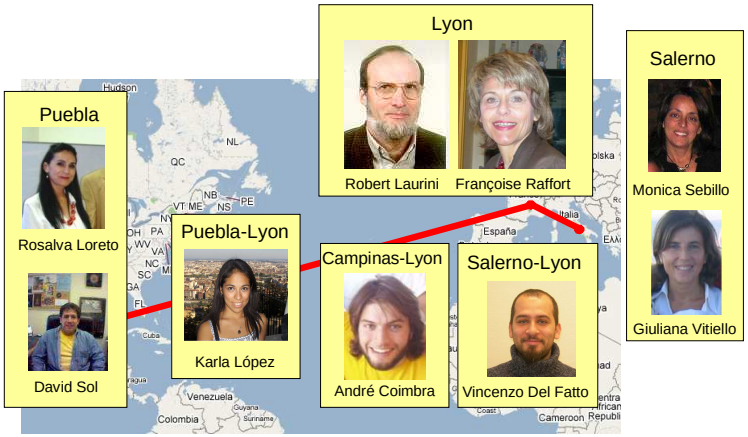
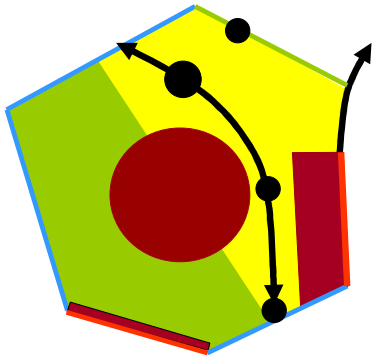
International Project



Two Visions of France

Official map with administrative levels



A chorem map emphasizing the existing problems

Visual Summaries of Geographic Databases

- 1 – Importance and Use of Summaries
- 2 – What are Chorems?
- 3 – Manual Design of Chorems
- 4 – Spatial Data Mining
- 5 – Architecture of a Chorem-based System
- 6 – ChorML
- 7 – Final Remarks

1 – Importance and Use of Summaries

- Global view of a DB
- Structure or contents?
- Important items
  - How to define them?
  - How to extract them?
  - What number?
- Textual or visual layout?
- Decision-makers love having global view and important items

Summaries

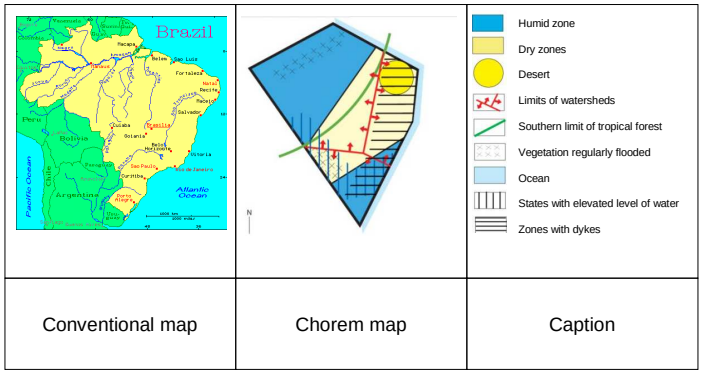
- Summaries for
  - Exploring
  - Understanding
  - Accessing
- For a GeoDB, why a textual summary?
- ➔ Visual summaries

2 – What are Chorems?

- Invented by Pr. Roger BRUNET (University of Montpellier)

