

# VISUALIZING ONLINE ACTIVITY

**T**he dot-com revolution established the e-channel as a critical component of corporation/customer interaction. A characteristic aspect of this channel is its rich instrumentation, whereby every visit, click, page view, purchasing decision, repeat visit, and other fine-grain detail describing online visitor activity can be captured automatically on the fly and stored for later analysis. The problem for business managers and site operators, as well as for Web site designers, is that the huge volume of relevant data overwhelms conventional analysis tools (such as spreadsheets and reports). To overcome this problem, I describe how my company, Visual Insights, and other organizations have developed information visualization tools for displaying Web site structure, navigation, paths, flows, and activity.

Information visualization is a research area in computer science that focuses on creating rich visual interfaces to help users understand and navigate through complex information spaces. Data sets are frequently large and time-varying and usually involve multi-dimensional structure. Unlike scientific visualization, where the research focus is creating 3D representations of physical phenomena, the fundamental problem for developers of information visualizations is that, because the data is nonspatial, it lacks natural physical representation. Thus, the information visual-

---

*Look and learn which page of perhaps thousands Web site visitors linger on before clicking yes and no, which drove them to the exit button, and many more actionable patterns, trends, and insights.*

*Stephen G. Eick*

---

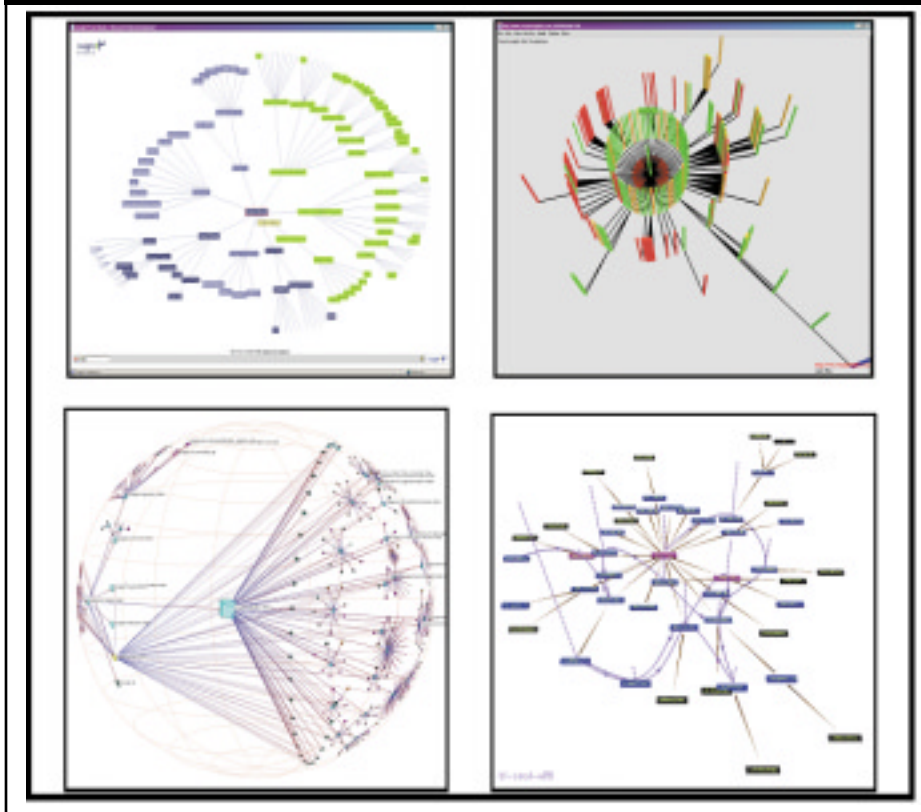
ization research challenge is how to invent new visual metaphors for presenting information and developing ways to manipulate these metaphors to make sense of the information.

