

Case Study: A Combined Visualization Approach for WWW-Search Results

Thomas M. Mann, Harald Reiterer

Computer and Information Science, University of Konstanz, D-78457 Konstanz, Germany

Thomas.Mann@uni-konstanz.de, Harald.Reiterer@uni-konstanz.de

Abstract

The idea of Information Visualization is to get insights into great amounts of abstract data. Especially document sets found by searching the World Wide Web are a special challenge. The paper gives a short overview on the variety of possible visualizations for this application area. Crucial factors for the success of visualizations are discussed. A combined approach is presented to use alternative simple visualizations, grouped around the traditional result-list, for usage with a local meta web search engine.

1. Introduction

The goal of Information Visualization (IV) is to support the exploration of large volumes of abstract data with computers. IV can be defined as "The use of computer-supported, interactive, visual representations of abstract data to amplify cognition [5]." Digging for information in the World Wide Web, as a huge, heterogeneous, unstructured load of data, represents a special challenge. With current search engines, users typically enter some keywords and get in return an ordered list of links, ranked and displayed in a linear way. This paper presents research results, how to facilitate the access of this information space for the user. Based on an examination of the literature about visualization approaches (chapter 2), and a compilation of crucial factors for the usefulness of visualizations (chapter 3), some ideas have been selected for a combined approach, called Synchronized Alternative Visualizations (chapter 4). The application domain discussed is focused on the presentation of search results from a local meta search engine.

This paper is primarily influenced by research activities within the EU-ESPRIT project INSYDER (Project No. 29232). The projects target is to supply small and medium size enterprises with business information from the internet by using a local meta search engine.

2. State of the Art

One of the first steps, when utilizing IV approaches for search results, is to get an overview on visualization possibilities and an orientation when to use which form. Different types of approaches to classify visualization ideas can be used: data, task/goal or phase oriented. Combina-

tions are possible. An example for a high level task approach is the four phase framework for information seeking by Shneiderman [18]:

- Formulation: expressing the search
- Initiation of action: launching the search
- Review of results: reading messages and outcomes
- Refinement: formulating the next step

We took this framework to decide which phases are possible candidates for support by visualization techniques, and tried to loosely assign the visualization candidates to the four different phases. We focus here on the result phase, which is the most interesting one from the "users point of view". Here he gets the suggestions to satisfy his information need. If a long list of URLs is displayed, it would be a good idea to help the user finding the needle in the haystack by applying adequate visualizations here.

On the **set level**, which means the representation of the whole set of results, it will be interesting to get an overview. Are there any trends, clusters, hot spots? Do the suggestions seem to satisfy the information needs at all? A lot of authors discuss here the use of interactive scatterplots in different forms. Examples are starfields [1] or systems using the third dimension [12]. Another group of approaches uses landscape or map metaphors, called themescapes or self-organizing maps [7] [14]. Other interesting ideas are the use of a wall metaphor [15], Venn diagrams [19], vertical columns of bars [21], spirals [8], or cone trees [17].

After zooming or filtering steps the focus may change from the set level to specific suggestions (commonly called hits) at the **document level**. The user has to decide for a specific URL if it is interesting enough to be followed. Here are visualizations needed to get an overview on the content and the relevance of a single document. For the document level much less ideas can be found. Approaches range from tilebars [11] to thumbnails [6].

3. Factors for choosing visualizations

As shown in the last chapter there are a considerable number of visualization ideas, guidelines and also some findings based on experiments and investigations. This leads to the question: Is there any best form of visualization? Or in the words of Washburne in the year 1927:

"Can the forms be given a rank order for general effectiveness? That is, is there any form which is more effective in all respects than any other form?" [22]. His findings based on an experimental study of various graphic, tabular and textual methods of presenting quantitative material are: "The answer is simply 'no'.". Same could be said looking at the research done in the following years. As implied in the classification schemes mentioned, there are some factors influencing usefulness and effectiveness of visualizations. Taking a closer look at the experiences with different visualization approaches, application areas, taxonomies and experiments, there are four main factors influencing the usefulness of a given visualization, subsequent named "4T-environment": Target user group, Type and number of data, Task to be done and Technical possibilities .

Target user group does not only mean a scientist before the screen or a blue-collar worker. There are also interpersonal differences in information perception and processing, which depend for example on the way people think in spatial dimensions.

The **type of data** to be displayed is essential for choosing a graphical representation. If there is e.g. a hierarchy in the data it makes a sense to exploit this for visualization. But it's not only the type, but also the number of data which influences the success of visualization.

The **task to be done** is also a very important factor influencing the effectiveness of a chosen visualization. There are a considerable number of attempts to classify or rate visualizations for different forms of tasks, with a wide variation of the level on which "tasks" are defined.

