

# A Visualization-Based Approach to Explore Geographic Metadata

R. Albertoni, A. Bertone, M. De Martino  
IMATI-CNR  
Via de Marini 6  
16149, Genova, Italy  
albertoni, bertone, demartino@ima.ge.cnr.it

## ABSTRACT

The paper presents a visualization-based approach to explore large databases of geographic metadata. The aim of the data exploration is to search and acquire expressive geographic data to create a geographic information system. Metadata repositories which provide information about available geographic data exist on the web and are usually queried to perform data acquisition task. We propose a visualization-based tool to assist the user in this metadata exploration. The approach is characterized by the integration of multi visualization techniques, interaction methods and data mining concepts. The use of multi visualization techniques and a dynamic interaction with them allow to discover new information among data and assists the user in his reasoning.

## Keywords

Visualization, geographic information, metadata exploration.

## 1. INTRODUCTION

A basic problem in the geographic information system design is the search for actual and expressive geo-data and their analysis. In recent time, there has been an explosive growth in our capabilities to both generate and collect geo-data; these advances have allowed to transform this data into "mountains" of stored data. Users who performs data acquisition usually does not have a complete knowledge about the data and their exploration goal can be vague.

These volumes of data, the intrinsic geo-data heterogeneity, and vague research goals compromise the results of the data acquisition task.

There is an urgent need for methods and tools able to support the user in data exploration both when user searches data in intranet repository and when he looks for new geo-data in Internet. At the moment the metadata concept is used to describe geo-data properties [ISO2002] and the geodata search is performed by exploring their metadata. The query

refinement process, consisting of a succession of queries dynamically built, is used to explore metadata. However, many queries may be needed to find the interesting data and the result can be not satisfactory. In this paper we propose an approach to assist the user in his data exploration based on the integration of different techniques such as data mining [Fayyad96], visualization [Wong97] and graphical interaction techniques [Keim02a]. The idea is to assist the user with a visual data mining approach ([Keim02a], [Keim02b]) to discover new information in the database. The tool aids the user in data acquisition task offering both an automated presentation of data to dynamically visualize the metadata and interactive functionalities to discover the relationships among the different metadata attributes.

## 2. VISUALIZATION-BASED APPROACH

We propose an exploration approach of geographic metadata based on visual data mining: it is characterised by the integration of multi visualization techniques, interaction functionalities and visual data mining concepts as a support for the query refinement process.

The query refinement process, consisting of a succession of queries dynamically built, is used to explore metadata. The main idea is that the use of multi visualization techniques and a dynamic

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

**WSCG POSTERS proceedings**

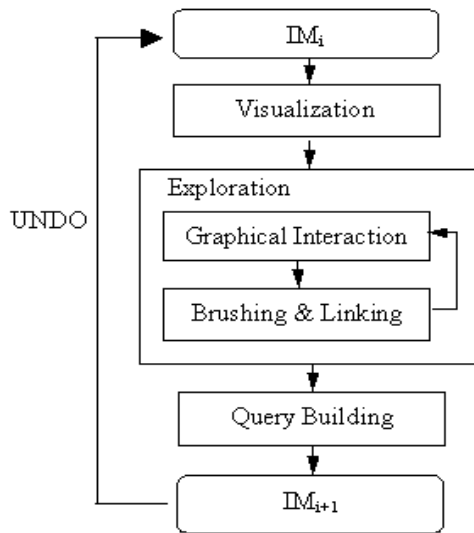
*WSCG'2003, February 3-7, 2003, Plzen, Czech Republic.*

Copyright UNION Agency – Science Press

interaction with them helps to discover new information among data and assists the user in the mining of the query criteria.

In particular the tool aims to provide the following typologies of information:

- quantitative information,
- qualitative information,
- relationship among the available geographic metadata.



**Figure 1: Visualization-based approach**

Figure 1 shows the visualization-based approach proposed in the paper.