- Schematized representation of a territory

Water problem in Brazil

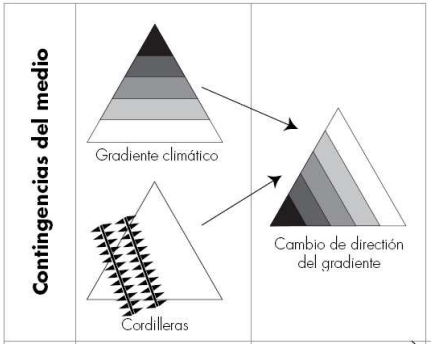


Conventional map	Chorem map	Caption

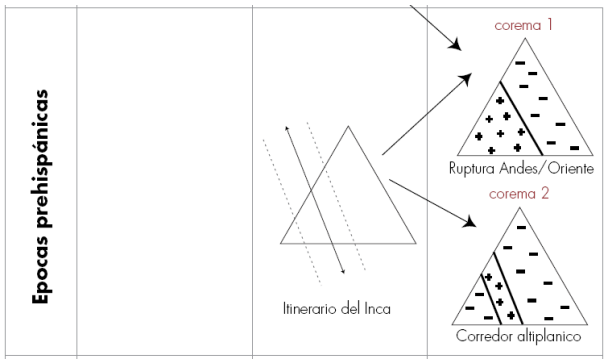
3 – Manual Construction

- Example Bolivia

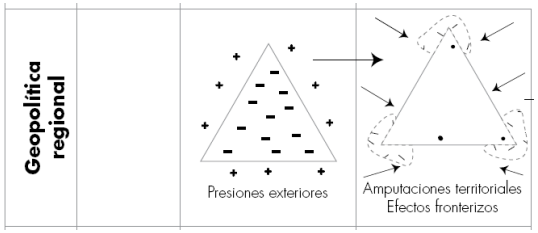
Environment



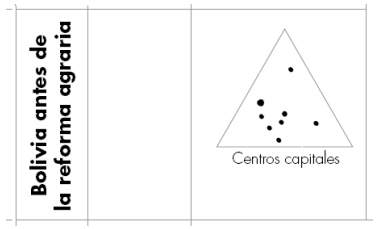
Prehispanic structure



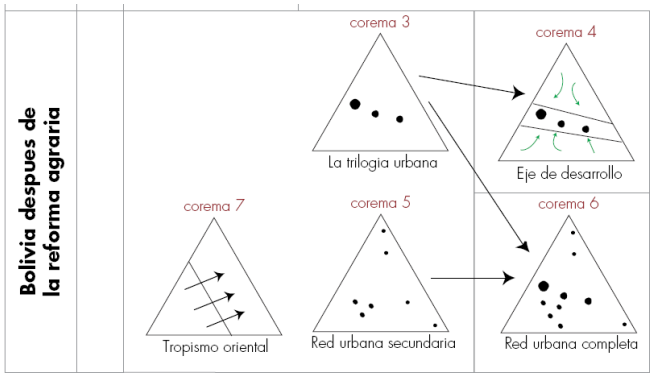
Regional geopolitics



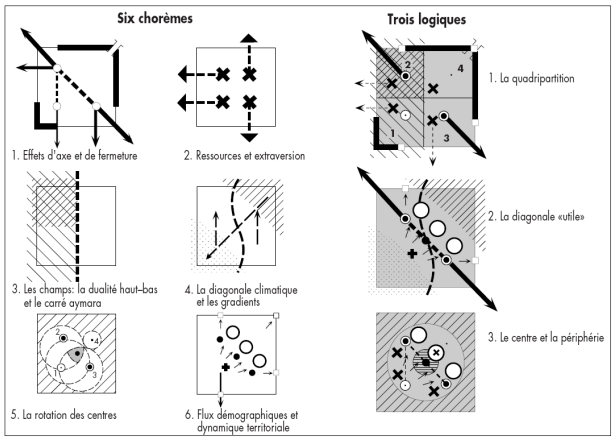
Before land reform



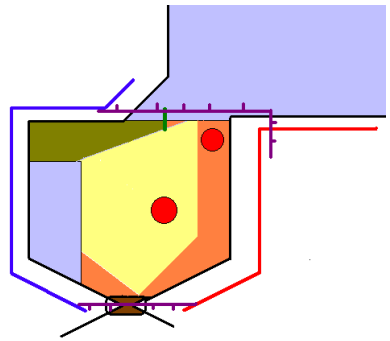
After land reform



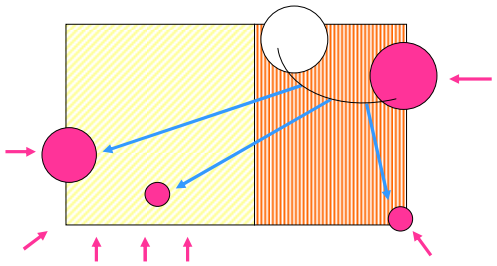
Another chorem of Bolivia



Guess!



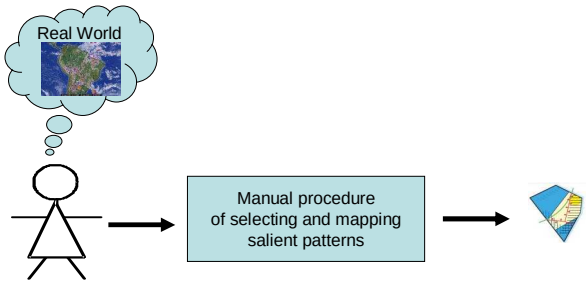
Guess!



