiello. Potentialities of chorems as visual summaries of geographic databases contents. In *Qiu, G., Leung, C., Xue, X.-Y., Laurini, R. (Eds.) Advances in Visual Information Systems*, pages 537–548, June 2007. VISUAL 2007, 9th International Conference on Visual Information Systems, Shanghai, China, 28-29 June 2007. Springer Verlag LNCS, Volume 4781.

[30] Vincenzo Del Fatto, Robert Laurini, Karla Lopez, Monica Sebillo, and Giuliana Vitiello. A chorem-based approach for visually synthesizing complex phenomena. *Information Visualization*, 7(3-4):253–264, 2008.

[31] Vincenzo Del Fatto, Luca Paolino, and Fabio Pittarello. A usability-driven approach to the development of a 3d web-gis eenvironment. *J. Vis. Lang. Comput.*, 18(3):280–314, 2007.

[32] Vincenzo Del Fatto, Luca Paolino, Fabio Pittarello, Monica Sebillo, and Giuliana Vitiello. Webmgisql 3d - iterating the design process passing through a usability study. In *Proceedings of the 20th Human Computer Interaction conference (HCI Engage!).*, pages 69–73, 2006. London, September 11-15, 2006.

[33] Vincenzo Del Fatto, Luca Paolino, Monica Sebillo, Genny Tortora, and Giuliana Vitiello. Spatial factors affecting user's perception in map simplification: An empirical analysis. In Michela Bertolotto, Cyril Ray, and Xiang Li, editors, *W2GIS, (volume 5373 of Lecture Notes in Computer Science)*, volume 5373 of *Lecture Notes in Computer Science*, pages 152–163. Springer, 2008.

[34] Hugo A.D. do Nascimento and Peter Eades. User hints for map labeling. *Journal of Visual Languages & Computing*, 19(1):39 – 74, 2008.

[35] Jeffrey S. Doerschler and Herbert Freeman. An expert system for dense-map name placement. In *Proc. Auto-Carto 9*, pages 215–224, 1989.

[36] G. Dorigo and Waldo Tobler. Push pull migration laws. *Annals of the Association of American Geographers*, 73:1–17, 1983.

[37] D. Dorling. *The Visualization of Spatial Structure.* PhD thesis, United Kingdom : Department of Geography, University of Newcastle upon Tyne, 1991.

[38] D. Dorling. Cartograms for human geography. *Visualization in Geographical Information Systems*, pages 85–102, 1994.

[39] D.H. Douglas and T.K. Peucker. Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. *Canadian Cartographer*, 10:112–122, 1973.

[40] Cécile Duchêne and Christophe Cambier. Cartographic generalisation using cooperative agents. In *AAMAS '03: Proceedings of the second international joint conference on Autonomous agents and multiagent systems*, New York, NY, USA. ACM. pages 976–977, 2003.

[41] Cécile Duchêne and Anne Ruas. Généralisation de données géographiques : Présentation des résultats du projet agentd. *Bulletin d'information de l'IGN*, (72), 2001.

[42] Jason Dykes, Alan M MacEachren, and Menno-Jan Kraak. *Exploring Geovisualization.* Oxford : Pergamon, 2004.

[43] Jason A. Dykes. Exploring spatial data representation with dynamic graphics. *Comput. Geosci.*, 23(4):345–370, 1997.

[44] M. Egenhofer and J. Herring. Categorizing binary topological relationships between regions, lines and points in geographic databases. *In M. Egenhofer and J. Herring (Eds.), A Framework for the Definition of Topological Relationships and an Approach to Spatial Reasoning within this Framework.* Santa Barbara, CA : National Center for Geographic Information and Analysis, 1991, pages 1-28.

[45] M. J. Egenhofer. A formal definition of binary topological relationships. In *3rd International Conference, FODO 1989 on Foundations of Data Organization and Algorithms.* New York : Springer-Verlag, 1989, pages 457-472,.

[46] Max J. Egenhofer. Query processing in spatial-query-by-sketch. *Journal of Visual Languages and Computing*, 8:403–424, 1997.

[47] Thomas Erlebach, Torben Hagerup, Klaus Jansen, Moritz Minzlaff, and Alexander Wolff. A new approximation algorithm for labeling weighted points with sliding labels. In *Proc. 22nd European Workshop Comput. Geom. (EWCG'06)*, Delphi. Delphi, Greece, 2006, pages 137-140.

