

Flash Forums and ForumReader: Navigating a New Kind of Large-scale Online Discussion

Kushal Dave*
IBM Research
One Rogers St.
Cambridge, MA USA 02142
617-693-0168
kdave@us.ibm.com

Martin Wattenberg
IBM Research
One Rogers St.
Cambridge, MA USA 02142
617-693-0168
mwatten@us.ibm.com

Michael Muller
IBM Research
One Rogers St.
Cambridge, MA USA 02142
617-693-0168
michael_muller@us.ibm.com

ABSTRACT

We describe a popular kind of large, topic-centered, transient discussion, which we term a *flash forum*. These occur in settings ranging from web-based bulletin boards to corporate intranets, and they display a conversational style distinct from Usenet and other online discussion. Notably, authorship is more diffuse, and threads are less deep and distinct. To help orient users and guide them to areas of interest within flash forums, we designed ForumReader, a tool combining data visualization with automatic topic extraction. We describe lessons learned from deployment to thousands of users in a real-world setting. We also report a laboratory experiment to investigate how interface components affect behavior, comprehension, and information retrieval. The ForumReader interface is well-liked by users, and our results suggest it can lead to new navigation patterns. We also find that, while both visualization and text analytics are helpful individually, combining them may be counterproductive.

Categories and Subject Descriptors

H.5.2. [Information interfaces and presentation (e.g. HCI)]: User Interfaces – *evaluation, graphical user interfaces*. H.5.3. [Information interfaces and presentation (e.g. HCI)]: Group and Organization Interfaces – *computer-supported cooperative work*.

General Terms

Human Factors

Keywords

Mass interaction, thumbnail interface, visualization, user interface, persistent conversations, large-scale conversations, collaboration, user study, prototype

1. INTRODUCTION

People have long used networks to hold large, distributed

asynchronous conversations online. These discussions usually occur in forums—long-lived places built around topics, goals, or communities—where coherent threads deal with particular issues. Usenet, with its thousands of newsgroups, is perhaps the most studied of this type; many mailing lists, Lotus Notes databases, and web-based discussions display the same structure.

Popular new web-based conversation tools, however, have characteristics different from traditional forums. The most notable example is “News for Nerds” portal Slashdot.org. Slashdot posts links to newsworthy web pages and provides a place for readers to comment during a limited period of time. Usually hundreds of messages are posted in the days before discussion is closed. This brain dump approach is also seen on the IBM intranet. In periodic “Jams,” employees are encouraged to post their thoughts on a particular topic over the course of a few days. Four Jams have been held, discussing the roles of consulting, managers, and values within the company. By the end, the forums contain hundreds or thousands of messages [12].

Slashdot discussions and the Jam forums have four characteristics in common: they are diffuse in authorship, large in size, focused in topic, and constrained in time. If threads are present, they are shallow, without the long strings of replies-to-replies that are common in Usenet. We term conversations that share these features *flash forums*, emphasizing their bursty nature by analogy to a flash mob, which is “a large group of people who gather in a usually predetermined location, perform some brief action, and then quickly disperse” [20]. A detailed discussion of what makes flash forums unique is provided in Section 3.

Flash forums are a powerful way of gathering a large number of viewpoints on a particular issue or problem in a short time, but they present significant challenges to users. With authorship and thread structure providing weaker cues than traditional forums, navigating the large number of messages in a flash forum can be difficult. Several current flash forums (including Slashdot and the IBM Jams) have implemented voting (or “moderation”) systems designed to let readers highlight particularly interesting messages. Although moderation systems are valuable, they are not a complete solution. Both the Jams and Slashdot also employ official moderators, but recent work by Lampe shows that Slashdot's sophisticated system may still fail to surface valuable posts in a timely manner [18].

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, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CSCW'04, November 6–10, 2004, Chicago, Illinois, USA.
Copyright 2004 ACM 1-58113-810-5/04/0011...\$5.00.

* Current address: Kushal Dave, Google, 1440 Broadway 21st floor, New York NY USA 10018, kushal@google.com

Table 2. Samples of forum types

(Usenet threading is conservative, Slashdot counts “Anonymous” once and is thresholded, ValuesJam statistics are approximate.)

	Posts	Authors	% Repeat authors	Mean posts / author	% Root posts	Median/mean post depth	Studied/total duration
Usenet – comp.os.linux.advocacy	890	174	81	5.11	9	4 / 6.27	3 days / Ongoing
Usenet – comp.programming	273	85	69	3.21	11	4 / 7.41	3 days / Ongoing
Slashdot – MyDoom Virus	442	254	43	1.74	21	1 / 1.8	3 days / 14 days
Slashdot – Dark Matter	777	387	50	2.01	17	2 / 2.7	3 days / 14 days
Values Jam – Role of Values	4,084	2,316	42	1.74	39	1 / 1.0	3 days / 3 days
Values Jam – Impact	858	634	26	1.36	60	0 / 0.6	3 days / 3 days

We therefore describe ForumReader, an interface specially tailored to flash forum participants. It combines a visual navigation tool (the basis of which we introduced in [39]) and automatic topic extraction technology to give users a multi-faceted overview of the conversation. The interface has been implemented in several prototypes, and we discuss two forms of evaluation. First, we discuss feedback from a wide-scale deployment to thousands of users within IBM. Second, we analyze results from an experiment on reading comprehension and behavior. The results show significant promise and shed light on ways in which users navigate flash forums differently from Usenet-style forums.

2. RELATED WORK

Many researchers have recognized the problems with current interfaces to discussions, though none have addressed the special needs of flash forums. We describe previous interfaces for online discussions, as well as work in text analytics and interface design relevant to ForumReader.

