

Exploring Blog Archives with Interactive Visualization

Indratmo, Julita Vassileva, and Carl Gutwin

Department of Computer Science

University of Saskatchewan

Saskatoon, SK S7N 5C9, Canada

{j.indratmo, julita.vassileva, carl.gutwin}@usask.ca

ABSTRACT

Browsing a blog archive is currently not well supported. Users cannot gain an overview of a blog easily, nor do they receive adequate support for finding potentially interesting entries in the blog. To overcome these problems, we developed a visualization tool that offers a new way to browse a blog archive. The main design principles of the tool are twofold. First, a blog should provide a rich overview to help users reason about the blog at a glance. Second, a blog should utilize social interaction history preserved in the archive to ease exploration and navigation. The tool was evaluated using a tool-specific questionnaire and the Questionnaire for User Interaction Satisfaction. Responses from the participants confirmed the utility of the design principles: the user satisfaction was high, supported by a low error rate of the given tasks. Qualitative feedback revealed that the decision to select which entry to read was multidimensional, involving factors such as the topic, the posting time, the length, and the number of comments on an entry. We discuss the implications of these findings for the design of navigational support for blogs, in particular to facilitate exploratory tasks.

Categories and Subject Descriptors

H.5.2 [Information Interfaces and Presentation (e.g., HCI)]: User Interfaces; H.4.3 [Information Systems Applications]: Communication Applications – *information browsers*.

General Terms

Design, Experimentation, Human Factors.

Keywords

Blog visualization, social interaction history, social navigation.

1. INTRODUCTION

Blogs have emerged as a relatively new medium for communication. According to a survey report [16], in the US, about 12 million people maintain blogs, and about 57 million Internet users read blogs. Blogs promote conversation between the bloggers and their audiences by allowing users to comment on

published entries, or to cite the entries of interest in their own blogs and then send notification to the source using the TrackBack protocol [27].

A blog archive contains a collection of entries arranged in reverse chronological order. Typically, only a few of most recent entries are displayed on the front page of a blog. Useful content, however, is not limited to most recent entries; there may be many old entries in a blog that are worth reading and offer valuable information. The value of tips about designing good visualization, for example, does not necessarily decrease with time. Therefore, users may find useful entries by browsing a blog archive.

Browsing a blog archive, however, is not well supported. Typically, blogs only provide links to monthly archives and a list of tags for selecting a subset of entries assigned with a specific keyword. This navigational support does not offer the users any cue where to find potentially useful entries—entries that have sparked lively discussion or have been read by many people. As a result, users have to rely on their own capability to select entries and navigate through a blog archive. This task becomes tedious when users browse a large collection of entries.

To support exploration of blog archives, we developed iBlogVis—an interactive visualization tool that offers a new way to browse a blog archive (see Figure 1). The key features of iBlogVis are twofold. First, it provides a rich overview of a blog to enable users to reason about the blog at a glance. Second, it visualizes the history of social interaction in a blog to help users identify potentially useful entries in the blog. The evaluation showed that the prototype was successful in realistic tasks.

The contributions of this research are the synthesis and application of existing visualization and interaction techniques to a new domain (blogs), and the evaluation of the resulting prototype. There are two generalizable principles embodied in the tool that can be applied to other domains. First, we show that visualizing social interaction history is useful for supporting exploratory tasks; second, our results add further evidence that providing an overview is valuable. Both quantitative and qualitative results of the study confirmed the utility of these principles. These principles can have immediate impact on real-world sites such as Blogger (www.blogger.com) and LiveJournal (www.livejournal.com) to complement their current user interface design.

In this paper, we discuss the design and implementation of iBlogVis and present the results of our usability study. We conclude the paper by discussing the implications of this study for the design of navigational support for blogs, especially to support exploratory tasks.

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.

AVI'08, May 28–30, 2008, Naples, Italy.

Copyright 2008 ACM 1-58113-000-0/00/0004...\$5.00.

