

Visual Information Retrieval for the Web

Harald Reiterer, Thomas M. Mann

Computer and Information Science, University of Konstanz, D-78457 Konstanz, Germany
Harald.Reiterer@uni-konstanz.de, Thomas.Mann@uni-konstanz.de

Abstract

This paper presents a visual information seeking system for the Web called INSYDER. The aim of INSYDER is to find business information from the Web. Different visualizations supporting the reviewing of search result are shown. The main design ideas and results of an evaluation with 40 users are presented.

1. Introduction

The number of documents published on the Web has been increasing dramatically. This brought the research about information retrieval systems into the focus of people, dealing with the Web. The main reason is the need for efficient tools to manage, retrieve, and filter information from this ubiquitous database. The main challenges posed by the Web can be divided in *problems with the data* itself and problems related to the *user and his interaction with the retrieval system*. This paper focuses on the problems regarding the user and his interaction with the retrieval system. There are basically two problems: *how to specify a query and how to interpret the answer provided by the system*.

To overcome these problems this paper presents a variety of visualizations supporting different parts of the information seeking process. To show the practical relevance of our ideas, we have developed and implemented a visual information seeking system for the Web called INSYDER¹. Chapter 2 introduces the main characteristics of INSYDER. Chapter 3 describes the different visual structures used for visualizing the search results. Chapter 4 presents evaluation results. Chapter 5 gives an outlook on future work.

2. INSYDER

It was not our intention during the development of the INSYDER system to realize new visual metaphors supporting the retrieval process. The idea was to select existing visualizations for text documents and to combine them in a novel way. Nowadays there are a lot of visualizations of search results in document retrieval systems available (a good overview can be found in Zamir 1998). Our selection of existing visualizations was based on the assumption to find expressive visualizations keeping in mind the target users (business analysts), their typical tasks (to find business data in the Web), their technical environment (typical desktop PC and not a high end workstation for extraordinary graphic representations), and the type of data to be visualized (text documents). The idea was to visualize additional information about the retrieved documents to the user in a way that is intuitive, fast to interpret and can scale to large document sets.

Another important difference of our INSYDER system with existing retrieval systems for the Web is the comprehensive visual support of different steps of the information seeking process. The visual views used in INSYDER supports the interaction of the user with the system during the formulation of the query (e.g. visualization of related terms of the query terms with a graph), during the review of the search results (e.g. visualization of different document attributes like date, size, relevance of the

¹ The research project INSYDER (INternet SYstème DE Recherche) was funded by a grant from the European Union, ESPRIT project number 29232.

document set with a scatter plot or visualization of the distribution of the relevance of the query terms inside a document with a TileBar), and during the refinement of the query (e.g. visualization of new query terms based on a relevance feedback inside the graph representing the query terms). For a detailed description of the visualizations supporting the query formulation and query refinement see (Mußler et al. 2000). This paper focuses on the visualizations used for the review of the search results.

The *multiple view approach* in combination with a *comprehensive visual support* of most of the information seeking activities is one of the strengths of the INSYDER system and is quite novel in the field of search systems for the Web.

3. Visualizations of the INSYDER System

Systems combining the functionality of retrieval systems with the possibilities of information visualization systems are called *visual information seeking systems*. The development of INSYDER was strongly influenced by principles for visual information seeking applications (Ahlberg et al. 1994). For the following discussion of the design decisions during the development of INSYDER we use an extended version of the reference model for visualization shown in Figure 1 (Card et al. 1999).

