

Exploring and Visualizing the History of InfoVis *

Daniel A. Keim, Helmut Barro, Christian Panse, Jörn Schneidewind, Mike Sips

University of Konstanz, Germany

{keim,barro,panse,schneide,sips}@inf.uni-konstanz.de

1 Motivation

The exploration and visualization of large information spaces is a challenging task. The provided contest data set for example contains more than 1000 authors of about 600 papers. The basic idea for an effective data exploration is to include the human in the data exploration process and combine the flexibility, creativity, and general knowledge of the human with the enormous storage capacity and the computational power of today's computers. The key concept of our visualization approach is to visualize the whole dataset to provide a first overview and to provide rapid, incremental, and reversible analysis actions. Our techniques follow the well-known Information Seeking Mantra: overviews first, zoom and filter, and details on demand. All visualizations, actions, and details are tightly coupled using the well-known linking and brushing concepts.

2 Data Cleaning

When processing and visualizing large data sets, data cleaning as part of data pre-processing is a very important step, since it directly influences the quality of the visualization. Since there were some inconsistencies in the contest data set, like ambiguous authors or different formats and spellings for the conference names, some data cleaning was necessary. Therefore we wrote some shell scripts, based on regular expressions, to correct these inconsistencies. Additionally we corrected the spelling of some author names manually. Another problem was, that for several attributes no values were recorded. An example are the keyword attributes, where for a lot of publications the keywords were missing. For these publications we extracted some keywords from their title. For other missing attributes we set their value to *not defined* and handled it as special cases in the visualization step.

3 Overall Concept

To find solutions for the 3 contest tasks, we applied different visualization approaches. One technique we used, is adapted from a popular visualization tool for movie data sets, called

FilmFinder [Ahlberg and Shneiderman 1994]. We applied a similar technique to the contest database and called our technique PaperFinder. The basic idea is to use a two dimensional layout, e.g. the euclidian plane and use the time attribute for the x-axis ordering and the dependent attribute for the y-axis ordering. Applied to the contest data set, the idea is to group the 600 papers and their authors together in a first overview. The papers are displayed in groups belonging to the publication year. Each paper is represented by a colored icon (rectangle). The color represents the thematic theme of each paper. Our results show that this technique is very useful to get insight in the development of research topics over time as well as over the number and research areas of each authors publications. To show the connection and collaborations of authors we used a graph drawing approach. If we assume, that the authors are the nodes of the graph and edges represent paper collaborations, the goal is to find strong connected components (SCC) in order to identify groups of authors, which published most papers together. Then we employed a spring embedder to find an appropriate graph layout for the computed SCC's. To visualize all co-authors for a single author over time, we used a technique which is based on Interrings [Yang et al. 2002],

4 Overview of the 10 years of InfoVis

To show the overview over InfoVis we displayed the number of publications belonging to each conference over time, using the PaperFinder. Additionally we visualized an ranking of authors, depending on their number of publications over the years. The colors represent the research topics, so that it is easy to see how the research areas developed. To show the publications for a single author and his co-authors, we applied the Interring technique and the PaperFinder technique.

4.1 TASK 1: Static Overview of 10 years of Infovis

To get an overview on the last 10 years of InfoVis, we used the PaperFinder. To see the development of topics over the years, we placed the time attribute on the x-axis and the number of co-authors on the y-axis. The color represents the paper categories as shown in the legend. We defined 5 categories to which we assigned all publications, depending on their keywords. The categories are Information Visualization, HCI, Data Analysis, Computer Graphics and Graph

*Acknowledgement: This work was partially funded by the Information Society Technologies programme of the European Commission, Future and Emerging Technologies under the IST-2001-33058 PANDA project (2001-2004).

Drawing. Figure 2 shows the development of these categories and the number of publications per conference over the last years.

