

MonkEllipse: Visualizing the History of Information Visualization

Tzu-Wei Hsu, Lee Inman, Dave McColgin, Kevin Stamper
{thsu, lee, davemail, kstamper} @cc.gatech.edu
Georgia Institute of Technology

ABSTRACT

In this paper, we describe the process and result of creating a visualization to capture the past 10 years of history in the field of Information Visualization, as part of the annual InfoVis Conference Contest. We began with an XML file containing data provided by the contest organizers, scrubbed and augmented the data, and created a database to hold the information. We designed a visualization and implemented it using Flash MX 2004 Professional with ActionScript 2.0, PHP, and PostgreSQL. The resulting visualization provides an overview of the field of Information Visualization, and allows users to see the connections between areas of the field, particular researchers, and documents.

CR Categories: K.2 [History of Computing]: People; H.5.2 [Information Interfaces and Presentation (HCI)]: User Interfaces --- Graphical User Interfaces (GUI); H.5.4 [Information Interfaces and Presentation (HCI)]: Hypertext/Hypermedia --- Navigation

Keywords: information visualization, history, text arc, document, text, articles, contributors, research areas, InfoVis conference

1 INTRODUCTION

The field of information visualization is rich with many different research areas. Our visualization attempts to characterize the field by displaying the titles of published articles from the past 10 years of the InfoVis conference and linking these articles to their authors, research areas, and references so that a user of the visualization can explore and discover relationships between articles, authors, research areas, and references in an intuitive and interactive manner.

Our design was inspired by the TextArc application, authored by W. Bradford Paley. The visualization displays article titles and research areas. Articles are listed around the form of an ellipse, and the center of the ellipse is filled with research areas. Users can search by keywords to highlight applicable articles, or browse the articles inside the visualization. By clicking on articles, users can see relationships between articles, research areas, and contributors.

2 APPLICATION

Our visualization was built in Macromedia Flash MX 2004 Professional, using ActionScript 2.0. It accesses the PostgreSQL database using PHP. PERL scripting was also used extensively to find additional content for the database, to develop categories, and classify authors and papers.

2.1 Visual Design

Our design orders the papers in chronological order along an ellipse. The derived topics/research areas are scattered in the center and are presented based on the number of documents that fall within each research area. For example, a research area with more documents appears in a larger font than a research area with fewer documents. This layout helps to characterize the research areas by emphasizing those research areas with the most articles. The advantage of this importance-by-size view gives users an overview of the most important research areas within the field.

Each document is clickable, as is each research area. If a user clicks on a paper title, all other papers referenced by that paper are highlighted, as well as all research areas to which the paper applies. When the user clicks on a particular research area, all papers associated with that research area are highlighted.

Once the user has clicked on a research area and sees the resulting highlighting on all papers within that research area, a list of authors contributing to that research area appears. The list is scrollable, with the authors sorted by their total contributions to the field. Thus, the research areas and authors are bridged, giving the user insight into both dimensions of a paper. This bridge serves to illustrate where a particular author/researcher fits within the research areas being studied by the user.

When the user clicks on a document, the pane below the ellipse shows the paper's title, abstract, authors, year, and a link to the entry in the ACM Digital Library, so that the user can find more information and access the entire paper.

2.2 Programming

We used PHP as middleware between Flash MX 2004 Professional and our PostgreSQL database. The PHP component takes query values from Flash and transforms them into PostgreSQL queries that can retrieve data from the database. The PHP then formats the query results back into a format Flash can use. In Flash, we used an ellipse equation to display the article titles around the outside of the ellipse.

3 ANALYSIS PROCESS AND TASKS

The overall goal of our visualization is to provide an overview of the field while still allowing users to discover interesting relationships within the data. This set of tasks was chosen from the list of contest tasks because of their wide appeal across a broad range of audiences. Students new to the field as well as established researchers can benefit from a visualization that addresses this set of tasks.

3.1 Overview of InfoVis

Through visual design, the system provides a static overview of the field of Information Visualization in one screen. Our goal for this particular task was to minimize the number of 'modes' needed to gain an overview of the field. Our approach of displaying articles, year indicators, and research areas on one screen, and allowing interaction between elements, produced a display that could statically show chronology, evolution, and the names and relative influence of our derived research areas.