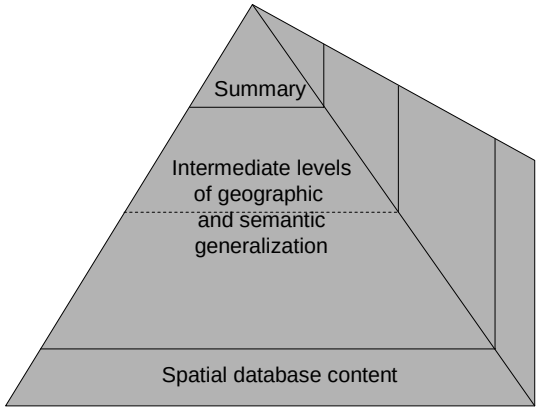
Criticisms

- Original Brunet's chorem list no in use
- Authors create their own chorems
- Geometry
  - Too much simplification perhaps
- Semantics
  - Points of view
  - What is important?
  - Legend

Conventional Manual Procedure



Database summaries



4 – Spatial Data Mining

- Differences between
  - Spatial data mining
    - Patterns which are "true" everywhere
    - *If lake + road to the lake ➔ restaurant*
  - Geographic data mining
    - Positioned patterns (spatial patterns with toponyms)
    - *Eastern coast of Spain is touristically saturated*
- In our case: Geographic Data Mining

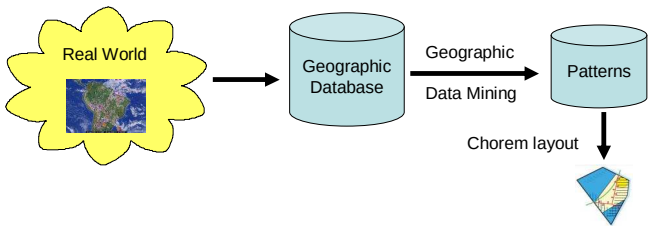
Geographic Data Mining (1/2)

- Lots of techniques have been developed
- Find a combination of techniques suited for geographic pattern discovery
- Pattern types
  - Important locations (located facts)
  - Flows
  - Cluster
  - Co-location relations

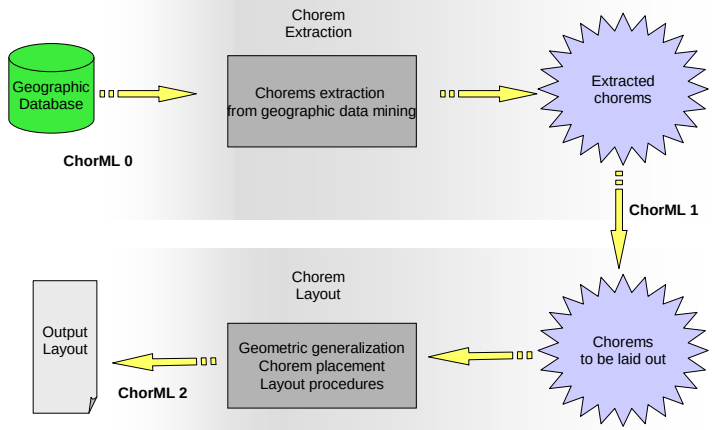
Geographic Data Mining (2/2)

- Starting from a geographic database
- Limited list of geographic patterns
  - Maybe 7 ± 2
- How to define the more important patterns?
  - Suppose you've found 10 000 geographic patterns: how to select 7 ± 2
- Encoding geographic patterns
  - XML, GML, KML, etc..

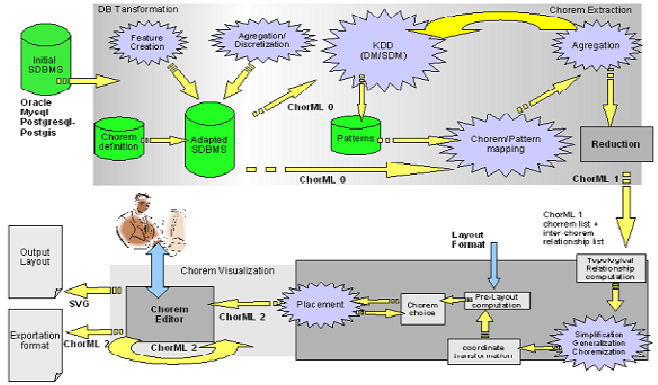
5 – Architecture



Proposed architecture



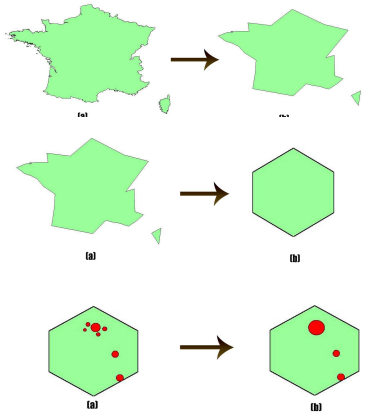
Detailled Architecture



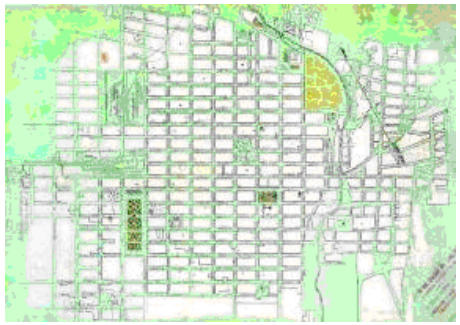
Chorem Layout

- Defining a library of elementary patterns (vector format)
- Defining rules for pattern placement
  - similarities with name placement
  - similarities with geographic generalization