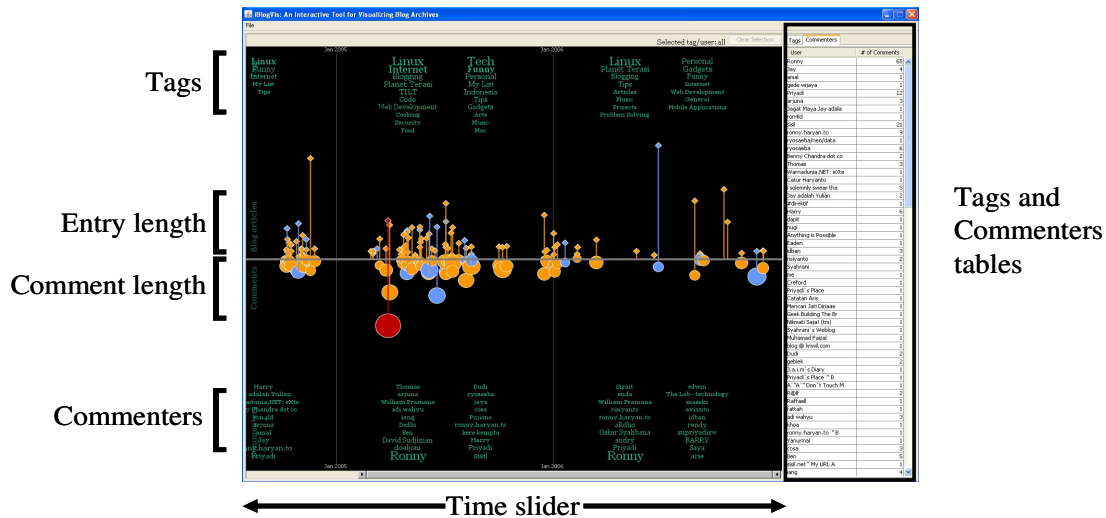


Figure 1. A screenshot of iBlogVis visualizing a collection of blog entries along a timeline.

2. RELATED WORK

Research on blog visualization currently focuses on analyzing and visualizing the link structure and the content of blogspace—a large-scale collection of blogs. Link analysis, for example, has been used to track information flow through blogspace [1, 9]. Gruhl and colleagues [9] developed a model to track the routes of topic propagation through individual blogs. Their model is similar to disease epidemic models in which a blogger gets “infected” with a topic and then spreads the topic further to his/her contacts. In this context, contacts refer to the audience of a blog. Similarly, Adar and Adamic [1] developed a technique to infer the source of information spread in blogspace based on the timestamps of entries and the link structure of blogs. They visualize the inferred routes as “infection trees” where the nodes and the edges represent blogs and propagation paths, respectively.

Herring and colleagues [14] analyze link structure of blogspace to examine the interconnectedness among blogs. They use social network visualization to plot information such as inbound and outbound links and whether the links are one-way or reciprocal. The visualization reveals clusters of topic-oriented blogs that are more interconnected and reciprocally linked than “A-list” (most popular) blogs. A-list blogs, however, have more one-way inbound links, and hence are more reachable than other blogs.

Using visualization, social network analysis, and a sense of community survey, Chin and Chignell [3] developed a model to discover communities in blogspace. Unlike other link analysis methods, their method does not use links mentioned in blog entries. Instead, it uses links provided by bloggers while leaving comments on other blogs (the links usually point to the bloggers’ personal blogs). Chin and Chignell consider such links more explicit in suggesting a social relationship among the bloggers.

As more and more people use blogs to express their opinions, mining blog content can yield useful information. Companies are interested in knowing what people say about their products or what the hot topics are in blogspace. Analytical tools and search engines have been developed to meet such demands, including BlogScope (www.blogscope.net), Digg (digg.com), and Technorati (www.technorati.com).

Tirapat and colleagues [26] developed an interactive tool to examine whether there is a correlation between the success of a movie and the “buzz” in blogspace. They use document-clustering techniques to analyze blog entries and construct a topic map, capturing associations between movies and blog entries. Based on the topic map, the tool generates multiple views to allow users to explore different aspects of the data set.

Harris and Kamvar [12] developed a tool that looks for the phrases “I feel” and “I am feeling” in blog entries and extracts human feelings from the entries along with information about the bloggers (age, gender, and location) and the local weather conditions while the entries were posted. This information is saved; the feeling is identified (e.g., happy, sad) and then visualized as a particle. The attributes of a particle (e.g., color) represent some encoded information. A particle can be clicked to display a full sentence that describes a human feeling. The visualization tool allows users to search and sort data by feeling, gender, age, weather, geographic location, and date.

Other related projects, not specific to blog visualization, include LifeLines [20], TagLines [6], and Timeline [25]. These systems provide an interactive environment or an application programming interface (Timeline) to visualize data sets along a timeline. Specifically, LifeLines visualizes personal histories such as medical records, whereas TagLines identifies most representative tags at Flickr (www.flickr.com) during a certain time period and visualizes the evolution of these tags.

Compared to the existing work on blog visualization, our project differs in the following way. Instead of analyzing and visualizing blogspace, our work focuses on supporting exploration of individual blogs. To some extent, our goal is similar to that of Box Grid [19]. Box Grid visualizes blog entries in a grid, where the position of an entry in the grid is determined by the entry’s category on the vertical axis against its posting date on the horizontal axis. Unlike our visualization tool, however, Box Grid neither visualizes the social interaction history preserved in a blog archive nor allows the users to filter the visual items. In the next section, we discuss the design and rationale behind iBlogVis.