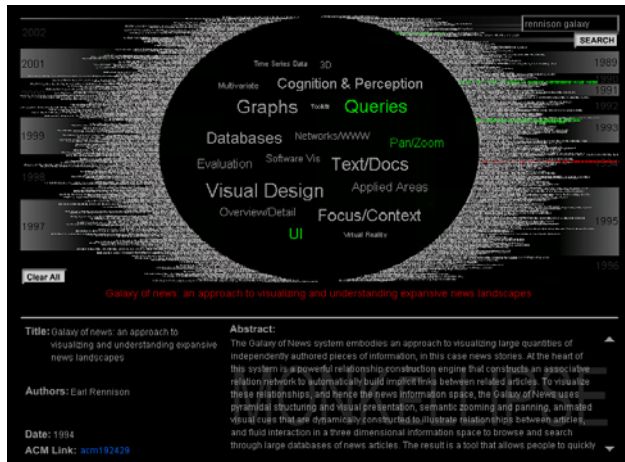


Fig. 1 InfoVis documents from the past 10 years are arranged chronologically in an ellipse with research areas displayed at the center and year indicators for context.

3.2 Research Areas and their Evolution

MonkEllipse allows users to explore the evolution of different research areas. When a user clicks on a research area, articles belonging to the topic are highlighted. Having the entire 10 year history in one view takes advantage of our ability to see clusters and trends in data. Visually following the ellipse clockwise, trends in research can be observed in the density of highlighted articles through the years.

3.3 Authors/Researchers in the Field

The task of discovering where a particular author/researcher fits within the field is difficult, as there are many contributors across a variety of topics. As our system stands now, searching an author's name is the simplest way to find him. The results of this search return all articles where the researcher was an author, was referenced, or was mentioned in the abstract. This method gives an idea of how long this particular researcher has been involved in the field. Clicking on a highlighted article reveals the research area(s) into which the article fits, helping to reveal the author's involvement in that research area. The user can also click on a topic to see which researchers are involved in that particular research area.

3.4 Search

The ever-present search box allows partial word matching, highlighting relevant authors and research areas.

3.5 Document Details

This lower pane of MonkEllipse provides the document's basic information as well as a direct link to the selected document in the ACM Digital Library, which allows the user to first determine if this article warrants further exploration by providing the article's abstract.

4 SYSTEM CRITIQUE

There are several tasks for which our system is not optimal. One issue is the scale of the dataset and number of articles being referenced. Currently, with 614 articles, it is impossible to see the name of each article in a readable font size. We have compensated for this by increasing the font size of the article in focus, and by displaying the title currently in focus below the ellipse as well. However, if the dataset were significantly larger, the system would have increased difficulty supporting the amount of information necessary to display on the screen.

Additionally, the system does not adequately support direct walk from one article to another. While the user can go from one article to its reference and then on to that article's references, the user ultimately loses the original context. One potential solution is to allow the user to retain highlighting from previous queries, but the system would run out of highlighting colors after a certain number of steps in the walk. Also, the additional highlight colors would be hard to track and would potentially make the screen cluttered and confusing.

The system in its current implementation can handle linking from an article to its references, but not the reverse operation, determining articles that reference one particular article. This would be a useful task because it would illustrate the influence of a seminal paper over the course of time.

5 CONCLUSION

Our system visualizes the past ten years of the field of InfoVis and enables users to explore and learn about the field. The visualization characterizes the different research areas and illustrates the documents that relate to each research area using a chronological listing. It also illustrates where a particular author/researcher fits within the different research areas, even if the author touches on multiple research areas in his or her work. Finally, the visualization illustrates the relationships between researchers, including coauthoring, referencing, and contributions to the same field at similar times.

REFERENCES

- Franklin, D., and Makar, J. Flash MX 2004 ActionScript. Macromedia Press: Berkeley, CA, 2004.
- Paley, W. B. TextArc: Showing Word Frequency and Distribution in Text, poster presented at InfoVis '02.
- Spence, R. *Information Visualization*. Pearson Addison Wesley/ACM Press: Essex, England, 2001.
- Stasko, John. "Text and Documents 2" College of Management, Georgia Institute of Technology. 18 Mar. 2004