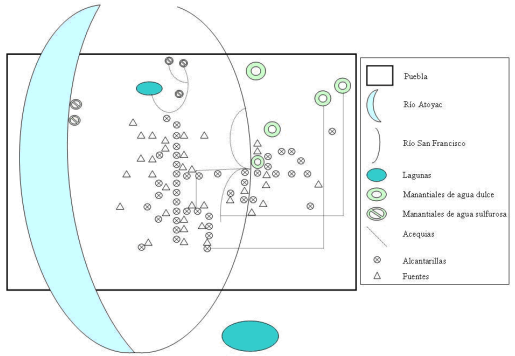
Geographic and Semantic Generalization

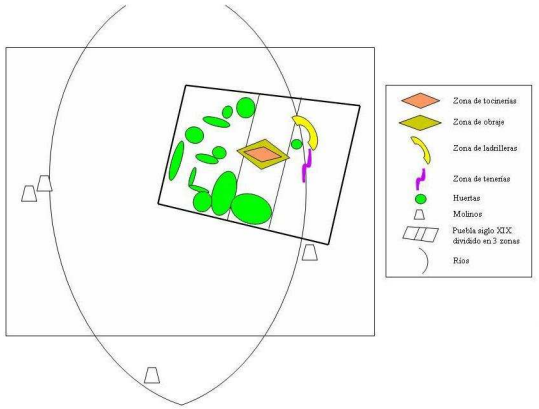


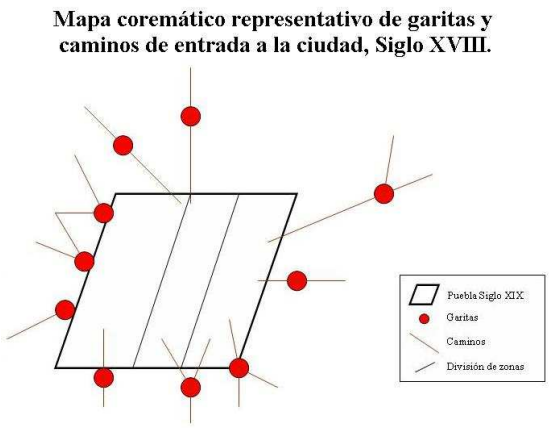
Example: Puebla XVIIth century



Chorem for water supply







6 – ChorML

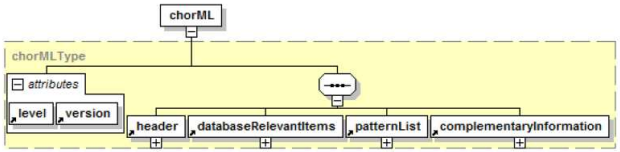
- An XML-based language dedicated to chorems
- Levels
  - level 0 corresponds to the initial database in GML (Geographic Markup Language) (See http://www.opengis.net/gml/ for details),
  - level 1 corresponds to the list of extracted patterns
  - level 2 is a subset of SVG [http://www.w3.org/Graphics/SVG/].

Example

Point representation through levels of ChorML.

Level	Representation
Level 0	Point with longitude/latitude - represents a city and its population.
Level 1	Points with longitude/latitude and its importance. If within a pattern.
Level 2	Point with pixel coordinates, represented by a circle with radius and color.

ChorML Grammar



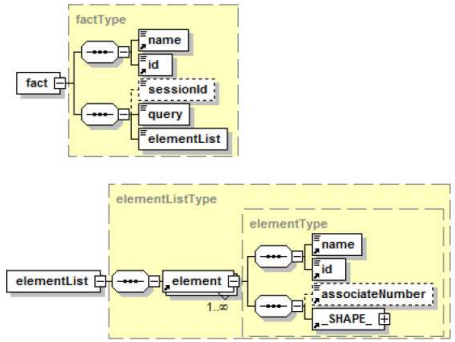
Only spatial constructs will be developed, not the whole grammar!

Spatial knowledge

- Spatial patterns
  - Facts
  - Cluster
  - Flows
  - Co-location patterns
- Spatial constraints
- Outer boundary

Facts

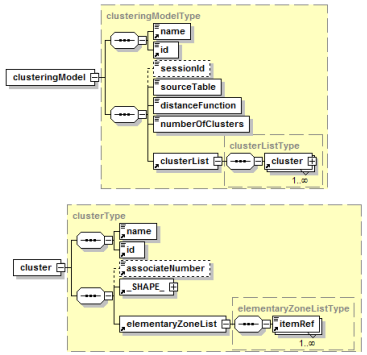
- Can be retrieved thru SQL queries



Cluster

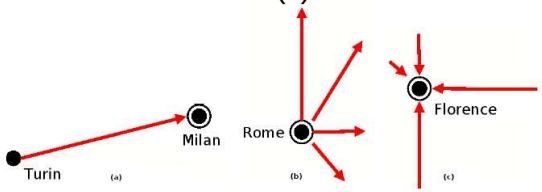
- Can be identified thru data mining
- Small number

