Filtering Mechanisms for the Visualization of Geo-referenced Information

Maria Beatriz Carmo, Sérgio Freitas, Ana Paula Afonso, Ana Paula Cláudio

Faculdade de Ciências da Universidade de Lisboa, Departamento de Informática

Campo Grande 1749-016 LISBOA, Portugal

Tel. +351 21 750 0509

bc@di.fc.ul.pt, sfreitas@lasige.di.fc.ul.pt, {apa, apc}@di.fc.ul.pt

ABSTRACT

This paper describes a prototype of a web tool for the visualization of geo-referenced information organized in several categories. The main features that distinguish this tool from others with the same purpose are the inclusion of filtering mechanisms based on semantic criteria and the use of multiple representations with different levels of detail. Filtering mechanisms contribute to reduce the amount of displayed data allowing the generation of intelligible representations. The user selects interactively the categories she/he is interested in. This is a basic filtering mechanism, but besides this, a degree of interest function is used to include semantic criteria. This function quantifies the user's interest in order to visualize the most relevant data, suppressing the less relevant data.

The definition of multiple representations with different levels of detail enables the reduction of the detail when small scales are being used. In addition less detailed representations are used to present less relevant data.

Categories and Subject Descriptors

H.5.2 [**Information Interfaces and Presentation**]: User Interfaces – *graphical user interfaces.*

H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval – *Information filtering, query formulation*.

General Terms

Design, Experimentation.

Keywords

Visualization, geo-referenced data, filtering mechanisms, degree of interest function.

1. INTRODUCTION

Visualization tools that display geo-referenced data over a map are essential to help the user to obtain the information he/she is looking for. This type of applications requires a user-friendly interface to enable querying the data interactively. Moreover, the displayed image must be intelligible and the user should be allowed to interact with the image to get more information about the displayed data.

GIR'05, November 4, 2005, Bremen, Germany.

Local Google [11] is an example, among others, of this kind of applications. Local Google is a Google search service that helps users locating businesses in a specific geographic location. The user specifies the desired category and the location in the *what* and *where* fields, respectively. All business items that satisfy the query are displayed over a map or satellite image. These maps and images are highly interactive and it is possible to zoom in, zoom out, or drag them to view adjacent areas. On the same page exists a textual list with links to get more information about the displayed items. This information can also be obtained clicking over the graphical symbol that represents the item, which is usually called "details on demand". Google Earth [10] is a new service that presents map information using an intuitive and interactive virtual globe. The interface provides a list of available categories. The user selects the categories he/she is interested in using a check box. However, if a lot of items satisfy the query, the image is cluttered. Besides, from this visualization the user only knows how many items satisfy the query and what is its geographical location.

Other search engines solve the problem of cluttered images by grouping elements that are geographically close. For instance, MetaCarta [12], a commercial system that organizes collections of documents according to geographical locations, uses three different types of icons to represent documents. A square is used to represent a single document, a square with a smaller inscribed square is used to represent documents that refer to spots very close, and a stack of squares is used to represent a set of documents that refer to the same location. These icons do not convey any semantic information about the documents.

Nevertheless, images will be easier to understand if improved filtering mechanisms are available. This means that fewer items should be visualized and the ones that are displayed are selected according to more accurate filtering criteria.

The goal of this work is to develop a web tool to display geo-referenced information organized in several categories. The main features that distinguish this tool from others with the same purpose are the inclusion of filtering mechanisms based on semantic criteria and the use of multiple representations with different levels of detail.

The user interactively selects the categories he/she is interested in. In addition, the user can zoom and pan over a map in order to select a geographical region. One filtering mechanism is to reduce the number of categories selected. But the most important feature is to reduce the volume of data displayed according to the users´ interest. To achieve this purpose we have adopted a degree of interest function [8]. In addition, the use of different representations according to the scale can also be considered a filtering mechanism, that is, it contributes to generate a more intelligible visualization. We adopted the approach described in [5] using an architecture presented in [7] that will enable the user to query geo-referenced information using a web browser.

