

Visualization Components for Persistent Conversations

Marc A. Smith, Andrew Fiore

September 21, 2000

Technical Report

MSR-TR-2000-98

Microsoft Research
Microsoft Corporation
One Microsoft Way
Redmond, WA 98052

Visualization Components for Persistent Conversations

Marc A. Smith, Andrew T. Fiore
Collaboration & Multimedia Group
Microsoft Research
One Microsoft Way
Redmond, WA 98052 USA
+1 425 936 6896
{masmith, t-afiore}@microsoft.com

ABSTRACT

An appropriately designed interface to persistent, threaded conversations could reinforce socially beneficial behavior by prominently featuring how frequently and to what degree each user exhibits such behaviors. Based on the data generated by the Netscan data-mining project [9], we have developed a set of tools for illustrating the structure of discussion threads like those found in Usenet newsgroups and the patterns of participation within the discussions. We describe the benefits and challenges of integrating these tools into a multi-faceted dashboard for navigating and reading discussions in social cyberspaces like Usenet and related interaction media. Visualizations of the structure of online discussions have applications for research into the sociology of online groups as well as possible interface designs for their members.

Keywords

Visualization, persistent conversation, asynchronous threaded discussions, Usenet, newsgroup, social cyberspaces

INTRODUCTION

In William Gibson's now-famous vision, cyberspace is

“A consensual hallucination experienced daily by billions of legitimate operators, in every nation ... a graphic representation of data abstracted from the banks of every computer in the human system. Unthinkable complexity. Lines of light ranged in the nonspace of the mind, clusters and constellations of data. Like city lights, receding.” (William Gibson, *Neuromancer*)

In practice, however, the landscape of cyberspace has been difficult to see. There is no single, unified representation of cyberspace or even of its major components. The tools used to connect to social cyberspaces leave us blind to a range of information that is readily visible in face-to-face interaction. Most tools display only the leaves – the messages, posts, texts, or pages – of the vast, branching patterns of relationships and dependencies in computer-mediated interaction. Few present the branches; none reveal the trees and the forest. Information about the number of people who are or have been present, how they are associated, their histories, and the patterns of activity of the group as a whole are missing entirely.

An ideal interface would allow the user to view simultaneously the set of messages in a given time span or newsgroup, the corresponding set of people, and the interactions within and between those sets. Giving users the power to see messages, people, and their intersection is not just a usability improvement; it also allows designers to shape social cyberspaces to beneficial ends. Visualization tools and novel interfaces built around them can explicate interpersonal relationships, revealing individuals' roles and importance. The pervasive availability of such information provides incentive for users to improve their social standing as presented in the visualization by helping others, answering questions, or performing other actions that contribute to the community. An appropriately designed interface could reinforce socially beneficial behavior by prominently featuring how frequently and to what degree each user exhibits such behaviors. The designer of the interface or visualization that reveals these data could therefore encourage the particular kind of interaction that will foster quality content and user satisfaction simply by exposing a metric of that interaction to all users.

This paper describes a collection of visualization components that illustrate patterns of activity and conversational structure in Usenet newsgroups and, potentially, any threaded, persistent messaging system. Each component highlights one of the many possible ways of viewing conversational structures like those found within Usenet. We integrated these components into a Web-based “dashboard” that provides new interfaces to support newsgroup discovery, thread exploration, and message

selection and review. We present descriptions of these components and the data sets that drive them, a discussion of principles and decisions that drove the refinement of the designs, and the results of a user study in which these components were used by experienced newsgroup users.

Usenet

Usenet is a global, distributed system for hosting, storing, and disseminating asynchronous, threaded conversations. Every message that a user contributes to the Usenet system must be directed to one or more newsgroups, which are repositories for discussions on particular topics. (For example, the newsgroup comp.lang.perl.misc is devoted to the discussion of the programming language Perl.) Within the newsgroup, the conversational unit is the thread, usually an initial message and all replies to that message. We offer a more rigorous definition of a thread below, in the description of our Thread Tree visualization.

The visualization components described here are all based on the results of data mining an extensive collection of Usenet messages. The Netscan project consumes an extensive Usenet feed and parses each message's header into a SQL database. Usenet is based on the Network News Transfer Protocol (NNTP) defined in RFC 977 [6] and the message format defined in RFC 1036 [5], which specifies the data structure of messages so that they can be identified uniquely and interconnected to form potentially complex tree structures. After significant data scrubbing and aggregations, the resulting database contains summary statistics on the size, activity, and structure of Usenet at multiple levels (from individual messages up to threads, newsgroups, and hierarchies, including Usenet as a whole) and resolutions (days, weeks, months, and years).

