

# INVISIP: Usage of Information Visualization Techniques to Access Geospatial Data Archives

Stefan Göbel<sup>1</sup>, Jörg Haist<sup>1</sup>, Harald Reiterer<sup>2</sup>, and Frank Müller<sup>2</sup>

<sup>1</sup> Fraunhofer Institute for Computer Graphics, GIS department,  
Fraunhoferstr. 5, 64283 Darmstadt, Germany  
{stefan.goebel, joerg.haist}@igd.fhg.de  
<http://www.igd.fhg.de/igd-a5/>

<sup>2</sup> University of Konstanz, Department of Computer & Information Science,  
Universitätsstr. 10, 78457 Konstanz, Germany  
{harald.reiterer, frank.mueller}@uni-konstanz.de  
[http://www.inf.uni-konstanz.de/iw\\_is](http://www.inf.uni-konstanz.de/iw_is)

**Abstract.** Complex application examples or scenarios such as site planning need a lot of information, especially spatially referenced data<sup>1</sup> in order to find a best place for a new building or an industrial area and to solve individual planning tasks such as generating ecological, environmental or traffic reports. Recent initiatives to geospatial data archives offer access to a wealth of distributed data covered by the widespread information spectrum of different geodata disciplines (e.g. environmental data, geologic data, cadastral data, remote sensing data or socio-demographic data), but offer only basic levels of interactivity and user assistance. The EU-funded project INVISIP (Information Visualization in Site Planning, IST-2000-29640) addresses this lack of usability and aims to develop new concepts and methods to support all involved parties within the different phases of the multi-step site planning process. This paper describes INVISIP and introduces GeoCrystal and SuperTable as new information visualization techniques to support users in this process.

## 1 Introduction

Concerning this information demand in complex applications such as site planning, in the last decade a lot of scientific and organizational effort has been spent to structure the geodata market and to establish information systems and global infrastructures enabling data suppliers to describe (-> eCommerce, metadata) and users to find appropriate data (-> information retrieval, data mining). Most popular peculiarities of such information systems are metadata information systems (MIS) or catalogue systems (CS) on regional, national or international basis, online-shops or web-portals to geospatial data archives. Recent initiatives offer access to a wealth of distributed data covered by the widespread information spectrum of different geodata disciplines, but

---

<sup>1</sup> Studies have shown that the amount of spatial information (spatially referenced data, geodata) is 80-85 percent relating to all new data produced by the actual information society.

offer only basic levels of interactivity and user assistance within the different steps of the information retrieval process.

The EU-funded project INVISIP<sup>2</sup> addresses this lack of usability and aims to develop new concepts and methods to support all involved parties within the different phases of the multi-step site planning process. INVISIP has been initiated by the authors of this paper as technical driven project within action line III.4.2 (Information Visualization) of the actual IST-2000 call. The technical basis (background) for INVISIP builds the InGeo IC [8] as technical information and communication platform of the InGeoForum<sup>3</sup> and results of the INSYDER project referring to WWW search result visualization [13].

## 2 State-of-the-art analysis

The more information and data are produced in the actual information society, the more important become mechanisms and systems which organize data and provide information where to find which data. Most popular peculiarities of such information systems are web-based search engines (e.g. AltaVista, Lycos and Yahoo), digital archives, MIS or CS for geodata. Opposite to commonly used search engines, MIS/CS are mostly theme-specific taking into account the multidimensional characteristics of geodata (title, abstract, keywords, reference date, spatial and temporal extent, coordinate reference system, lineage and quality information such as resolution code, scale or accuracy, distribution formats, fees, etc.). Usually, these MIS for geodata provide both thematic and geographic access variants in addition to traditional keyword and full-text searches. Göbel and Jasnoch [4] give an overview of existing MIS and CS for geodata and point out technical strengths and weaknesses of existing approaches.

Most important results of a comprehensive state-of-the-art analysis concerning metadata-based search and visualization techniques within the different information retrieval phases [5] are:

1. All analyzed approaches offer some kind of keyword search and geographic search. Additionally, some systems offer temporal search mechanisms. Thematic access is realized by categories and domain values corresponding to metadata formats and well-defined terminology used in the special geodata application area.
2. GUI components consist of text fields, attribute/keyword lists and sometimes maps which are implemented as click-able image maps or web-based GIS (geographic information systems).
3. Search results are presented as textual result lists. There are first approaches which show the relevance score of results, but in general the comparison and interpretation of result sets is hardly possible, because there is no visual feedback referring to search parameters and its effects to search results. Besides, this also negatively influences query modification.