Applied to e-business data, information visualization designers face three significant issues:

**Scale.** There are a vast number of Web sites, each of which may be arbitrarily complex. The largest of them in terms of numbers of page views accommodate millions of visitors per day. The combination of complexity and volume overwhelms traditional approaches to calculating and analyzing Web site activity. For effective analysis, information visualization systems have to manipulate and process exceedingly large, real-time data sets.

**Dimensional complexity.** It is now technically feasible and cost-effective for any organization operating a Web site to collect rich, fine-grain, online data sets. Integrating online and offline activity further increases the dimensionality of the problem. Moreover, many of the most interesting and important analysis problems involve correlating site activity with enterprise data.

**Figure 1. Four visualizations of Web sites: Inxite Software's Hyperbolic Tree (top left), showing site structure; Visual Insights' site map (top right), showing Web page use; Silicon Graphics' site manager (lower left); and a visitor's path through a Web site (lower right).**



- Visualizing site structure as a visitor navigation aid;
- Showing paths and flow through the site to help designers build a more effective site; and
- Monitoring the site's real-time activity to help site operators run their businesses more efficiently.

For each related problem, researchers have developed new visual metaphors, interactions, and domain-specific applications that help user/viewers understand e-business data.

### Visualizing Site Structure

A classic information visualization problem involves visualizing the structure of an information space. Typically, the pages are treated as nodes in a graph and hyperlinks or directory

*Range of analyses and navigation tasks.* Consider, for example, the goals of a site visitor. From this perspective, important tasks include navigating without getting lost, finding a recently accessed page, and understanding the thousands of apparently legitimate hits that are returned from an insufficiently specified search. From the site author's perspective, typical tasks involve identifying which specific pages on the site are viewed most frequently, where on the site visitors are most likely to find interesting content, pages with high dwell times, favorite entry pages, and most likely exit pages. From the site operator's perspective, it is important to be able to quickly identify and repair site errors, find pages with broken links, fix server problems, and ensure the IT infrastructure is operating properly. From a marketing and business perspective, it is critical to be able to understand how visitors found the site, which promotions are most effective at driving profitable traffic to the site, which content visitors find most appealing, and the relationships between online and offline marketing strategies.

I focus here on three areas in which information visualization makes a significant contribution to understanding online visitor activity:

structure as edges. For site visitors, visualizations of site structure (Web site maps) serve as navigation aids. For site designers, these maps can be used to ensure the site has an easy-to-understand structure. For researchers, the visualizations can help gain insight into commonly used site designs.

A popular way to visualize the structure of hierarchically organized sites employs a tree metaphor (see Figure 1). The trees are frequently laid out radially, with the root page in the middle and pages at various depths positioned in circles around the root node. The radial layout utilizes screen real estate more efficiently than a vertical layout. As the number of nodes increases, horizontal tree layouts use  $\log(n)$  vertical and  $n$ -squared horizontal screen real estate. For larger trees, the aspect ratio becomes very wide and does not match normal displays. Lines between the nodes show tree structure. The radial layout works well, since many sites are designed hierarchically, with progressively more detail and levels of finer information. Sites with more complex structures can be coalesced into a tree by eliminating back links and ignoring the links that skip hierarchical levels. Ignoring symbolic and back links is effec-

# THE INFORMATION VISUALIZATION RESEARCH CHALLENGE IS HOW TO INVENT **NEW VISUAL METAPHORS** FOR PRESENTING INFORMATION AND DEVELOPING WAYS TO MANIPULATE THESE METAPHORS TO MAKE SENSE OF THE INFORMATION.

tive for understanding site structure but ineffective for navigation tasks.

Straightforward implementations of tree layout algorithms work well for relatively small sites with perhaps tens to hundreds of pages. The information visualization research challenge for visualization software developers involves scaling the technology to sites with hundreds to thousands or even tens of thousands of nodes. For these sites, screen real estate limitations make it impossible to show every page simultaneously. A general strategy for aiding users navigating a site is to distort the layout, so the pages close to the visitor's current location are shown in full detail and distant pages on the periphery are shown in less detail or not at all.