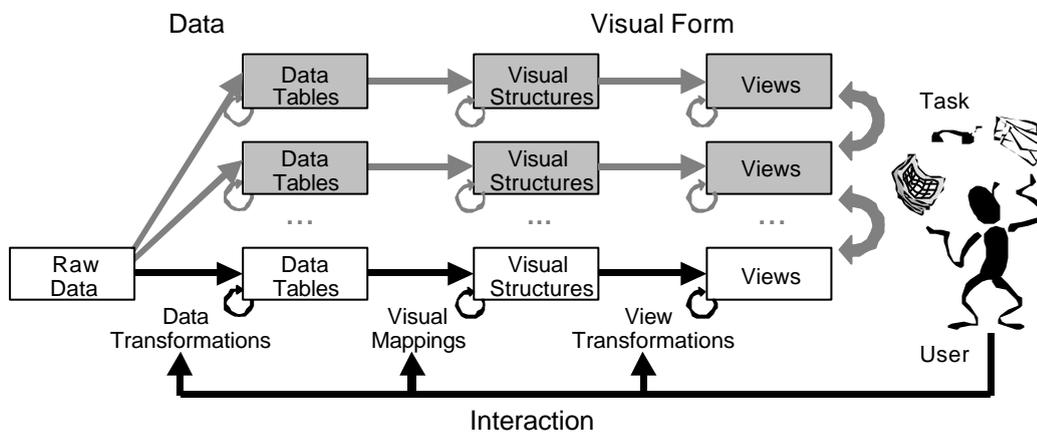


Figure 1: Reference model for visualization, extended for Multiple Coordinated Views

After the visual supported formulation of the query Web documents (our **raw data**) are analyzed, classified, and ranked with the help of different retrieval agents. Our first decision was to use a multi-agent-based metasearcher architecture (Reiterer et al. 2000). The results of these agents are the retrieved, classified, and ranked search results of the query.

The next design decision was to **transform** and save all search results and their characteristics in a local repository (RDBMS) with a specific data schema. The data schema for each document is described in **data tables** and represents additional information about the retrieved documents. There are two categories of additional information that could be visualized: visualization of *document attributes*, and visualization of *inter-document similarities*. Our approach is mainly focused to the visualization of document attributes. One important effect is the possibility to group documents that share similar attributes together. We have used two different kind of document attributes for the visualization: visualizations (Scatterplot, Table, List) that show how the retrieved documents relate to *predefined document attributes* (e.g. title, relevance, date, size, document type, server type), and visualizations (Bargraph, TileBar, Stacked Column) that show how the retrieved documents relate to each of the terms used in the query (*query terms` distribution*).

The next step in the development process was to find **visual mappings** of the data tables to good visual structures. The default presentation of the search results is a so-called *Result Table*. The main difference to the traditional result list of search engines is that all available attributes of each document are shown in different columns of the table. Each row shows one document. The user has the possibility to sort each document attribute in an increasing or decreasing order or to customise the table to his personal preferences (e.g. to show only the attributes he is interested in or to rearrange the order of the columns). Figure 2 shows an example of the Result Table for the query *visual data mining* and its results. On the left there are the different Spheres of Interest (SOI). Each SOI represents an information need of the user. The user might have a sphere-of-interest called "Competitors", one "Development tools", "Technology" and maybe one "Recruitment". In each of these spheres the user can define searches, watches and portals. A typical example for a search in the competitors sphere would be to look for new competitors, while a typical watch activity would be to monitor the Web-Site of a distinctive competitor. A web portal presents interesting links to different competitors WWW-sites. The advantage of SOIs is in accordance to the user's information needs in a structured way. This will help the user to navigate between and keep different information interests at the same time, easily. On the right of the window shown in Figure 2 the user sees the Result Table and an integrated browser which shows a preview of the actually selected document. The visual mappings of Web documents we have chosen are text in 1D (Bargraph, TileBar, Stacked Column) and text in 2D (Scatterplot). This final selection of the visual structures was based on suggestions of a field study we made with potential users, an extensive study of the state of the art in visualizing text documents (Mann 1999), and the design goal, to orient our visual structures as far as possible on typical business graphics. Figures 3 to 5 show the different visual structures for the example query *visual data mining*.

Another important design decision was to use a **multiple view** approach offering the user the possibility to choose the most appropriate visualization view for his current demand or individual preferences.

In all different views we have made extensive use of different **interaction** techniques (e.g. direct manipulation, details-on-demand, zooming, dynamic queries, sorting) to give the user control over the mapping of data to visual structures.