---

<sup>2</sup> INVISIP: Information Visualization in Site Planning, IST-2000-29640, <http://www.invisip.de>

<sup>3</sup> Information and Cooperation Forum for Geodata, see <http://www.ingeoforum.de>

With regard to search result presentation, results of the INSYDER approach have shown that users need a support when expressing their information need and reviewing and refining their search results [10]. The notoriously low precision of web search engines coupled with the ranked list presentation make it hard for users to find the information they are looking for. Surveys have shown that users have problems with the current paradigm of information retrieval systems for Web search simply presenting a long list of results. These long lists of results are not very intuitive for finding the most relevant documents in the result set. Here, empirical findings motivated us to develop a new type of user interface for Web retrieval that supports the user in the information seeking process by providing special visualisations (e.g. result table, scatterplot, bargraphs, segment view) in addition to the traditional result list. For this a dynamic search system with an on-the-fly ranking, a visualisation of the query and a relevance feedback option are used in the INSYDER system. The current redesign of INSYDER includes a new component – the SuperTable (see section 4.2). In this approach a new data model is used to maintain an independency from the field of application. The system is no longer restricted to the visualization of WWW search results, but provides the possibility to represent for example medical or geographic metadata.

Based on these facts the following lacks of usability respectively derived requirements to improve current approaches become obvious:

Graphic-interactive components should offer easy thematic and geographic as well as application-driven access variants to interact with geospatial data archives. Concerning semantic aspects such as usage of promising keywords within the query formulation process, the latter access variant is very important.

Visual feedback should facilitate locating appropriate data and should be integrated into all different steps of the information retrieval process. Especially search result visualization should enable users to understand and interpret results. This includes transparent ranking, comparison of hits and hints referring to query modification.

The necessity for these components underlines the current action line “III.4.2 Information Visualization” by the EU which aims to facilitate data handling and visual data mining in (often unfamiliar) information spaces. Here, special attention is paid to visual feedback and metaphor concepts in order to support all phases of the information retrieval process. The INVISIP approach exactly addresses these key issues and obstacles of existing systems and intends to establish a framework to access geospatial data archives.

### **3 INVISIP Framework**

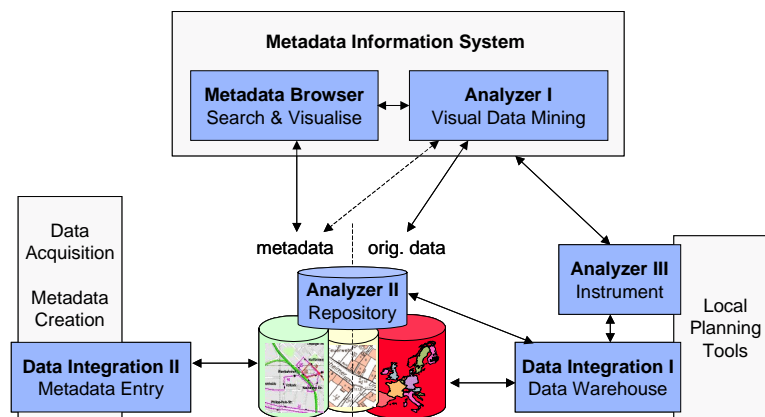
INVISIP provides a technical platform as an aid to facilitate information access and data handling for the site planning process (time-saving, intuitive analysis).

The core of the system is a metadata browser, which facilitates user interaction and helps planners to search for appropriate topic-related geographic data and to visualize search results in an intuitive, transparent way. Here, new mechanisms for

query formulation and query refinement as well as presentation respectively visualization of search results will be developed.

Secondly, a knowledge base –in the form of a repository- will contain several visual data mining techniques to analyze received data and will provide context info, which helps users to understand data and to recognize semantic relationships between various geographic data domains.

A data integration component based on web-based GIS components will be the basis for the dissemination of the project results and transfer of technology to other countries and application domains.