Let suppose that  $i^{\text{th}}$  step of query refinement process has been performed and the set of selected geodata  $IM_i$  still contains too many element. User needs to reduce  $IM_i$  to a subset  $IM_{i+1}$  using criterion that he does not know. He is able to mine new selection criteria using the proposed approach that is characterized by three tasks: visualization, exploration, query building.

### Visualization

The first task is characterized by a visualization phase to provide different representations of metadata and to show its attributes and values in a compact and human understandable way. Different visualisation techniques can be applied at the same time. They are classified according to the number of attributes that they can display: single attribute and multiattribute visualization. Hence, a subset of attributes of  $IM_i$  is selected by the user as candidate to express the query constraints.

### Exploration

The visualizations previously displayed are analyzed to explore metadata and extract knowledge about metadata: single attribute visualizations provide

knowledge of the available values and quantitative information of metadata attributes, whereas multiattribute visualizations provide a knowledge on metadata attributes and the existing relationships.

The exploration task is performed using:

- interaction functionalities with the element displayed in a visualization view,
- brushing and linking process [Keim02a] to combine different visualization methods: the selection of data is performed and linked to other views of the data.

The result of the exploration task helps the user to choose the subset of attributes that has to be used in the query.

### Query Building

The criteria are completed and the query is generated. Finally the attribute values are graphically selected to express the query. The query on the current data set is performed and  $IM_{i+1}$  is obtained. As soon as the query is performed, all of the displayed visualizations are updated showing the new set of data  $IM_{i+1}$ . Moreover, if it is necessary after a query refinement step, a new step of the process can be performed starting from the previous visualizations or activating new visualizations. Otherwise, if the results obtained does not satisfy the user need, it is possible to delete some performed selections and returns to an “old” set  $IM_j$  where  $j \leq i$ . Note that as the value for each attribute is directly chosen on the queries generated by the graphical visualisation, all queries that are generated satisfy the property that, given a certain constraint, for each pair attribute-value, the set of values is really present in the available data  $IM_i$ . This avoids the empty set as result of the query.

## 3. TOOL DESCRIPTION

A tool based on the proposed approach has been developed within European project INVISIP (IST-2000-29640) [Albertoni02]. The system is based on graphical interfaces developed in Java. It works as an applet, and it is connected to a database of geographic metadata. The system is characterized by three components: visualization techniques, graphical user interface (GUI) and interaction functionalities.

The visualisation component provides both single attribute and multiattribute techniques to visualize data. An example of single attribute visualization is the histogram technique. The histogram displays the values that an attribute assumes (see Figure 2). It provides complete and immediate information of the available values and their occurrences. The quantitative information of a metadata field can help the user to better address his choices. An example of multiattribute visualization is the Parallel Coordinates

Plot (Figure 3). It displays many attributes at the same time. It provides a complete view of the existing correlations among the attributes previously selected. Whereas Histogram focuses on the quantitative characteristics of metadata, Parallel Coordinates Plot provides the “quality” information, that is, the set of characteristics (attributes and values) of a metadata set.

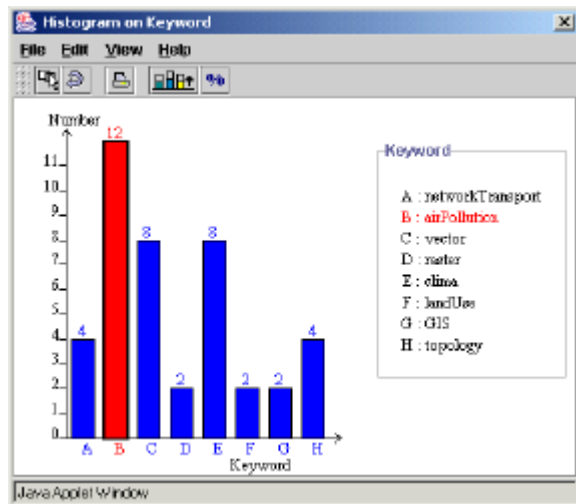


Figure 2: Histogram on Keyword

The graphical user interface is characterized by a tool to select the attribute(s) of metadata and the visualization techniques to display them, and different visualization windows.

