

Visual Information Retrieval with the SuperTable + Scatterplot

Peter Klein¹, Frank Müller¹, Harald Reiterer¹, and Maximilian Eibl²

¹ *University of Konstanz, Department of Computer & Information Science,
Universitätsstr. 10, 78457 Konstanz, Germany*

{peter.klein, frank.mueller, harald.reiterer}@uni-konstanz.de

² *IZ Social Science Information Centre, Schiffbauerdamm 19, 10117 Berlin, Germany
eibl@berlin.iz-soz.de*

Abstract

We present a new visualization approach for metadata combining different visualizations into a so-called SuperTable accompanied by a Scatterplot. The goal is to improve user experience during the information seeking process. Our new visualizations are based on our experiences developing a visual information retrieval system called INSYDER to supply small and medium size enterprises with business information from the Internet. Based on extensive user tests the original visualizations have been redesigned in two different design variants. Instead of offering multiple visualizations to choose from the SuperTable + Scatterplot combines them in a new way. Therefore, the user has the feeling that he is working with one single visualization in different states. Further the SuperTable solves a problem which seemed to be imminent to visualizations in document retrieval: the change of modalities.

1. Introduction

Conventional document retrieval systems return long lists of ranked documents that users are forced to sift through to find relevant documents. The majority of today's web search engines follow this paradigm. 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. These 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 visualizations in addition to the traditional result list. Systems combining the functionality of retrieval systems with the possibilities of information visualization systems [2] are called *visual information retrieval systems*.

This paper presents our main design ideas developing such a visual information retrieval system. Section 2 summarizes our experiences developing and evaluating the first version of our visual information retrieval system called INSYDER. Based on a comprehensive empirical user test with 40 users we have made a redesign of the

original INSYDER visualizations. Section 3 presents in detail the redesigned visualizations, the so-called SuperTable supplemented by a Scatterplot. We are currently developing two different versions of the SuperTable using different strategies showing the granularity of details (levels of details versus a smooth change of granularity approach). Section 4 discusses related work that has influenced our visualizations. Conclusions and an outlook are given in section 5.

2. A Visual Information Retrieval System for the Web

The first implementation of the INSYDER¹ system includes five visualizations for the presentation of search results [8, 11, 12]: a traditional list (mainly for evaluation purposes), a ResultTable, a ScatterPlot, a BarGraph, and a SegmentView with two modes: TileBars and StackedColumns. The primary intention for the use of different visualizations was to present additional information (metadata) about the retrieved documents to the user in a way that is intuitive, may be quickly interpreted, and can scale to large document sets.

An extensive evaluation done with 40 users [9] has been focused on the different visualizations used to present the search results in the result phase of the search process. The primary goal of this summative evaluation was to determine the usability of the visualizations. A second goal was to detect problems with the visualizations used in the INSYDER system, and to collect suggestions for improvements. The usability evaluation part of the study was focused on the added value of the visualizations (ScatterPlot, BarGraph, TileBar, StackedColumn) in terms of their effectiveness (accuracy and completeness with which users achieve task goals), efficiency (the task time users took to achieve task goals), and subjective satisfaction (positive attitudes towards the use of the visualization) for reviewing Web search results.

¹ The project was funded by the European Commission under the Fourth Framework of the ESPRIT Program, Project No. 29232. www.insyder.com

The evaluation results indicated some difficulties of user interaction with the system, e.g., more than 50% of the users voted for the ResultTable when asked with which visualization they performed best. Other visualizations were helpful as an addition to the ResultTable, but not as primary tools. When studying the expected value of the visualizations, it can be said that in the visualization (e.g. Scatterplot, BarGraph, SegmentView) plus ResultTable conditions where the user had the possibility to decide which component to use both components were used in the majority of cases. When analyzing usage times in these conditions, the ResultTable was the favorite component of the users. It was used in all three user interface test conditions with ScatterPlot, BarGraph, and SegmentView for more than 50% of the overall task time. Interpreting usage time as an indicator of expected value, the expected value of the ResultTable seemed to be higher than that of the other components for the users. Switching between completely different visualizations confused the users. Therefore, we tried to find a possibility to combine the regular table view with other views like the BarGraph or the SegmentView.