2.1 User Goals and Models

Users’ goals when interacting with a large discussion vary with the range of users and settings. Previous work on persistent asynchronous conversations [25], surveys of participants in the Jam discussions, and our own study results show some common themes: readers are looking to find ideas or information, receive support, take the pulse of a community, and meet people.

A major barrier to the success of these discussions is scale. Jones et al. note that mass interaction in what they call “virtual publics” is limited by cognitive load, and that factors including volume and interactivity can increase this load and cause changes in participation strategies [16]. Jones et al focus on Usenet and suggest that different technologies will produce different behaviors. This was borne out in our study of flash forums.

User navigation through large bodies of text can be understood through models such as locally optimal “satisficing” and global “information foraging” [26],[27]. Along the way, users must make sense of the conversation [30]. But gauging performance on synthesizing and skimming tasks is difficult. Traditional approaches use a single document or multi-document corpora, measuring reading comprehension using detailed questions or summaries, or looking at time and accuracy in information retrieval tasks.

2.2 Discussion Interfaces

Usenet has inspired much research on interfaces to large discussions. Sack combines visualization with text analytics and social structure in Conversation Map [31]. Smith focuses on the representation of threads and authors in Netscan [32], as does Neustaedter with Grand Central [22]. The Loom project displays thread structure and emotional content [4]. Research has also been done on threads in email, notably [17],[24],[37]. These strains of work rely heavily on thread structure and authorship, features that are less salient in flash forums.

2.3 Text Analytics

Natural language processing (NLP) can assist users when interacting with any large corpus, particularly by providing intelligent search capabilities or grouping related documents. In addition to their use in Conversation Map, these technologies have been tailored to online discussions for generating summaries of threads [38] and clustering related posts [2]. We do not introduce new techniques here, instead using the standard clustering and similarity determination provided by the eClassifier text analytics package, described in [34].

2.4 Thumbnail Interfaces

One common method of visualizing textual data is what we call a *thumbnail* approach, in which a large document or set of documents is shown as if seen from a distance. Connecting navigation in these thumbnails to a detail view produces an overview+detail interface [15],[19]. One of the best-known thumbnail interfaces is Eick’s SeeSoft and SeeSys, which use text thumbnails, annotated with search results, authorship, or software-development metadata, as navigation tools [6]. The Reader’s Helper is a similar exploration [10]. More abstract, iconic thumbnails are investigated in TileBars [13] and Context Lenses [3], among others.

3. FLASH FORUMS

Flash forums are sufficiently different from traditional forums that conventional newsreader interfaces are not appropriate. In this section, we describe in detail four key distinguishing characteristics of flash forums: lack of importance of authorship information, large size, tight focus overall with overlapping topics between threads, and a short timeframe for the conversation. As far as we know, characterizing a forum in terms of such dynamic features is new. Related work on Jams considers them as

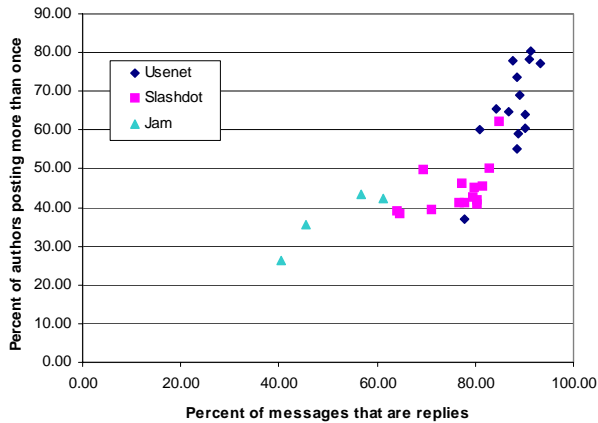


Figure 1. Differences in repeat authorship and interactivity between forum types.

Massively Parallel Conferences [35] or large-scale distributed meetings [21], and Slashdot has been cast as a virtual public [11]. In keeping with traditional typologies of online discussions, these characterizations focus on purpose, membership, or medium [36]. Besides Slashdot and the Jams, flash forums occur on other community sites, such as kuro5hin.org or plastic.com. Discussions attached to Yahoo! News stories, weblog posts, and products on shopping sites often display similar behavior, too.

Various studies have sought to quantify the types of discussions that take place in Usenet, Notes databases, and mailing lists [1]. To compare our examples of flash forums (Slashdot and IBM Jams) with Usenet, we randomly selected 14 Slashdot discussions and matched them with 14 active Usenet forums of similar content or size, looking at 3-day time windows in each. Together with the four 3-day Jams we had available, we had 32 data points to which we applied one-factor Anovas ($F_{2,31}$), and we used Tukey HSD tests to compare the means. All results reported here were significant for $p < .05$ or better. Slashdot and Usenet discussions did not differ in terms of the total number of messages or total number of authors. Portions of this data are shown in Table 1, and one key behavior is depicted in Figure 1.

3.1 Authorship

The small number of repeat authors makes authorship a poor proxy for salience, a point also made in user feedback discussed below. Slashdot forums had fewer mean postings per author than Usenet forums (1.83 vs. 3.25) and, concomitantly, a lower percentage of repeat authors (45% vs. 66%). Jams had even lower repeat authorship (37%). Even across Slashdot discussions, repeat authorship is low—in the 14 forums we looked at, only 18% of the 3,286 distinct authors participated in more than one discussion. Repeat authors produced significantly more of the content in Usenet (81%) than in Jams (52%) or Slashdot (57%). Flash forums more often display “drive-by” postings than person-to-person conversation, where posters post once and are unlikely to post again, if they return to the discussion at all. Many of these posts occur as replies; some flash forums even have larger threads on average than Usenet newsgroups on similar topics. In both email and Usenet, authorship information is considered highly salient by users [8], but in flash forums, there are few recognizable individuals in the community, though other

information about authors, such as reputation or position, may be valuable.