This paper is organized as follows: section 2 explains how the degree of interest function was adapted to the prototype; section 3 describes the organization of the representations; section 4 describes user interface; section 5 presents the system architecture and, finally, section 6 concludes and points at future work directions.

2. DEGREE OF INTEREST FUNCTION

The value of a degree of interest function (DOI), as defined in [8], in a point x depends on the *a priori* importance of the point, API(x), and on the distance between x and the current focus, y: DOI(x|y)=API(x)-D(x,y). Only the points whose DOI exceeds a given threshold are displayed. Varying the threshold will change the number of elements displayed. The elements that are suppressed are the less interesting for the user.

In our prototype the focus is always the centre of the display area. When the operation mode pan is enabled, the point selected by the user will move to the centre of the display, so the user can change the focus easily.

To apply the DOI function we have to define an *a priori* importance function and a distance function. The data is organised in several categories and each category can have several attributes. For each category, the value of the *a priori* importance function corresponds to one of its attributes. For instance, if the user is looking for a high quality hotel, the stars of a hotel can be chosen as the *a priori* importance attribute for the category hotels. In this case, the *a priori* importance of a 5-star hotel is 5, a 4-star hotel has *a priori* importance 4 and so on. This means that the user is giving more importance to a hotel with more quality. The *a priori* importance function expresses a semantic criterion and can be seen as a way of giving context according to the user's purpose [14]. However the value of the degree of interest function depends also on the distance to a selected focus. This means that a 4-star hotel that is closer to the focus may have a higher degree of importance than a 5-star hotel that is further from the focus.

The distance function calculates the Euclidean distance between the location of a data element and the focus. However another distance function can be used, for instance, a function that takes into account the travelling distance between the location of a data element and the focus.

As the distance and the *a priori* importance values belong to different ranges, we have normalized the result of each function in [0,1]. So, DOI result belongs to [-1,1]. As stated before, only the points whose DOI exceeds a given threshold are displayed. So, if DOI threshold is –1, all the data about the selected categories in the selected region are displayed. To reduce the number of displayed elements, the user must reduce the range of DOI results, this means that the user has to choose a higher DOI threshold.

Other authors have also defined functions to quantify the interest of a data element in a particular query. In [13] it is defined a relevance factor that quantifies the distance of each element in a database to the conditions stated in a query. Weighting factors are used to combine the distance calculated for the different conditions. In [15] it is defined a relevance function that is applied to the visualization of event queries in mobile environments. This function depends on a spatial distance, a topical distance and a time distance. The spatial distance is the Euclidean distance. The time distance is the difference in minutes between the time of the event and the current time. The topical distance takes the following values: 0, if the event does not belong to the category searched in the query; 0.5, if the event belongs to the category searched in the query but the type of the event does not match; 1, if the category and type of the event match the query. The total relevance of each event is the sum of the normalised values of each function (spatial, temporal and topical).

Although there are similarities between the relevance and the DOI function, we can point out the following differences, which make the DOI function more suitable for our application. The DOI function does not take into account the time distance, but this is not significant in our application. Besides that, there are differences between the *a priori* function and the topical distance: as the category is selected interactively by the user, its type is not included in the DOI function value, but the main difference between these two functions is that the *a priori* importance function expresses more closely the actual value of the attribute that determines the importance of a data element than the topical distance. In [15] the relevance function is used to associate distinct levels of opacity to symbols that represent the events. In our application, the user controls the amount of data elements represented varying the threshold for the minimum DOI value.

3. GRAPHICAL REPRESENTATION

Multiple representations can be organized in different ways. In [3] a detailed representation is defined and less detailed representations are derived using generalization operators. Another approach consists of defining several representations that can be organized in a hierarchical structure [1, 6, 16] or in a graph [18]. In the last case, if more than one representation is available for the same range of scales, the user selects interactively the representation that must be used. A hierarchical structure is more suitable when we intend to aggregate objects in smaller scales. In [2] a hierarchical structure is also used in order to aggregate information in lower levels, however in this case the hierarchy does not store graphical representations, but the number of observations in lower levels. In the GiMoDig project [17] a Multiple Representation/Resolution Database is used to store multiscale representations. Links between multiscale representations are created in the generalisation process.