[48] Martin Ester, Hans peter Kriegel, and Jörg Sander. Spatial data mining: a database approach. Berlin : Springer-Verlag, 1997, pages 47-66. (Lecture notes in computer science. no. 1262.

[49] Usama Fayyad, Gregory Piatetsky-shapiro, and Padhraic Smyth. From data mining to knowledge discovery in databases. *AI Magazine*, 17:37–54, 1996.

[50] Jacky Fontanabona. Ebauche d'une grille de concepts spatiaux utilisables lors de l'analyse et de la construction de cartes. *Mappemonde*, 1:1–5, 1994.

[51] M. Gahegan. On the application of inductive machine learning tools to geographical analysis. *Geographical Analysis*, 32(2):113–139, 2000.

[52] Mark Gahegan, Monica Wachowicz, Mark Harrower, , and Theresa-Marie Rhyne. The integration of geographic visualization with knowledge discovery in databases and geocomputation. *Cartography and Geographic Information Science*, 28(1)(1):pp. 29–44(16), January 2001.

[53] M. T. Gastner and M. E. Newman. From the cover: Diffusion-based method for producing density-equalizing maps. *Proc Natl Acad Sci U S A*, 101(20):7499–7504, May 2004.

[54] E. Gbei, B. Moulin, I. Cosma, N. Jabeur, and N. Delval. Conception d'un prototype de service web golocalis appliqu l'industrie rcrotouristique. *Revue Internationale de Gomatique. European Journal of GIS and Spatial Analysis*, 13:pp.375–395, 2003.

[55] Michael F. Goodchild. Geographical data modeling. *Computers and Geosciences*, 18(4):401–408, 1992.

[56] Michael F. Goodchild, Paul A. Longley, David J. Maguire, and David W. Rhind. *Geographical Information Systems: Principles and Technical Issues.* New York : John Wiley & Sons, 1999.

[57] M. Granados. *The Use of Analytical Models to Improve the Exploratory Visualization of Interregional Flow Data.* PhD thesis, Columbus : Geography Department, Ohio State University, Columbus., 2000.

[58] C. Grataloup. *Lieux d'histoire. Essai de geohistoire systematique.* Montpellier : La Documentation française [diff.],, 1996.

[59] Daniel A. Griffith and Larry J. Layne. *A Casebook for Spatial Statistical Data Analysis: A Compilation of Analyses of Different Thematic Data Sets.* USA : Oxford University Press,, 1999.

[60] L. Harrie. The constraint method for solving spatial conflicts in cartographic generalization. *Cartography and Geographic Information Science*, 26(1):55–69, 1999.

[61] Jeffrey Heer, Fernanda B. Viégas, and Martin Wattenberg. Voyagers and voyeurs: Supporting asynchronous collaborative information visualization. In *CHI '07: Proceedings of the SIGCHI conference on Human factors in computing systems.* New York, NY, USA : ACM, 2007, pages 1029-1038.

[62] John Hershberger and Jack Snoeyink. Speeding up the douglas-peucker line-simplification algorithm. In *Proc. 5th Intl. Symp. on Spatial Data Handling*, pages 134–143, 1992.

[63] David C. Hoaglin, Frederick Mosteller, and John W. Tukey, editors. *Understanding Robust and Exploratory Data Analysis.* New York : Wiley-Interscience, 2000.

[64] Robert E. Horn. *Visual Language: Global Communication for the 21st Century.* Island, Wash. : Macrovu Inc., April 1999.

[65] INPG IGN. Generalisation modelling using an agent paradigm. Technical report, INPG - IGN, 2003.

[66] Eduard Imhof. Die Anordnung der Namen in der Karte. In *International Yearbook of Cartography.* Bonn : Kirschbaum, 1962. pages 93-129.

[67] Eduard Imhof. Positioning names on maps. *The American Cartographer*, 2(2):128–144, 1975.

[68] N. Jabeur. *A Multi-agent System for On-the-fly Web Map Generation and Spatial Conflict Resolution.* thèse de doctorat en informatique, Canada : Université Laval, 2006.