3. DESIGN AND RATIONALE

3.1 Design Objective

The main design objective of iBlogVis is to facilitate exploration of individual blog archives. To achieve this objective, iBlogVis provides an overview of a blog and visualizes not only the content of the blog, but also the history of user interaction, thereby providing cues for social navigation [5]. We expect that users will be able to answer the following questions:

- Content: What is the blog about? Does the blogger post entries regularly? Do the blogger's interests change over time?
- Social interaction history: Did the blog receive many comments from the audience? Which entries received many comments? When did the blog start getting popular? Who are the regular commenters?

Without a tool, to gain such information normally requires one to follow a blog actively: by reading entries and comments regularly, observing interests of the blogger, and identifying active users in the blog. Even so, some of this information is an aggregate of values over a long time, which cannot be estimated easily.

3.2 Visual Design

Blogs display entries in reverse chronological order. To reflect this organizational structure, iBlogVis arranges visual items along a timeline (see Figure 1). The visualization panel consists of two main parts. The upper part (above the timeline) visualizes the content of a blog: entries and the associated tags. The font size of a tag represents the tag's popularity during a certain year. The larger a tag, the more frequently the tag is used during the corresponding year. A blog entry is represented by a diamond shape and a line. The diamond shape provides an interface to view the content of an entry, while the length of a line represents the number of characters in an entry.

The lower part (below the timeline) displays the social interaction history contained in a blog archive. The line length and the font size have similar meanings as above, except that these visual items represent comments and commenters, respectively. The length of a line represents the total number of characters in all comments received on a particular entry, while the area of a circle represents the number of comments in that entry.

The right panel contains two tables. One displays a list of all tags and the number of entries labeled with each tag. Another table displays a list of all commenters and the number of comments they have written on the blog. These tables can be sorted alphabetically (by tag or commenter) or numerically (by the number of entries or comments). These sorting functions are to ease retrieval and reveal the popularity of tags and commenters. Users may select an item from these tables to filter items displayed in the visualization panel.

Visualization of blog content allows users to get an overview of the blog's subjects and the length of the entries. Tag visualization is supported because tagging is currently the main method for classifying entries. Tag visualization is useful for assessing if a blog matches a user's interests, while the length of entries can serve as a retrieval cue. Furthermore, the decision to select which messages to read (in Usenet newsgroups) is influenced strongly by

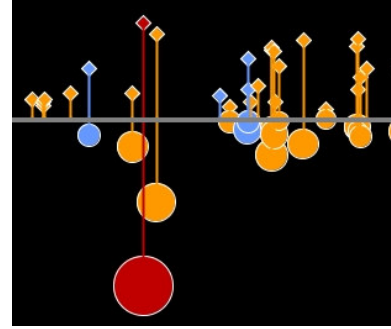


Figure 2. Read wear and social navigational cues.

the subject and the size of a thread [8]. Tags and the length of entries and comments in blogs are similar to the subject and the size of a thread in Usenet, and hence may be influential as well in the exploration of a blog.

Visualization of social interaction history can serve as social navigational cues. As shown in Figure 2, entries that receive many comments can be spotted easily by looking at the size of the circles. Since other users have left many comments on these entries, the entries may offer useful information. With hundreds of entries available in a blog, such social navigational cues can help users select potentially interesting entries without spending too much effort on skimming over every entry in the blog.

To further facilitate the browsing activity, iBlogVis uses the idea of read wear [15] to help users keep track of entries that have been read (blue), have not been read (orange), or the one that is currently being read (red) (see Figure 2). Read wear and its variants have been shown to improve navigation through information spaces [23, 30].

3.3 Interactive Components

The interactive components in iBlogVis are designed based on heuristic guidelines on information visualization: "overview first, zoom and filter, then details-on-demand" [22]. These components also serve as a dynamic queries interface [2], which allows users to formulate queries dynamically and get feedback immediately by clicking on the items in the visualization panel (e.g., tags, commenters), selecting items from the tables in the right panel, or adjusting the time slider (see Figure 1). Highly interactive interfaces are engaging and support exploratory tasks [18], and hence fit the characteristics of browsing a blog archive.

The visualization tool starts by providing an overview of a blog archive (see Figure 1). It displays all entries and comments, most popular tags, and most frequent commenters along a timeline. The popularity of tags and commenters is aggregated on a yearly basis. This overview reveals temporal posting patterns of the blogger and enables the viewers to scan the content of a blog quickly and to reason about its general structure and community dynamics.

Using iBlogVis, users can filter entries by tag, commenter, and posting time. Displaying a subset of entries labeled with a particular tag (filtering by tag) or those commented on by a particular person (filtering by commenter) can be done by clicking on the tag or the commenter, correspondingly. To perform this task, users can click on a tag or a commenter either in the visualization panel or in the right panel (listed in tables). These

filters exist to allow users to limit their search space by removing irrelevant or uninteresting items. Furthermore, since current blogging tools use tagging as the main mechanism to label, organize, and retrieve blog entries, users would expect to be able to explore a blog archive by tag.