Last but not least the **technical possibilities** are a determining factor for utilization and success of a visualization idea. Example for such a technical factor is the choice to use a web browser based user interface.

4. Synchronized Alternative Visualizations

Knowing that there is no "best visualization", and that the success of a specific visualization depends on the user, his current task and the data, we decided to use a combined approach. It offers the user the possibility to choose the most appropriate visualization for his current demand. Different systems in other application domains follow also this approach [2], [13], [20].

But there are also some drawbacks in this approach: The user interface of the system becomes more complex and therefore will be harder to use. The user can choose an inappropriate visualization for a specific situation.

To intercept the possible drawbacks a number of guidelines have been considered. Only six simple visualizations have been chosen. Feedback from real users has been used to make the final choice and to improve the selected visualizations. The visualizations are adapted to each other in color, orientation and the overall style. The visualizations are synchronized in a way that a selection

in one representation of the result set will be updated immediately in the other representations too. The chosen visualizations are grouped around the traditional result list, which is the default view, because it is the most familiar one for many users. The visualizations are presented in a web-browser and could be selected with the help of icons. They are in a order with increasing level of detail information from left to right, with the list positioned in the middle of the icon row (see figure 1).

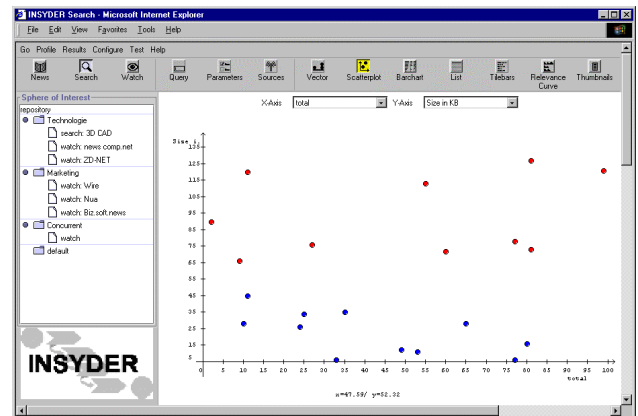


Figure 1. INSYDER GUI

4.1 Document Vector

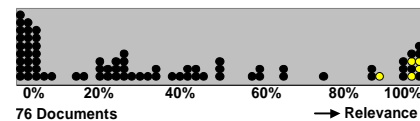


Figure 2. Document Vector (Relevance)

The aim of the **document vector** view is to give the user a simple overview on a larger number of suggestions made by the search system. It is laid out in one dimension. Each document is represented by a black dot (the light dots are the result of some user interaction and will be explained below). If there is more than one point at a column of the scale, the document is displayed in a second row and so on. The type of data displayed can be chosen by the user from a drop down list box. Examples are relevance or date of the documents. In figure 2 it can be detected at a glance that there is only a small number of highly relevant documents.

To give one example for the process of choosing the visualizations, we will explain here the reasons why using a "document vector". User expressed in informal evaluations (using mockups), done at the beginning of the project, that they are interested in easy understandable overviews of the document set found by the INSYDER system. The document spiral idea by Cugini et.al. [8] was one of the ideas presented. To reach the goal to have really simple understandable visualizations, the original document spiral idea was simplified in two ways. First the

representation of the document was explained just to be an icon (no use of colors, the little bars ...). Second, when changing the weight of a keyword, we explained the users, the icons will be rearranged on the spiral, in a way that the now most important documents for this high weighted keywords will tend to go into the middle of the spiral and be concentrated there. Cugini et al. used elevation to separate user weighting from the relevance calculated by the system. We didn't explain the elevation idea, and explained instead the sliders will change the ranking of the system, because it seemed to be easy understandable by a larger group of users.

Our simplified version was really welcomed by most users (even we just had a small sample). But there was also criticism from power users, who remarked that there is a tendency to think documents with close positions to each other will be in some close relation, but in fact a document on the opposite side of the spiral, can be much "closer" than a nearby document on the next "ring". This possible "misinterpretation" was latter also reported by Cugini et.al. [9]. It is easy explainable using Gestalt principles. We decided to transform the spiral to a simple vector, which is not so good in using screen space, but without the problem, of misinterpreting closeness. To avoid the problem of having to many dots on one point, because the straight line is much shorter than the spiral, we "stack" them, as shown in Figure 2.