3.2 Size and Focus

The volume of text to be examined in flash forums is disorienting—often as much as a megabyte, the size of a novel—and finding areas of interest and navigating to them is difficult. Although a Usenet newsgroup may see more posts in a given period than a Slashdot forum, the more appropriate comparison is between a Slashdot forum and a single thread in Usenet. In both cases the constrained topic means that if the germinating post is of interest, any of its descendants might also be worth reading. The concentrated topic leads to significant overlap in themes across sub-threads. In extreme cases, nearly identical posts might start two separate threads about the same issue at approximately the same time. This situation is exacerbated by thread drift and users who are unsure where to post their replies. Perhaps as a result, threads in flash forums are significantly shallower than Usenet threads. A smaller percentage of Slashdot posts were replies (76.4% vs. 87.7%) and there were fewer extended discussions, as measured by the median depth of posts (1.20 vs. 2.86). Usenet threads routinely exceed a depth of 10, while—in our data—Slashdot never did. Jams were even more extreme, with 51 percent replies and a median thread depth of .5.

3.3 Time Limits

The final defining characteristic of a flash forum is that it has a time limit. Typically online conversations are permanently open, although particular threads may die out over time. Flash forums often have explicit time limits built into the software (14 days in the case of Slashdot, 3 days in the case of IBM Jams) as well as implicit time limits set by the subject under discussion (Slashdot, for example, often discusses breaking news, with discussion trailing off after as little as a day). Furthermore, in the case of the Jams—intended to produce a series of action items or summary documents—the discussion must have closure and be analyzable as a self-contained archive. Time limits have the powerful effect of spurring people to post sooner, knowing that posts late in the discussion are unlikely to elicit responses or be read as the forum approaches closure. As the Jam facilitators explain, “A Jam’s authenticity is derived from the fact that it’s a real-time and finite event, and that there are real, often serendipitous ‘knowledge accidents’ among participants that emerge because of the time constraint imposed” [5]. This pressure to post quickly is another reason that flash forum interfaces should allow users to navigate as rapidly as possible. Despite this intensity, the asynchronous medium makes it more coherent than threaded chat [32].

4. FORUMREADER DESIGN

4.1 Goals

We set out to design a tool that would permit easy and intuitive interaction with flash forums, based on the characteristics described in Section 3. Although threads and authorship would be de-emphasized, this information would still be surfaced to establish context and link directly-related material. Orientation and movement would be facile both globally and within a thread.

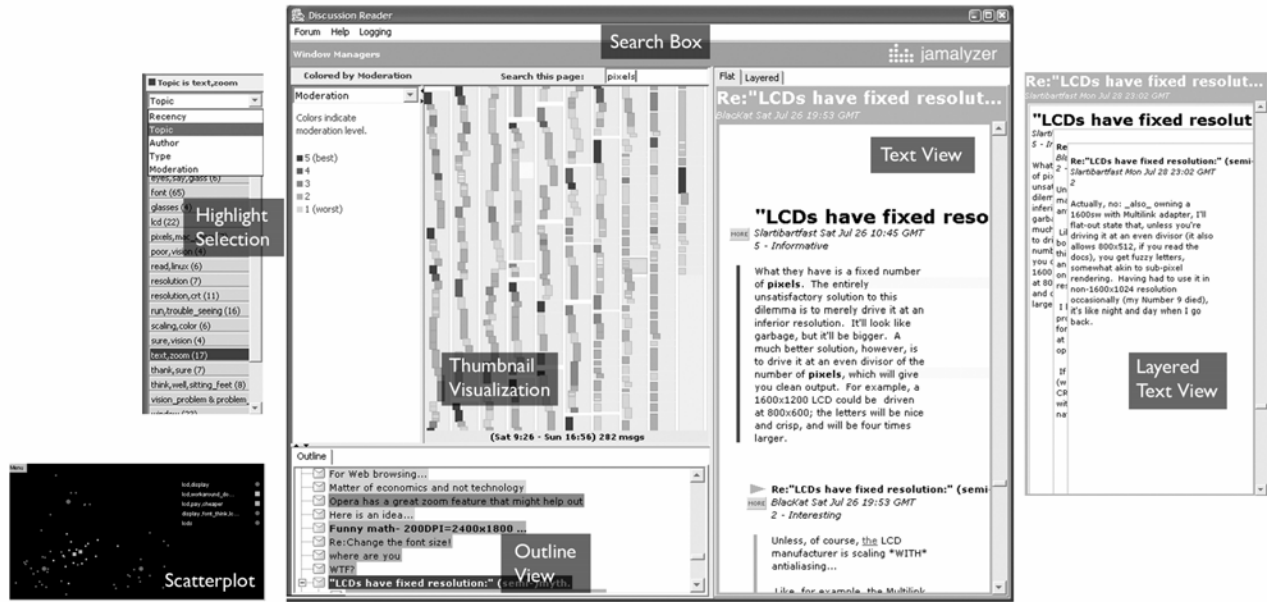


Figure 2. ForumReader, showing the first of the discussions used in lab testing. Bright lines on the thumbnail show lines matching the search for “pixel.” The scatterplot was only included in the Jamalyzer version, and the layered text view was only available in the lab tests.

We adapted the classic thumbnail layout, creating, we believe, the first system to incorporate thumbnails specifically for use with discussions, taking advantage of the additional structure provided by conversational threads. And although eClassifier had been deployed in previous Jams, its use with our visualization generated a completely different experience.

After several iterations with talk-alouds and critiques, it became clear that navigation was an idiosyncratic process, alternating between browsing, searching, revisiting, and synthesizing. We therefore decided to create an interface with multiple tightly-coupled components.