Filtering by posting time is supported by a time slider, located at the bottom of the visualization panel (see Figure 1). The slider is positioned across the visualization panel to give users a good sense of the possible query range, as the range corresponds directly to the width of the visualization. This design allows users to derive the query range visually, simply by looking at the timeline in the visualization. Users, therefore, can pay their attention to the visualization while adjusting the time slider. The time slider is also used to zoom in on the area of interest in the visualization. As the visible time range decreases, the level of details in the visualization increases: more tags and commenters are displayed whenever there is more space on screen.

Finally, users can view the content of an entry through a pop-up window by clicking on a diamond shape (representing an entry) or a circle (representing the number of comments on the entry) in the visualization (see Figure 2). When users view an entry, iBlogVis changes the color of the entry to indicate its read wear status.

4. METHODOLOGY

Browsing a blog is an exploratory activity in which people wander around the information space to find the entries of interest. To match the nature of this activity, we conducted an exploratory study to evaluate the usability of our visualization tool. Our tool offers a new way to browse individual blog archives, and to the best of our knowledge, there was no other blog browser that we could use for comparison. Thus, we focused on soliciting feedback on the visualization design and assessing the utility of our approach to exploring blog archives. Due to the similar purpose of the tool (i.e., facilitating exploration of an information space) and the characteristics of user tasks (i.e., exploratory search), the study method was designed based on [17, 28].

The usability of iBlogVis was evaluated using both objective and subjective performance measures. The objective measure was the error rate of the completed tasks, while the subjective measure was the user satisfaction with the tool. The user satisfaction was measured using a set of tool-specific questions adapted from [17, 28] and a short version of the Questionnaire for User Interaction Satisfaction (QUIS 7.0) [4, 11]. We included only relevant QUIS items from the following categories (the number indicates the number of questions used in that category): overall user reactions (6), screen (3), terminology and system information (5), learning (4), and system capabilities (4). Besides using Likert-scale items to measure user satisfaction, we also asked the participants to explain their ratings. These exploratory questions were aimed to shed some understanding of user practices in exploring a blog and to get constructive feedback about the tool.

Since iBlogVis offers an alternative way to browse a blog, the decision whether to use the tool depends much on the user satisfaction with the tool, which makes this criterion important. Quantitative results from this measurement, combined with observation and comments from the participants, can give valuable feedback for improving the visualization tool.

The tasks in the study were designed to evaluate three aspects of the visualization tool: (1) how effective the tool was in giving the users an overview of the content and community dynamics of a blog; (2) how well the tool was in helping the users do things that they could do with typical blogs; and (3) how effective the tool was in giving the users information about community dynamics in a blog (e.g., revealing regular commenters and popular entries).

4.1 Participants

Nineteen students (13 males, 6 females) from the University of Saskatchewan participated in the study. Subject ages ranged from 23 to 37 years old. Each participant received a \$10 honorarium. As a prerequisite, the participants had to be familiar with browsing the web and have some experience reading blogs. On a scale of one (beginner) to five (expert), the participants self-rated their computer skills as an end-user at level three or above (two at level three, six at level four, and eleven at level five). On average, twelve participants browsed the web between one to five hours daily, while the rest spent more than six hours daily.

4.2 Apparatus

iBlogVis was implemented as a desktop application using Java and the prefuse toolkit [13]. It has a pre-processing module that computes the aggregate values required by the visualization and transforms the data structures of a blog into tables and the GraphML format [24]. The data set used in the study contained approximately 100 entries and 300 comments posted from October 2004 to December 2006.¹ The blog attracted regular audience and received comments regularly. All information presented in the visualization was publicly available.

4.3 Procedure

Participants were introduced to the purpose of the study and asked to sign an informed consent form. After that, they were introduced to the features and the meaning of the visualization using a hypothetical data set. This demonstration took approximately five minutes. Then participants were given an opportunity to familiarize themselves with the visualization tool.

After a short period of practice, participants were given a blog data set and a set of tasks. There was no time limit for completing these tasks. There were 16 tasks divided into three categories:

- Overview tasks: getting a sense of the time of the year, identifying the main topics of the blog, recognizing regular commenters, and getting a sense of popularity of the blog.
- Typical browsing tasks: finding the most recent entry, filtering entries by tag (by browsing and selecting a tag from the tag cloud or the tag table), and browsing monthly archives.
- Social navigational tasks: identifying popular entries and finding entries commented on by specific persons.

During the data collection session, we observed how the participants use the visualization tool and took notes of comments and difficulties faced by them.

¹ Thanks to R. Haryanto for providing the data set.

Table 1. Summary of the results of the tool-specific questionnaire (n = 19).