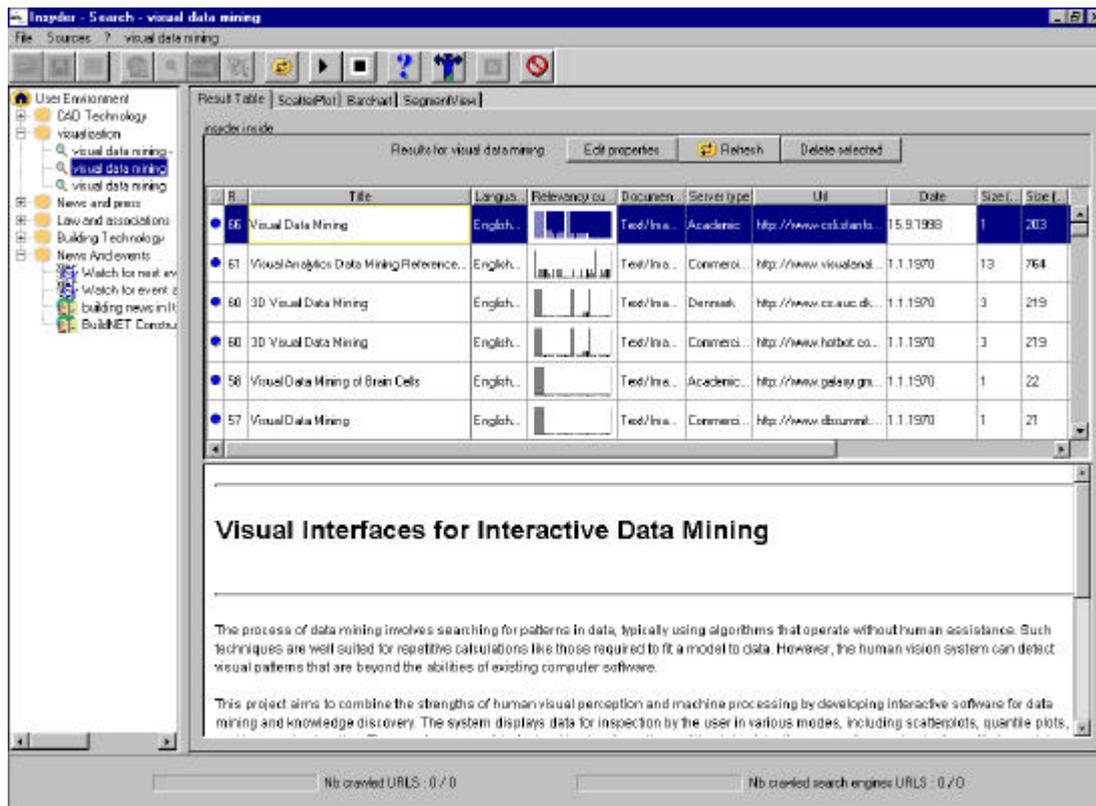


Figure 2: Result Table

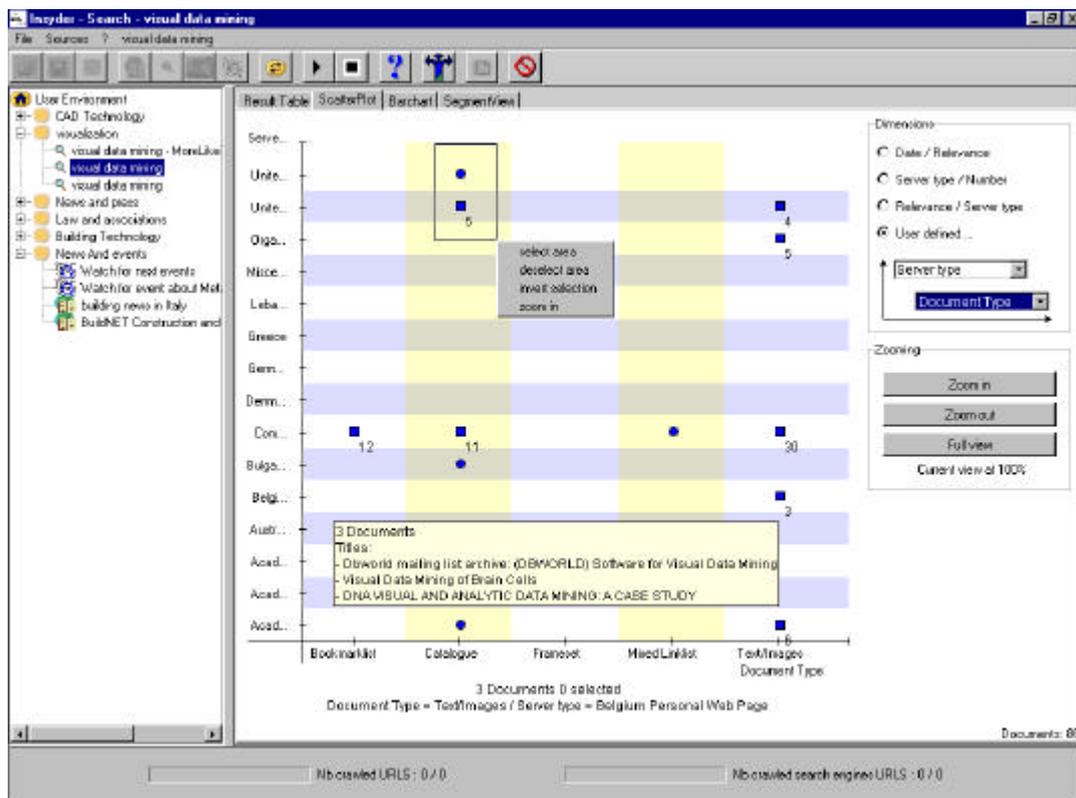


Figure 3: Scatterplot

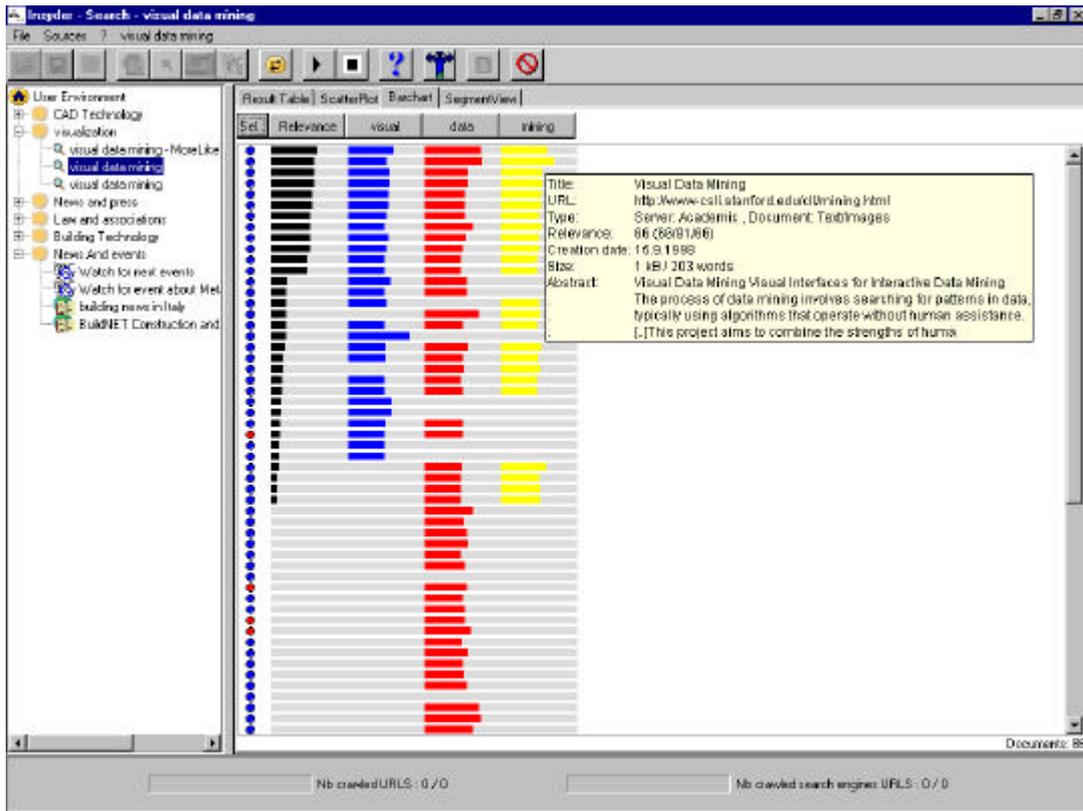


Figure 4: Barchart

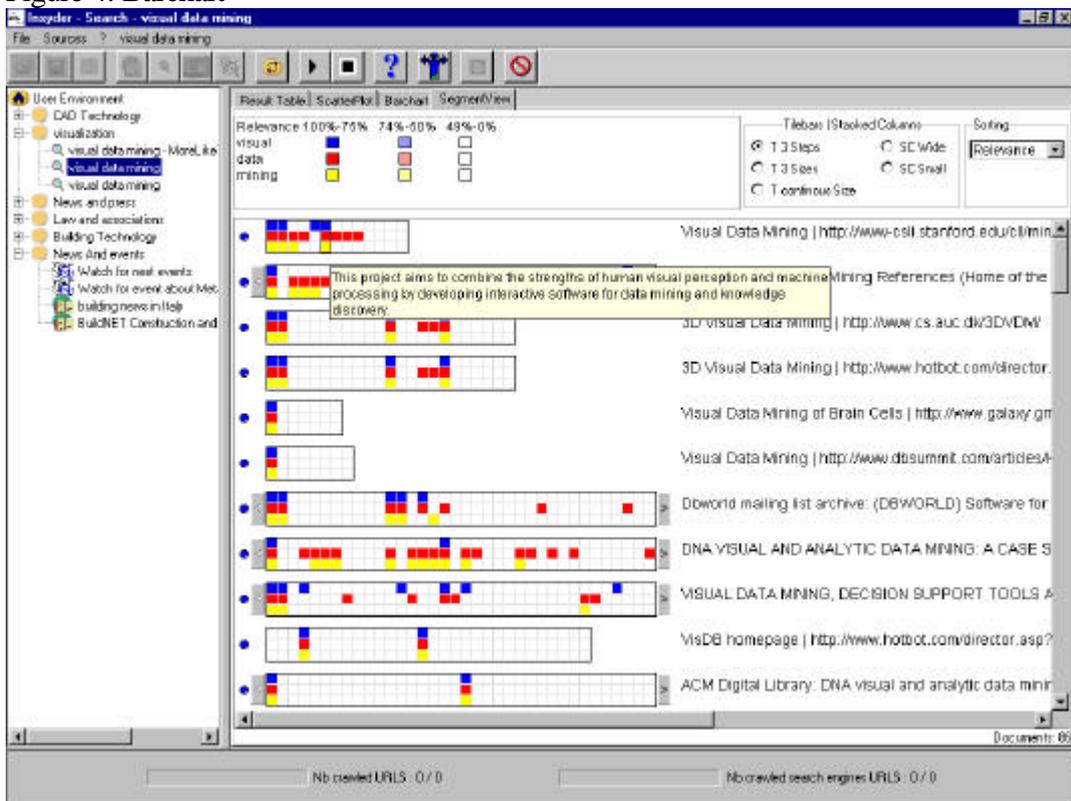


Figure 5: Tilebar

4. Evaluation

During the project three user tests with users from small and medium size enterprises were conducted to test the overall usability of the user interface. After these formative evaluations we have made a summative evaluation at the University of Konstanz. The primary goal of this summative evaluation was to measure the added value of our visualizations in terms (dependent variables) of **effectiveness** (accuracy and completeness with which users achieve task goals), **efficiency** (the task time users expended to achieve task goals), and **subjective satisfaction** (positive attitudes to the use of the visualization) for reviewing Web search results. Knowing advantages of the multiple view approach documented in user studies (North et al. 1999), we didn't intend to measure the effects of having Scatterplot, Bargraph and TileBar/Stacked Column (also called SegmentView) *instead* of the List and Table. We wanted to see the added value of having these visualizations *in addition* to the Table and List.