In our prototype, different representations are defined both for the background maps and for the elements that satisfy the query. In the results area, a map is displayed with a level of detail according to the current scale. Over this map, the data elements corresponding to the selected categories are displayed. As aggregations are not considered, the representations corresponding to data elements have a linear organization. This means that each representation has an associated range of scales, and one representation replaces other when the current scale changes to a different range. The scale of representation is automatically adjusted with zoom operations. Each category has its own list of representation. Actually we are using only symbolic representations to present the data elements belonging to the different categories. We assume that for higher scales less symbols are displayed and a more complex representation can be used. To emphasize the degree of interest of each data element, the representation is also selected taking into account the value of the degree of interest function, *i. e.*, for less relevant information a less detailed representation is chosen.

A way to achieve this goal is to divide the range of values of the degree of interest function [minDOI, maxDOI] by the number of available representations. Objects belonging to the upper interval will be displayed with the representation associated with the current scale. Objects belonging to lower intervals will be displayed using less detailed representations [4].

The worst case occurs when the current scale is associated with the simplest representation: all the objects of the same category will be visualized with the same representation. On the other hand, when the current scale corresponds to the most detailed representation, all the available representations can be used.

Let us assume that we have **n** representations, **1** through **n**, with the $\mathbf{n}^{\mathbf{th}}$ representation corresponding to the largest scale and the $\mathbf{1}^{\mathbf{st}}$ to the smallest one. The range of values of the degree of interest function will be split into **n** intervals, $\mathbf{s_1}$ through $\mathbf{s_n}$. If the representation associated with the current scale is **k**, then objects belonging to the upper interval, $\mathbf{s_n}$, will be displayed with representation **k**, objects in the interval $\mathbf{s_{n-1}}$ will be displayed with representation **k-1,** and so on. Objects in the intervals $\mathbf{s_{n-k+1}}$ to $\mathbf{s_1}$ will be displayed with representation **1**, the one with less detail.

So, to obtain the representation of a data element, we have to consider its DOI value: if **DOI=maxDOI**, the current representation is used; if **DOI** <**maxDOI**, we have to determine first the subinterval, **subint**, where DOI belongs

subint = (DOI-minDOI) div ((maxDOI-minDOI) /n)+1

And then the distance, **delta**, between this subinterval and the subinterval with the highest values

delta = n – subint

The selected representation is

Max (1, **current representation – delta)**

To illustrate the proposed approach, four symbols were chosen for each category (Fig. 1). The simplest representation uses a letter that identifies the category. The remaining three symbols are obtained by drawing up to three concentric circumferences around the basic symbol. Our goal is to have symbols with different complexity to illustrate the concept, although better solutions can be derived taking into account visual design principles pointed out in [15].

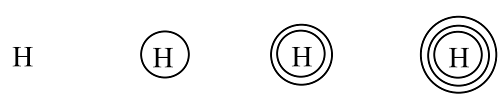

Figure 1. Symbolic representations

In the query result shown in Fig. 2, the data elements represented with more circumferences correspond to those elements that have a higher degree of importance and are a better answer for what the user is looking for: either because they have a higher *a priori* importance or because they are closer to the focus.

4. INTERFACE

The interface window (Fig. 2) is divided into four main areas: a results area, where the data is visualized; a reference area, where a global map is displayed with a rectangle that shows the area enlarged in the results area; a navigation area to select the operation mode (zoom in, zoom out, pan); and a categories area with the list of categories that the user can select. The categories area also includes a box with arrows to increase or decrease the degree of interest function threshold.

