Research Agenda:
Visual Overviews for Exploratory Search

Ben Shneiderman
Human-Computer Interaction Lab &
Department of Computer Science
University of Maryland
College Park, MD 20742
ben@cs.umd.edu

ABSTRACT
Exploratory search is necessary when users knowledge of the domain is incomplete or when initial user goals do not match available data or metadata that is the basis for search indexing attributes. Such mismatches mean that users need to learn more in order to develop a better understanding of the domain or to revise their search goals. Exploratory search processes may take weeks or months, so interfaces that support prolonged exploration are necessary. The attraction of exploratory search is that users can take on more ambitious goals that require substantial learning and creative leaps to bridge the gaps between what they know and that they seek.

Author Keywords
Information visualization, network visualization, exploratory search, user interface

ACM Classification Keywords
H.5. Information interfaces and presentation (e.g., HCI). H.2 DATABASE MANAGEMENT

INTRODUCTION
The success of information visualization stems, in part, from its capacity to provide overviews that offer deeper understanding of important phenomena [Card, Mackinlay & Shneiderman 1999]. These overviews reveal distributions, clusters, gaps, and outliers that may lead to insights by domain experts. These insights may indicate erroneous or missing data, important relationships, or surprising patterns. Determining the efficacy of visual overviews is difficult, but we can study human performance in the process of making known-item searches, exploratory information seeking inquiries, and insight discovery events [Shneiderman & Plaisant 2006].

Prepared as a working paper for National Science Foundation workshop on Information Seeking Support Systems (ISSS), June 26-27, 2008, Chapel Hill, NC. Workshop organizers: Gary Marchionini and Ryen White.

http://www.ils.unc.edu/ISSS/

The tools that support search, browsing, and visualization have dramatically improved in the past decade, so there is value for the information retrieval community to re-examine recent work and consider what future opportunities there are for integrating exploratory search technologies with interactive information visualization [Shneiderman et al., 2000; Kules & Shneiderman 2008].

As Turing award-winner Richard Hamming, wrote: "The purpose of computing is insight, not numbers." I might paraphrase with "The purpose of visualization is insight, not pictures." Successful tools support a process of information-seeking that leads to important insights for individual users, organizational teams, and larger communities. The term *insights* makes clear that exploratory search is a human experience, made possible by well-designed tools that support discovery [Saraiya, North & Duca 2004, 2005; North 2006].

This research agenda suggests that visual overview for web searches of many kinds could dramatically accelerate human performance in understanding the distribution of search results [Chen 2005; Thomas & Cook 2005]. This could lead to more effective search outcomes in a shorter amount of time, even by those with less domain knowledge and less search experience.

The emphasis here is on network visualizations for citations among results sets [Shneiderman & Aris 2006; Aris & Shneiderman 2007, 2008; Perer & Shneiderman 2006, 2008a, 2008b], but treemaps and other overview methods also offer opportunities [Kules et al., 2008]

ACKNOWLEDGMENTS
Thanks for partial support from National Science Foundation grant IIS-0705832) iOPENR--A Flexible Framework to Support Rapid Learning in Unfamiliar Research Domains (8/1/2007-7/31/2010), Principal Investigator Bonnie Dorr, also includes the author, Judith Klavans, Jimmy Lin, and Dragomir Radev.