Cluster gammar

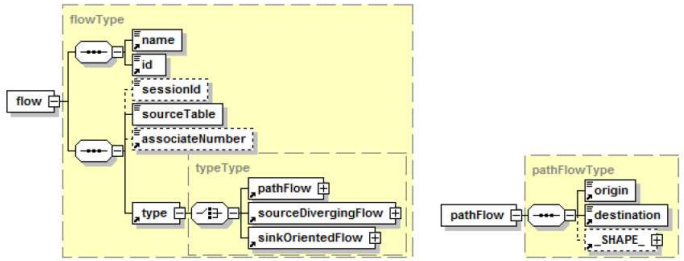


Flows

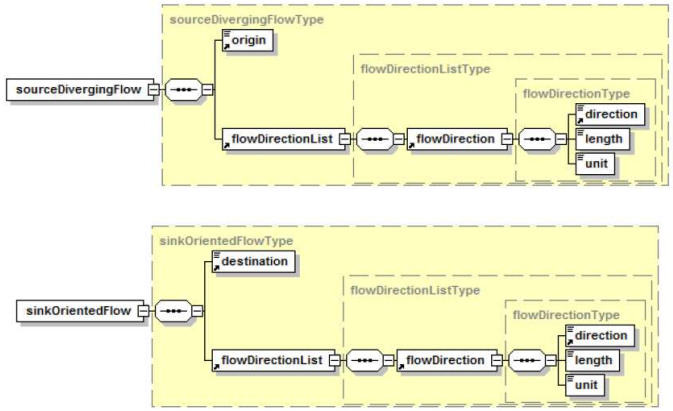
- Can be identified thru spatial data mining
- Various kinds of flows
  - Path flow (a)
  - Source-diverging flow (b)
  - Sink-oriented flows (c)



Flow grammar (1/2)



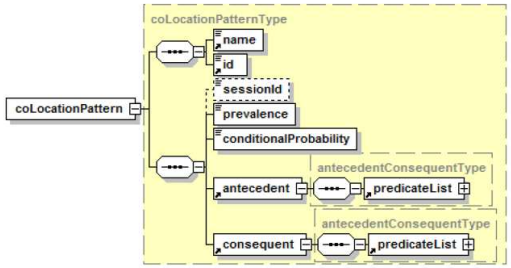
Flow grammar (2/2)



Co-location patterns

- Spatial ***co-location patterns*** represent the subsets of features whose instances are frequently located together in geographic space
- Examples:
  - *"If there is a square in the city center there is a church nearby"*
  - *"Most big cities in Canada are close to the Canada-U.S. border"*
- *Must be retrieved thru powerful SDM procedures*

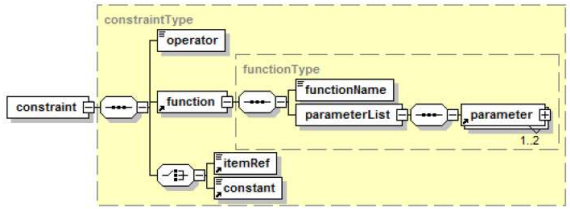
Co-location pattern grammar



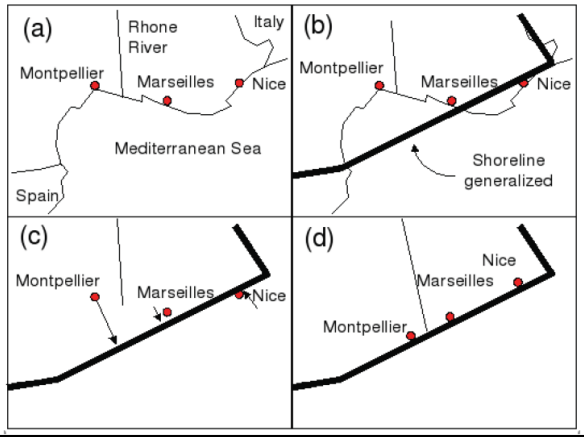
Spatial constraints

- Topological constraints that must be followed especially after generalization or important feature displacement
- Examples
  - Harbors must be always on "land"
  - Outer boundary cities must remain outside
  - River must go to see
  - Etc.
- Egenhofer relations