**Fig. 1.** INVISIP framework – proposed architecture: Components and Interfaces.

Figure 1 shows the proposed INVISIP framework (architecture/scenario) including components and interfaces. Simultaneously, these components represent the most important expected results:

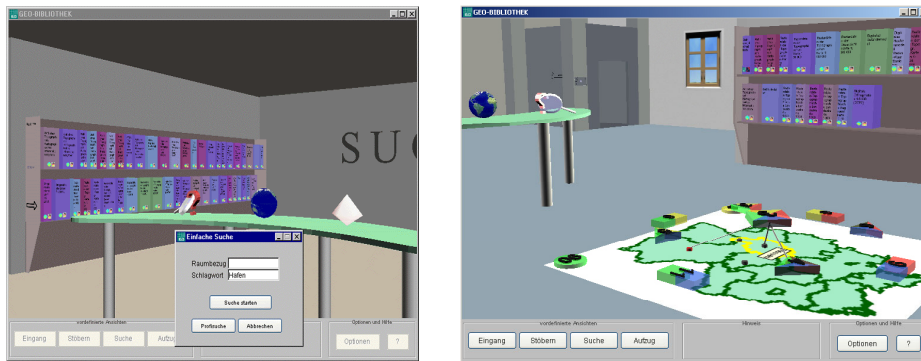
1. A *Metadata Browser* contains mechanisms and information visualization techniques to support the search for appropriate geodata within the site planning process. From the technical point of view, the metadata browser is the most important and innovative component of the INVISIP framework as graphical user interface between users and the “geodata world”. It is enhanced by information visualization techniques [1] and metaphors [12] to access geodata. Aside from metadata concepts (ISO standard [7], semantic network for geodata [3]) as information and navigation instrument to describe and locate appropriate geodata, information visualization techniques and metaphors are used to improve the usability of the metadata browser. Whereas information visualization techniques offer visual feedback in all steps of the data mining process, metaphors enable users to navigate in information spaces using intuitive spatial representations such as maps or well-known symbols settled in the library world like the book metaphor, information rooms or the complete library as 3D information and navigation space [14]. A visualization component facilitates browsing and comparing of geodata.
2. The interface to the (metadata-based) visual data mining component of the *Analyzer* enables users to navigate in unfamiliar information spaces and to determine

semantic relationships between necessary geodata. The main part of the analyzer builds the analysis instrument, which could be used as add-on for local planning tools. Thus, planners can use the instrument to contact a (regional, national or European wide) metadata information system and search for appropriate data. Further on, an analysis (context) repository will be established, which contains both metadata and original data, e.g. guidelines or laws, which have to be taken into account during the multi-step planning process.

3. The *Data Integration* component consists of two parts: Firstly, a data warehouse component that builds the interface between planning tools and original geodata. Similar to the analysis instrument it could be used as add-on for various planning tools already in use in different partner countries. Secondly, a Metadata Entry Tool will be established to enable geodata suppliers to describe their original data and to enter it into the infrastructure of INVISIP (MIS of INVISIP).

## 4 GeoLibrary

The GeoLibrary has been modeled in VRML and developed as 3D information and navigation space (using Java 3D) to improve orientation in the heterogeneous and multi-faceted geodata world. Different geodata disciplines and its terminology are used to structure the GeoLibrary. Analogue to a thematic classification of geodata application areas (environmental geodata, geo-basis and geo-scientific data and earth observation data) there are different floors: The first floor contains geo-basis and geo-scientific data, environmental data is located in the second floor and earth observation data is placed in the third floor below the roof. Users can poke around the GeoLibrary and graphic-interactively explore the geodata world and figure out semantic relationships between geodata of different application areas.



**Fig. 2.** GeoLibrary and GeoCrystal: Metadata based 3D information visualization technique to access geodata archives using metaphors – query formulation and search result presentation.

On the other hand the GeoLibrary offers an information desk which enables users to carry out “quick searches” similar to web-based search engines such as AltaVista or Yahoo. Thus it is not necessary for users to walk around in the library in order to

find individual geodata sets (presented by books) or to compare different datasets (search results). With regard to both search result presentation and the comparison phase of results a graphic-interactive information room is used as information kiosk. This information kiosk represents the central place of the GeoLibrary. Entering the GeoLibrary users directly walk to the information desk (in front of the information kiosk), enter their tasks and goals and decide how to use the library. Different symbols on the information desk indicate various access variants:

1. A magnifying-glass initiates a search mode offering different easy (keyword search, spatial search by geographic names) and complex (based on metadata formats) search forms for casual and advanced users.
2. The globe symbolizes the “geodata world” and represents an entry point to dive into the wide range of application disciplines. Thus users can follow the different navigational hints and terms listed on blackboards in front of the elevator and bookshelves on the three floors of the library.
3. Triggering the rotating Crystal (symbolized as information pyramid) switches on GeoCrystal and fades in a map on the ground of the information room and the GeoCrystal elements in the 3D space above that map.
4. Additional thematic and application-driven access modes are symbolized by a tree-view diagram.