Critique of existing interfaces

The process of browsing Usenet newsgroups has long been counterintuitive. In fact, the term "browsing" is a misnomer; a user of traditional news-reading software cannot pore over the extensive hierarchy of newsgroups the way he or she might browse through a library's selection because Usenet's organizational hierarchy is based solely on a nomenclature that has become largely meaningless (e.g., *alt.alien.vampire.flonk.flonk.flonk*) as Usenet has grown.

The principle of visibility in user-interface design suggests that users can make a selection from many choices far more easily if they are given a list of possibilities than if they must know what choices are valid and type the one they want; this is clear in the superior usability of a graphical menu system to a command-line interface for most users. Modern news-reading software indeed provides a list of all newsgroups available on the local news server, but in practice, such listings are too massive and disorganized to help a user who does not know already which newsgroup he or she wants to read. A semantically meaningless hierarchy of tens of thousands of newsgroups, however nicely the

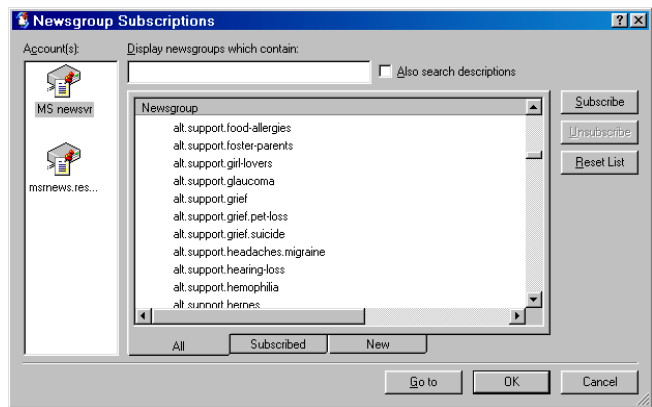


Figure 1. Microsoft Outlook Express's newsgroup selection interface

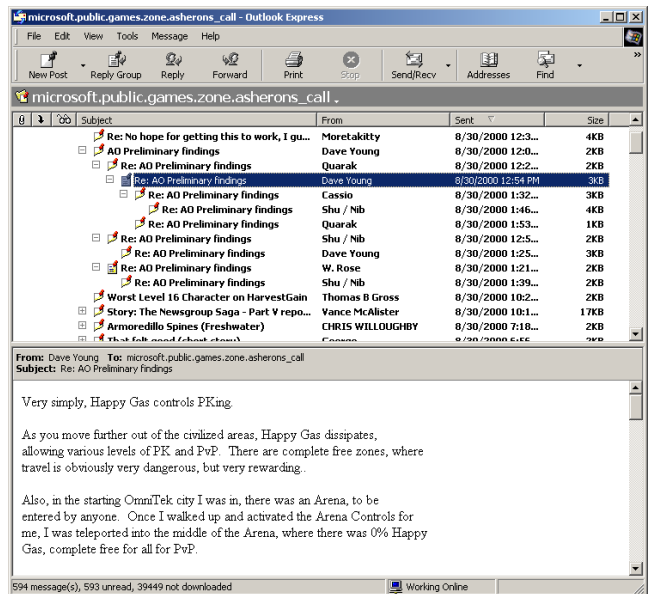


Figure 2. Microsoft Outlook Express's newsgroup browsing interface

software displays it (e.g., Figure 1), offers no insight into the groups' activity, structure, interconnection, or content. Such a listing might as well be a blinking command prompt, especially for novice users.

Once the user has selected a newsgroup to read, current software offers a list of discussion threads in the newsgroup with replies arranged hierarchically beneath each thread's initial message (Figure 2). Users can sort such listings by date, subject line, author, and length to glimpse the conversation's structure, but the whole picture remains elusive. For example, the user can determine who is the most prolific author in a newsgroup by sorting the posts by author and counting, but this is cumbersome and leaves unanswered the more important question: in which threads does the most prolific author participate? The answers to such questions should be obvious at a glance but at present are not.

Although one of the few recent interface innovations in Usenet, the full-text index (for example, the service provided by Deja.com [1]), does a remarkable job of retrieving specific messages of interest, it does a poor job

of conveying the contexts of their creation and the characters and relationships of their authors. This omission of context erodes any sense of social connection and obligation, reducing the likelihood that people will feel compelled to return value to the community from which they just took information.

The current difficulty users have constructing participants' histories undermines their ability to assess one another, develop trust, and cultivate common goals. Although people can eventually develop a refined sense of the rhythms, leaders, and fools in a particular social cyberspace, this information does not come easily or transfer readily to other spaces. Many of the resources and practices that constrain abuse and sustain trust in social relationships are missing in social cyberspaces, thus limiting the effectiveness of established methods of encouraging and maintaining collective projects [7]. This blackout of social context may be responsible for the decline in recent years of Usenet's once-respectable signal-to-noise ratio.