[69] Nafaâ Jabeur and Bernard Moulin. A multiagent-based approach for progressive web map generation. pages 99–108. Heidelberg : Springer Berlin, 2005. (Lecture Notes in Computer Science, vol. 3762).

[70] Bin Jiang and Zhilin Li. Geovisualization: Design, enhanced visual tools and applications. *Cartographic Journal, The*, 42:3–4(2), 2005.

[71] Michel Journot. Schmatisation et modlisation cartographiques en france: des pratiques scolaires en dbat. *Cahiers de Gographie du Qubec*, 43(120):495–515, 1999.

[72] Sudhir Kaushik, Elke Rundensteiner, Sudhir R. Kaushik, and Elke A. Rundensteiner. A spatial exploration environment based on a directmanipulation paradigm. *IEEE Transactions on Knowledge and Data Engineering*, 13:654–670, 2001.

[73] D. A. Keim and H. P. Kriegel. Visdb: Database exploration using multidimensional visualization. *Computer Graphics and Applications, IEEE*, 14(5):40–49, 1994.

[74] Nikhil S. Ketkar, Lawrence B. Holder, and Diane J. Cook. Subdue: Compression-based frequent pattern discovery in graph data. In *OSDM '05: Proceedings of the 1st international workshop on open source data mining.* New York, NY, USA : ACM, 2005, pages 71-76.

[75] Alexander Klippel, Heike Tappeb, Lars Kulikc, and Paul U. Leed. Wayfinding choremesa language for modeling conceptual route knowledge. *Journal of Visual Languages & Computing*, 16(4):311–329, August 2005.

[76] Krzysztof Koperski and Jiawei Hah. Discovery of spatial association rules in geographic information databases. pages 47–66. Berlin : Springer-Verlag, 1995. (Lecture notes in computer science, vol. 951).

[77] Menno Jen Kraak. Visualization viewpoints: Beyond geovisualization. *Computer Graphics and Applications, IEEE*, 26(4):6–9, 2006.

[78] Menno Jen Kraak. Geovisualization illustrated. *ISPRS Journal of Photogrammetry and Remote Sensing*, 57:390–399(10), April 2003.

[79] Ruas A. Demazeau Y. Jackson M. Mackaness W. Weibel R. Lamy, S. The application of agents in automated map generalisation. In *Proceedings of 19th ICA meeting.* Ottawa, Canada, 1999.

[80] T. Lang. Rules for the robot draughtsmen. *The Geographical Magazine*, 42:50–51, 1966.

[81] Robert Laurini, Vincenzo Del Fatto, and Karla Lopez. First findings in chorem discovery and layout. In *Horst Kemers, 2nd International Symposium on Generalization of Information.* Berlin : CODATA-Germany, 2007, pages 85-104.

[82] P. U. Lee, S. Fabrikant, D. R. Montello, and J. Bateman. The cognitive conceptual approach as a leitmotif for map design. In *Reasoning with Mental and External Diagrams: Computational Modeling and Spatial Assistance. Papers from the AAAI 2005 Spring Symposium. March 21-23*, 2005. Menlo Park, Calif. : AAAI Press, 2005, pages 90-95.

[83] Maozhen Li, Sheng Zhou, and Christopher B. Jones. Multi-agent systems for web-based map information retrieval. In *GIScience '02: Proceedings of the Second International Conference on Geographic Information Science.* London, UK : Springer-Verlag,2002 pages 161-180.

[84] Chor Pang Lo and Albert K. W. Yeung. *Concepts and Techniques of Geographic Information Systems (2nd Edition) (Ph Series in Geographic Information Science).* Upper Saddle River, NJ, USA : Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 2006.

[85] M. Lonergan and C. B. Jones. An iterative displacement method for conflict resolution in map generalization. *Algorithmica*, 30(2):287–301, December 2001.

[86] Paul A. Longley, Michael F. Goodchild, David J. Maguire, and David W. Rhind. *Geographic Information Systems and Science.* Chichester : John Wiley & Sons, August 2001.

[87] Alan M. MacEachren, Mark Gahegan, William Pike, Isaac Brewer, Guoray Cai, Eugene Lengerich, and Frank Hardisty. Geovisualization for knowledge construction and decision support. *IEEE Comput. Graph. Appl.*, 24(1):13–17, 2004.

[88] Alan M. Maceachren and Menno jan Kraak. Research challenges in geovisualization. *Cartography and Geographic Information Science*, 28:3–12, 2001.