## Visitor Navigation

Figure 1 shows four information visualizations using flavors of the tree-layout technique. Developed at Xerox PARC, the hyperbolic tree (upper left) organizes pages in a tree laid out in hyperbolic space [1]. Since hyperbolic space provides an exponential amount of room, there is a natural “impedance match” between the exponential growth in number of nodes and the space available as tree depth increases.

The hyperbolic tree can be used as a site map and browser navigation tool. As browsers click on any node, the associated page is displayed in the main browser frame. This visualization technique provides an overview of the site structure, shows a visitor's current location in relationship to the site, and is highly scalable. Effective use of screen real estate allows it to handle sites with thousands of pages.

*Site usage patterns.* Figure 1 (upper right) shows the landscape visualization of my company's corporate Web site (see [www.visualinsights.com/](http://www.visualinsights.com/)). Each page is represented as a 3D vertical, whose color, height, and shape encodes page attributes, including page type (such as html and asp), page errors, average dwell time, number of page views, content, and even number of visitors currently accessing the page. The glyphs are arranged in a 3D landscape and positioned radially, with the position showing the location of the corresponding page in the site hierarchy. A particularly

useful encoding (not shown) ties the number of page views to glyph size. This encoding causes the pages viewed most frequently to stand out visually.

Further layout improvements might make better use of screen real estate. For example, in a data-driven layout, the difference in radius between one level of the hierarchy and its descendants at the next level could be made to depend on the number of nodes in the lower level. This layout would position the nodes more uniformly and thus make more efficient use of screen space.

Landscape visualizations of site activity are especially useful for helping site operators understand usage patterns. Tying statistics to glyph attributes (such as color, so pages with many request errors are red) makes this display effective for quickly identifying site problems, showing frequently accessed pages, and identifying and isolating unaccessed content.

*Site authoring and management.* Figure 1 (lower left) shows the H3 hyperbolic site viewer, developed by Tamara Munzner while at Stanford University [2, 3]. Using a sophisticated two-pass algorithm to organize pages in hyperbolic space, it results in pages laid out on a hemisphere using a non-Euclidean distance metric, ensuring there is exponential room to place nodes and enabling it to cope with large Web sites. The H3 viewer is also interactive; rotating the sphere gives the viewer a fixed target frame rate to maintain interactive performance. Moreover, in practice, it is highly scalable and easily handles sites with tens of thousands of nodes.

Silicon Graphics employs this technique in its site manager tool, which provides a visual directory showing the hyperlink structure of a Web site in a 3D sphere. Users rotate the sphere's structure and “zoom in” on a point of interest to examine a specific document's link hierarchy, as well as how this closeup fits into the entire Web site. It is also possible to animate the statistics from access logs to show a graphical representation of visitors' paths through a site.

## Paths and Flows

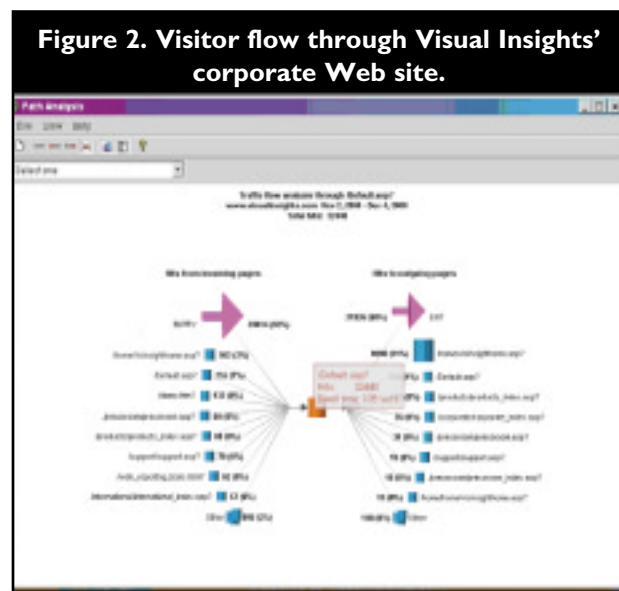
The goal of a path analysis is an understanding of the sequences of pages (URLs) traversed by a visitor within a site. By understanding typical visitor brows-