4.2 Components

As shown in Figure 2, the interface consists of a navigable thumbnail of the conversation, a panel for choosing how this visualization is highlighted, a detail window for displaying message text, a search entry box, and a tree widget. In the Jam deployment, a scatterplot depicting eClassifier clustering of messages was included, as was a method for paging through large forums. In our testing version, we added a view showing a message’s parents as layered panels in the text view. We also removed certain views in some test conditions. All major views can be resized using sashes to suit user preference.

4.2.1 Thumbnail View

The thumbnail visualization is intended to counter the disorientation users often experience in flash forums, taking advantage of the fact that these forums are coherent, self-contained units. It uses rectangles to represent each message in the discussion, drawing them in depth-first order with indentation indicating thread depth and rectangle height corresponding to message length. Rectangle widths are the same for each message, and chosen to fill the available space. This view is a thumbnail of

the full discussion as it might be read on the web, wrapping from the bottom of one column to the top of the next.

Because it becomes difficult to visually distinguish threads when the conversation grows large, we draw a faint border around each thread and add white space below each proportional to its size. Tool tips indicate thread subject. Users navigate with the thumbnail by clicking or dragging a selection rectangle in any direction; the text view changes to show the corresponding portion of the discussion. Our implementation can reasonably show up to 1,000 messages where thread depth does not exceed 10 when the software is used on a 1024x768 screen.

4.2.2 Highlighting

In flash forums, threads are inadequate for the role they traditionally serve—breaking the discussion into easily-parsed chunks of varying interest. We counteract this by providing cross-cutting highlights. The messages in the thumbnail may be colored with different brightnesses to highlight attributes (e.g. moderation level in a Slashdot discussion) selected by the user. Highlighting can be either binary—showing all of the messages from a given author, for example—or continuous—such as shading more recent or highly moderated messages darker. Users may choose which attribute to highlight by using a drop-down box to select a class of highlights. For continuous attributes, the space below the drop-down displays a simple key; for binary attributes, it contains buttons allowing the user to switch between the possible values to highlight.

eClassifier, which uses NLP to put messages into clusters (corresponding to distinct “topics” of conversation) based on their content, provides the basis for two modes of highlighting. One allows users to select a cluster topic and highlight all of its member messages, while the other provides a graded coloring of

other messages based on their similarity to a selected post. Similarity in both cases is determined by the weighted sum of automatically derived “keywords” they have in common.

4.2.3 Text View

The text view component displays the full text of the messages in the discussion. To support meandering navigation of flash forums, we preserve some of the feel of a typical web discussion interface, allowing users to quickly skim through content using a scrollbar, a mouse wheel, or standard arrow and page keys.

Message text is indented by an amount proportional to its depth in the thread tree, allowing easy identification of reply relationships. This indentation method, however, can force messages in a deep thread off the side of the window. To address this problem, the text view component automatically scrolls sideways as the user scrolls vertically, thereby centering the messages in view.

At the start of each message is a header showing the subject, author, time, and forum-specific metadata (such as moderation or author’s job title). Clicking on the author will search for posts by him or her. Alongside each message body is a rectangle colored to the match the message’s brightness on the thumbnail. We emphasize the start of each thread with white space and a larger subject line font.

The currently selected message is indicated with an orange triangle next to the header. As the user scrolls, the selection follows along. We also display the selected message’s header in a fixed location at the top of the screen, giving the eye a fixed location to focus on during rapid scrolling.

4.2.4 Search

A unique feature of ForumReader is the capacity to perform real-time visual text searches, providing another way to quickly sense what is present in unwieldy flash forums. Users may either type a text phrase into a search box, or click on any word in the text view to search for other messages containing that word. At each keystroke or mouse click, the entire corpus is searched (in a time under 100 milliseconds) and the thumbnail is updated with visual highlights to show where in the discussion matches occur. The exact line where a match occurs is colored yellow on the thumbnail, and the rest of the message where it occurs will be a darker yellow. In the text view, the line will be highlighted in the bright yellow with the word in bold.

4.2.5 Tree View

A tree widget is provided as a familiar and easy way to navigate between and within threads. Unlike the thumbnail, the tree provides textual information about nearby messages and clear indication of thread depth.

4.3 ValuesJam Version

In deploying ForumReader during IBM’s ValuesJam (discussed further below), we encountered some special opportunities and constraints. Since the size of the forums exceeded the natural limits of the visualization, with the largest reaching 4,000 messages, we split forums into pages by time, providing links to earlier intervals. Switching between pages took several seconds. This version incorporated the scatterplot generated by eClassifier. The plot depicts each message as a small dot in a projection of a high-dimensional term-frequency space, so that messages featuring similar words are placed closer together. Each dot is

colored to reflect the topic it has been placed in, and clicking on one brings up the text of the message. The eClassifier topics themselves were manually tuned and labeled by a human expert. This version also allowed users to highlight the thumbnail based on author metadata such as location, title, or division.

4.4 Slashdot Version

Based on feedback from the ValuesJam deployment, we added two new features to the version used in lab testing. One was *read wear* [14]. As a user spent more time over a given portion of the map, the background faded to white, giving the user a representation of areas they had already been and providing useful landmarks. We also colored the text in the tree widget with the same scheme used in the thumbnail view to separate the effect of the highlighting from that of the visualization. The items in the tree were given the same brightness as in the thumbnail, and items matching a search showed bold subject lines and yellow icons.

This version also made the brightness options appropriate to Slashdot—the default coloring was by moderation, and could be switched to color all messages deemed to be of a given type (funny, insightful, etc.).

4.5 Implementation

Both ForumReader prototypes were implemented in Java 1.4. They load the entire discussion or page into memory, along with information about line breaks, in order to rapidly highlight lines of interest. With corpus sizes up to several thousand messages, simple linear search is sufficiently responsive for finding matches.