No	Questionnaire items	Not at all easy	Not easy	Easy	Extremely easy
1.	Getting a sense of the time of the year was ...	0	3	9	7
2.	Identifying the main topic of the blog was ...	0	0	8	11
3.	Identifying regular commenters in the blog was ...	0	0	5	14
4.	Getting a sense of popularity of the blog was ...	0	2	12	5
5.	Finding the most recent entry in the blog was ...	0	0	5	14
6.	Finding blog entries tagged by a specific keyword was ...	0	0	4	15
7.	Finding blog entries posted in a specific month was ...	0	4	9	6
8.	Finding popular entries in the blog was ...	0	0	8	11
9.	Finding blog entries commented on by a specific user was ...	0	1	7	11

After performing each task, the participants rated their satisfaction with the tool. Some of the questionnaire items were open-ended questions asking the participants to explain their ratings. The participants were also asked about their most favorite and least favorite features of the visualization tool, desired functions that did not exist at the time, whether they would be interested in using the tool if it were integrated into blogs, and whether they have privacy concern with the information presented in the visualization. At the end of the data collection session, the participants completed a short version of QUIS 7.0 [4, 11].

5. RESULTS

Each data collection session took up to one hour. In general, the participants did not have difficulties in learning to use iBlogVis and completing the tasks. After listening to a brief introduction about iBlogVis, all participants spent less than five minutes to feel comfortable in using the tool and ready for the tasks.

Overall, iBlogVis received positive reviews from the participants. Subjective user satisfaction was high, supported by a low error rate of the completed tasks. Quotations included in this section were from written comments from the participants.

5.1 Error Rate

Out of 16 given tasks, the number of errors made by the participants ranged from zero to three. Seven participants performed all tasks correctly; six made one mistake; five made two mistakes; and one participant made three mistakes. The average error rate was 6.25% (1 out of 16 tasks).

The most common mistake occurred when the participants were asked to find the most recent entry in the blog. Among 19 participants, five performed this task incorrectly. As some of the posting dates of the entries were close to one another, there were some overlapping visual items on the overview of the blog. Mistakes occurred when the participants simply selected a visual item on the overview that seemed to be at the right most position of the timeline. The selected item happened to be the second most recent entry in the blog. The more accurate way to perform this

task was to zoom in on the area of interest to separate the overlapping items before selecting the most recent entry.

5.2 User Satisfaction

The tool-specific questionnaire for measuring user satisfaction used forced-choice Likert-scale items (using four scales, e.g., not at all easy, not easy, easy, and extremely easy). Table 1 provides a summary of the results. Overall, the participants expressed their high satisfaction with the tool. Most items were rated easy or extremely easy. The results of this questionnaire, however, also indicate room for improvement (e.g., problems related to the timeline—see statement 1 and 7 in Table 1).

Besides evaluating task-specific functions above, the participants gave their overall reactions to the visualization tool. They rated how effective the tool was in giving overviews of the content and community dynamics of the blog. Most participants thought that the tool presented an overview of the blog content effectively. Four participants rated it highly effective; thirteen rated it effective; and two rated it not effective. In terms of providing an overview of the community dynamics, iBlogVis was rated highly effective by four users and effective by fifteen users.

The participants also rated how well iBlogVis helped them do things that they could and could not do with typical blogs. Compared to what the participants could do with typical blogs, iBlogVis was rated extremely well by seven users and well by twelve users. Additional functions of iBlogVis (things that the participants could not do with typical blogs) were rated extremely well by nine users and well by ten users.

All participants agreed that having access to the visualization tool would affect their choices of which entries to read (ten strongly agreed, nine agreed). The most common reason was that, by looking at the visualization, they would be able to see the popularity and the length of entries:

“I would like to know both how long an entry is and how popular it is before reading. Since, I would prefer to read short entries that are thought provoking/noteworthy.”

“I’d probably just read the most popular stories first to get a sense of the blogger’s style and focus. Then I’d read a few of the least popular entry for comparison.”

All participants thought that, while exploring a blog, having access to an overview of the blog was useful (eleven rated it extremely useful, eight rated it useful):

“It lets me know what the blog is about overall, the author’s evolving interests over time, and how popular it is based on how many readers it got.”

Most participants also thought that having access to the community dynamics in a blog was useful (nine rated it extremely useful, eight rated it useful, and two rated it not useful):

“It gives me an idea if the entry is worth reading. My experience is that most of the entries I am interested in reading are usually heavily commented. Also, I enjoy reading the comments but prefer many short comments to a few long comments.”

Most of the participants did not have privacy concern regarding information presented by the visualization tool, as the information was already in the public domain, and users had a choice to use pseudonyms while leaving comments. However, two participants indicated their privacy concern, and one was not sure about her attitude to this issue.

The most favorite features of iBlogVis included (1) the visualization of the length of entries and comments; (2) the lists of tags and commenters; and (3) the visualization of the number of comments on blog entries.