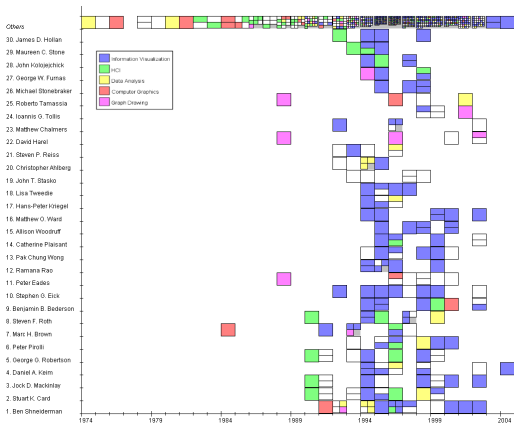


Figure 1: Top 30 authors, based on number of publications

4.2 TASK 2: Characterize the research areas and their evolution

Figure 2 shows how the research areas developed over time. For example, it turns out that the number of papers submitted under HCI topic decreased while the number of Visualization papers increased. It is easy to see, that there were more and more publications from other research topics than visualization, published at the InfoVis Conference (more yellow, green boxes). For last years InfoVis, most keywords were missing in the database (white rectangles).

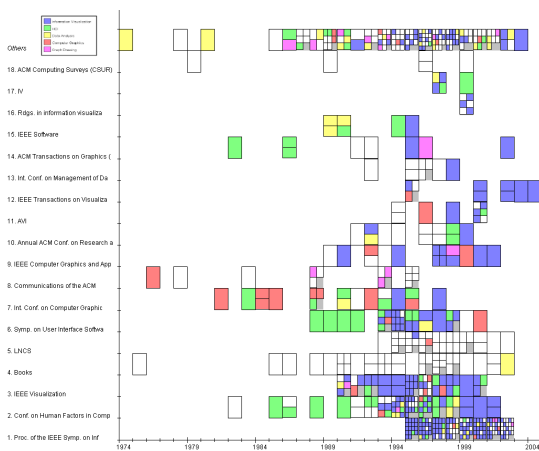


Figure 2: Complete Overview of publications per conference, Conferences are ranked by number of publications, color represents research topic

4.3 TASK 3: The people in InfoVis

To visualize the information about the people contained in the contest data set, we used the PaperFinder and the Interrings to visualize the 30 authors with most publications. As shown in Figure 1, B. Shneiderman has the highest number of publications (more than 18), followed by S. K. Card (12), J. D. Mackinlay (10) and D. A. Keim (10). The colors show the categories to which the publications belong. To show the publications of a single author we used the interring technique. The examples in Figure 3 show the co-authors of Daniel A. Keim (left) and George Robertson (right). Each co-author is represented by a different color. As shown in the right figure, G. Robertson wrote most of his papers together with S. K. Card and J. D. Mackinlay. In Figure 4 the co-authors of publications from G. Robertson and their research topics are visualized using PaperFinder.

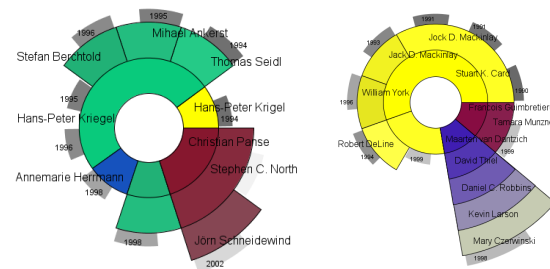


Figure 3: Interring showing co-authors of D. A. Keim (left) and G. Robertson (right) over the years

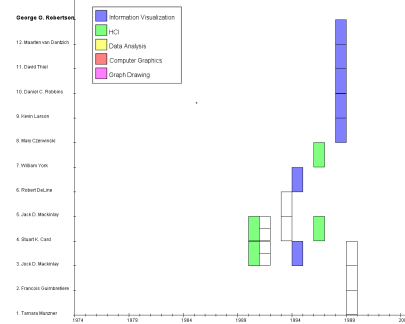


Figure 4: Co-authors of G. Robertson and their publication topics over the years

References

AHLBERG, C., AND SHNEIDERMAN, B. 1994. Visual information seeking using the filmfinder. In *ACM CHI'94 Conference Companion*, 433–434.

YANG, J., WARD, O. M., AND RUNDENSTEINER, E. A. 2002. Interring: An interactive tool for visually navigating and manipulating hierarchical structures. In *Proceedings of the IEEE Symposium on Information Visualization*, 77.