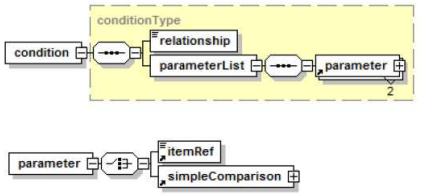
Spatial Constraint Grammar



Spatial constraints



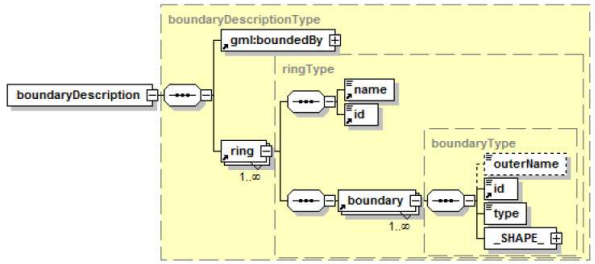
Spatial Constraint grammar



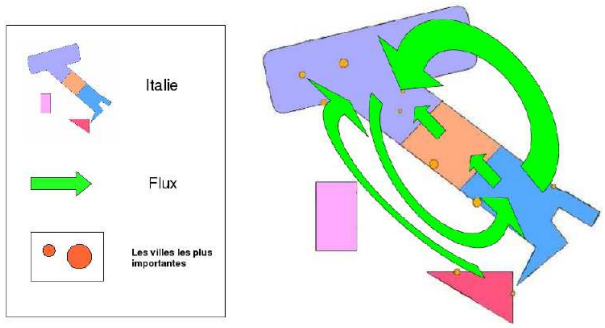
Boundary

- Usually, a Geo-DB includes features inside a territory, but rarely information outside
- Examples
  - Names of seas, Names of neighboring countries, cities, etc.
- Notion of rings
  - Italy rings: mainland, Sicily, Sardinia, etc.

Boundary Grammar

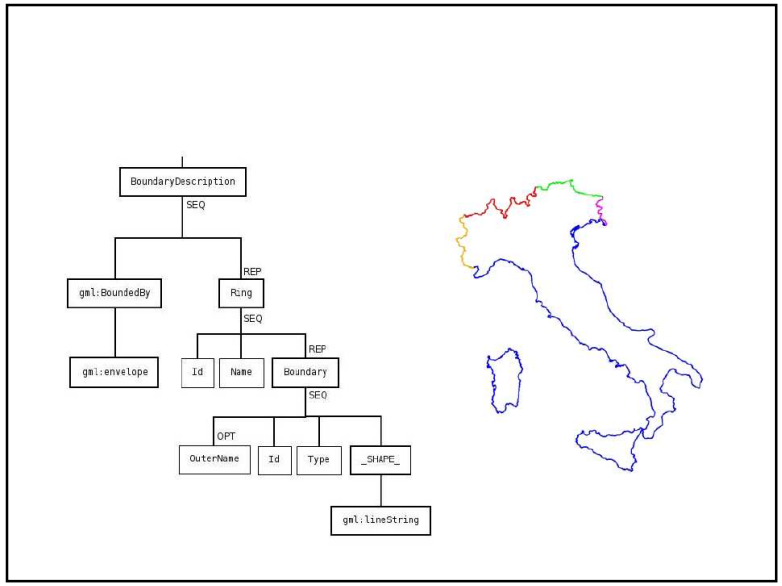


Example: Italy



Example of the Header of the Chorem of Italy

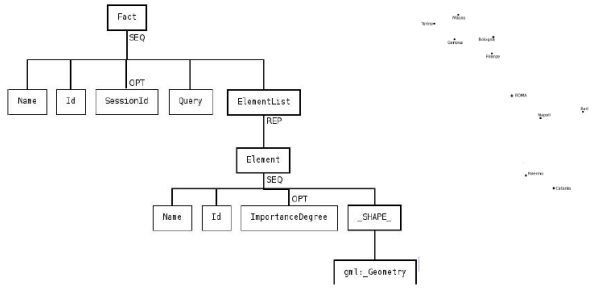
```
<Header>
        <Metadata>
                <Identifier>ItalyChorem0</Identifier>
                <Title>Migration in Italy</Title>
                <Subject>Italy, Chorem, Migration, Displacements</Subject>
                <Coverage>Migration in Italy between the years 2001-03</Coverage>
                <Originator>Del Fatto, Lopez</Originator>
                <Contributor>Coimbra</Contributor>
                <Date>23-04-2008</Date>
                <Source>/chorems/ItalyLevel0.xml</Source>
                <Rights>Copyleft</Rights>
                <Language>EN</Language>
        </Metadata>
        <gml:_ReferenceSystem>
                <srsName>urn:ogc:def:crs:EPSG:6.6:4326</srsName>
        </gml:_ReferenceSystem>
        <SessionList>
                <Session>ItalyDatabase.15-04-2008.18:35.Session1</Session>
                <Session>ItalyDatabase.15-04-2008.23:06.Session3</Session>
        </SessionList>
</Header>
```



