Information Visualization at the Turn of the Century:

Bleeding Edge to Leading Edge?

by Steven Anderson

New Internet-based technologies are exploding on the information scene almost daily. Some are not quite ready for the market, while others hold the potential to change dramatically the ways in which we seek and retrieve information. Information visualization software, although still in an experimental phase, belongs in the latter category.

Law librarians must become familiar with information visualization technology for several reasons: to suggest new applications for corporate intranets; to stay current with new Internet technologies for our institutions' Web sites; and to be able to evaluate the strengths and weaknesses of proposed software applications. It certainly seems possible that library catalog vendors and Internet search engines will employ this technology before too long; law librarians had best understand it before our users surpass our own knowledge base.

Fig. 1. *Xerox PARC's Information Visualizer Cone Tree (Image courtesy of Xerox Corp. and Inxight Software, Inc.)*

Information visualization can perhaps best be described as a system of "seeing" information without resorting to the use of text. It has its roots soundly based in scientific data visualization schemes, such as graphs, scatter plots, and pie charts. The difference is that, by necessity, information visualization takes on a more ambitious task: representing types of information in documents AND representing groups of similar documents. It turns out to be much more complex than plotting X and Y coordinates on a graph.

There are two primary reasons why an information architect might want to go to the trouble of designing a search system that incorporates an information visualization scheme. First, most graphical user interface designers have come to realize that "a picture is worth a thousand words," and that its corollary, "time is money," is also quite true. The faster you can process huge amounts of information, the more efficient you are. Second, few text-based information displays "scale" well to the Internet. Even though you might narrow your relevant hits from 10,000 to 230 documents, you still need to sift through the remainder. With information being produced and consumed at faster rates, the problems associated with filtering through information increase exponentially. At this point, visual information processing tools look increasingly attractive.

Information visualization software once was a mere toy for many computer scientists, but as the Web has exploded, so too has advancement on the visualization front. The first commercial information visualization software packages are just now reaching the market. This article will summarize both some of the proposed experimental models as well as currently marketed applications.

Xerox's Palo Alto Research

continued on page 6

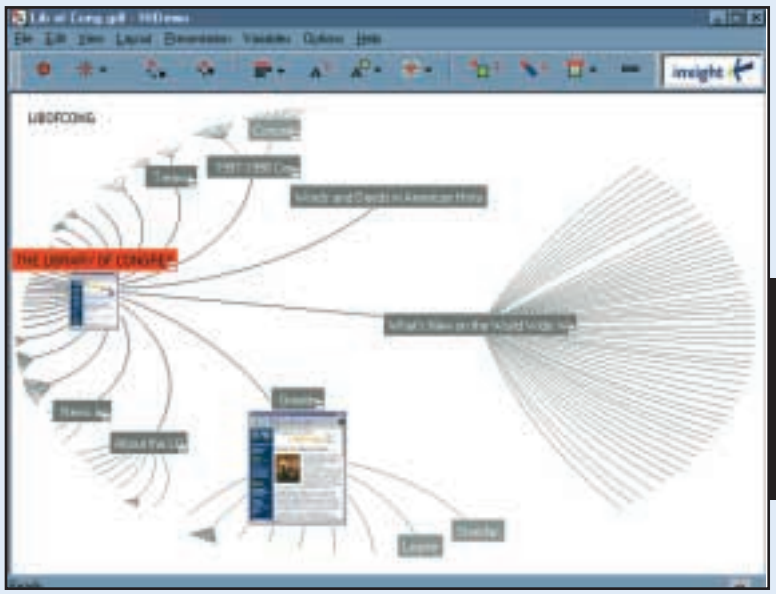


Fig. 2. *Inxight's Java-based Hyperbolic Tree (Image courtesy of Xerox Corp. and Inxight Software, Inc.)*

Center (http://www.parc.xerox.com/istl/projects/uir/) has been at the forefront of much of the visualization experimentation. In 1996, Xerox spun off Inxight Software, Inc. (http://www.inxight.com), with a small product line of advanced user interface designs. Inxight's Web site includes about 50 demonstrations, including, of all things, maps of the Library of Congress Web site and the Kennedy Family Tree.

