Visualizing the Crowds at a Web Site

Nelson Minar
MIT Media Lab
E15-305 20 Ames St.
Cambridge, MA 02139 USA
http://www.media.mit.edu/~nelson/
nelson@media.mit.edu

Judith Donath
MIT Media Lab
E15-428 20 Ames Street
Cambridge, MA 02139 USA
http://judith.www.media.mit.edu/
judith@media.mit.edu

ABSTRACT
A visualization of the crowds of people visiting a web site is developed. Visitors are drawn as icons on a map of the web site; the animation of people's movements conveys the crowd dynamics of visitors. The visualization combines three pieces: a map of the web site, an iconic representation of individual visitors, and an interpretation of the dynamics of visitor patterns to show crowd phenomena. The effect is to make a web site look like a social, active space.

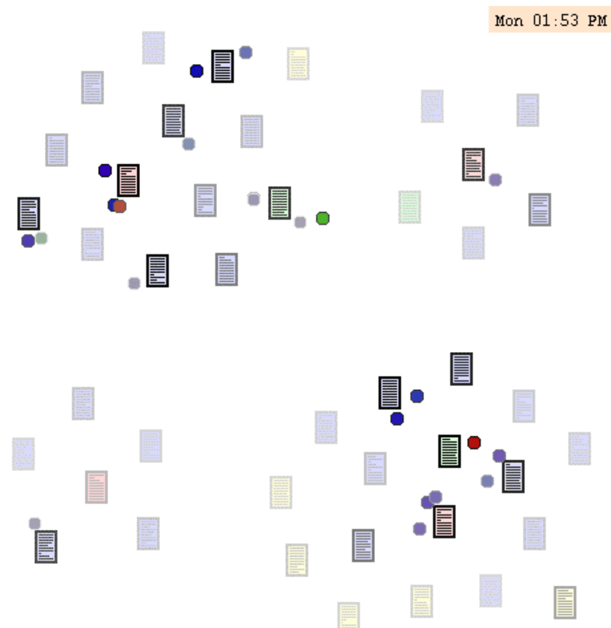
Keywords
Web visualization, crowds, dynamic visualization.

INTRODUCTION
Our everyday social spaces are rich environments, full of people milling around, loitering, walking purposefully from place to place. By contrast, World Wide Web sites today feel like empty, lonely places. When we visit a web site, or even when we own one, we have no sense of the people visiting the site. Where are people going? Where do they linger, how do they move through the site, how is the space inhabited?

The urban environment is a source of inspiration for visualizing and understanding the dynamics of crowds of people [2, 4]. The camera makes depicting the real world crowd relatively easy; time-lapse movies of crowd formation and dispersal make quite compelling visualizations of crowd dynamics.

What would be an equivalent visualization for the crowds of people visiting a web site? The viewer should be able to easily see what parts of the site are popular, what paths people take while navigating the site, as well as get a feel for the overall activity on the site.

VISUALIZATION
Effectively visualizing the crowd on a web site requires solving three distinct problems. A map of the site must be created to provide spatial structure, a representation of individual people is needed to show the visitors, and an animated display is necessary to show the crowd dynamics. This work concentrates primarily on the latter problem, crowd dynamics, but all three components are important.



Crowd Visualization

The figure above shows the implemented visualization of a large, active web site at the MIT Media Lab. Each document icon represents a collection of logically related pages. Circles represent people near the documents they are currently visiting. This image is one still frame from a dynamic visualization. The full dynamics can be seen as a Java applet on the authors' web site.[1]

Site Map
For the purpose of this work, a simple map was constructed based on site-specific knowledge. Pages on our web server can be broken up into natural groupings. For example, all of the author's pages can be grouped as /people/nelson, and all of the Software Agents group's pages are under /groups/agents. For this visualization, a single icon represents all pages in one group. The icon's color represents the kind of group it is (research group page, personal page, project page, etc.). The page placement

[1] http://www.media.mit.edu/~nelson/research/crowdvis/

reflects a hand-constructed concept of how the site is organized.