#### **4.1 GeoCrystal**

GeoCrystal [2] has been developed as new information visualization technique combining algebraic mechanisms (InfoCrystal [15]) with library (book, bookshelves, floors, etc.) and spatial metaphors (2D maps respectively 3D spatial information landscapes). The global aim of GeoCrystal is to improve understanding of search results by offering familiar metaphors settled in the field of set theory (algebra) combined with the provision of illustrated visual parameters such as spatial references of result sets. Hence the structure of GeoCrystal consists of two parts:

On the ground of the information room there is a vector-based map realized as GIS component, which both enables users to specify spatial queries and supports spatial orientation and navigation.

Above the map the result sets of a query are visually presented by block symbols indicating GeoCrystal elements, whereby the result sets are classified by the number of matching parameters corresponding to the query.

Result sets matching all query parameters are located in the center of the map at a high position (directly in the viewpoint of the user). Other result set classes (matching some, but not all query parameters) are positioned on spherical levels between the top of GeoCrystal and the base-map. Similar to InfoCrystal each query parameter is mapped to a color and represented by a circle at the corners of the GeoCrystal (on ground level closely placed to the interactive map). Result set classes are visually mapped to other symbols such as rectangular (matching two parameters), triangular (matching three parameters) or square blocks (matching four parameters) and are colored according to the colors of matching parameters. Additionally, each element of the GeoCrystal contains a number of matching result sets. Hence, users easily get an

idea about the numerous distribution of results referring to special query parameters and how to modify the query in order to get better results.

With regard to the detailed analysis and comparison of result sets, users can interact with the elements of GeoCrystal and initiate several events. These events result in the provision of detailed textual and graphical presentations of the result sets, which are placed on two boards at the front side of the information room as well as on bookshelves placed at the over walls. Initial query search results are visualized as result lists indicating a check box, the global rank, title (with a hyperlink to the full metadata set) and iconic presentations of the metadata format used to describe the original dataset and the location of the geodata server respectively geospatial data archive of the matching dataset as well as links to contextual tooltips and to GeoCrystal. Activating the hyperlink below the title offers users the full meta dataset presented as HTML or XML-based file on the left board. Activating the GeoCrystal symbol initiates highlighting of the GeoCrystal element (class) containing the selected dataset. In order to compare several matching hits, user can select the checkboxes in first column of the result list and activate a “comparison button”. Thus, another visualization is dynamically created offering detailed ranks such as a spatial rank, a thematic rank or a temporal rank to visualize the precision of individual search parameters (distance between a value of the geodata attribute/metadata element and the query condition of a search parameter). At this, GeoCrystal provides visualizations such as tile bars, scatter plots or proportional symbols.

Further on, picking of GeoCrystal elements in the 3D space causes the creation of an interactive list showing all datasets of that result class (e.g. datasets matching all query parameters at the top element of GeoCrystal). Then the selection of an individual dataset causes the creation of relations between a result, the map and the dataset itself represented as book on bookshelves:

The spatial extent of a dataset is visualized as bounding box (west, north, south and east bounding coordinates) or highlighted area on the map.

The locations of the geodata archive containing that dataset and other responsible parties involved in that dataset (place of the geodata originator, supplier, content provider, distributor, metadata provider, etc.) are visualized as point-features (filled circles) on different layers on the map.

Optionally those points and highlighted areas are connected by pointers with the 3D elements of GeoCrystal.

## 4.2 SuperTable

The main goal of the SuperTable [9] is the Integration of Table Lens as seen in [11], ResultTable, HTML-List, BarGraph, and SegmentView into one component. There would be different predefined zoom-levels. Starting for example with a BarGraph view in a first zoom level, revealing document details in a one-line-per-document-mode for the whole table in a second level, a several-line-per-document-mode in a third level, and document information revealing segment details in a fourth level. In addition, tooltips or lens mechanisms can allow easy inspection of details for single documents. Power users can use multiple focus possibilities for comparisons. The

SuperTable as an integrated component will be more complex than the former single components. With adequate design a subject who is starting to use the system may not be aware of this complexity, and therefore would not be distracted. The integration will minimize context-switching effort and is able to allow a smooth learning curve from beginner mode to power user mode.