Inxight's premier software package is its $4995 Site Lens Studio (available at reduced cost to educational institutions). Site Lens Studio enables a site designer to "map" various documents and Web pages housed within a site. This "map," which is really a Java applet, is then portrayed in what Inxight calls a "Hyberbolic Tree." Users can manipulate the tree in order to examine tree "branches" more carefully; clicking on a particular "node," for example, will bring subsequent branches into a more detailed view. Users retrieve a mapped Web page or document by clicking on the branch's label.

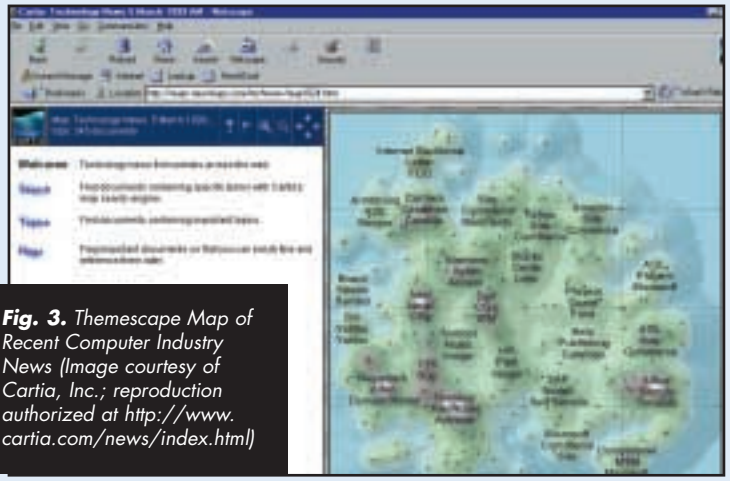


Fig. 3. *Themescape Map of Recent Computer Industry News (Image courtesy of Cartia, Inc.; reproduction authorized at http://www.cartia.com/news/index.html)*

As you can see from the image reproduced here (Fig. 2), the Hyperbolic Tree excels at reproducing the contents of a given Web site in a straightforward, visually dynamic way. However, what it cannot always tell the user is the relevancy of any of the documents hanging from the tips of the Tree's branches. Another current limitation of the Hyperbolic Tree is that it incorporates only the title of the HTML document; it does not scan the text of the entire document. The result is that since few Web designers are accustomed to paying detailed attention to HTML page titles, many Tree branches could end up being poorly named. In order to minimize some of these shortcomings, Inxight also offers its new "Summary Server," a server-side application that can be paired with Site Lens Studio. Summary Server enables a user to click on a branch of a Hyperbolic Tree and retrieve short, extracted bits of information about the document shown by the Tree branch.

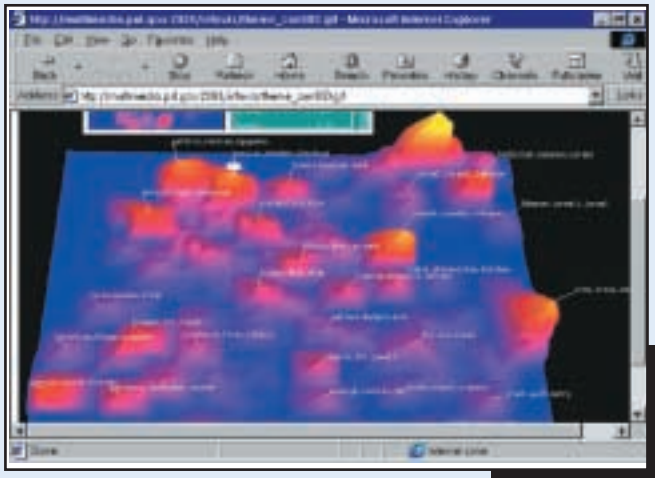


Fig. 4. *PNNL's SPIRE-based ThemeView (Image courtesy of Battelle Memorial Institute and Pacific Northwest National Laboratory)*