This kind of map is limited and site-specific; building a map of a web site is a large research topic in its own right. Many options are possible, from something as simple as basing a map on the literal hyperlink graph structure of the site to sophisticated approaches such as using self-organizing maps [1] or examining visitors' navigation patterns [3]. Future versions of this visualization would benefit from an algorithmically constructed site map.

Individuals

Much like people briefly seen in a crowd, there is little to distinguish a single visitor to a web site. The main information available about a visitor is the IP address of the visitor's computer. For this visualization, we simply assume each unique IP address is a visiting individual, and represent the person with a small dot colored according to their top-level domain name. While this representation is a gross simplification, it leaves the way clear to concentrate on crowd dynamics.

Crowd Dynamics

The most important part of this social visualization is the dynamics, the animation of the crowds of visitors. The goal is to make it possible to see the web site as an inhabited, social space, full of visitors passing through.

With this visualization, popular groups of pages on the web site are visibly crowded with people jostling around as they move from page to page. Seldom-visited portions of the site are easily identifiable, being relatively empty and faded. The track of an individual person through the site can be followed, revealing a particular person's interest in the site. Group dynamics are also visible in the aggregate movements of all visitors. For example, heavy traffic between the pages for the Agents group and the Agents' advisor make the connection between these pages visually apparent.

The dynamic visualization is implemented as a multithreaded Java applet. Two threads run concurrently: an animation loop and a web log reader. The web log reader reads in data from the web server logs, parses out relevant information about visitors, and updates the data structures that the animation loop uses for display. The reader plays back the web server log data in accelerated time. Displays of 50x real-time are appropriate for the moderate traffic level of our site. A clock in the corner shows the current time in the log file.

Every entry in the log file yields data about a person visiting a particular page. If the person has never mentioned in the log file before, then a new icon (a colored circle) is created to represent him or her and placed next to the page they first visited. If the person has been to the site before and is fetching another page in the same page group, then the icon is jittered a bit around the page to convey activity. Finally, if the person is going to a page in a different group, then the icon is animated along a path from the old page to the new one, showing the motion of the visitor.

The animation loop renders people's visits to the web site. In addition, the display is also aged to show the timeliness of the information. If a visitor to the site is not heard from in awhile, his or her icon is faded out until it disappears. If a person has not been heard from in five minutes, they are assumed to have left the site. Similarly, if a document has not been visited recently it is partially faded so that more active parts of the web site are emphasized. The combination of animation and aging creates a dynamic, timely display.

CONCLUSION

Three elements are sufficient to visualize the behavior of crowds on a web site: a map of the site, a representation of visitors, and a carefully constructed animation to show crowd dynamics. Future work will concentrate on improving the site map, making it less site specific by using developed web site mapping techniques. More work can be done in making people's movements more naturalistic, by using a higher frame rate and making sure individuals do not stand on top of each other. Finally, these visualization techniques should be applied and tested on sites with heavier traffic. For example, a crowd visualization of a major news web complex could be quite revealing about the dynamics of people's interest in current events.

ACKNOWLEDGMENTS

We would like to thank Alan Wexelblat and Adriana Vivacqua for their critique of this work.

REFERENCES

1. Honkela, Timo, et al. Self-Organizing Maps of Document Collections. In proceedings of WSOM'97, Workshop on Self-Organizing Maps, Helsinki University of Technology, Finland, June 4-6, 1997, pp. 310—315.
2. Milgram, S. 1969. The drawing power of crowds of different size. In *The Individual in a Social World, 2nd edition.* New York: McGraw Hill
3. Wexelblat, Alan and Maes, Pattie. "Footprints: History-Rich Tools for Information Foraging." CHI '99 Proceedings, ACM Press 1999.
4. Whyte, William H. 1988. *City: Rediscovering the Center*. New York: Doubleday.