The redesign of the INSYDER system combines the two approaches of SuperTable and ScatterPlot into a single window. *Level 1* represents an overview over all documents. All rows are as small as possible, so that in the best case all documents fit on the available space. Corresponding to the number of documents, the height of the rows can vary. Usually the rows will be too small to hold text, so only bars will be displayed. The length and the position of the bars encode various characteristics of the document, depending on the type of data they represent. The length of the bars (representing numeric data like size, relevance, ...) are equivalent to their numeric values. Nominal attributes can also be represented by a bar. For a few attributes we can code their values through position; for example the language, which can be english or french in our application (the left half of the cell means "english" and the right one means "french"). If there are too many different nominal values a visualization would be too confusing, which is for example the case with the title. Therefore this column will be empty. Different colors can additionally encode different search terms.

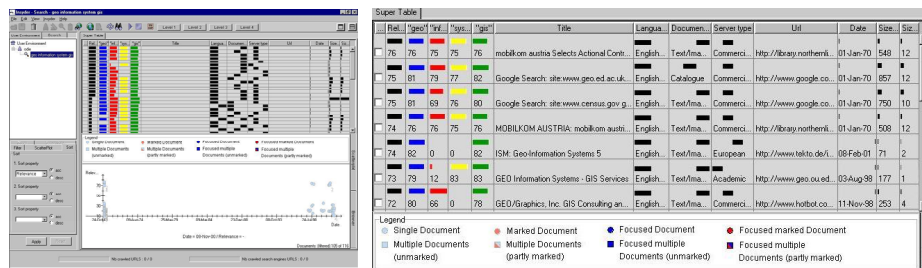


Fig.3. SuperTable (left: *Level 1*, right: *Level 2* zoomed in).

In *Level 2* more information will be visible in form of text, completing the visual representation of the multicolored bars. Now numeric values add detailed information about the bar displays from the initial, graphics-only display. Title and URL are now readable, but only upto the width of the respective column. All wider texts become truncated, clarified by three dots. *Level 3* provides the opportunity to read the whole text of those characteristics, which had to be abbreviated because of their size. Visualizations were cut off to gain space, so that title, url and abstract are now completely visible. In addition a new column is introduced, the so called "Relevance Curve". It represents a two-dimensional chart of the whole document, by dividing the document into a number of segments, e. g. sentences, subordinate clauses, etc. The height of single bars encodes the overall relevance for each individual segment. *Level 4* (Figure 4, right) displays only the (in our opinion) most important values of a document: Title, abstract and an extension of the relevance curve, the so called "Segment View", which uses tile bars as seen in [6]. In this case, every segment has the same length,



not varying from one document to another. Consequentially the length of the Segment View differs from document to document, always corresponding to the real length of the document. The Tilebars are colored according to the colored search terms in the former levels. So it is easier to discover the segments, where all search terms can be found, not just a few of them. You can spot segments, that include only one or two search terms, and discern these from parts including all terms. Some terms may be seen more important than others, so a segment with a lower relevance can be important although not all terms are included.

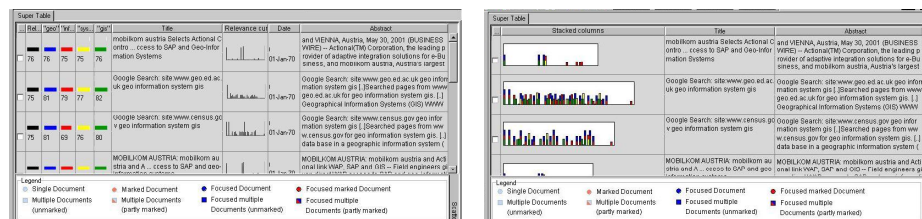


Fig. 4. SuperTable (left: *Level 3* zoomed in, right: *Level 4* zoomed in).

## 5 Conclusion

This paper describes the INVISIP approach which aims to support users in the data mining process of complex application areas such as site planning and to facilitate graphic-interactive access to geodata archives. Here, the idea is to provide metadata based information visualization techniques in the different information retrieval phases to locate appropriate geodata which is necessary to solve planning tasks, e.g. generation of ecological, environmental or socio-demographic reports.