4.2 Scatterplot

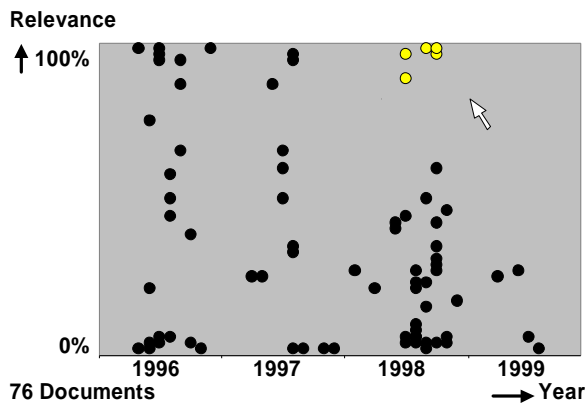
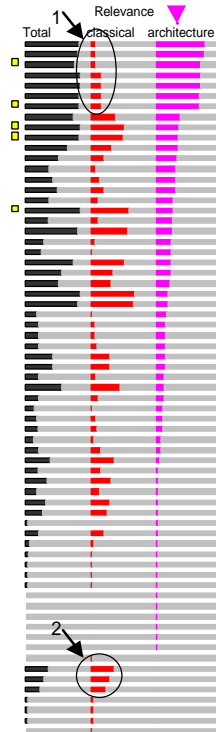


Figure 3. Scatterplot Relevance and Year

In the **scatterplot** view two variables can be shown at the same time. There is also the idea to give the user the possibility to choose which categories to display. This could be overall relevance and year but also relevance for keyword A and keyword B or others at the two different axes. In figure 3 it can be seen that there is a small cluster of highly relevant documents from last year. The group can be marked with the mouse. The selection will now be highlighted in this and all the others views, including the traditional list. The selection can be changed in all views.

4.3 Bargraph

The **bargraph** view shows overall relevance and single relevance for each entered keyword (in figure 4 "classical" and "architecture"). The dots on the left side represent the highlighting described in the scatterplot chapter.



The list can be ordered by each column. In the example the ordering by "architecture" shows that there is a group of documents, marked by circle 1, that are highly relevant for this keyword and are also highly relevant overall, but obviously not very relevant for the keyword "classical". Another group, marked by circle 2, with average relevance for "classical" and overall seems to be not relevant for "architecture". The original idea from [21] is here adapted in several ways. First it is rotated 90 degrees to have the same way of displaying the documents like in the other views where document details are given: top down instead of from right to left. Second the impression of a document as an entity is emphasized using Gestalt principles, without disturbing the keyword orientation too much.

Figure 4. Bargraph

4.4 Tilebars

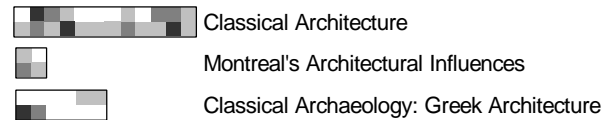


Figure 5. Tilebars and Document Titles

The following visualizations are showing more details on the document level. The first one is the **tilebar** view [11]. Three documents with different lengths are shown in figure 5. Each row of tiles stands for one keyword. Every vertical group of tiles represents a part of the document. The darker a tile, the more relevant the keyword is for this part of the document. In the example the query consists of two keywords: "classical" in the first row and "architecture" in the second. Looking at the third document, it can be easily detected that there are no parts of this document dealing with "architecture" AND "classical". It deals with "architecture" at the beginning and with "classical" at the end. Looking at the title of the document "Classical Archaeology: Greek Architecture" this interpretation makes sense. To give the user the possibility to prove his impressions, the tilebars offer a possibility to jump to a certain

part of a document by clicking on the appropriate tile.

4.5 Relevance Curves

The same goal as behind the use of tilebars, but with a different mapping to visual structures, can be found in using relevance curves. The original idea from [4] is enhanced by using the length of the curve to show the length of the document (in the original curve length is independent from document length), and using color coding for the impact of the different keywords.

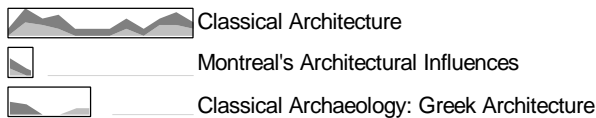


Figure 6. Relevance Curves and Document Titles

4.6 Thumbnail views

The last visualization on document level is the representation of the documents as **thumbnails**. Probably it will be combined with other visualizations on document level and some textual information. The idea behind the representation of a document as a thumbnail is mainly to give users, who often work in the same document spaces, some hints about probably known documents. There may be also some support getting a first impression for unknown documents. A known problem offering thumbnails are the crawling times (text and images).

5. Conclusions and Outlook

The idea discussed in this paper is to embed synchronized alternative visualizations in a user interface of a local meta search engine. Main ideas are to group simple visualizations around the traditional result-list view for searching the web (in two directions: with decreasing and increasing level of detail information), and to use the browser as a familiar interface for the user.