Example of the Boundary Description of the Chorem of Italy

```
<BoundaryDescription>
        <gml:boundedBy>
                <gml:Envelope>
                        <gml:lowerCorner>30.70 11.60</gml:lowerCorner>
                        <gml:upperCorner>45.70 22.00</gml:upperCorner>
                </gml:Envelope>
        </gml:boundedBy>
        <Ring>
                <Id>Ring1</Id>
                <Name>Sicily </Name>
                <Boundary>
                        <OuterName>Mediterranean</OuterName>
                        <Id>Ring1.Boundary1</Id>
                        <Type>Sea</Type>
                        <_SHAPE_>
                                <gml:LineString gml:id="p1">
                                        <gml:posList dimension="2">
                                                45.67 18.56 35.56 15.44
                                        </gml:posList>
                                </gml:LineString >
                        </_SHAPE_>
                </Boundary>
        </Ring>
        <Ring>
                <Id>Ring2</Id>
                <Name>Sardinia</Name>
                <Boundary>
                        <OuterName>Mediterranean</OuterName>
                        <Id>Ring2.Boundary1</Id>
```

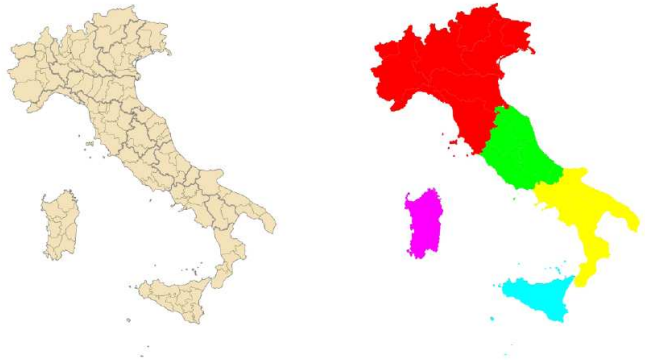
Facts

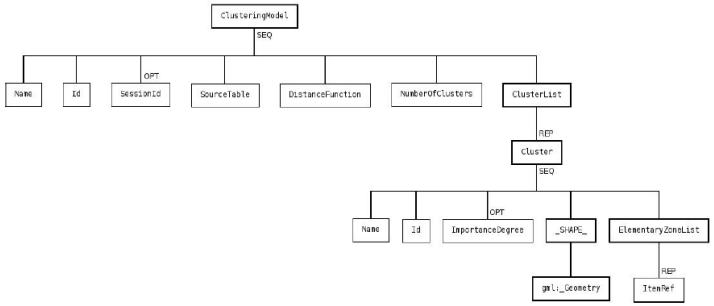


Example of important cities in the chorem of Italy

```
<Fact>
        <Name>Important cities in Italy</Name>
        <Id>Fact1</Id>
        <SessionId>ItalyDatabase.15-04-2008.23:06.Session3</SessionId>
        <Query>SELECT * FROM (SELECT * FROM ville ORDER BY POP2001 DESC)
                WHERE ROWNUM between 1 and
                (select (count(*)*10)/100 from ville)
        </Query>
        <ElementList>
                <Element>
                        <Name>Turin</Name>
                        <Id>Fact1Element1</Id>
                        <ImportanceDegree>2</ImportanceDegree>
                        <_SHAPE_>
                                <gml:Point gml:id="p31">
                                        <gml:pos dimension="2">45.67 88.56</gml:pos>
                                </gml:Point>
                        </_SHAPE>
                </Element>
                <Element>
                        <Name>Milan</Name>
                        <Id>Fact1Element2</Id>
                        <ImportanceDegree>2</ImportanceDegree>
                        <_SHAPE_>
                                <gml:Point gml:id="p31">
                                        <gml:pos dimension="2">45.67 88.56</gml:pos>
                                </gml:Point>
                        </_SHAPE>
                </Element>
```

Cluster





Example of a Clustering Model of the Chorem of Italy

```
<ClusteringModel>
        <Name>Italy in 5 macro regions</Name>
        <Id>ClusteringModel1</Id>
        <SessionId>ItalyDatabase.15-04-2008.18:35.Session1</SessionId>
        <SourceTable>ItalyDatabase.ItalyAdministrativeDivision</SourceTable>
        <DistanceFunction>Euclidean</DistanceFunction>
        <NumberOfClusters>5</NumberOfClusters>
        <ClusterList>
                <Cluster>
                        <Name>South Region</Name>
                        <Id>ClusteringModel1Cluster1</Id>
                        <_SHAPE_>
                                <gml:LineString gml:id="p1">
                                        <gml:posList dimension="2">
                                                45.67 18.56 35.56 15.44
                                        </gml:posList>
                                </gml:LineString >
                        </_SHAPE_>
                        <ElementaryZoneList>
                                <ItemRef>Region3</ItemRef>
                                <ItemRef>Region4</ItemRef>
                                <ItemRef>Region5</ItemRef>
                                <ItemRef>Region6</ItemRef>
                        </ElementaryZoneList>
                </Cluster>
                <Cluster>
                        <Name>Center Region</Name>
                        <Id>ClusteringModel1Cluster2</Id>
                        <_SHAPE_>
```