The interaction functionalities are properly related to the visualizations and can be distinguished into interaction with the element displayed in a single visualization window, and interaction among elements displayed in different visualization windows (brushing and linking).

### Illustrative Example

In this paragraph an example of the tool application is shown. Let suppose the following scenario: the user has to acquire geo-data. He performs a geographical data searching by exploring a geographical metadata repository. In the repository, metadata are characterized by several attributes such as keywords related the typology of data (clima, economy, air pollution ...), language of the data, map format, URL where data are stored, map resolution, etc. Let suppose that the user addresses his data search to some of these attributes: for instance, he analyses keyword, language and format attributes. And let consider the user has only a vague idea of how these attributes can be combined. For example he is interested in data having MapInfo format, in English and dealing with the keyword emission data, but he does not know all the typologies of data.

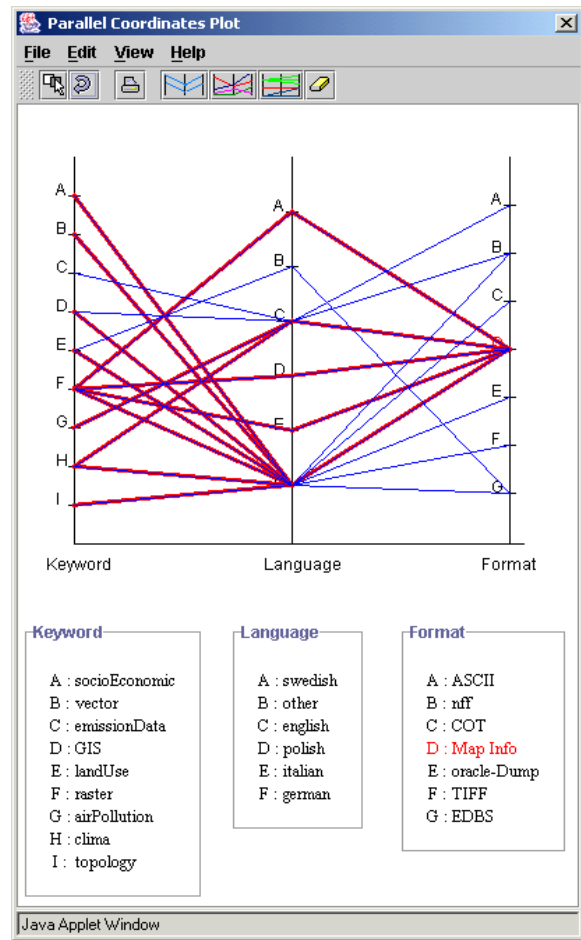


Figure 3: Parallel Coordinate Plots on Keywords, Language and Format

A hypothetic scenario of the exploration activity can be characterized as follows:

The below query is performed

$$(Language = English) \wedge (Format = MapInfo) \wedge (Keyword = emissionData)$$

If the result of the query is an empty set and the user has performed it using a textual query tool then he has to try new query to find a set of data. This task may be extremely boring and more attempts may be needed.

On the contrary, if the visualization-based technique is applied for instance using a Parallel Coordinates Plot (Figure 3), the user immediately understands that there are no data about emission data in MapInfo format. He also realizes that exists another type of data that could anyway satisfy his search: air pollution.