Another software vendor's entry into the information visualization market also represents information in the form of maps. **Cartia, Inc.** (http://www.cartia.com), a new venture relying on technology created for the intelligence community by the Federal Pacific Northwest National Laboratories, uses a type of "topographic map" to demonstrate the quantity of, and relationships among, documents. Cartia's "Themescape" software scans the full text or abstracts of documents and then "maps" them—that is, places them near each other when the content is similar or separates them when the content is disparate. Each group of similar documents makes a peak, while widely divergent documents make valleys. You can easily see (Fig. 3) how a map like this might aid in sifting through discovery documents to find all references relating to, say, Dr. Smith's dental surgery deposition. Those documents with the greatest cluster of similar terms would be located at the top of the peak, while tangential references would be located around the base of the mountain. Cartia users can run searches for keywords in documents, with results appearing as numbers superimposed on the map. Therefore, by running a search for "Bill Clinton" against the current day's "NewsMap," you can discover, for example, that the President spoke out on the Yugoslav crisis, but said nothing about peace in the Middle East. The suggested price for the software license is $20,000—steep enough to place Themescape only in the largest corporate intranets and Web sites.

Other research experiments that use the same root technologies developed for Cartia's Themescape are also promising. The **Federal Pacific Northwest National Laboratory** (PNNL, at http://www.pnl.gov/infoviz/) has been working on other information visualization schemes for a number of years and has produced a variety of other programs with such high-tech names as "Galaxie," "Topic Islands" (Fig. 5), and "Cosmic Tumbleweed." Many of them are based on the Lab's SPIRE Project (Spatial Paradigm for Information Retrieval and Exploration) (Fig. 4). Although these applications are still in the experimental phase, they promise much of the same versatility as Cartia's finished product.

Manning & Napier (http://www.mnis.net) is the only information content

continued on page 8

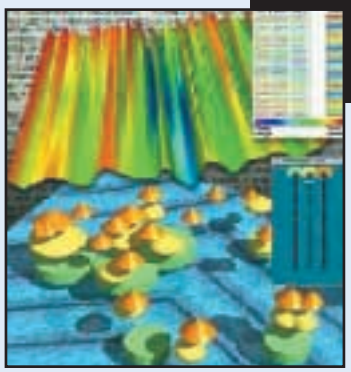


Fig. 5. *PNNL's Topic Islands (Image courtesy of Battelle Memorial Institute and Pacific Northwest National Laboratory)*

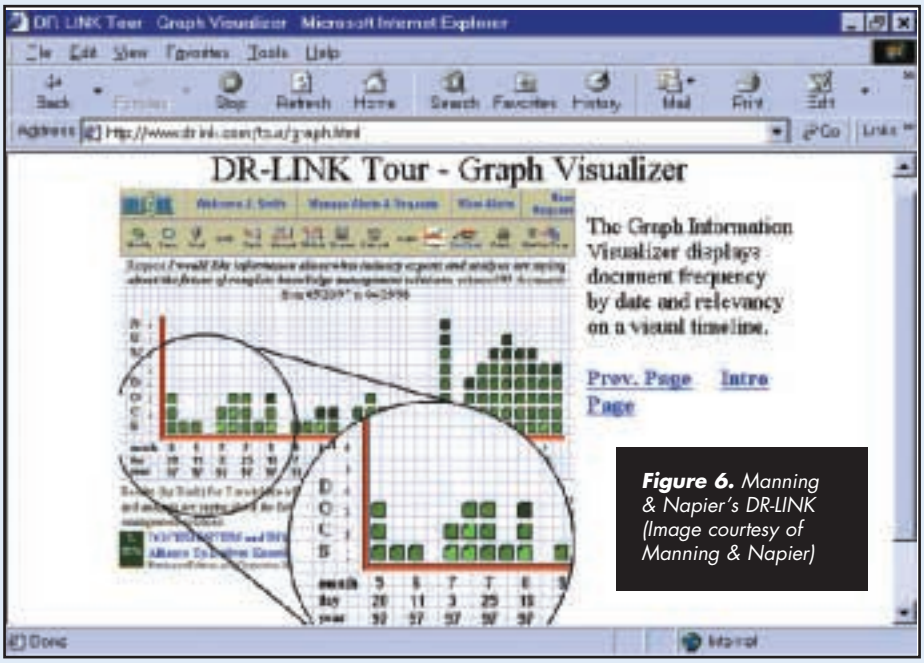
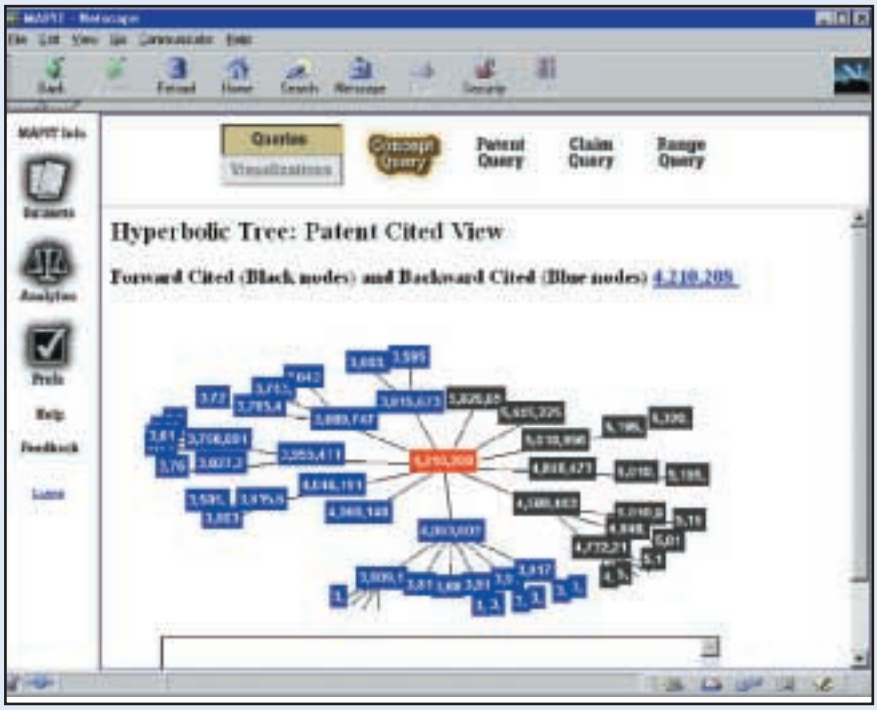