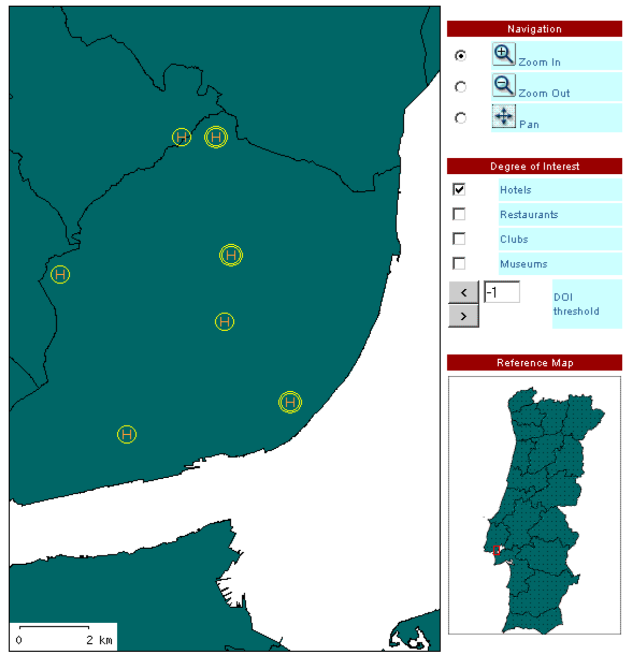

Figure 2. Interface

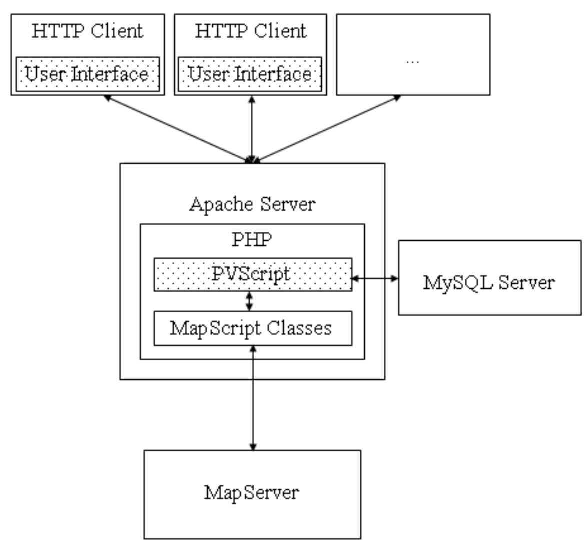

Figure 3. Architecture

5. ARCHITECTURE

The architecture of the system (Fig. 3) follows the Internet client/server model and is composed by the following components:

- Apache Server (version 2.0) with PHP support (version 3.0);
- MapServer (version 4.2) - Server that renders the maps to display in the user interface;
- MapScript (version 4.2) - Library of PHP classes that establishes the communication between Apache Server and MapServer;
- MySQL database – Component used to store a common database with the category data, and graphical representations range scales;
- PVScript (version 1.0) – PHP classes that make the personalization of the information accordingly to the user input and the stored information;
- Graphical user interface – User interface written in PHP and HTML.

6. CONCLUSIONS AND FUTURE WORK

The paper describes a web tool that supports the visualization of geo-referenced information organized in categories. The main contribution of this tool is a semantic filtering mechanism, a feature that enables the user to express his degree of interest in each particular category.

Due to this feature, intelligible graphical displays are produced: uninteresting information is suppressed and all the items are displayed with a level of graphical detail directly proportional to the interest the user has in it and to the scale value in use.

To implement this feature, a degree of interest function has been used along with multiple representations with different levels of detail for each data category.

Future work will be focused on improving user's capabilities to express his degree of interest in a particular category. To attain this, the interface will allow the user to choose which attribute corresponds to the *a priori* importance and their order of increasing importance. For instance, the attribute "number of accommodations" of a hotel can be used as *a priori* importance selecting large hotels as the less important hotels or selecting large hotels as the most important hotels. In the first case the user is interested in a quiet hotel and, in the second one, the user is looking for a cosmopolitan atmosphere hotel.