ing patterns, or the hyperlinks that are actually followed, and on which pages visitors linger, Web site designers can create more usable sites that are more likely to increase visitor satisfaction. Flow analysis aggregates the navigation experience and site behavior of individual users to understand the overall traffic patterns within a site.

**Path visualization.** One way to show paths for an individual visitor superimposes a trace of the pages visited on top of a site map. This idea is similar to showing how an automobile driver might progress through a road network by highlighting the roads traveled on a road map. The software tool VISVIP, which visualizes data on Web-based applications as part of the WebMetrics tool set, developed by John Cugini, a visualization researcher at the U.S. National Institute of Standards and Technology ([zing.ncsl.nist.gov/webmet/](http://zing.ncsl.nist.gov/webmet/)), implements this page-tracing technique for path and timing data [4]. Figure 1 (lower right) shows a visitor's path using a curved purple line connecting each page visited. Pages are positioned out as a directed graph on a 2D plane using a

times varying by some predefined amount of time from the average. The Figure shows flow into and out of the page designated in the center (Default.asp), the company's home page. The large pink arrows labeled Entry and Exit show the number (and percentage) of visitors entering (92%) the site at this page and exiting (60%) the site at this page. Thus, 8% of the visitors entered through bookmarked pages or other links pointing to internal pages on the site. The symbols listed vertically show other pages from which traffic flows into and out of Default.asp in decreasing order. The lower pane in the interface is a html viewer that displays the designated page. Clicking on any page re-centers the analysis to show the flow into and out of that page and displays the page in the html viewer. By progressively clicking on pages or selecting from a list (not shown), the interface enables user/viewers to track flows and navigate through the site by tracing interesting and important flow patterns.

A useful feature of the information visualization in Figure 2 is its ability to transition between flow and path analyses. For example, clicking on the list icon activates a list selection control, and mousing on a page lets a user adjust the display to show lists of the most common paths from or to any page (not shown). The paths, which are sorted in decreasing frequency, show average dwell times.



force-directed graph layout algorithm. The dotted lines above certain pages indicate those on which the visitor lingered.

**Flow analysis.** Unlike path analysis, flow analysis aggregates over individual paths to find the flows within a site. Flow analysis identifies the most frequently clicked-on links, entry and exit points, and bad links and other problems within the site.

Figure 2 shows a visual interface for tracking the flow on Visual Insights' corporate Web site, identifying common entry and exit pages, the most frequently followed paths, and pages with unusual dwell

### Real-Time Monitoring of Site Activity

For many decision-making processes involving site activity, promotions, offers, and inventory management, traditional daily, weekly, and monthly reporting logs are inadequate for corporate decision makers. IT organizations and providers have long appreciated the need for real-time monitoring and management of networks and other critical infrastructure. E-organizations also need similar real-time business intelligence. For example, in Web-based marketing campaigns, companies can and do change content, adjust banner ads, modify email messaging, and change site content throughout the day. They have to understand campaign productivity as it is happening and make adjustments on the fly. This nimble acquisition, manipulation, and reporting of business data involves measuring how different stimuli influence site traffic flows, site stickiness, and entry and exit points. For e-commerce sites, it involves relating this activity directly to buying behavior. There is no point, for example, in stimulating more demand for a promotion if inventory is running low, if the site is experiencing technical problems, or if a weather pattern will delay product shipments.