GeoCrystal and SuperTable are introduced as new information visualization techniques to support users during the different information retrieval phases, especially search result presentation and comparison of search results (hints for query modification). Currently (May 2002) both techniques are integrated into a GeoLibrary as 3D information and navigation space using familiar metaphors such as books, bookshelves, an information desk or spatial landscapes (maps) within the first development phase of the INVISIP project.

From the application-oriented point of view, future work will be investigated to extend the GeoLibrary by application-driven access variants. From the technical point of view further research work will consider different technologies in the field of artificial intelligence and knowledge management to improve the INVISIP MIS as basis for the GeoLibrary. Much effort is spent on the visual support within all information retrieval phases: query formulation, search result presentation and query refinement. Especially concerning the query formulation innovative concepts are developed taking into account not only information visualization techniques but also the ISO 19115 standard as structured format to describe geographic data. Besides, further informa-

tion visualization techniques could be integrated within the different information retrieval phases to facilitate data mining (of original geodata) in general.

## References

1. Card, S.K., Mackinlay, J.D., Shneiderman, B. Readings in Information Visualization. Using Vision to Think. Morgan Kaufmann Publishers, Inc, San Francisco, CA, 1999.
2. Göbel, S., Haist, J., Goebel, C. GeoCrystal : Graphic-Interactive Access to Geodata Archives. Proceedings SPIE 2002 – Visualization and Data Analysis, San Jose, CA, 2002.
3. Göbel, S., Jasnoch, U. Development of a semantic network for environmental geodata. Published in Sustainability in the Information Society, 15<sup>th</sup> International Symposium Informatics for Environmental Protection, Zurich, 2001, edited by Lorenz M. Hilty, Paul W. Gilgen. Metropolis Verlag, Marburg, 2001, pp. 679-684. ISBN 3-89518-370-9.
4. Göbel, S., Jasnoch, U. Metadata Information Systems for Geospatial Data, S.283-296 in Environmental Information Systems in Industry and Public Administration, edited by Claus Rautenstrauch, Susanne Patig. Idea Group Publishing, Hershey, USA, 2001. ISBN 1-930708-02-5.
5. Göbel, S., Jasnoch, U. Visualization Techniques in Metadata Information Systems for Geospatial Data. *Proc. ISESS'2000*. Zell a. See, Austria, 2000.
6. Hearst, M. TileBars: Visualization of Term Distribution Information in Full Text Information Access. *Proc. ACM SIGCHI'95*, pp. 59-66, Denver, CO, 1995.
7. ISO. Draft International Standard on Metadata for Geographic Information, ISO 19115-Metadata (status: DIS level), see <http://www.statkart.no/isotc211/>.
8. Jasnoch, U., Göbel, S., Balfanz, D. InGeo IC: The portal to Geodata. Proceedings e2000, Madrid, Spain. Published in Stanford-Smith, Kidd: E-business. Key Issues, Applications, Technologies. IOS Press, Amsterdam, 2000.
9. Klein, P., Müller, F., Reiterer, H., Eibl, M.: Visual Information Retrieval with the SuperTable + Scatterplot, Accepted Paper at the 6<sup>th</sup> International Conference Information Visualization, London, England, 2002.
10. Mann, T.H.: Visualization of Search Results from the World Wide Web, PhD Thesis, University of Konstanz, 2002.
11. Rao, Ramana; Card, Stuart K.: The Table Lens. Merging graphical and symbolic representations in an interactive focus + context visualization for tabular information. In: Adelson, B.; Dumais, S.; Olson, J. S. (Eds.): CHI 1994: Conference Proceedings Human Factors in Computing Systems. Conference: Boston, MA, April 24-28 1994. New York (ACM Press) 1994. p. 318-322. 1994
12. Rauber, A., Bina, H. A Metaphor Graphics based Representation of Digital Libraries on the World Wide Web: Using the libViewer to Make Metadata Visible. *Proc. DEXA '99*, Florence, Italy, 1999.
13. Reiterer, H., Mußler, G., Mann, T., Handschuh, S.: *INSYDER – An Information Assistant for Business Intelligence*, Proceedings of the 23 Annual International ACM SIGIR 2000 Conference on Research and Development in Information Retrieval, ACM press, 2000, pp.112-119.
14. Robertson, B. biz viz gets real. *Computer Graphics World*, April 1999, S. 29-34.
15. Spoerri, A. InfoCrystal: A Visual Tool for Information Retrieval. *Proc. IEEE Visualization '93*, San Jose, CA, 1993, pp. 150-157.