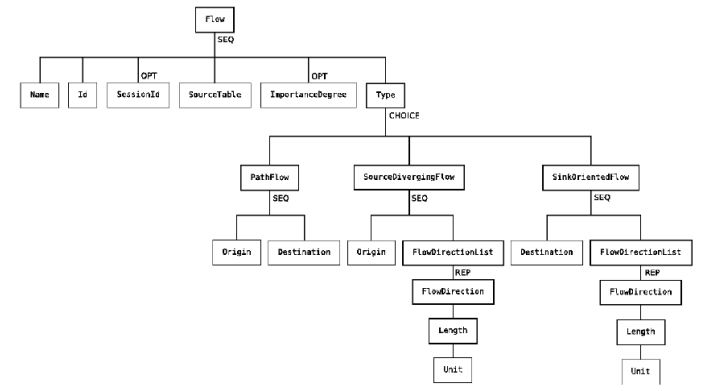


Figure: Path Flow.

Figure: Path Flow.

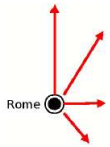


Figure: Source Diverging Flow.

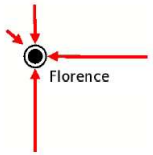


Figure: Sink Oriented Flow.

Example of Flows of the Chorem of Italy

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

Egenhofer Relations

Objects	Relation
Point-Polygon	Contains/Within
Point-Polygon	Touch
Polygon-Polygon	Contains/Within
Polygon-Polygon	Overlap

Figure: Center Region contains Rome.

Figure: Genoa touches Ring3Boundary1.

Example of Conditions of the Chorem of Italy

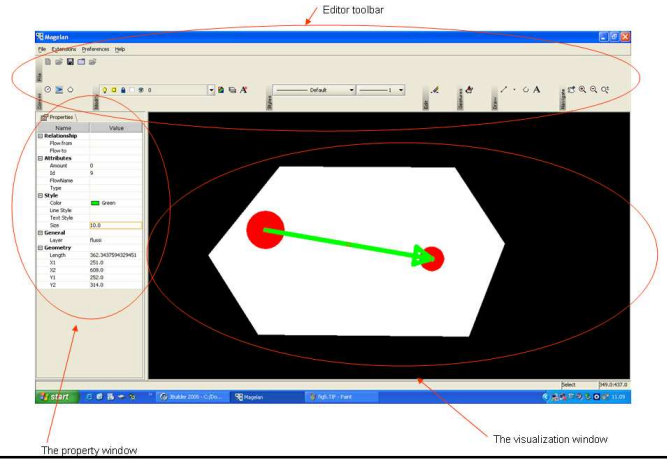
```
<Condition>
        <Relation>Contains</Relation>
        <ParameterList>
                <Parameter>
                        <ItemRef>ClusteringModel1Cluster3</ItemRef>
                </Parameter>
                <Parameter>
                        <ItemRef>Fact1Element1</ItemRef>
                </Parameter>
        </ParameterList>
</Condition>
<Condition>
        <Relation>Touch</Relation>
        <ParameterList>
                <Parameter>
                        <ItemRef>ClusteringModel1Cluster3</ItemRef>
                </Parameter>
                <Parameter>
                        <ItemRef>Ring3.Boundary1</ItemRef>
                </Parameter>
                <Parameter>
                        <ItemRef>Fact1Element3</ItemRef>
                </Parameter>
        </ParameterList>
</Condition>
<Condition>
        <Relation>Contains</Relation>
        <ParameterList>
                <Parameter>
```

(GML) Geographic Markup Language	Describe, model and store geographical information.
(PMML) Predictive Model Markup Language	Encode datamining models.
(SPMML) Spatial Predictive Model Markup Language	Extension of PMML to support spatial data mining
(MathML) Mathematical Markup Language	Describe mathematics.
(SVG) Scalar Vector Graphics	Two-dimensional and graphical applications.

Spatial Database Summaries

- Geographic generalization applied to geometric shapes
- Semantic generalization applied to non-spatial contents

Chorem editor



7 – Final Remarks (1/3)

- Chorems:
  - Visual representation of geographic knowledge
  - Visual summaries for decision-making
  - Gradual access to geographic databases

Final Remarks (2/3)

- Prototype based on ORACLE 10g
- Using ORACLE data mining procedures
- Several results
  - Database on Italy
  - Database on the city of Puebla, Mexico

Final Remarks (3/3)

- Future works
  - Finalizing ChorML
  - Finalizing process of pattern discovery
  - Cognitive studies