ALTERNATIVE INTERFACES TO SOCIAL CYBERSPACES

The collective result of millions of people using Usenet is a tangled territory of thousands of inter-linked newsgroups, messages, and participants. How can this complex terrain be grasped and visualized? Recent work in this area serves as a useful guide to the variety of forms that are possible.

Donath et al. created Loom [3], a system that illustrates reply sequences in threaded discussions in Usenet newsgroups. Loom presented information about the temporal distribution of each participant's contributions to an individual newsgroup in a format that resembled a "piano roll," a strip of paper with notches that represent the score of a piece of music for a player piano. Temporal and authorial patterns were encoded in rows for each poster that intersected columns that marked off the passage of time. The interface allowed the user to toggle the display of lines that connected posts to those others to which they were replies, thus illustrating the threaded structure of the conversation over time. A relatively simple analysis of message content allowed Loom to color-code nodes based on emotional content.

Sack's conversation maps [10] take a related approach, displaying small glyphs that represent all the threads in a newsgroup over a period of time and their semantic and social interrelationships. Like Loom, these maps are interactive, allowing the user to call up more detailed maps as well as the content of individual messages.

Like these researchers, our goal is to present the social information embedded but obscured in these spaces to help

people navigate and make sense of the discussions that take place there. Following Donath's concept of "social visualization," we have created a system that provides access to the existing message bases of a large newsgroup feed – about 91,500 newsgroups – via an interactive Web interface composed of multiple visualization components. Drawing from a range of prior work in information and data visualization, our designs focus on highlighting the social and temporal patterns present in the data.

NETSCAN VISUALIZATION DASHBOARD

Usenet has many facets and can be visualized at many scales. The challenge lies in finding ways to present these data that connect them closely with the process of reading and contributing to newsgroups. We have created a set of Web components that constitute a first step toward expanding the range of interfaces to conversational spaces like Usenet. Our initial efforts focused simply on delivering information in tabular format from the Netscan database through the Web [9].

This paper documents our effort to move beyond text-only representations of the Netscan data. Using the Vector Markup Language [13], a Web browser-based vector graphics XML schema developed by Microsoft, we developed a set of visualization components that illustrate the structure of individual message threads, the temporal activity of authors, and the roles of authors within threads. This dashboard also included a textual pane that identified the selected newsgroup and reported the thread's subject line, start and end dates and number of active days within that period, total number of messages and authors, and maximum depth (number of generations from root to leaf) and breadth maximum (number of siblings at the same generation).

User Study

To evaluate the efficacy of the components we created, we studied nine people using early implementations of our technologies. The participants, seven males and two females, identified themselves as frequent or expert users of Usenet newsgroups.

In a questionnaire, subjects reported that they checked newsgroups a least a few times every week and that they spent on average between 15 and 30 minutes per session. Their dominant activity involved reading technical support newsgroups followed by discussing news and current events. In these sessions, users mostly read other participants' contributions and rarely posted messages of their own. They identified the subject line, date, and the author as features of messages that were important in deciding to read specific messages and threads.

