2011

The visualization of database search results

James A. Van Fleet
Bucknell University, vanfleet@bucknell.edu

AC 2011-529: THE VISUALIZATION OF DATABASE SEARCH RESULTS

James A. Van Fleet, Bucknell University

James A Van Fleet has been Librarian for the Sciences and Engineering at Bucknell University and a member of ASEE ELD since 1989. He has served the Engineering Libraries Division as Membership Directory Editor and Newsletter Editor.

The visualization of database search results

Abstract:

Edward Tufte describes the visual presentation of quantitative data as "envisioning information."[1] A number of databases, from Google to specialized subject databases, offer a graphical representation of research results. Visualization is offered as an alternative method for evaluating search results. A survey of the current mechanisms or tools for providing a visual or graphical presentation of database information shows that this trend may be on the decline.

Introduction:

Edward Tufte describes the visual presentation of quantitative data as "envisioning information." The third book of his famous trilogy is titled Visual Explanations, and best lays out his thesis that "information displays should be documentary, comparative, causal and explanatory, quantified, multivariate, exploratory."[2] These qualities are, for the most part, lacking from the standard display of a bibliographic database search result! Although a list of references or citations may be numbered and sorted, the display of this information offers limited guidance to the user in determining the best source to choose from the list. Within the last five or six years, a number of databases, from Google to specialized subject databases, have offered a graphical representation of research results. These visualization tools are usually offered as an alternative method for evaluating search results, through a sidebar menu or button. This trend deserves closer scrutiny, since the question of how database users evaluate their search results is central to the success of their library research.

Diagramming Relevance:

When the result of a database search is a long list of citations; to journal articles, books in a catalog, or web sites, the cues for selecting the "best" sources from the list are limited. Most databases sort by relevance or by date, and users have come to rely on this feature to assume that the first half-dozen results on the list are the best. More sophisticated users might look for key search words in the title or link. Only an experienced researcher might scan a citation list and select items based on familiarity with the authors or their research institution, which journal an article is published in, number of times cited, or other visual cues. When scanning the results of a Google search, similar cues are used (eg: is the URL a dot.com or a dot.gov?), and familiar web sources such as Wikipedia often appear on the first page of search results.

In the last decade Groxis developed the visual search engine Grokker,[3,9] creating graphical representations of search results that looked like something between a Venn diagram and a bubble chart. This sorting and grouping of search results did appear to make narrowing a search easier. The diagrams showed references in a broader context, and offered alternative avenues for research, and their relative weight based on the size of the graphics. For a time,

some EbscoHost databases offered Grokker visual search results as an option, but not a prominent feature of their databases. Groxis had "ceased operations" as of 2009.[4]

Ebsco databases still offer a "visualization" tool to show how search refinements narrow results. However, this is not much different from a breadcrumb trail in the menu bar of a web browser. It does not highlight alternative avenues of research, or show context for the selected references.

Web 2.0:

Tag clouds or weighted lists, introduced by Flickr and del.icio.us, are databases built on individually assigned metadata "tags." These information displays are wonderfully intuitive, highlighting the most popular or common tag words at a glance. Flickr users can search tags, and even georeference tags on maps. Thumbnail photographs display the results of an image database search, but do not organize the search results in any meaningful way. Yahoo's "delicious" search results are displayed as text hit lists. Both fail to take full advantage of the possibilities of tag cloud as search result. Although individual users can create their own personal tag cloud, database searchers cannot see a cloud generated by a specific search, collating the collective entries of many users. The wisdom of crowds is not translated into clouds.

IBM's Many Eyes is an interesting experiment in user-generated information visualization. This site "allows users to upload data and then produce graphic representations for others to view and comment upon."[5] The database can be searched successfully by broad topics; the availability of specific information that a researcher might be seeking is entirely dependent on what has been uploaded by other users. Not strictly a bibliographic database, Many Eyes is close to some of the ideals put forward by Tufte for the visual presentation of information.