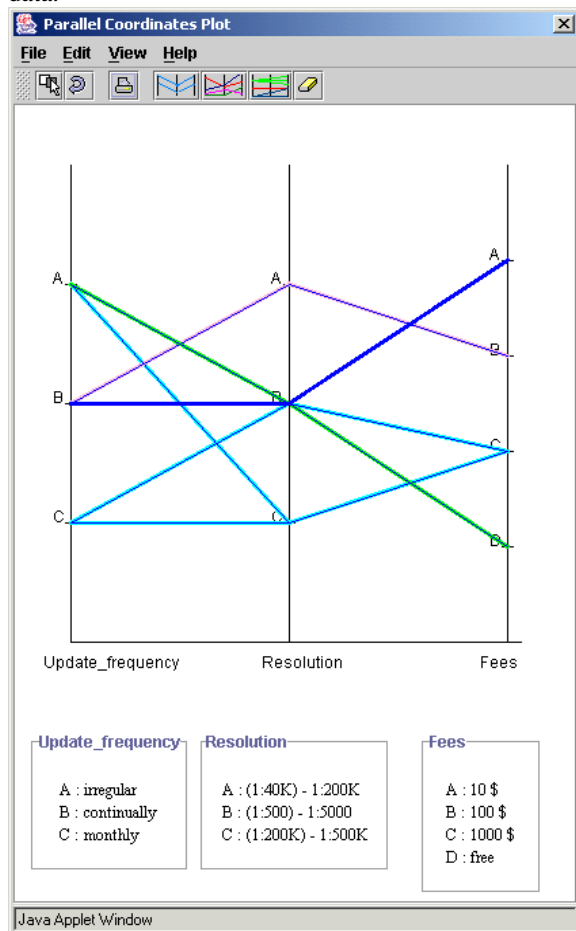
Now interacting with Parallel Coordinate Plot he tries the query

$$(Language = English) \wedge (Format = MapInfo)$$

The obtained situation is visualized by Histogram of Keyword (Figure 2). User selects the air pollution keywords but there are still too many data having this attribute.

Visualizing a Parallel Coordinates Plot on update frequency, resolution and fees attributes (Figure 4), he performs his choices: for instance, the data set that costs 100\$ and is continually updated rather than a free data set that is irregularly updated.

Data exploration is facilitated by “brushing and linking” and “undo” functionalities: every time the user performs a selection on a visualization window, the other views are automatically updated to show the available data and related to the selected subset of data.



**Figure 4: Parallel Coordinate plot on Update Frequency, Resolution and Fees**

The combined use of these functionalities lets the user explore data, improve the query refinement process: interacting with different visualization it is possible to mine relationships among data that are useful hints about how to come to an arrangement when the “ideal” data are not available. That improves user knowledge about data and finally better addresses his choices.

## 4. CONCLUSION

The paper proposes a visualization approach to explore large metadata repository containing information of geographic data via web and assist the user in the data selection. Multivisualization techniques and a dynamic interaction with them achieve two goals: to help the user in the mining process to ease data exploration and to provide an intuitive way to formulate queries by graphical instruments. The tool is open to new visualization techniques and database. In the future work, the tool is going to be also applied for the geographic data analysis: data mining techniques will be integrated to extract new information from data.

## 5. ACKNOWLEDGMENTS

The research described in the paper is part of a research activity which is performing within the European project INVISIP, supported by the European Commission. The authors acknowledge the support of the European Commission, IST Program. They thank all partners of the INVISIP project who contributed to this study: a special thank goes to KTH Team in Stockholm, UoC Team in Krakow and D’Appolonia Team in Genoa for the important discussion.

## 6. REFERENCES

- [Albertoni02] Albertoni R., Bertone A., De Martino M., Information Visualisation And Interactive Geo-Data Mining In Site Planning Process, Eurographic Italian Chapter, 2002.
- [Fayyad96] Fayyad, U. M., and Piatetsky-Shapiro, G. and Smyth P. From Data Mining to Knowledge Discovery: An Overview, Advances in Knowledge Discovery and Data Mining, AAAI Press / The MIT press, pp. 1-34, 1996.
- [Keim02a] Keim, D. Information Visualisation and Visual Data Mining, IEEE Transaction on Visualisation and Computer Graphics, NO 1, 2002.
- [Keim02b] Keim, D., and Müller, W., and Schumann, H. Visual Data Mining in Eurographic STAR proc., Saarbrücken, 2002.
- [ISO02] Draft International Standard on Metadata for Geographic Information, ISO 19115 (Status: DISlevel), see <http://www.statkart.no/isotc211/>, 2002.
- [Wong97] Wong, P. and Bergeron, R. 30 Years of Multidimensional Multivariate Visualization Scientific Visualization - Overviews, Methodologies and Techniques, G. M. Nielson, H.. Hagan, and H. Muller, IEEE Computer Society Press, pp. 3-33, 1997.