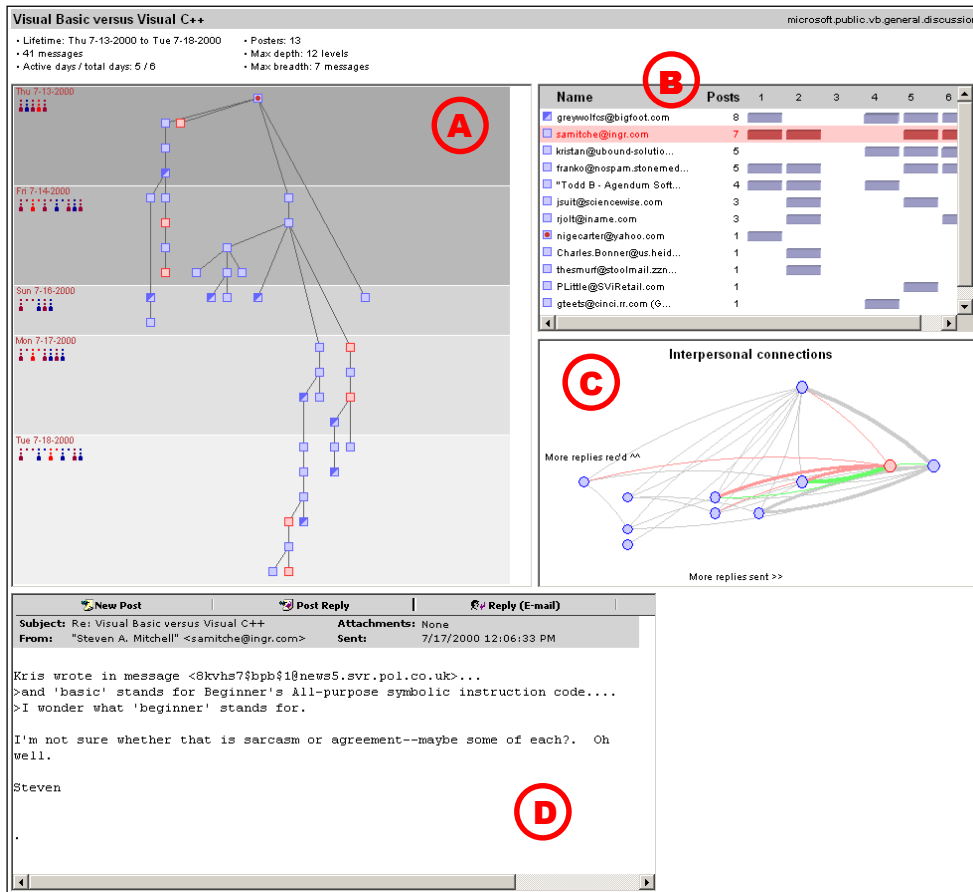


Figure 3. The Netscan dashboard combines the thread tree, piano roll, reply-based sociogram views, message display, and header information.

(A) The thread tree visualization presents information about the structural and temporal history of the selected thread. Selecting a message displays its contents in the message pane (D) below.

(B) The piano roll component displays a list of all posters who are present in the selected thread ordered by the number of posts they contributed. Columns for each day the thread was active contain bars if the poster posted on that day. When the user mouses over each poster, the related posts are highlighted in the thread tree and the interpersonal connection component.

(C) The interpersonal connection component displays a sociogram that relates users with those they reply to and who reply to them. Posters are located based on the number of responses they send (the x axis) and the number they receive (the y axis).

(D) The message display pane presents the contents of a selected message and controls for replying via Usenet or email.

Each participant identified a favorite newsgroup and answered a series of questions about his or her browsing habits, including frequency of use, extent of participation in newsgroups as opposed to lurking, and types of use (e.g., socialization, technical support, commerce).

The participants then selected a newsgroup from the *microsoft.public* hierarchy to explore. Netscan allows users to work with any Usenet newsgroup, but because our data-warehousing software was in transition to a new version during this study, we had to restrict them to *microsoft.public* to ensure the availability of all the necessary thread data.

Their first task was to browse through their chosen newsgroups with Microsoft Outlook Express, a news browser whose interface serves as a reasonable standard for current, threaded news-browsing software, and to describe what they liked and disliked about the experience. Additionally, they identified people and messages they found to be important in these newsgroups.

We then asked the participants to investigate their newsgroups with Netscan's social accounting search engine, which provides access to statistical data about posts and posters in specific newsgroups. (The usability data for the social accounting search engine is part of a separate study.) Next, participants examined their newsgroups with the Netscan visualization dashboard. With written and spoken

questions, we instructed them to find specific pieces of data about the newsgroup – for example, with whom the top poster in a thread interacted the most or how many days the selected thread was active. After each participant had used each tool, we asked him or her to rate whether it was useful and whether it was easy to understand on a five-point scale from “Strongly Disagree” (value=1) to “Strongly Agree” (value=5).

We will present the results of the user study as we describe the visualization components in the following sections.

Thread Trees

Typical news-browsing software obscures information about threads' temporal patterns, population, and structure, particularly the development of parallel branches of conversation. In a Usenet newsgroup, a thread is a collection of messages associated with one another through a pattern of replies. We may conceive of a thread as an n -ary tree with the initiating message at the root. Usenet news servers, however, do not distribute messages in such a coherent package but rather as distinct, unordered units that require extensive analysis to reconstruct their underlying tree structure.

According to the specification for Usenet messages [5], each message receives a globally unique message identifier when it is posted. Furthermore, when an author posts a reply to a previous message, the new message indicates its

parent message – the message to which it is a reply – by citing the parent’s message ID in its own *References* header. If that parent is itself the child of an earlier message, the new message also cites its grandparent and, often, earlier ancestors in its *References* line. Thus, the recent ancestry of a message can usually be determined by examining its references. Using these references, we can reconstruct from Usenet messages of common ancestry an *n*-ary tree.