3. The SuperTable + Scatterplot

Based on the empirical findings of the evaluation we have decided to integrate the ResultTable, BarGraph, and SegmentView into one visualization called „SuperTable” and to improve the ScatterPlot.

In the user test the users requested a number of features for the BarGraph and the SegmentView already implemented in the ResultTable. All this could also be implemented in the ResultTable. Therefore, the proposed *SuperTable* integrates the concept of a distortion-based table, the BarGraph and the SegmentView (with TileBars and StackedColumns) in a way that allows easy manipulation of the table.

The redesign of the INSYDER visualizations combines the SuperTable + ScatterPlot into one single window offering different brushing techniques between them. Therefore the *ScatterPlot* will supplement the SuperTable by giving the user a quick overview of all search results, and offering the user a variety of controls (e.g. defining own views, zooming, selecting, filtering) to reduce the amount of hits to a smaller group of interesting documents. These documents can then be selected by the user and analyzed in more detail in the SuperTable.

The enhanced Scatterplot with additional lens mechanisms (e.g., a magic lens for filtering operations [4]), distortion techniques in both X- and Y-dimensions (allowing a smooth transition between focus and context) and the radial MDP visualization tightly coupled with the SuperTable and a document browser (showing the detailed document with keyword highlighting) are our main redesign ideas.

3.1 Discrete Change of Modalities: the Level Concept

The first design variant of the SuperTable with the *level concept*, which is combined with the Scatterplot and a Browser window on a single panel is shown in Figure 1.

At first sight, the SuperTable doesn't differ from the original ResultTable: Every row represents the metadata of a web search result - in that case a html-page. The columns describe all characteristics concerning this document, e.g. title, url, date, size and so on. The main feature of the initial view is that no text at all will be seen in the table. There are just a number of multicolored bars representing numeric and textual data.

The distinction between the origin ResultTable and the SuperTable consists of four different views representing the "level of detail" for documents. We call this method of looking closer and closer at the details of a document the "Focus of Interest". The more you want to know about a document, the deeper you have to look, i.e. the higher the level has to be. It is possible to focus single documents, several documents or the whole list of documents.

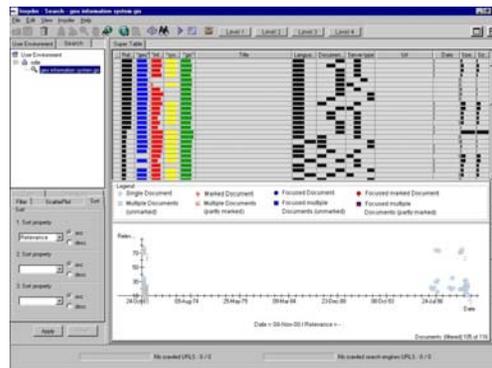


Figure 1: SuperTable Level 1

Level 1 represents an overview over all documents (Figure 1). 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) is 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 may be English or French in our application (the left half of the cell means "English" and the right one means "French"). If there were too many different nominal values, 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.

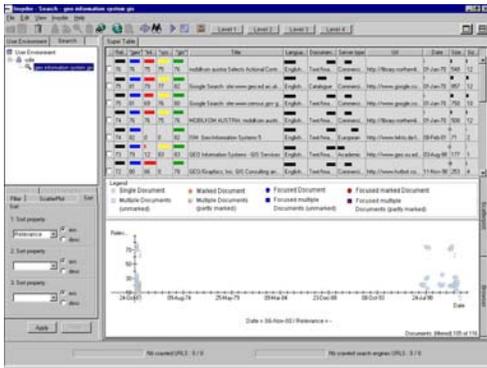


Figure 2: SuperTable Level 2

On Level 2 (Figure 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 up to the width of the respective column. All wider texts become truncated, marked by three dots.