eBizLive from Visual Insights is a system that col-

lects and presents real-time displays of site activity. Unlike network monitoring systems that focus on errors, low-level packet counts, and link availability, eBizLive looks to visualize visitor and site activity; eBizLive Portal (see Figure 3) extends the sophisticated application framework of Microsoft Commerce Server 2000 for building e-commerce Web sites. Using an administrative console, groups of related pages and URLs are combined into “watch lists.” Each cylinder represents a watch list being monitored, including the referring URL, promotion type, catalog entry, or item. Server errors and purchases are shown on the far left and right, respectively. The height of



each cylinder represents the number of page views during the last sample. The trailing graph associated with each cylinder shows historical trends. Visitors moving through the site are metaphorically represented by animating a glyph that “jumps” between the pages. The size of the animating glyph encodes the number of visitors moving between the pages.

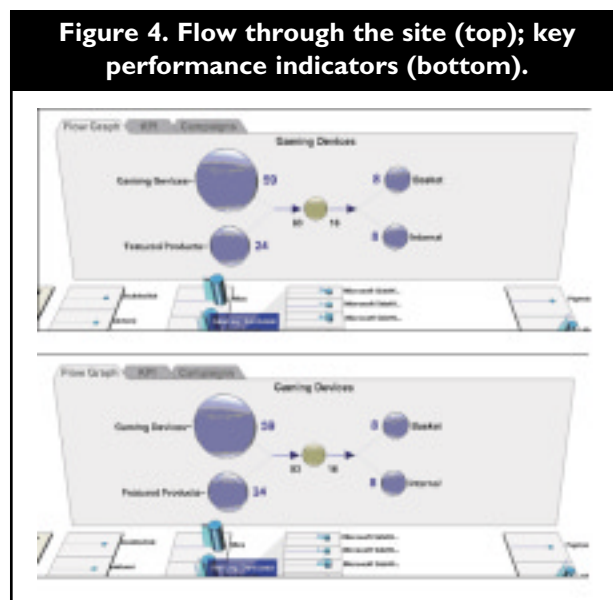
The display in the figure is organized into a “floor” and a “back wall.” On the floor, the cylinders corresponding to the referrals typically represent visitors entering the site prompted by marketing campaigns. Catalog entries and products are organized hierarchically, like aisles in a store. Clicking on a catalog entry expands the aisle to show the products within that aisle. The figure shows that the user has expanded the catalog entry for “Gaming Devices” to view activity by device type. This drill-down capability can be organized by category or by actual product pages. Each cylinder in the “Buy Pipeline” sector shows activity by stage, including “Basket,” “Ship To,” “Ship Item,” and “Payment Type.” Organizing data hierarchically yields the added benefit of increasing the scalability of eBizLive.

A site overview provides instant value, showing

where visitors entering the site are coming from, what they do while on the site, and how long they stay. Instantly apparent are site traffic flows, errors, and problems. Problems can therefore be fixed and opportunities addressed immediately, rather than days later when reports are published.

The tabbed pane on the back wall of the display organizes three displays: Flow Graph, Key Performance Indicator (KPI), and Campaigns. The Flow Graph shows traffic into and out of the selected watch list (see Figure 4, top). The size of the disks represent traffic flow to and from the selected watch list; the numbers show the actual counts. The KPI tab (bottom), displays KPIs using a time-series graph and textual displays. Common KPIs for e-commerce sites are page views, purchases, number of visitors, dwell time, and conversion rates. The Campaigns tab presents the effectiveness of any ongoing marketing promotion.

Brushing over any item with the mouse causes detailed information about that item to be displayed using a transparent pop-up text field. Using the mouse and keyboard accelerators, a user can zoom in and out and move around the scene. Clicking on the Left, Home, Wall, and Right buttons at the bottom moves the scene to fixed viewpoints. Catalog entries



and items along the flow can be sorted and filtered.

The site activity shown in Figure 3 is the Web site analogue of store activity. For example, in large department stores, managers often work out of a second-floor office that provides a bird’s eye view of flow patterns among the departments, high-activity areas, aisle obstructions, and underutilized sections of their stores. They see overall store activity that frequently

correlates with transaction activity. In the same way, eBizLive's site-activity display provides a gestalt of crucial site activity.