5. EVALUATION

5.1 Deployment with ValuesJam

5.1.1 Deployment Details

In August 2003, IBM conducted ValuesJam, an internal, web-based, company-wide discussion about the values of the company, bringing in over 22,000 unique participants. (IBM has been holding such online discussions for various parts of the company since 2001.) Four forums brought in a total of 8,973 posts from 4,614 unique participants.

A simple web interface was provided to all participants in the company. This interface provided various ways of navigating, the most basic of which was paging through threads in reverse chronological order. The two most popular ways of browsing were lists of posts rated highly by other participants, and lists by topic, determined using human experts and eClassifier.

A version of ForumReader, branded Jamalyzer, was provided as an experimental complement to reading and navigating on the web, and it was downloaded over 6,000 times. The contents were updated periodically throughout the 3 days of the event.

5.1.2 Feedback

The Jam team randomly surveyed approximately 2,500 participants, and 1,248 responded (50%). Sixteen percent reported that they successfully used the Jamalyzer tool as a way of finding areas of interest in the Jam. We infer from download statistics and open-ended feedback that most of the other 84 percent did not see links to the tool or did not download it, and that others downloaded the tool but encountered technical or usability problems. Among those using the tool, the Jamalyzer received substantially higher scores than a simpler tool in an earlier Jam

Measure	F(1,12)	p<
Bullet items	1.95	n.s.
Mindmap items	6.57	.03
Reasons	5.20	.04
Expert rating	4.86	.05

Table 2. Text-analytics by visualization interaction effects for five dependent variables

that included only the eClassifier scatterplot. Jamalyzer scored 3.5 out of 5 in importance (vs. a score of 2.8 for the earlier tool) and 3.2 out of 5 in satisfaction (vs. 2.7 for the earlier tool).

The survey provided space for open-ended comments about Jamalyzer. One of the most frequent was an appreciation of its ability to provide global orientation. One user explained, “It was easy to navigate and allowed rapid scanning of themes and comments.” Another noted, “It was valuable to get an overall idea of the jam sessions.” Some users felt that the visualization reinforced the text analytics: “I could see patterns of responses based on themes or words.” In contrast, users called the web interface “overwhelming” and “cumbersome.”

Some users encountered technical troubles or found certain components (the scatterplot and, to a lesser extent, the thumbnail) difficult to understand. Others lamented the clumsy paging interface as well as glitches in the message data. Several asked for the rating data, unavailable in Jamalyzer for technical reasons.

Jamalyzer served the goals expressed by users in other ways, however. Seventy-four percent of respondents said they wanted to find interesting discussion threads, while 64 percent wanted to find well thought out individual posts. And while users were very interested in ideas (78 percent), vision (90 percent), and depictions of reality (89 percent), only 49 percent were interested in meeting people. By using threads but emphasizing cross-cutting searches, and by focusing on content instead of people, Jamalyzer was a good match for this flash forum.

Many users expressed a strong feeling of excitement about the prototype. Spontaneous posts to the Jamalyzer intranet site showed much enthusiasm. “Amazing. To be able to locate commonalities, etc., and analyze the worth of this VALUABLE effort IS GREAT!” Exclamation points were common in the comments, and other users termed it “fantastic,” “easy,” and “so convenient.” Several people asked about whether the technology might be available for their own projects. This level of excitement suggests that even in prototype form the interface is providing a genuinely useful solution to the problem of reading flash forums.

5.2 Lab testing with Slashdot

The Jam provided valuable experience and user feedback on ForumReader as a whole, but we also wished for a more controlled test. We were particularly curious to learn which of the many features in ForumReader were beneficial, and how they might help users navigate flash forums. A secondary goal was to examine how users approach these discussions in general, and what they believe to be the most helpful cues for navigation. To examine these issues more closely, we conducted a laboratory study with a small group of users.

One challenge we faced in designing the study was that time and resource constraints made it impossible to test all combinations of

ForumReader features. We considered examining the efficacy of just a single feature, but decided that would shortchange the multi-faceted nature of the ForumReader interface. We therefore settled on a 2x2 design focusing on two key features, the thumbnail visualization and the eClassifier text analytics.

5.2.1 Study Design

We selected two discussions from Slashdot, both on subjects that are representative of the expertise of Slashdot’s readership. One dealt with window managers¹, the other with databases². After removing the lowest-moderated posts (following the current web interface’s default behavior to hide spam, flames, and trolls), the first contained 282 messages, and the second had 362. The order of these discussions was balanced across users in a 2x2 study design, using factors of presence/absence of text analytics, and presence/absence of thumbnail.

A total of 16 people participated (13 men and 3 women, ranging in age from 19 to 31, recruited by advertisements to local universities and technology companies). All were regular Slashdot readers, but none had read these two discussions. Each user had at least five years of experience with computers.

Each participant experienced only one of the four user interface configurations (i.e. baseline (T-V-), text analytics (T+V-), thumbnail visualization (T-V+), and text analytics with visualization (T+V+)). Thus, four participants experienced each of these four conditions. The order of databases was balanced across conditions: two subjects in each condition saw the database discussion first, and the other two the one about window managers; this assignment was random.

We wanted to understand how the interface affected a user’s ability to make sense of the discussion—quickly identifying topics and salient points about them. To measure this, we gave users various tasks while logging their actions. We began by showing users the interface with a sample forum and gave them a few minutes to get comfortable using it. For each discussion, users then recorded their interest in the stated topic and what they might find. Users then read the discussion for ten minutes while noting the major areas of disagreement. This was followed by two minutes to structure these areas of disagreement into an outline or mindmap, similar to [28]. Next was a recall and retrieval task, where users indicated how often they thought 10 different topics were discussed. In the final task users spent 5 minutes arguing for or against a statement related to the discussion. Subjects concluded the session by giving their impressions of the discussion they had just read, and by evaluating the features of the version of the browser that they had experienced.