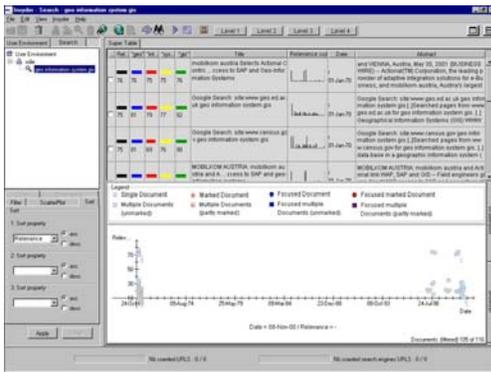


Figure 3: SuperTable Level 3

Level 3 provides the opportunity to read the whole text of those characteristics, which had to be abbreviated because of their size (Figure 3). 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) displays only the most important values of a document. The title, an abstract and an extension of the relevance curve, the so-called Segment View which uses stacked columns or TileBars [6]. Which kind of query term distribution visualization will be used de-

pends on the user preferences. In this case, every segment has the same length, not varying from one document to another. Consequentially the length of the SegmentView differs from document to document, always corresponding to the real length of the document. The stacked columns or 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 which 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. Per default, instead of the Scatterplot a preview browser is offered showing the selected document or per default the first document of the SuperTable. This offers the user the greatest level of detail showing him the text of the document (without images) enhanced by keyword highlighting.

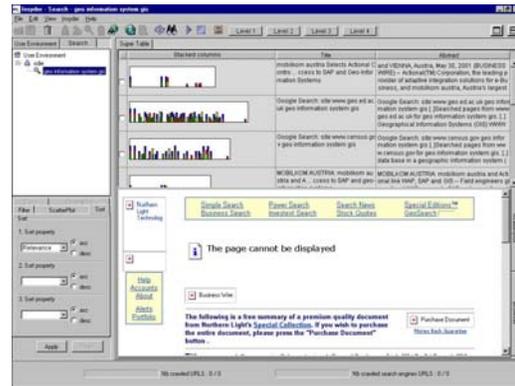


Figure 4: SuperTable Level 4

3.2. Smooth Change of Modalities: the Granularity Concept

Visualizations in document retrieval have one common flaw: the change of modalities during a search. The query is formulated in a textual way, for example by a simple list of search terms, a Boolean combination of search terms, a form, a SQL-statement and so on. Only a few systems allow a mere graphical input. In a next step a graphical version of the result set is presented. Here the user can see the effects of his query on the result set. Some visualization even allows manipulating the result set. Finally, the documents themselves are presented textually, usually in a selectable list of titles.

Thus, the user has to switch at least twice between modalities. The first time in order to get from his query formulation to the visualization and the second time in order to get from the visualization to the actual documents. Nevertheless, in practice, he will use iterative retrieval

improve the advantages of this combination in a more detailed way.

After the implementation of both versions is finished², we will perform an evaluation (user tests) to compare these two redesign ideas and then we will decide, which approach will be preferable. Additionally, a highly sophisticated data model will enable us to adapt INSYDER to a wide range of application domains where metadata play an important role (e.g. digital libraries, web, geodata archives). Therefore, INSYDER should become a visualization framework for different application domains visualizing metadata.

We are currently adapting our visualization framework to a new application domain in a project called INVISIP (Information Visualization in Site Planning)³. The aim of this project is to support users in the retrieval process of complex application areas such as site planning and to facilitate graphic-interactive access to geodata archives. Here, the idea is to provide 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. The SuperTable + Scatterplot will be introduced in a 3D GeoLibrary [5] as one new information visualization technique to support users during the different information retrieval phases, especially search result presentation and comparison of search results (hints for query modification).