Figure 6. *Manning & Napier's DR-LINK (Image courtesy of Manning & Napier)*

provider to enter this market thus far. One of its products, DR-LINK (*D*ocument *R*etrieval Using *L*inguistic *K*nowledge), generates familiar-looking graphs and bar charts to better indicate the relevance of returned documents (Fig. 6). A more visually centered offering, however, is MAPIT (*M*anagement and *A*nalysis of *P*atent *I*nformation *T*ext), which discovers the relationships between patents by showing a hyperbolic tree, similar to Inxight's. Each branch of the hyperbolic tree represents a different patent, presenting an easily recognizable representation for the patent researcher (Fig. 7). Pricing for these products is available from the vendor.

Information visualization is a quickly-growing field, necessitated by the tremendous increase in Web-based information. The systems sampled above represent only a relatively small portion of all the experiments being performed in university computer science departments and industry laboratories. Law librarians can turn to one of the sources listed below to find out more about information visualization products and research. Perhaps one day in the near future, a picture really WILL say a thousand words, and our jobs will be just that much easier ... and more interesting.

Fig. 7. *Manning & Napier's MAPIT for Patents (Image courtesy of Manning & Napier)*



Steven Anderson *(sanderson@gfrlaw.com) is Librarian at Gordon Feinblatt Rothman Hoffberger & Hollander in Baltimore, Maryland.*

To Learn More...

The Big Picture: Visual Browsing in Web and Non-Web Databases (http://www.public.iastate.edu/~CYBERSTACKS/BigPic.htm).

Stuart K. Card et al, *Readings in Information Visualization: Using Vision to Think,* Morgan Kaufman, 1999, ISBN: 1-55860-533-9.

Marti A. Hearst, "Interfaces for Searching the Web," *Scientific American,* March 1997 (http://www.sciam.com/0397issue/0397hearst.html).

Information Visualization Resources on the Web (http://graphics.stanford.edu/courses/cs348c-96-fall/resources.html).

Louis C. Vroomen's *Visualization Information & Report Links* (http://www.crim.ca/~vroomen/mainPages/visual/info.html).

Mark Dubin's *Bookmarks: Data Mining/Visualization* (http://spot.colorado.edu/~dubin/bookmarks/b/500.html).

"Ten Trends for the Post-PC World: What's Next in the New Era of Ubiquitous Computing," *Red Herring,* December 1998 (http://www.redherring.com/mag/issue61/trends.html).