5.2.2 Reading Performance

In general, we were interested in users’ performance using each of the four experimental versions of the Forum Reader. We used a 2x2 Anova to examine the following performance measures:

- number of areas of disagreement identified
- number of topics in the mindmap or outline
- number of reasons cited in arguing for one position in an assigned disagreement

¹ <http://ask.slashdot.org/article.pl?sid=3/07/23/1546244>

² <http://slashdot.org/article.pl?sid=02/09/23/1652243>

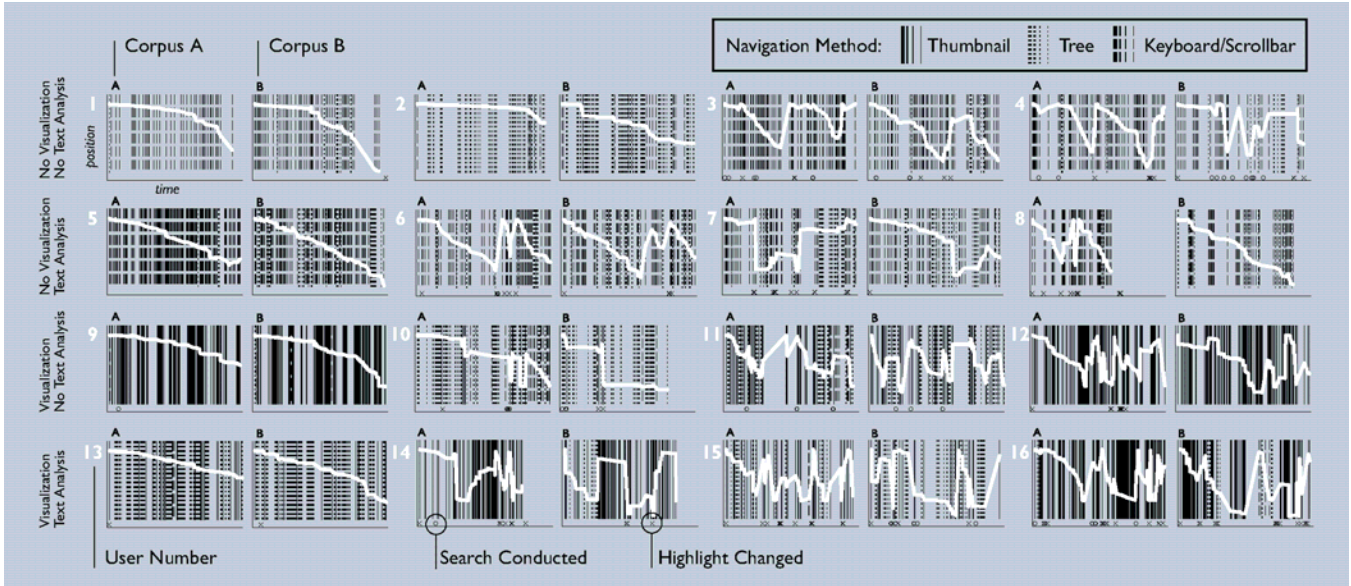


Figure 3. Graphs show paths of users through the discussions during the 10-minute browsing period. Vertical lines are used to indicate navigation events, and glyphs show highlight changes and searches. Although behavior is highly idiosyncratic, we see that the thumbnail was used by most of the people who had it, and that it seems to encourage more nonlinear navigation.

In addition to these countable performance measures, we also conducted one evaluation of the quality of responses:

- expert rating of answers (as determined by two expert reviewers, in a blind-scoring procedure; reviewers discussed ratings until they were in agreement)

We averaged the results for each measure across the two corpora into a single measure to increase the stability of each performance estimate. The most frequent significant effect across these measures was an interaction between text analytics and visualization, as shown in Table 2. Surprisingly, the form of the interaction was always some variation of cross-over. The consistent pattern of the crossover was that adding visualization to a no-text-analytics condition *improved* performance, whereas adding visualization to a text-analytics condition *reduced* performance (or, equivalently, adding text analytics to a no-visualization condition *improved* performance, but adding text analytics to a visualization condition *reduced* performance).

5.2.3 Navigational Behavior

The logs of user navigation gave us a window into how the different interface components affected user reading behavior. Figure 3 shows one simple but suggestive analysis of how subjects moved through conversations. For each subject, we plot time (x) versus position in the discussion (y). A reader who moves sequentially through a discussion, one line at a time, would thus produce a steady diagonal line, while a reader who jumps back and forth would produce a series of zigzags. We see that in each of the two conditions where the thumbnail visualization was present (T-V+ and T+V+), many graphs show dramatic, discontinuous leaps. (Note that subject 13, whose graphs are largely monotonic, used the map infrequently.) In the other conditions, the graphs more often show long, nearly monotonic segments, although the text analytics-only condition shows slightly more nonlinearity.

There were clear patterns from user to user as to how they approached the discussions. Initial linear periods ranged from the full 10 minutes to none at all, and sometimes second or third passes would still be strongly linear. Subject 9 used the thumbnail exclusively but linearly, while others used it in combination with other techniques. Some used the scrollbar only for fine-grained local movements, while some did not. In general, map users seemed to read more of the messages than other users (as defined by lingering for more than 2 seconds).

From a foraging perspective, users who navigated linearly probably did not consider the cost of finding clues as to where to go next adequately compensated for by the benefits. Instead, they read down until they lost interest, demonstrating a strong bias toward the start of the discussion. Interestingly, some users did search by both typing in queries and clicking on words in the text view during the exploratory period, reinforcing survey responses showing that users did have specific and (differing) questions they hoped the discussion might answer, even when their reading had no explicit purpose.