The least favorite features of iBlogVis were mostly related to the timeline visualization and the use of a time slider to filter visual items by date. Some participants expected to see a clearer boundary between periods (e.g., a monthly boundary). They also wanted to be able to quickly select exact dates or periods by having predefined filters (e.g., by year, month, or week). Other desired features mentioned by the participants included filtering entries by multiple criteria, searching by keywords, and better highlighting the currently selected tag or commenter.

When asked whether they would be interested in using the tool if it were integrated into blogs, all participants showed interest in the tool (ten were extremely interested, nine were interested).

After completing a tool-specific questionnaire, the participants filled out a short version of QUIS 7.0 using a nine-point scale. Table 2 presents the mean score of each QUIS category used in the study and the lower and upper limits of the mean at 95%

Table 2. QUIS results using a nine-point scale (1: low satisfaction, 9: high satisfaction, n = 19).

QUIS Category	Mean	95% Confidence Interval	
		Lower	Upper
Overall	7.670	7.216	8.124
Screen	7.860	7.295	8.425
Terminology	8.316	7.988	8.644
Learning	8.434	8.091	8.777
Capabilities	8.083	7.687	8.479

confidence level. Based on [21], the midpoint scale (five) can be used to represent mediocre user satisfaction. Compared to this value, the results show that, in all categories, the user satisfaction with iBlogVis was significantly higher than mediocre.

6. DISCUSSION

iBlogVis was designed based on a working hypothesis that providing an overview of a blog and revealing social interaction histories would help users explore a blog archive. As presented in the previous section, both the quantitative and the qualitative responses from the participants supported this hypothesis. A common reason was that an overview and visualization of social interaction history enable users to learn about a blog quickly and to identify popular entries in the blog:

“I can get a quick overview about the blog.”

“I can easily know when the blogger posted entries frequently and find which entries are more popular.”

“It helps me to choose the entry that might interest me most. There might be hundreds of entries with the tag I want, and the community dynamics can help me to filter them.”

Two participants, however, did not perceive having access to social interaction histories as useful. One participant wrote:

“I’m not overly concerned with comments a blog gets. I usually make up my own mind, but it’s sometimes useful to know what the most popular/contentious entry was.”

While all participants acknowledged the usefulness of having an overview of a blog, the usefulness of visualizing social interaction histories depends on the kind of blogs and what the users look for in the blog. Visualization of social interaction histories is particularly useful when the users want to *explore* an information space. That is, they do not have specific information to retrieve, but want to learn about an information space—its content and social dynamics within it—while hoping to find useful or interesting information within the space. In such cases, social interaction histories can provide navigational cues for the users so that they can follow the crowd to find entries that have attracted a lot of attention in the community:

“Usually the topics receiving many comments are ‘hot.’ Thus most likely they will be interesting for me too.”

Visitors to blogs containing information that is easily out-of-date may receive less benefit from the visualization of social interaction histories compared to those visiting topic-oriented blogs. For example, consider a diary-kind of blog used for sharing news with friends. Information contained in this kind of blogs may become out-of-date or irrelevant quickly. Knowing that a friend was visiting our city last week is no longer useful, as we could not meet up with him/her. Furthermore, there is little need to revisit old entries in such blogs. The readers may just want to follow the most recent entry in the blog. From their perspective, comments from other people on old entries are not important.

The content of topic-oriented blogs (e.g., programming tips, education, and research) does not easily become outdated. Old entries may still contain relevant information that is worth reading. Comments from the audience can enrich the discussion, as they may offer different perspectives or add new content to the entry. In this kind of blogs, visualization of social interaction

histories can provide guidance for visitors to select which entries to read and to explore the blogs effectively.

6.1 Design Implications

Responses from the study participants revealed that the decision to select which entries to read was affected by factors such as the posting time, the topic, the length of entries and comments, and the number of comments on entries. To facilitate exploration of blog archives, blogs should provide navigational support that allows users to search for particular entries using these criteria. Relying only on time- and topic-oriented navigation is not enough.

The most favorite feature of iBlogVis was the visualization of the length of entries and comments. Some participants mentioned that these aspects influenced their decision whether to read an entry—an initial finding that was similar to the results of a Usenet study [8]. Some users preferred to read a short, popular entry to a long one, while others might have different preferences.

A challenge for designers is how to present various attributes of blog entries effectively to users. Users should be able to see, compare, and analyze entries from different aspects simultaneously. Simply providing additional sorting functions is insufficient, as there are multiple factors involved, and it is hard to maintain all contextual information that is important to the users while the entries are rearranged based on different criteria. Moreover, users do not always want the most popular entry. What they want may be recently published entries that are not too long and receive many short comments. Formulating such queries is complicated because the criteria are vague: recently published, not too long, many short comments. Requiring users to come up with exact criteria, however, will increase their cognitive load, and hence is not a desirable solution.