Netscan *thread trees* (Figure 3a, Figure 4), the centerpiece and most fully developed component of the Netscan visualization dashboard, present a novel way to browse message threads and understand their structure. In a thread tree, the message that begins a thread appears as the root of the tree, from which successive replies branch down. Square glyphs – small, graphical units – represent the messages. We divide the vertical space of the tree view into gray bands that represent calendar days on which messages are posted during the thread’s lifetime. The size of each band is proportional to the number of generations of messages exchanged on that day; days on which no messages were posted do not appear at all but are suggested by a more dramatic shift in shading in the succeeding band.

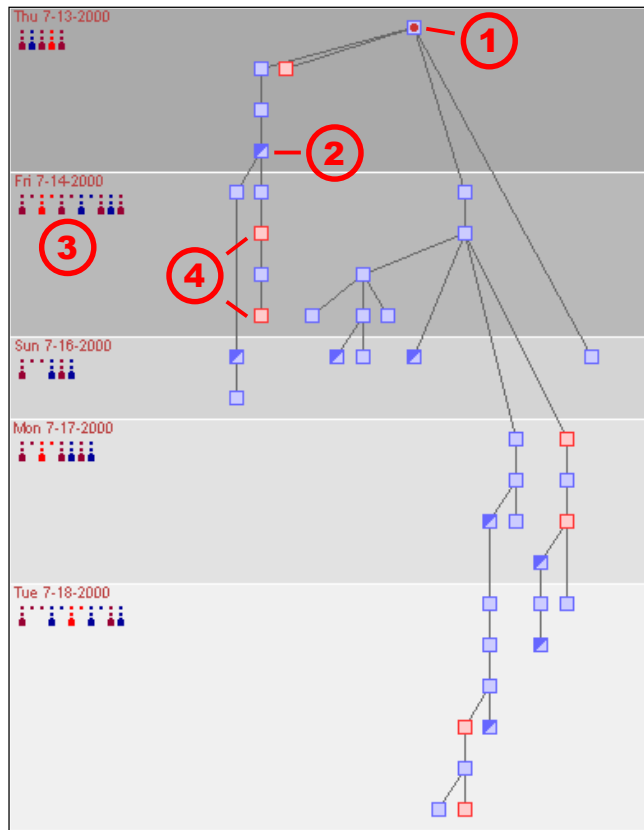


Figure 4. The Netscan thread tree for the thread with the subject “Visual Basic versus Visual C++” from the *microsoft.public.vb.general.discussion* newsgroup starting on July 13, 2000. In this view, boxes represent messages. A line between two messages indicates that the more recent message is a reply to the earlier message.

Because the sizes of the day bands vary, the vertical scale cannot be used linearly to infer time differences between messages.

Thread trees reveal information about message sequencing and ancestry that is difficult or impossible to gather from traditional news-browsing interfaces, especially when threads split into several sub-threads as time passes. The separate, parallel development of sub-threads is essentially invisible in any other view.

Our implementation has several additional graphic indicators and interactive behaviors (pictured in Figure 4 and described here) that enhance the basic tree:

1. *A dotted-box glyph for messages written by the author of the root post.* This allows users to determine easily whether the first poster ever revisited the thread. Such information is particularly important in question-and-answer or technical support newsgroups, because experts seeking questions to answer and novices seeking answers might like to know at a glance whether the root poster’s question was successfully answered or is still unresolved, requiring him or her to post follow-up messages. Reading the last response from the root poster is one way to determine if the question is resolved. (In Figure 4, the initial poster never posted to the thread again.)
2. *A half-shaded-box glyph for messages written by the most prolific poster(s).* The user who posts the most in a thread clearly figures prominently in the discussion, although post quantity alone does not reveal whether that user’s contributions were positive or negative.
3. *A daily histogram showing posts, posters, and posts per poster.* Small humaniform glyphs at the left end of each day’s gray band indicate how many people posted to the thread on that day. The dots above and to the right of each person-glyph show many messages that person posted. Viewed together, these form a histogram of daily activity.
4. *Highlighting the author’s contributions.* Rolling over any node in the tree highlights all the other posts in the thread by the same author and highlights the poster’s row in the piano roll component.
5. *Message display.* Clicking a message node loads the content of the selected message into the message display pane. (Not visible in Figure 4. See Figure 3d.)

The thread tree evolved significantly over time as we attempted to layer multiple metrics in a single view. Informational density is key to a useful display, but we sought to balance our desire to display as much data as possible with the need for simple, elegant design that would not overwhelm the user [12]. For example, in early prototypes of the thread tree, the nodes representing messages were significantly larger than in the current version. The large nodes were easy to click, but threads

with many messages quickly cluttered the display, so we chose a smaller default size for the nodes and added an algorithm to scale them dynamically when rendering unusually dense threads.