[89] Alan M. Maceachren and Menno-Jan Kraak. Exploratory cartographic visualization: Advancing the agenda. *Computers & Geosciences*, 23(4):335–343, May 1997.

[90] William A. Mackaness. An algorithm for conflict identification and feature displacement in automated map generalization. *Cartography and Geographic Information Science*, 21(4):pp. 219–232(14), October 1994.

[91] Map-Labeling. The map-labeling bibliography [online,] website, `http://i11www.iti.uni-karlsruhe.de/~awolff/map-labeling/bibliography/`, accessed 01 march 2009.

[92] Duane F. Marble, Zaiyong Gou, Lin Liu, and James Saunders. Recent advances in the exploratory analysis of interregional flows in space and time. *Innovations in GIS 4*, pages 75–88, 1997.

[93] Kim Marriott and Bernd Meyer, editors. *Visual Language Theory.* New York : Springer-Verlag, New York, NY, USA, 1998.

[94] B Mayer. Beyond icons: Towards new metaphors for visual query languages for spatial information systems. In *Proceedings of International Workshop on Interfaces to Database Systems (IDS 92)*, pages 113–135., 1992.

[95] Robert B. Mcmaster and Stuart K. Shea. Cartographic generalization in a digital environment: A framework for implementation in a geographic information system. In *Proceedings of GIS/LIS '88.* San Antonio, Texas, USA, volume 1, pages 240-249, 1988.

[96] Ted Mihalisin, John Timlin, and John Schwegler. Visualizing multivariate functions, data, and distributions. *IEEE Comput. Graph. Appl.*, 11(3):28–35, 1991.

[97] F. J. Ormeling. Brunet and the revival of french geography and cartography. *The Cartographic Journal*, 29:20–24, 1992.

[98] Gilles Palsky. *Des chiffres et des cartes.* Paris : CTHS (Comit des Travaux Historiques et Scientifiques), 1996.

[99] Luca Paolino. *Visual and Textual Languages for GIS QUerying.* PhD thesis, Italy : University of Salerno, 2003.

[100] Luca Paolino, Fabio Pittarello, Monica Sebillo, , Genoveffa Tortora, and Giuliana Vitiello. Webmgisql - a 3d visual environment for gis querying. In *Proceedings of International Conference on Visual Languages and Computing (VLC03).* Miami, Florida, USA, 2003, pages 294-299.

[101] Luca Paolino, Genoveffa Tortora, Monica Sebillo, Giuliana Vitiello, and Robert Laurini. Phenomena: a visual query language for continuous fields. In *GIS '03: Proceedings of the 11th ACM international symposium on Advances in geographic information systems.* New York, NY, USA : ACM,pages 147-153.

[102] Palatio M. Pech, Martinez D. Sol, and J. Gonzalez. Adaptation and use of spatial and non-spatial data mining. In *International Workshop Semantic Processing of Spatial Data, GEOPRO 2002.* Mexico City, Mexico, December 3-4, 2002.

[103] M. P. Peterson. Active legends for interactive cartographic animation. *International Journal of Geographical Information Science*, 13:375–383(9), 1 June 1999.

[104] R. M. Pickett and G. G. Grinstein. Iconographic displays for visualizing multidimensional data. In *Systems, Man, and Cybernetics, 1988. Proceedings of the 1988 IEEE International Conference on.* volume 1, pages 514-519, 1988.

[105] Vincent Piveteau and Sylvie Lardon. Chorèmes et diagnostics de territoire: une exprience de formation. *Mappemonde*, 4:1–6, 2002.

[106] U. Ramer. An iterative procedure for the polygonal approximation of plane curves. *Computer Graphics and Image Processing*, 1(3):244–256, November 1972.

[107] Nicolas Regnauld. Constraint based mechanism to achieve automatic generalization using agent model. In *Proceedings GIS Research UK GISRUK.* University of Glamorgan, 2001, pp. 329-332, 2001.

[108] K. Reumann and A.P. Witkam. Optimizing curve segmentation in computer graphics. *International Computer Symposium.* Amsterdam : North Holland Publishing, 1974, pages 467-472.