## Conclusion

Information visualization helps users, managers, and executives cope with the rich data sets routinely collected by online systems. Perhaps the most significant difference between traditional channels and the new self-service e-channel is the e-channel's richer instrumentation. For example, in the old less-information-oriented economy, a department store matching credit card receipts might know that a customer bought shoes in the men's department and a tennis racket in the sports department. In the new economy, an online magazine publisher knows which ad impressions and marketing messages are directed at which individual customers, which ones they clicked on, the content they browsed, which articles they actually read, and how long it took them to read each article; it might even provide buttons to let readers rate the quality of each article.

The e-business challenge for executives and managers is how to make sense of this exceedingly rich new information source. The first problem for any kind of e-business—collecting the clicks and building a Web warehouse providing an integrated view of both online and traditional customer activity—is technically difficult but can be handled through existing technology. Much more difficult is making sense of the information and gaining insights that lead toward actionable results.

The information visualization challenge is to create better interfaces, more effective visual metaphors, and new analysis techniques, addressing three information management goals: site structure (from the perspectives of the visitor, operator, and author); path and flow analysis; and live site monitoring.

However, before these techniques become mainstream, visualization tool developers need to address three much deeper research challenges:

**Scalability.** The largest Web sites include thousands of pages and attract millions of visitors. Information visualization software tools need to increase the scalability of the techniques discussed here by two orders of magnitude to display large sites.

**Support action.** Information visualizations do not by themselves create value but lead to valuable insights. Turning insights into decisions and activities that add value requires an action step. For example, a flow visualization might show high exit rates during evening hours from a particular page with complex graphics. In another, eBizLive might show many buyers abandoning the buying process due to an out-of-

stock inventory problem. Reducing graphics complexity involves integrating the visualization software with the system used to create the site; solving an inventory problem might involve supply-side integration between procurement and shipping. Tight linking of visualization software with support for action is necessary for the information visualization to yield value.

*Taxonomies that define the appropriate analysis problems an information visualization might address.* Besides the three examples considered here—visualizing site structure, showing paths and flow through a site, and monitoring e-commerce activity—other problems involved in generating actionable insight include:

- Visualizing visitor segmentations;
- Understanding site profitability, especially in money-losing dot-coms;
- Correlating site activity with promotional campaigns; and
- Displaying normal and abnormal browsing patterns.

The research agenda for developers of visualization tools is to partition the subproblems of creating visual metaphors to represent data into orthogonal categories, create taxonomies, understand which tasks are common and which unique, and build visualizations that fill holes in the taxonomies. **C**

## REFERENCES

1. Lamping J. and Rao, R. Laying out and visualizing large trees using hyperbolic space. In *Proceedings of the ACM Symposium on User Interface Software and Technology (UIST)* (Marina del Rey, CA, Nov. 2–4). ACM Press, New York, 1994, 13–14.
2. Munzner, T. *Interactive Visualization of Large Graphs and Networks*, Ph.D. Dissert., Stanford University, June 2000; graphics.stanford.edu/papers/munzner\_thesis.
3. Munzner, T. Exploring large graphs in 3D hyperbolic space. *IEEE Comput. Graph. Appl.* 18, 4 (July/Aug. 1998), 18–23.
4. Cugini, J. and Scholtz, J. VISVIP: 3D visualization of paths through Web sites. In *Proceedings of the International Workshop on Web-based Information Visualization (WebVis'99)* (Florence, Italy, Aug. 30–Sept. 3). IEEE Computer Society Press, Los Alamitos, CA, 1999, 259–263; in conjunction with the 10th International Workshop on Database and Expert Systems Applications (DEXA'99) (Florence, Italy, Sept. 1–3, 1999).

---

STEPHEN G. EICK (eick@visualinsights.com) is the chief technology officer of Visual Insights, Naperville, IL.

---

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.