User Study Results

All of the participants found the concept of the thread tree interesting, but some questioned its value for basic news reading. However, their mean rating of its usefulness for reading news was 3.3 on the five-point scale. One user commented that the ability to see the branching degree of the thread was particularly valuable.

The location of messages within the gray day bands in the background confused several participants. Although no one had trouble understanding that each band was a different day, some thought that the vertical position of messages within the bands indicated the time of day when the message was posted. In fact, a message's vertical position within a day band depends solely on how many messages need to fit above and below the message inside the band. Because time records are so often incorrect in newsgroup messages – users' clocks and time zones are frequently wrong when they post – we chose not to use time at all in our visualizations rather than risk displaying faulty data.

Piano Roll

Following the design of Donath et al's Loom, our piano roll display (Figure 5) presents the intersection between authors, activity, and time. Each row of the piano roll display is dedicated to one author, all of whom are listed in order of the total number of messages they contributed to the current thread. Each column contains a histogram representing the number of messages contributed by each author on each day of the thread's existence.

The piano roll component is linked with both the thread tree and the interpersonal connections component. When the user mouses over an author in the piano roll, all of the author's messages in the thread tree are highlighted, as are his or her node and links to others in the interpersonal connections pane.

This display reveals each participant's patterns of activity and the days that a thread was most active.

User Study Results

Users liked the ordered list of posters in the thread, but our early implementation of the piano roll confused some. One user thought the histogram bars were buttons that would show the author's messages from that day. The vague labeling of the columns confused others. We used ordinal numbers because full dates took up too much space, but this design compromise turned out to be a mistake. Those users who understood the piano roll, however, found the intersection of the temporal and authorial dimensions interesting and useful.

Interpersonal Connections

The raw total of posting activity only roughly indicates the

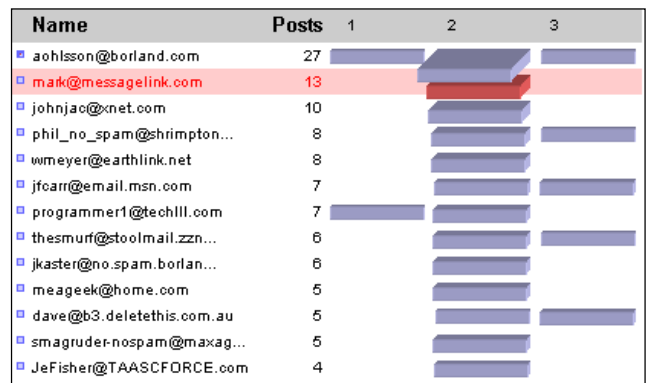


Figure 5. The Netscan piano roll display presents a ranked list of authors by their total activity and an indication of how much they posted on each day over the lifetime of a specific thread.

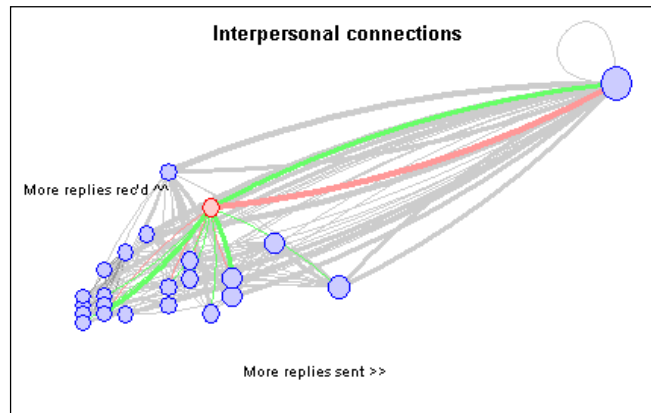


Figure 6. The Netscan interpersonal connections pane displays the authors in a thread and an indication of how much they replied to and were replied from all other participants over the lifetime of a specific thread.

role of a particular author within a thread or newsgroup. The interpersonal connections display (Figure 6) plots authors, represented by circular nodes, in a two-dimensional space according to the number of replies they post, which determines their position along the x-axis, versus the number of messages that are posted in reply to their messages, which determines their y-axis position. Lines connect those authors who have replied to each others' messages; the thickness of the lines varies directly with how many replies each pair exchanged.

The interpersonal connections component is interactive. When the user points the mouse at an author, lines indicating one or more replies by that author to another are highlighted in red. The lines representing replies by another author to the selected one are highlighted in green. Subsequent versions of this visualization will include arrows at the ends of the lines indicating the direction of the reply traffic.