[109] David Rhind. The incubation of gis in europe. *In Timothy W Foresman, The History of Geographic Information Systems: Perspectives from the Pioneers.* Upper Saddle River, NJ : Prentice Hall PTR, 1998, pages 293-306.

[110] William Ribarsky, Jochen Katz, Frank Jiang, and Aubrey Holland. Discovery visualization using fast clustering. *IEEE Comput. Graph. Appl.*, 19(5):32–39, 1999.

[111] Philippe Rigaux, Michel Scholl, and Agnès Voisard. *Spatial databases with application to GIS.* San Francisco, CA, USA : Morgan Kaufmann Publishers Inc., 2002.

[112] Arthur H. Robinson, Joel L. Morrison, Phillip C. Muehrcke, A. Jon Kimerling, and Stephen C. Guptill. *Elements of Cartography.* 6th ed. New York : John Wiley & Sons, 1995.

[113] Anne Ruas. A method for building displacement in automated map generalisation. *International Journal of Geographical Information Science*, 12(8):789–803, 1998.

[114] Anne Ruas. *Modèle de généralisation de données géographiques a base de contraintes et d'autonomie.* PhD thesis, Paris : Marne la Vallée University, COGIT laboratory, 1999.

[115] Anne Ruas and Cécile Duchêne. A prototype generalisation system based on the multi-agent system paradigm, chapter 14. In William A. Mackaness, Anne Ruas, and L. T. Sarjakoski, editors, *Generalisation of Geographic information: cartographic modelling and applications.* Boston, MA : Elsevier, 2007. pages 269–284.

[116] Anne Ruas and Jean-Philippe Lagrange. Data and knowledge modelling for generalization. *J C Muller; J P Lagrange; R Weibel, GIS and Generalization: Methodology and Practice.* London : Taylor & Francis, 1995, pages 73-90.

[117] Shashi Shekhar and Sanjay Chawla. *Spatial Databases: A Tour.* London : Prentice Hall, June 2002.

[118] Ben Shneiderman. *Designing the User Interface.* Addison-Wesley Publishing Company, 1997.

[119] Tycho Strijk, Bram Verweij, and Karen Aardal. Algorithms for maximum independent set applied to map labelling. Technical Report UU-CS-2000-22, 2000.

[120] Daniel Thalmann, editor. *Scientific Visualization and Graphics Simulation.* New York, NY, USA : John Wiley & Sons, Inc., 1990.

[121] Hervé Théry. Chronochorèmes et paléochorèmes: la dimension temporelle dans la modélisation graphique. *In Modèles graphiques et représentations spatiales,.* Paris : Anthropos, 1990, pages 41-61.

[122] W. R. Tobler. A computer movie: Simulation of population change in the detroit region. *Economic Geography 46*, 46 (20):234–240, 1970.

[123] Waldo Tobler. *Map Transformations of Geographic Space.* PhD thesis, Washington, USA : University of Washington, 1961.

[124] Waldo Tobler. Geographic area and map projections. *The Geographical Review*, 53:59–78, 1963.

[125] Waldo Tobler. A model of geographical movement. *Geographical Analysis*, 13:1–20, 1981.

[126] Waldo Tobler. Experiments in migration mapping by computer. *Cartography and Geographic Information Science*, 14:155–163(9), April 1987.

[127] Waldo Tobler. Geographical movement, [online] presentation `http://www.geog.ucsb.edu/~tobler/presentations/shows/Movement_files/v3_document.htm` , accessed 01 march 2009., 2001.

[128] Waldo Tobler. Movement mapping. Technical report, Santa Barbara ,USA : Center for Spatially Integrated Social Science of the Geography Department of the University of California at Santa Barbara, 2003.

[129] Waldo Tobler. Thirty five years of computer cartograms. *Annals of the Association of American Geographers*, 94(1):58–73, 2004.

[130] Edward R. Tufte. *Envisioning Information.* Cheshire (Conn.) : Graphics Press, May 1990.

[131] J. W. Tukey. *Exploratory Data Analysis.* Reading (Mass.) : Addison-Wesley, 1977.

[132] USGS. "geographic information systems", an information brochure. reston,va. : United states geological survey.

[133] Steven van Dijk. *Genetic Algorithms for Map Labeling.* PhD thesis, The Netherlands, Utrecht : Department of Computer Science, Utrecht University, 2001.