6. References

- [1] Ahlberg, Christopher; Shneiderman, Ben: *Visual Information Seeking: Tight Coupling of Dynamic Query Filters with Starfield Displays*. 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. 313-317.
- [2] Card, Stuart; Mackinley, Jock; Shneiderman, Ben (Eds.) (1999). *Readings in Information Visualization. Using Vision to Think*. San Francisco.
- [3] Eibl Maximilian: *Visualisierung im Document Retrieval*. Forschungsberichte Band 3, IZ Informationszentrum Sozialwissenschaften, Bonn, 2000
- [4] Fishkin, Ken; Stone, Maureen C.: *Enhanced Dynamic Queries via Movable Filters*. In: Katz, Irvin R.; Mack, Robert L.; Marks, Linn et al. (Eds.): *CHI 1995: Conference Proceedings Human Factors in Computing Systems. Conference: Denver, CO, May 7-11 1995*. New York (ACM Press) 1995. p. 23-29.
- [5] Göbel, S., Haist, J., Goebel, C. *GeoCrystal: Graphic-Interactive Access to Geodata Archives*. Proceedings SPIE 2002 – Visualization and Data Analysis, San Jose, CA, 2002.
- [6] Hearst, Marti A.: *TileBars: Visualization of Term Distribution Information in Full Text Information Access*. In: Katz, Irvin R.; Mack, Robert L.; Marks, Linn et al. (Eds.): *CHI 1995: Conference Proceedings Human Factors in Computing Systems. Conference: Denver, CO, May 7-11 1995*. New York (ACM Press) 1995. p. 59-66.
- [7] Nowell, Lucy T.; France, Robert K.; Hix, Deborah et al.: *Visualizing Search Results: Some Alternatives to Query-Document Similarity*. In: Frei, Hans-Peter; Harman, Donna K.; Schäuble, Peter et al. (Eds.): *SIGIR 1996: Proceedings of the 19th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. Conference: Zürich, Switzerland, August 18 -22 1996*. New York (ACM Press) 1996. p. 67-75.
- [8] Mann, T.H., Reiterer H.: *A Combined Visualization Approach for WWW-Search Results* in: Proceedings of the IEEE Visualization 1999 (Vis '99), San Francisco
- [9] Mann, Thomas M: *Visualization of Search Results from the World Wide Web*, University of Konstanz, 2002, <http://www.ub.uni-konstanz.de/kops/volltexte/2002/751/>
- [10] 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.
- [11] Reiterer, Harald; Mußler, Gabriela; Mann, Thomas; Handschuh, Siegfried: *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
- [12] Reiterer, Harald; Mußler, Gabriela; Mann, Thomas M.: *Visual Information Retrieval for the WWW*, in: Smith M.J. et al. (eds.), Proc. HCI International 2001, New Orleans, Lawrence Erlbaum, 2001, pp. 1150-1154
- [13] Spence, Michael; Beilken, Christian; Berlage, Thomas: *FOCUS: The Interactive Table for Product Comparison and Selection*. In: *UIST 96: 9th ACM Symposium on User Interface Software and Technology* New York (ACM Press) 1996. p. 41-50.
- [14] Spoerri, Anselm (1994a). *InfoCrystal: Integrating Exact and Partial Matching Approaches through Visualization*. In: RIAO'94 Proceedings, Oct.11-13, 1994 New York (NY), S.687-696.
- [15] Veerasamy, Aravindan; Navathe, Shamkant B.: *Querying, Navigating and Visualizing a Digital Library Catalog*. In: *Digital Libraries 1995: The Second Annual Conference on the Theory and Practice of Digital Libraries. Conference: Austin, TX, June 11-13 1995*.

² We are currently implementing the redesign variants in Java using the shown mockups (html- and paper-prototypes) as starting points.

³ The project is funded by the European Commission under the Fifth Framework of the IST Program, Project No. IST-2000-29640, www.invisip.de