Participants who appear in the upper right quadrant are frequently playing the role of *major discussants*; they are heavily involved in a give-and-take conversation. Those in the lower left quadrant participate similarly but are less prolific, so we call them *minor discussants*. People in the

lower right quadrant respond to other people's messages more than others respond to theirs. These may be answer people or experts in some capacity who respond to the queries of others, but without further analysis of message content we can say only that they are *reply sources*. Finally, those in the upper left quadrant receive more replies to their messages than they write to others' messages, so we call them *reply sinks*. We conjecture that people become reply sinks when they post interesting questions or controversial messages that garner many replies. (Simply posting the initial message in an active thread is not sufficient to be a strong reply sink. Only immediate replies, not replies to the replies, play into this measure. A reply sink, therefore, is a poster who regularly receives large numbers of direct responses to his or her messages.)

An alternative way to convey the information in the interpersonal connections component might be to indicate each person's role via distinct glyphs in the thread tree and piano roll. This approach would discard much of the visualization's detail, but it would make a more coarsely grained version of its information available at a glance.

User Study Results

As with the piano roll display, users who understood the interpersonal connections diagram, either on their own or after coaching from the usability researcher, found it informative. Their average rating of its usefulness was 3.3 on the five-point scale. However, even those who thought the interpersonal connections view was useful found the rough implementation confusing, as we expected. Users gave the component the lowest rating (2.1 on the five-point scale) for ease of understanding of any of the Netscan dashboard components. We believe, however, that an improved version of this component would prove not only useful but also usable to many users. (It would be useful to some classes of users, at least – we found a negative correlation [-0.699, $p=0.036$] between using newsgroups to buy and sell things and perceived usefulness of the interpersonal connections display. Clearly, only those using newsgroups for socialization or discussion would find the interconnections between participants interesting.)

FUTURE DIRECTIONS

The tools presented here address only some of the stages of interaction with persistent conversation systems like Usenet. Additional components are needed to address important parts of the browsing experience, particularly the discovery of newsgroups and the monitoring of an already selected set. We have begun exploring visualizations for these tasks.

Box Plot

One step toward this goal is an effort to represent large segments of Usenet, up to the entire system, in a single, comprehensible image. Given the amount of data involved, we had to carefully select the metrics to show in the graphic. We chose to focus on the hierarchical structure of

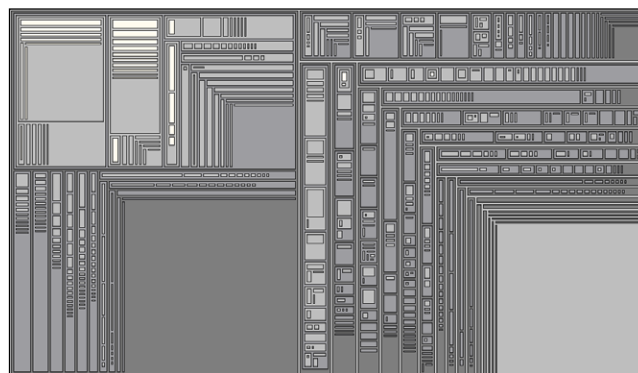


Figure 7. The Netscan hierarchy display presents a view of the entire Usenet in which newsgroups, represented by rectangular regions, are nested within their hierarchies and have volumes equal to their cumulative numbers of posts. In the interactive version, pointing the mouse at a box brings up a text label identifying the newsgroup or hierarchy.

Usenet and the size of each group or hierarchy as measured by the number of messages posted to it. (The hierarchy of newsgroups is merely lexical, based on the groups' names. For example, in the hierarchy, *comp.sys.mac.games* is the parent of *comp.sys.mac.games.action* and *comp.sys.mac.games.adventure*.)

Using tree map techniques [11] that have been applied to the visualization of file systems [2], we created an interactive, Web browser-based rendering of the entire Usenet (Figure 7). To the best of our knowledge, this is the only representation to date of Usenet as a whole.

A tree map approach reveals the dense, detailed hierarchical structure of Usenet by representing newsgroups as boxes sized proportionally to the number of messages each group receives and nested according to the lexical hierarchy of newsgroup names. Although we still must address serious performance issues – Figure 10 took 30 minutes to render on a fast machine – such an interface is effective at conveying the relative proportions of various segments of Usenet.

The relative sizes of the boxes in this visualization need not depend on the number of posts to the group. Usenet's landscape would look considerably different from Figure 7 if instead we let the number of posters, the ratio of posts to posters, or the average message length determine the area of each group's rectangle. Of particular value might be an indicator of percentage change in message traffic from the previous week to the current one; the largest newsgroups, then, would be those whose activity has dramatically increased.

Content Analysis

In this paper, we have focused exclusively on the structural properties of conversations to the exclusion of their content and meaning. In part, we have been constrained by the severe processing demands that textual content analysis would place on our computer systems, given that our news feed carries approximately 18 million messages every

month. In addition, a preliminary exploration of the literature indicated that the textual similarity and summarization analyses that we would like to use are not yet mature.