[134] Corn P. J. M. van Elzakker. *The Use of Maps in the Exploration of Geographic Data.* PhD thesis, The Netherlands, Utrecht : Faculteit Geowetenschappen Universiteit Utrecht / ITC., 2004.

[135] Marc van Kreveld, Étienne Schramm, and Alexander Wolff. Algorithms for the placement of diagrams on maps. In *GIS '04: Proceedings of the 12th annual ACM international workshop on Geographic information systems.* ACM : New York, NY, USA, 2004, pages 222-231.

[136] Frank Wagner and Alexander Wolff. An efficient and effective approximation algorithm for the map labeling problem. In Paul Spirakis, editor, *Proc. 3rd Annu. European Sympos. Algorithms (ESA'95), 25-27 September 1995*, volume 979, pages 420–433, 25–27 September 1995.

[137] Colin Ware. *Information Visualization: Perception for Design.* 2 edition. Amsterdam : Morgan Kaufmann, 2004.

[138] J. M. Ware, I. D. Wilson, and J. A. Ware. A knowledge based genetic algorithm approach to automating cartographic generalisation. *Knowledge-Based Systems*, 16(5):pp. 295–303(9), July 2003.

[139] Robert Weibel and G Dutton. Generalising spatial data and dealing with multiple representation. In Paul A. Longley, Michael F. Goodchild, David J. Maguire, and David W. Rhind, editors, *Longlet P., Goodchild M. F., Maguire D.J. Geographical Information Systems : principles, techniques, management, and applications*, chapter Generalising spatial data and dealing with multiple representation. Hoboken, N.J. : Wiley, 1999. 1999, pages 125-155.

[140] M. Wessel and V. Haarslev. Visco: Bringing visual spatial querying to reality. *IEEE Symposium on Visual Languages, 1998.*, pages 170–177, Sep.

[141] Ian Wilson, Marrk Ware, and Andrew Ware. Reducing graphic conflict in scale reduced maps using a genetic algorithm. In *Fifth workshop on progress in automated map generalisation, ICA, commission on map generalisation, Paris*, 2003.

[142] Alexander Wolff, Michael Thon, and Yinfeng Xu. A simple factor-2/3 approximation algorithm for two-circle point labeling. 12(4):269–281, 2002.

[143] Michael Wooldridge and Nicholas R. Jennings. Agent theories, architectures, and languages: a survey. *Intelligent Agents*, 890:1–39, 1995.

[144] Worldmapper. Worldmapper website, `http://www.worldmapper.org/index.html`, accessed 01 march 2009.

[145] Shin T. Wu and Mercedes Rocio Gonzales Marquez. A non-self-intersection douglas-peucker algorithm. *sibgrapi*, 00, 2003.

[146] Missae Yamamoto, Gilberto Camâara, and Luiz Antonio Nogueira Lorena. Tabu search heuristic for point-feature cartographic label placement. *GeoInformatica*, 6(1):77–90, 2002.

[147] Missae Yamamoto and Luiz Antonio Nogueira Lorena. A constructive genetic approach to point-feature cartographic label placement. In *Proc. Fifth Metaheuristics Internat. Conf. (MIC'03).* Kyoto, Japan, 25-28 August, 2003, pages 84/1-7.

[148] Pinhas Yoeli. The logic of automated map lettering. *The Cartographic Journal*, 9:99–108, 1972.

[149] Feibing Zhan and David M. Mark. Conflict resolution in map generalization: A cognitive study. In *Eleventh International Symposium on Computer-Assisted Cartography.* Minneapolis, Minnesota, USA, 1993pages 406-413.

[150] Binhai Zhu and Chung Keung Poon. Efficient approximation algorithms for two-label point labeling. *Computational Geometry and Applications*, 11(4):455–464, 2001.

[151] Steven Zoraster. Integer programming applied to the map label placement problem. *Cartographica*, 23(3):16–27, 1986.

[152] Steven Zoraster. The solution of large 0-1 integer programming problems encountered in automated cartography. *Operations Research*, 38(5):752–759, 1990.

[153] Steven Zoraster. Expert systems and the map label placement problem. *Cartographica*, 28(1):1–9, 1991.

[154] Steven Zoraster. Practical results using simulated annealing for point feature label placement. *Cartography and GIS*, 24(4):228–238, 1997.