5.2.4 Feature Comparisons

The final task in each session was for users to evaluate the features that they had experienced in the version of the browser that they had used. We asked a total of 14 questions and provided a five-point scale for rating the “usefulness” of each feature (1=very useless, 5=very useful). However, each set of four users experienced a different version of the browser (T+V+, T+V-, T-V+, T-V-). Therefore, some features were not present for some users. (If robust analysis of feature comparisons had been our primary goal, we would not have used a 2x2 design, but even given the design of the main experiment asking about user preferences seemed worthwhile and cheap.) In summary, there were 11 features experienced by all users, and three additional features that were experienced by various subsets of users. All 14 features were experienced by only the users in the T+V+ condition.

We analyzed these feature comparisons in two different ways. The first analysis used data from all 16 users, but was restricted to comparing the 11 features experienced by all 16 users. The second analysis used data from only the four users who had experienced all 14 features. Both analyses used a repeated measures Anova, and both showed significant differences—for the 16-person/11-feature analysis $F(10,150)=15.507$, $p<.000001$; for the 4-person/14-feature analysis, $F(13,39)=5.967$, $p<.000001$. For both analyses, we conducted detailed comparisons of ratings on individual items using Tukey’s HSD test based on the Studentized range statistic, at $p<.05$. For simplicity, we present the results in Figure 4 with standard error bars.

The feature comparisons in Figure 4 provide preliminary benchmarks on user valuations of the various components of ForumReader, as well as a sense of how users generally navigate flash forums. The most valued features include some obvious choices, such as the window with the actual text of articles and the scrollbar used to navigate that window. More interesting is that no significant differences were found between these critical display elements and certain new features: coloring by moderation and (among the subset of user who experienced it) the thumbnail component. By contrast, many other new elements were rated significantly lower, such as the “layered” text view and coloring by criteria such as author (again suggesting the diminished importance of threads and authorship in flash forums).

These preference differences agreed with the navigation log data, since both the thumbnail and coloring by moderation were heavily used, while the layered view and the coloring by author were seldom used. Other frequently-used features included coloring by moderation type (funny, informative, etc.), suggesting that simple numerical moderation is not always sufficient, and searching by typing, which was particularly useful when users were asked to see how much a given topic was discussed. The preservation of threading information was useful for chunking messages, allowing users to quickly read the starts of threads and to mouse over threads without having to refocus their navigation.

6. DISCUSSION

The Jamalyzer deployment and the lab tests provided a number of insights into user views of flash forums, as well as the ForumReader interface, but also presented a puzzle.

6.1 Getting Oriented

Readers of flash forums seemed to value information about authors much less than in other, superficially similar, situations. The subjects in our experiment rated author information significantly lower than moderation information. In the Jamalyzer deployment, we read reports of people using the color-by-author feature—but primarily to look for their own posts! These results suggest that flash forums present a unique interface problem, and that designers cannot simply transfer Usenet or email interfaces.

Users might have been looking for their own posts to establish an entry point into the discussion. Read wear was equally useful because it provided landmarks showing where the user had been. (This ability to know what has already been read is so important that one study participant in the condition without the thumbnail carefully collapsed threads in the tree view as she read them.) The thumbnail helped orient users amidst heterogeneous authorship and homogeneous threads, but entry points were still needed.

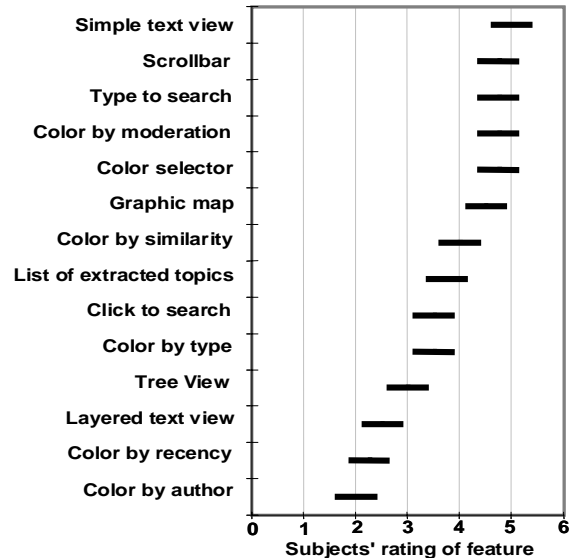


Figure 4. Feature evaluations based on data from 4 users who experienced all 14 features. Note that coloring by moderation is as valued as the text view, scrollbar, and search.

Starts of discussions, starts of threads, topic matches, and moderation information all served as entry points, and this is reflected in the graph of reading behavior shown in Figure 5.

6.2 Providing Cues

Once oriented, though, users must decide where to go next. A lesson drawn from both the Jam deployment and the experimental evaluation is that one of the most useful kinds of data is about how much others find a message valuable. The Slashdot moderation values were considered important by almost all of the subjects. Convergently, with Jamalyzer, the ability to see voting data, analogous to Slashdot moderation, was one of the most frequently requested features. These complementary results suggest that social recommendation and reputation information is highly valued when navigating a flash forum, possibly even more so than GroupLens found it to be for Usenet[22].

The results from navigation logs suggest that the thumbnail visualization encourages a non-sequential navigational style. (At the same time, the case of subject 9’s sequential navigation shows that the Forum Reader does not *force* a navigational style: user choice clearly plays a role.) This result is somewhat surprising because the linear navigation was often accomplished using the outline view, which on its surface seems to provide easy non-sequential navigation up and down the thread tree.