Such analyses are clearly a useful future direction for this work, however. For example, n-gram analysis [8] could be used to generate an indication of topic drift, the relative similarity of responses to the chain of messages from which they descend. Color-coding message nodes in the thread tree according to their similarity to the root message could give users an indication of where within a thread the topic has drifted to an entirely different subject despite the persistence of the earlier topic's subject line.

Active User Participation

The data used in this project are generated by data mining the existing messages in newsgroups. From these data – messages and the patterns of replies in threads – we may conjecture about or infer certain characteristics, such as the importance of a message within a thread or the role of a poster in a newsgroup. These characteristics are implicit in components like the interpersonal connections visualization, whose placement of authors on the plane reveals their likely role in the conversation.

Relying only on the existing message data to infer such traits makes our system automatic and passive from a user's perspective; we do not need the rating systems or feedback scores used by Web sites like Amazon.com, the online bookstore, and eBay, the auction Web site. However, as we build novel interfaces to persistent conversations, it may prove useful to solicit additional information from users about which messages they found interesting or helpful [4]. Such feedback could be combined with the measures we derive from the message data itself to generate detailed histories and reputation scores with greater accuracy than we can achieve with message data alone. We are currently investigating unobtrusive ways to include this functionality in new interfaces for persistent conversations.

DISCUSSION AND CONCLUSION

The diverse facets of online discussions – the messages themselves, their temporal and logical sequence, the relationships of their authors – do not integrate easily with each other. Thus, grasping the nature and extent of interaction in a complex conversation from just one kind of interface is difficult or impossible. It is sensible, then, to provide multiple ways to view the conversational space in a single dashboard. Netscan, with its massive data warehouse of Usenet messages and techniques for extracting meaning from such data, has the potential to provide a common platform for the creation and integration of multiple visualizations and interface components by many authors. The components we have presented here begin to suggest the possibilities for novel conversational interfaces. Such interfaces would only benefit from the plethora of developers whom a publicly available

platform would attract.

The components we have described present conversational behavior in the aggregate and, complementarily, detailed information about particular users and messages. If interfaces based on such tools become widespread, we believe they will make persistent conversations easier to navigate and even improve the quality of discussion by giving users a clear sense of the local population and its patterns of activity.

ACKNOWLEDGMENTS

We would like to thank Duncan Davenport, J.J. Cadiz, Dave Barger, Steven Drucker, Rebecca Xiong, Microsoft Research Usability Recruiting and our usability subjects.

REFERENCES

1. Deja.com. Available at <http://www.deja.com/>.
2. DiskMapper software from MicroLogic. Available at <http://www.miclog.com/>.
3. Donath, J., K. Karahalios, and F. Viegas. "Visualizing Conversation," *Proceedings of the Hawaii International Conference on System Sciences 32*, January 1999.
4. Hill, W., and J. Hollan. "History Enriched Data Objects: Prototypes and Policy Issues," *The Information Society*, Volume 10.
5. Horton, M. and R. Adams. "Request for Comments 1036: Standard for Interchange of USENET Messages." Available at <ftp://ftp.isi.edu/in-notes/rfc1036.txt>.
6. Kantor, B. and P. Lapsley. "Request for Comments 977: Network News Transfer Protocol." Available at <ftp://ftp.isi.edu/in-notes/rfc977.txt>.
7. Kollock, P. and M. Smith. "Managing the Virtual Commons: Cooperation and Conflict in Computer Communities," *Computer-Mediated Communication*. S. Herring (ed.). Amsterdam: John Benjamins, 1996.
8. Miller, Ethan L., Dan Shen, Junli Liu, and Charles Nicholas, "Performance and Scalability of a Large-Scale N-gram Based Information Retrieval System," *Journal of Digital Information*, January 2000, <http://www.cse.ucsc.edu/~elm/Papers/telltale.JODI.pdf>
9. Netscan Web site. Available at <http://netscan.research.microsoft.com>.
10. Sack, Warren. "Discourse Diagrams: Interface Design for Very Large-Scale Conversations," *Proceedings of the Hawaii International Conference on System Sciences 33, Persistent Conversations Track*, January 2000.
11. Shneiderman, B. "Treemaps for space-constrained visualization of hierarchies." Available at: <http://www.cs.umd.edu/hcil/treemaps/>.
12. Tufte, Edward. *Envisioning Information*. Cheshire, CT: Graphics Press, 1990.
13. Vector Markup Language specification. Available at <http://www.w3.org/TR/NOTE-VML>