Information visualization is a viable solution to the problem above. Designed properly, a visualization tool can present multiple attributes of blog entries simultaneously while allowing users to compare, analyze, and select entries that match their search criteria intuitively without having to formulate complex query statements. Contextual information can be maintained by providing an overview of a blog and enabling users to interact with the information space through a dynamic queries interface [2]. The power of information visualization relies on the fact that human vision is excellent at comparing, extracting, and recognizing patterns [29].

Despite its potential, information visualization is not a panacea for all navigational problems in blogs. For lookup tasks [18] such as fact retrieval, using search engines or keyword-based retrieval is more appropriate than using visualization tools because queries can be formulated easily in these tasks, and there is no or little need to compare the query results. Visualization, therefore, should be seen as a supplement to the current navigational support for blogs, especially to ease exploratory tasks.

From our observation, several participants tried to click on items displayed in the visualization panel, such as months, when the given tasks were relevant to the items. For example, when participants were asked to find entries posted in a specific month, some of them tried to highlight the entries by clicking on the corresponding month in the timeline. Repeated attempts to click on items in a visualization panel were also observed in a Usenet

study [28]. This observation implies that users expect that each item in a visualization panel, whenever relevant to their task, can be clicked to help them perform the task at hand.

6.2 Critical Reflection

Scalability in general is a challenging issue for visualization, and applies to our tool as well. In practice, however, our experiences suggest that users want to deal with a manageable data set at a time. Using our tool, users can narrow down a larger data set by filtering entries by tag, posting time, or commenter, and then zoom in on the area of interest to reduce visual occlusion. Another approach is to provide another level of overview such as a miniature view—a global overview on a small scale [10]. When the main visualization panel cannot show all entries, a miniature view would allow users to maintain contextual information of their position in a blog archive and then select a smaller set for exploration.

As elaborated by Ellis and Dix [7], empirical evaluation of visualization tools inherits several limitations, such as the absence of standard data sets and user tasks. Furthermore, people usually use a visualization tool to perform exploratory tasks, making it even more difficult to come up with a set of standardized tasks and reliable, objective performance measures. While our study also inherited these limitations, we have tried to incorporate Ellis and Dix's suggestions into our study methods as follows.

Our study used both quantitative and qualitative methods. The questionnaires gave quantitative results of the study; that is, a set of numbers indicating how satisfied or dissatisfied the participants were with our visualization tool. The observation and open-ended questions, however, produced insightful information beyond these numbers. This qualitative data gave some explanation for *why* users rated certain features of the visualization tool as useful or not useful. This explanation shed some knowledge of user practices in exploring blogs and contributed to understanding in which context revealing social interaction histories contained in blog archives is perceived to be useful.

Our study used a single data set. Although taken from a real blog, the data set might not be representative. There is no guarantee that the participants would give similar ratings if the visualization tool was used to visualize different data sets, particularly those having different characteristics (e.g., photo blogs). Therefore, the utility of iBlogVis is currently limited to a certain class of blogs. That is, topic-oriented blogs that contain mostly textual entries, have medium posting frequencies (a few entries per week), and receive comments from the audience regularly. Visualization of social interaction history is one of the main features of iBlogVis. Without using a data set that contains social interaction history, iBlogVis would not be able to deliver its full functionality, which consequently could affect the ratings of its utility.

Finally, most participants considered themselves to be advanced computer users. Most have background in computer science. Their ability to learn and use iBlogVis might not represent the average user's ability. On the positive side, their expertise was valuable in providing constructive feedback about the system.

7. CONCLUSION

We discussed the design, implementation, and evaluation of iBlogVis—an interactive visualization tool for facilitating

exploration of blogs. The tool was evaluated using various methods. First, the design rationale was explained and justified based on existing research on human-computer interaction and information visualization. Second, the usability of the tool was evaluated using both subjective and objective performance measures. The results of these measures showed that the user satisfaction was high, and the average error rate of the given tasks was low. Third, the study explored the reasons behind the user satisfaction ratings qualitatively, using observation and comments from the participants. These qualitative responses have added to the understanding of blog reading behavior and how to apply visualization techniques to ease exploratory tasks in blogs.

Comments from the participants indicated that the decision to select which entries to read was affected by multiple factors. Besides the topic and the posting time of an entry, the length and the number of comments on the entry also influenced the decision. The important role of these factors reflected in the participants' responses about their most favorite feature of iBlogVis: the visualization of the length of entries and comments. Thus, to facilitate exploratory tasks, blogs should provide additional support beyond the current time- and topic-oriented navigation.

There are several directions to follow up on our initial research.² From the development perspective, there are various features that can be refined or added to our prototype, such as advanced search and bookmarking facility. The prototype can be developed further as a web-based application and be used to visualize a blog in real time. Then a field study can be conducted to observe how people actually use the visualization tool to explore a blog archive.