Information foraging theory, however, provides a plausible explanation for this effect. The theory predicts that the likelihood of a forager moving to a new “patch” of information will increase when the visibility of more promising patches increases. Some users developed a routine of rapidly scrolling the tree up and down to look for matches, but this was still more expensive than glancing at the thumbnail. Since the visualization makes the entire discussion visible at once without scrolling, it would be logical for foragers to make more “inter-patch” moves. The text analytics condition might also be expected to increase patch visibility

slightly, since clicking on a topic will highlight a “patch” of related messages.

6.3 Feature Interaction

Our most puzzling findings are the interaction effects from the lab study between the visualization and the text analytics—i.e., that each feature was beneficial to performance, but that the two features together were associated with *worse* performance. It may be that although both methods are useful in encouraging nonlinear navigation, the two tools may be at odds. If topics do not seem to align with threads, or if the effort needed to consider this new level of semantics on the thumbnail and switch between facets is too great, users have difficulty determining where to go. A second explanation might be that there is a tradeoff between navigational features and learnability—perhaps learning two new navigational modes at once is overwhelming for users. In fact, some study participants noted that the interface seemingly provided too many options. The interaction effects pose questions for designers of interfaces that allow multiple navigational paths. It is also unclear how much the text analytics would have benefited from being more tuned to the discussion.

7. FUTURE DIRECTIONS

Although various types of online conversation have been researched extensively, flash forums present a relatively unstudied phenomenon. We suggest further research into their dynamics and how best to structure such discussion, as well as investigations into how tone, length, and other characteristics of posts vary across mediums.

The ForumReader interface presents a rich opportunity for exploration, as well. This includes highlighting and sorting by multiple attributes simultaneously, and overlaying text on the thumbnail visualization. Many users requested the ability to hide posts that were not of interest at the moment. It might also be useful to add other ways of exploiting thread data, since even in flash forums threads provide useful context. Finally, users suggested we apply parts of the tool to other corpora, such as e-mail, software code, or even Usenet.

A successful flash forum interface could radically alter the way users interact. One user in our study noted, “It would be more interesting to have the opposite of what you think you’re looking for”—the search features made it easy to find information, but tended to help readers confirm their beliefs. One possibility would be to highlight messages dissimilar to what has already been read.

Another possibility would be to support search at composition time, similar to the Remembrance Agent [29]. This could help authors of new posts place their posts in the most appropriate possible thread and with maximal awareness of relevant predecessors in a forum.

8. CONCLUSION

Flash forums are a powerful method of quickly bringing thousands of people’s thoughts to bear on a specific problem or issue and represent a potential evolution of the public sphere [1]. Whether it’s critiquing research before publication (as *Jane’s Military Weekly* once did on Slashdot), collecting solutions to computer problems (as Slashdot does routinely), or gathering plans for new corporate values (as IBM did in ValuesJam), the broad participation, large size, tight focus, and time constraints

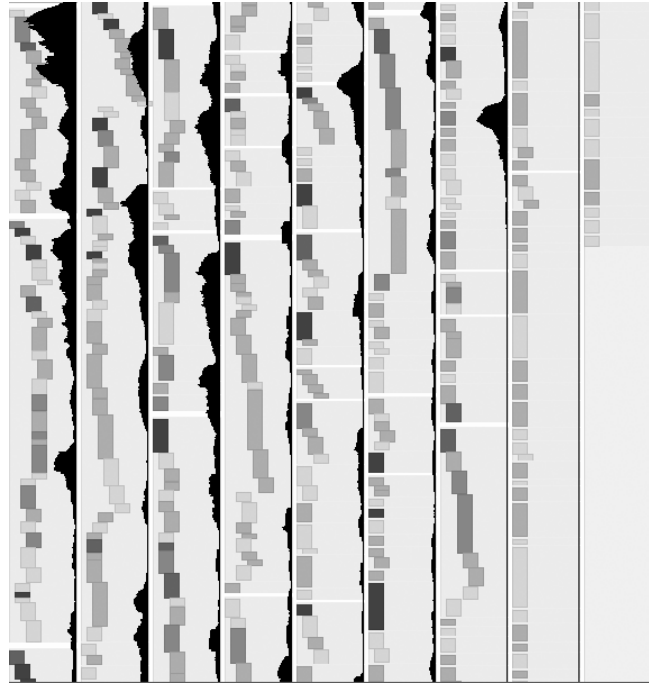


Figure 5. Length of black bars corresponds to the mean time (up to a threshold) test subjects spent on a given line during exploration. In general, highly-rated (darker) regions and the beginning of the discussion do well.

make them much more intense than community forums, while the asynchronous nature encourages more thoughtful posting than chat rooms.

One impediment to using such forums optimally is the absence of an appropriate interface. The excitement surrounding ForumReader should not be discounted. This tool helps highlight what is important in forums of this size and quality.

Much more work remains to be done on understanding what tasks users are doing when they navigate flash forums and how best to support them. With thousands of people routinely volunteering their expertise and opinions in forums such as these, it would be a shame to see their thoughts lost in a sea of words.

9. ACKNOWLEDGMENTS

This work would not have been possible without Scott Spangler’s work integrating Eclassifier and preparing Jamalyzer for deployment. We are also grateful to James Newswanger and the Jam team for allowing us to trial and collect feedback about Jamalyzer. Bernard Kerr made many important contributions to the design of the tool and the visualization of study results, and the rest of the CUE group provided valuable encouragement and ideas. The lab studies happened only due to the assistance of Joe Scherpa and Eileen Driscoll. Most importantly, we thank the users and study participants who provided feedback on the tool and IBM and Slashdot for generating such vibrant discussions.

10. REFERENCES

- [1] Baoill, A. Ó. Slashdot and the Public Sphere. *First Monday* 5, 9 (2000).