The following independent variables had been varied: Different configurations of the **user interface**: List only, Table only, Scatterplot + Table, Bargraph + Table, SegmentView + Table. From the five factors influencing the success visual structures of the user interface (Mann, Reiterer 2000) we decided to vary target user group, type and number of data, and task to be done. The fourth and fifth factors, the **technical environment** and the **training** was identical for all tests done. Probable training effects caused from the familiarity with certain forms of presentation are discussed latter. To see possible influences caused by the **target user group** we used two different groups of users: "beginners" and "experts". To see possible influences caused by the **type and number of data**, we used queries with three different numbers of keywords and two different sizes of result sets. To see possible influences caused by the **task to be done**, we decided to use "specific fact finding" and "extended fact finding" information seeking tasks.

The evaluation was done with 40 users divided into 5 groups using for all the same already prepared searches for 12 topics, with the same keywords and number of hits in the same order. For each query the users had to answer a question representing the intended information-seeking task. The difference between the five groups was the visualizations the user could use to answer the specific question.

Effectiveness: The effectiveness of the visualisation is measured with the help of the degree of fulfilling the test tasks. E.g., if 8 out of the 12 tasks were solved, the effectiveness is 66,6% out of a maximum of 100%. As it can be seen in the left part of Figure 6 (Effectiveness) there was no significant advantage of using a specific visualisation combination. All visualisations performed nearly as good as the static list, which had been used for reference purposes.

Efficiency: The efficiency of the visualisations has been defined as the effectiveness divided by the time the test persons needed to fulfil a test task. As no absolute minimum or best time exists for this test setting, the values derived are only comparable to each other. In Figure 6 (middle part) it can be seen that the Barchart combination performed second of all visualisation combinations. If we take into account that the Static List is something familiar to the user (well known from search engines), the Barchart has an outstanding role. Surprisingly it performs worst when looking at the effectiveness, but as the values are in a small interval, we do not give too much strength to this effect. Taking the efficiency values, a first interpretation would be that training effects influence the use of the Barchart the least. This will have to be evaluated in a next step.

User Satisfaction: The user satisfaction is derived from the final questionnaire based on a Likert-scale (-2 to +2). Therefore positive and negative values occurred. For the user satisfaction an overall value has been calculated. Figure 6 shows that this general impression of the visualisation was satisfying. This means that the majority of the test persons thought that none of the visualisations are dis-

pensable. They also had the impression that the visualisations helped them to solve a task. The subjective impression of the Scatterplot was the worst.

Users might have performed better, if they would have had more training time for the use of the Scatterplot and by performing better, it is likely that they have a more positive attitude towards a distinctive visualisation. Interestingly, most of the test persons were in a better mood after using INSYSYDER (positive mood before the test 92,5%, after the test 97,5%).