8. REFERENCES

- [1] Adar, E. and Adamic, L.A. Tracking information epidemics in blogspace. In *Proc. Web Intelligence*, 207-214, 2005.
- [2] Ahlberg, C., Williamson, C., and Shneiderman, B. Dynamic queries for information exploration: an implementation and evaluation. In *Proc. CHI*, 619-626, 1992.
- [3] Chin, A. and Chignell, M. A social hypertext model for finding community in blogs. In *Proc. HYPERTEXT*, 11-22, 2006.
- [4] Chin, J.P., Diehl, V.A., and Norman, K.L. Development of an instrument measuring user satisfaction of the human-computer interface. In *Proc. CHI*, 213-218, 1988.
- [5] Dieberger, A., Dourish, P., Höök, K., Resnick, P., and Wexelblat, A. Social navigation: techniques for building more usable systems. *interactions* 7(6), 36-45, 2000.
- [6] Dubinko, M., Kumar, R., Magnani, J., Novak, J., Raghavan, P., and Tomkins, A. Visualizing tags over time. In *Proc. WWW*, 193-202, 2006.
- [7] Ellis, G. and Dix, A. An explorative analysis of user evaluation studies in information visualization. In *Proc. BELIV*, 1-7, 2006.
- [8] Fiore, A.T., LeeTiernan, S., and Smith, M.A. Observed behavior and perceived value of authors in usenet newsgroups: bridging the gap. In *Proc. CHI*, 323-330, 2002.
- [9] Gruhl, D., Guha, R., Liben-Nowell, D., and Tomkins, A. Information diffusion through blogspace. *SIGKDD Explor. Newsl.* 6 (2), 43-52, 2004.
- [10] Gutwin, C., Roseman, M., and Greenberg, S. A usability study of awareness widgets in a shared workspace groupware system. In *Proc. CSCW*, 258-267, 1996.
- [11] Harper, B., Slaughter, L., and Norman, K.L. Questionnaire administration via the WWW: a validation and reliability study for a user satisfaction questionnaire. In *Proc. WebNet*, 808-810, 1997.
- [12] Harris, J. and Kamvar, S. <http://www.wefeelfine.org/>
- [13] Heer, J., Card, S.K., and Landay, J.A. prefuse: a toolkit for interactive information visualization. In *Proc. CHI*, 421-430, 2005.
- [14] Herring, S.C., Kouper, I., Paolillo, J.C., Scheidt, L.A., Tyworth, M., Welsch, P., Wright, E., and Yu, N. Conversations in the blogosphere: an analysis "from the bottom up." In *Proc. HICSS-38*, 2005.
- [15] Hill, W.C., Hollan, J.D., Wroblewski, D., and McCandless, T. Edit wear and read wear. In *Proc. CHI*, 3-9, 1992.
- [16] Lenhart, A. and Fox, S. Bloggers: a portrait of the internet's new storytellers. http://www.pewinternet.org/PPF/t/186/report_display.asp, 2006.
- [17] Marchionini, G. From overviews to previews to answers: integrated interfaces for federal statistics. http://ils.unc.edu/~march/bls_final_report_99-00.pdf, 2000.
- [18] Marchionini, G. Exploratory search: from finding to understanding. *CACM* 49 (4), 41-46, 2006.
- [19] Phiffer, D. Box Grid. <http://phiffer.org/projects/box-grid/>
- [20] Plaisant, C., Milash, B., Rose, A., Widoff, S., and Shneiderman, B. LifeLines: visualizing personal histories. In *Proc. CHI*, 221-227, 1996.
- [21] Quant QUIIS: information about quantitative analysis. <http://lap.umd.edu/quis/QuantQUIIS.htm>
- [22] Shneiderman, B. The eyes have it: a task by data type taxonomy for information visualizations. In *Proc. IEEE Symposium on Visual Languages*, 336-343, 1996.
- [23] Skopik, A. and Gutwin, C. Improving revisitation in fisheye views with visit wear. In *Proc. CHI*, 771-780, 2005.
- [24] The GraphML file format. <http://graphml.graphdrawing.org/>
- [25] Timeline. <http://simile.mit.edu/timeline/>
- [26] Tirapat, T., Espiritu, C., and Stroulia, E. Taking the community's pulse, one blog at a time. In *Proc. ICWE*, 169-176, 2006.
- [27] TrackBack Technical Specification. http://www.sixapart.com/pronet/docs/trackback_spec
- [28] Viégas, F.B. and Smith, M. Newsgroup Crowds and AuthorLines: visualizing the activity of individuals in conversational cyberspaces. In *Proc. HICSS-37*, 2004.
- [29] Ware, C. *Information Visualization: Perception for Design*. Morgan Kaufmann Publishers, 2000.
- [30] Wexelblat, A. and Maes, P. Footprints: history-rich tools for information foraging. In *Proc. CHI*, 270-277, 1999.

² Thanks to an anonymous reviewer for this feedback.