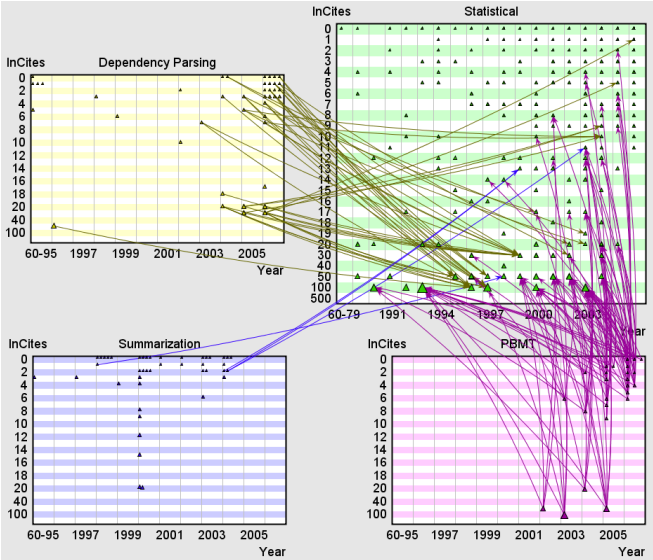

Figure 1: A semantic substrate shows regions for three small research topics (Dependency parsing, summarization, Phrase-based Machine Translation (PBMT)) and one large research topic (Statistical methods in natural language processing). The links show only citations that go from the small to the large topic, revealing strong linkage for some topics and weak ones for others. This figure was created using the Network Visualization with Semantic Substrates (NVSS) (http://www.cs.umd.edu/hcil/nvss).

REFERENCES

1. Aris, A. and Shneiderman, B., Designing semantic substrates for visual network exploration, *Information Visualization Journal 6*, 4 (2007), 1-20. doi: 10.1057/palgrave.ivs.9500162
2. Aris, A. and Shneiderman, B., A node aggregation to reduce complexity in network visualizations using semantic substrates, University of Maryland Technical Report, Dept of Computer Science (February 2008).
3. Card, S. K., MacKinlay, J. D., Shneiderman, B., *Readings in Information Visualization: Using Vision to Think*, Morgan Kaufmann Publishers, San Francisco, CA (1999).
4. Chen, C., Top 10 Unsolved Information Visualization Problems, *IEEE Computer Graphics & Applications* (July/August 2005), 12-16.
5. Kules, B. and Shneiderman, B., Users can change their web search tactics: Design principles for categorized overviews, *Information Processing and Management* (to appear 2008).
6. North, C., Toward Measuring Visualization Insight., *IEEE Computer Graphics & Applications 26*, 3 (May-June 2006), 6-9
7. Perer, A. and Shneiderman, B., Balancing systematic and flexible exploration of social networks, *IEEE Symposium on Information Visualization* and *IEEE Transactions on Visualization and Computer Graphics 12*, 5 (October 2006), 693-700.
8. Perer, A. and Shneiderman, B., Systematic yet flexible discovery: Guiding domain experts during exploratory data analysis, *Proc. ACM Conference on Intelligent User Interfaces* (January 2008a).
9. Perer, A. and Shneiderman, B., Integrating statistics and visualization: Case studies of gaining clarity during exploratory data analysis, *ACM CHI2008 Conference*, ACM, New York (April 2008b), 265-274.
10. Saraiya, P., North, C., and Duca, K., An Insight-Based Methodology for Evaluating Bioinformatics Visualization, *IEEE Trans. Visualization and Computer Graphics 11*, 4 (July/Aug. 2005).
11. Saraiya, P., North, C., and Duca, K., An Evaluation of Microarray Visualization Tools for Biological Insight", *IEEE Symposium on Information Visualization 2004 (InfoVis 2004)* (2004), 1-8.
12. Shneiderman, B., The eyes have it: A task by data-type taxonomy for information visualizations, *Proc. Visual Languages* (Boulder, CO, Sept. 3–6). IEEE Computer Science Press, Los Alamitos, CA (1996), 336–343.
13. Shneiderman, B. and Aris, A., Network visualization with semantic substrates, *IEEE Symposium on Information Visualization* and *IEEE Transactions on Visualization and Computer Graphics 12*, 5 (October 2006), 733-740.
14. Shneiderman, B., Feldman, D., Rose, A., and Ferre, X. A., Visualizing digital library search results with categorical and hierarchical axes, *Proc. 5th ACM International Conference on Digital Libraries,* ACM, New York (June 2000), 57-66.
15. Thomas, J.J. and Cook, K.A. (eds.), *Illuminating the Path: Research and Development Agenda for Visual Analytics*, IEEE Press (2005).