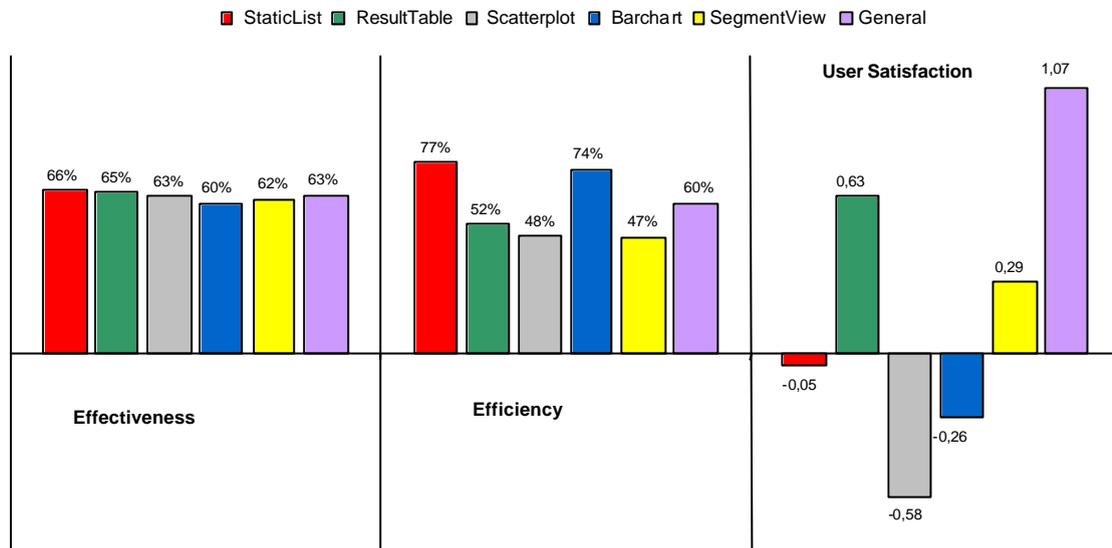


Figure 6: Evaluation results

5. Conclusion and Outlook

The results of the evaluations have motivated us to go ahead. Most of the users make use of our synchronized multiple visual views and saw in them a nice enabling technology to find the most relevant documents in the search result. But there is room for improvements. Mainly based on the experiences of our evaluations and improving our multiple coordinated visualization views, we are preparing a redesign of some parts of the INSYSYDER system. The primary goal of the redesign is to come to a stronger integration of the different visual structures, but still offering the user multiple synchronized views.

References

- (Ahlberg et al. 1994) Ahlberg, C.; Shneiderman, B.: Visual Information Seeking: Tight Coupling of Dynamic Query Filters with Starfield Displays. In: ACM CHI'94 pp. 313-317.
- (Card et al. 1999) Card, S.K.; Mackinlay, J.D.; Shneiderman, B. (Eds.): Readings in Information Visualization. Using Vision to Think. Morgan Kaufmann Publishers, Inc, SF, CA, 1999.
- (Mann 1999) Mann, T. M.: Visualization of WWW-Search Results. In: DEXA'99, pp. 264-268
- (Mann, Reiterer 2000) Mann, T.; Reiterer, H.: Evaluation of Different Visualization of WWW Search Results. Aus: Tjoa, A.; Wagner, R.; Al-Zobaidie, A. (Eds.): Proceedings 11th International Workshop on Database and Expert Systems Applications. Conference: September 4-8, 2000, Greenwich, London, United Kindom. Los Alamitos, CA (IEEE Computer Society) 2000. S. 586-590.
- (Mußler et al. 2000) Mußler, G.; Reiterer, H.; G.; Mann, T.; Handschuh, S.: INSYSYDER - Information Retrieval Aspects of a Business Intelligence System. In: Proceedings of the 7. Internationales Symposium für Informationswissenschaft, UKV Universitätsverlag Konstanz, Konstanz, 2000, pp.127-143

- (North et al. 1999) North, C.; Shneiderman, B.: Snap-Together Visualization: Coordinating Multiple Views to Explore. University of Maryland, CS-TR-4020 June 1999
- (Reiterer et al. 2000) Reiterer, H.; Mußler, G.; Mann, T.; Handschuh, S.: INSYDER – An Information Assistant for Business Intelligence. In: Proceedings of the 23 Annual International ACM SIGIR 2000 Conference on Research and Development in Information Retrieval, ACM press, 2000, pp.112-119
- (Zamir 1998) Zamir, O.: Visualization of Search Results in Document Retrieval Systems. General Examination's Paper, University of Washington, <http://www.cs.washington.edu/homes/zamir>