Besides this, more complex DOI functions should be conceived to combine several attributes at the same time.

Finally, we will explore the integration of this tool as part of the geo-search engine that is being developed by the XLDBGroup [9].

7. REFERENCES

[1] Bederson, B., Meyer, J , Good, L., Jazz: an Extensible Zoomable User Interface Graphics Toolkit in Java, *Proceedings of UIST ´00*, S. Diego, USA, Nov. 2000

[2] Burghardt, D., Purves, R. S., Edwardes, A. J., Techniques for on the-fly generalisation of thematic point data using hierarchical data structures, *Proceedings of the GIS Research UK 12th Annual Conference,* Norwich, UK, 2004

[3] Camossi, E., Bertolotto. M., Bertino, E., Guerrini, G., A Multigranular Spatiotemporal data Model*, Proceedings ACM GIS'03*, 2003

[4] Carmo, M.B, Cunha, J.D., Cláudio, A.P., Visualization of Large Volumes of Information Using Different Representations, *Proceedings of IEEE Information Visualization '97*, pp. 101-105, London, England, August 1997

[5] Carmo, M.B, Cunha, J.D., Cláudio, A.P., IVPrototype – an Information Visualization Prototype, *Proceedings of the 6th International Conference on Information Visualization*, IEEE Computer Society Press, pp. 159-164, London, England, July 10-12, 2002

[6] Frank, A. U., Timpf, S., Multiple Representations for Cartographic Objects in a Multi-Scale Tree - An Intelligent Graphical Zoom, *Computer and Graphics, vol.18, nº6*, 1994

[7] Freitas, S., Carmo, M. B., Afonso, A. P., A Personalized Visualization Tool for Geo-referenced Information, *student poster in Proceedings of ACM ITiCSE'05*, pp. 398, Monte da Caparica, June 2005

[8] Furnas, G., Generalized Fisheye Views, *Proceedings of Computer Human Interaction'86*, pp. 16-23, 1986

[9] GREASE Project, XLDB Group of LASIGE (Large Scale Informatic Systems Laboratory), http://xldb.fc.ul.pt/grease

[10] http://earth.google.com/

[11] http://local.google.com/

[12] http://www.metacarta.com/

[13] Keim, D. A., Kriegel, H.-P., VisDB: Database Exploration Using Multidimensional Visualization, *IEEE Computer Graphics and Applications,* vol 14, nº5, 1994

[14] Nivala, A.-M. and L. T. Sarjakoski, Need for Context-Aware Topographic Maps in Mobile Devices. In: Virrantaus, K. and H. Tveite (eds.), *ScanGIS 2003 - Proceedings of the 9th Scandinavian Research Conference on Geographical Information Science*, Espoo, Finland, pp.15-29, June 4-6, 2003

[15] Reichenbacher, T., Mobile Cartography - Adaptive Visualisation of Geographic Information on Mobile Devices, Verlag Dr. Hut, München, ISBN 3-89963-048-3, 2004

[16] Sarkar, M., Snibbe, S. S., Tversky, O.J., Reiss, S.P., Streching the Rubber Sheet: A Metaphor for Viewing Large Layouts on Small Screens, *Proceedings UIST'93*, ACM Press, pp. 81-91, Atlanta, November 3-5, 1993

[17] Sester, M., Sarjakoski, T., Harrie, L., Hampe, M., Koivula, T., Sarjakoski, T., Letho, L., Elias, B., Nivala, A.-M., Stigmar, H., Real-time generalisation and multiple representation in the GiMoDig mobile service, Project report, 2004, http://gimodig.fgi.fi/deliverables.php

[18] Stonebraker, M., Chen, J., Nathan, N., Paxson, C., Su,A. Wu, J., Tioga: A Database-Oriented Visualization Tool, *Proceedings Visualization'93*, Nielson, G.M., Bergeron, D., (eds), 1993