The next step is to evaluate the system with users in a systematic way. Typical questions for this evaluation: Will be the additional cognitive overload using a more complex interface compensated by the expected value to get better insights in shorter time? Which visualizations fit best to users needs? Are there differences between user groups? Based on the empirical experiences of the evaluation, it could be discussed to support the choice of alternative visualizations in certain cases automatically by using a knowledge base.

6. References

[1] Ahlberg, C.; Shneiderman, B.: *Visual Information Seeking: Tight Coupling of Dynamic Query Filters with Starfield Displays*. In: Proc. ACM CHI'94 pp. 313-317.

- [2] Ahlberg, C; Wistrand, E: *IVEE: An information visualization and exploration environment*. In: Proc. IEEE Information Visualization 95, pp. 66-73.
- [3] AltaVista, <http://www.altavista.com> [1999-02-04]
- [4] Arisem S.A., <http://www.arisem.com> [1999-06-20]
- [5] Card, S.K.; Mackinlay, J.D.; Shneiderman, B. (Eds.): *Readings in Information Visualization. Using Vision to Think*. Morgan Kaufmann Publishers, Inc, San Francisco, CA, 1999.
- [6] Card, S.K.; Robertson, G.G.; York, W.: *The WebBook and the WebForager: An Information Workspace in the World Wide Web*. In: Proc. ACM CHI'96, pp. 111-117.
- [7] Chalmers, M.: *Using a Landscape to represent a corpus of documents*. In: Proc. COSIT'93, pp. 377-390.
- [8] Cugini, J.; Laskowski, S.; Piatko, C.: *Document Clustering in Concept Space: The NIST Information Retrieval Visualization Engine (NIRVE)*. <http://zing.ncsl.nist.gov/~cugini/uicd/cc-paper.html> [1998-09-10]
- [9] Cugini, J.; Laskowski, S.; Sebrechts, M. M.: *Design of 3-D Visualization of Search Results: Evolution and Evaluation..* <http://www.itl.nist.gov/iaui/vvrg/cugini/uicd/nirve-paper.ps.gz> [1999-09-13]
- [10] Fowler, R.H.; Fowler, W.A.; Wilson, B.A.: *Integrating Query, Thesaurus, and Documents through a Common Visual Representation*. In: Proc. ACM SIGIR'91, pp. 142-151.
- [11] Hearst, M.A.: *TileBars: Visualization of Term Distribution Information in Full Text Information Access*. In: Proc. ACM CHI'95, pp. 59-66.
- [12] Hemmje, M.: *LyberWorld - Eine 3D-basierte Benutzerschnittstelle für die computerunterstützte Informationssuche in Dokumentmengen*. In: GMD-Spiegel, Vol. 1, 1993, pp. 56-63.
- [13] Henninger, S.; Belkin, N.J.: *Interface Issues and Interaction Strategies for Information Retrieval Systems*. In: Proc. ACM CHI'96, pp. 352-353
- [14] Lin, X.; Soergel, D.; Marchionini, G.: *A Self-Organizing Semantic Map for Information Retrieval*. In: Proc. ACM SIGIR'91, pp. 262-269.
- [15] Mackinlay, J.D.; Robertson, G.G.; Card, S.K.: *The Perspective Wall: Detail and Context Smoothly Integrated*. In: Proc. ACM CHI'91, pp. 173-179.
- [16] MediaLab B.V., <http://www.medialab.nl> [1999-02-24]
- [17] Robertson, G.G.; Mackinlay, J.D.; Card, S.K.: *Cone Trees: Animated 3D Visualizations of Hierarchical Information*. In: Proc. ACM CHI'91, pp. 189-194.
- [18] Shneiderman, B.: *Designing the User Interface. Strategies for Effective Human-Computer Interaction*. Addison-Wesley, Reading, Massachusetts, 1998.
- [19] Spoerri, A.: *InfoCrystal: A Visual Tool for Information Retrieval*. In: Proc. IEEE Visualization 93, pp. 150-157.
- [20] Stenmark, D.: *To Search is Great, to Find is Greater: a Study of Visualization Tools for the Web*. <http://w3.informatik.gu.se/~dixi/publ/mdi.htm> [1999-01-11]
- [21] Veerasamy, A.; Navathe, S.B.: *Querying, Navigating and Visualizing a Digital Library Catalog*. In: Proc. DL'95. <http://www.csdl.tamu.edu/DL95/papers/veerasamy/veerasamy.html> [1999-03-24]
- [22] Washburne, J.N.: *An Experimental Study of Various Graphic, Tabular and Textual Methods of Presenting Quantitative Material*. In: The Journal of Educational Psychology, Vol. 18 Num. 6, 1927, pp. 361-376.