Searching with Pictures:

The most successful visualizations are incorporated into databases which index graphical information, from photographs to molecular drawings. Chemical engineers recognize the value of searching a graphical database such as SciFinder, and the power of being able to draw part of a chemical structure to search for related information. Displaying search results as chemical structure diagrams with accompanying text makes this a powerful research tool, but one that requires specialized knowledge and training to use well.

Concept Mapping:

The Web of Knowledge citation databases have attempted to create maps that show the relationships between citations.[6] These literal graphical webs do demonstrate the interrelated nature of research publications. They can demonstrate broadly whether publication on a topic is growing or waning, and at their best, they can take the researcher on a journey of discovery

through the literature. The appearance of the web can be ordered by color, by subject category, or many other combinations. The concept maps combine the graphic with a separate text display of citation information.

Credo Reference uses a technique called hyperbolic browsing to show relationships between ideas.[7] These concept maps provide general outlines of related topics to explore. The Google "Wonder Wheel" also uses this approach to visualize relationships between concepts. Unfortunately, these tools tend to be peripheral menu options that are usually overlooked by searchers, and not often included in library user education sessions.

An idea that has come and gone?

There are questions about the commercial viability of visualization tools. I began exploring the topic of graphical search results in 2005. Since then, Groxis and their patented visualization tool Grokker have ceased business, and Yahoo, the new owner of delicious, has de.emphasized tag clouds. None of the commercial database providers has attempted to make their visualization tool the default for search results. Judy Luther and her colleagues point out that data visualization may become more of a challenge when users are accessing information on smaller and smaller screens![7]

For text citation databases, visualization tools may "display large data sets in an easily digestible way and provide a concise picture of trends and issues."[8] However, they may not be providing the guidance that researchers want. The interpretation of visual or graphical search results seems to be less intuitive than database search results presented as a list of citations or URL links. Library researchers have had years of informal and formal training in recognizing and evaluating the cues presented by a text list of references. They also rely on their chosen research tools, such as Google, to do this analysis for them, and offer up the "best" references first. Librarians may need to spend some time studying visual interfaces, and training library researchers in their use and interpretation.

The average Google user can interpret and evaluate a list of search results to their own satisfaction. The results are relevant and "good enough" for their purposes. Each of the visualization schemes discussed above requires a certain amount of training or practice to allow interpretation of the results. They require more interaction with the search results, and more effort in selecting a path to follow to the best results. These are the behaviors we seek to instill in our library researchers, and would seem to offer a distinct advantage over the text list. However the necessity to work harder to obtain useful results, and the trial and error nature of following many possible paths, seem to be significant barriers to the successful visualization of database search results. These questions offer possibilities for future investigation, such as usability studies and end-user assessment.

References

1. Tufte, E. R. (1990). *Envisioning information*. Cheshire, Conn.: Graphics Press.

2. Tufte, E. R. (1997). *Visual explanations : Images and quantities, evidence and narrative*. Cheshire, Conn.: Graphics Press.

3. Foenix-Riou, B. (2006). When search engines play at maps: Visualization technologies. *Online*, *30*(2), 29-32.

4. Marcinko, R., & Arnold, S. (2009). *Grokker mystery : Beyond search.* Retrieved 2/18/2011, 2011, from http://arnoldit.com/wordpress/2009/08/22/grokker-mystery/

5. IBM Research, & IBM Cognos Software Group. (2010). *Many eyes.* Retrieved 2/22/2011, 2011, from http://www-958.ibm.com/software/data/cognos/manyeyes/

6. Simboli, B. D. (2008). Web of science's "citation mapping" tool. *Issues in Science & Technology Librarianship*, (54), 5-5. http://www.istl.org/08-summer/electronic-1.html

7. Luther, J., Kelly, M., & Beagle, D. (2005). Visualize this. *Library Journal*, *130*(4), 34-37.

8. Mort, D.ANALYSIS: Visualisation boosts power of STM databases. *Research Information*, (April/May 2006), 12/3/2010. http://www.researchinformation.info/riaprmay06analysis.html

9. Dursteler, J. C. (2004). Grokker, or visual navigation. *Inf@Viz! the Digital Magazine of InfoVis.Net*, (138). Retrieved 3/7/2011, from http://www.infovis.net/printMag.php?num=138&lang=2