Deliberation in the Wild: A Visualization Tool for Blog Discovery and Citizen-to-Citizen Participation

Candida Tauro, Sameer Ahuja, Manuel A. Pérez-Quiñones, Andrea Kavanaugh, Philip Isenhour
Center for Human Computer Interaction
Department of Computer Science
Virginia Tech
Blacksburg VA, USA 24061-0106
(540) 231-6931
{ctauro, sameer.ahuja, mperezqu, kavan, isenhour}@vt.edu

ABSTRACT
Web logs (or blogs) have become a means for citizens to share opinions and deliberate on local issues. However, the large number of blogs makes finding and exploring content of interest relatively difficult. This discovery problem presumably also limits participation by interested citizens. We present a tool to display a representation of citizen-to-citizen discussion in blogs that reveals similarity across blog entries. Through association and content analysis, blog entries are linked to each other to form clusters of related local content. Users can navigate and explore online discussions by manipulating the graph, filtering content, and clicking on a blog title to go directly to a given blog in order to read further. The visualization of online discussion can promote participation by highlighting popular topics and laying out the structure of conversations. We conducted a case study on regional Southwest Virginia blogs to validate the tool’s usability and capacity for facilitating citizen-to-citizen discussion, as well as government awareness of diverse voices in the local community. In this paper we present the tool design, its functionality, the usability evaluation and summarize the results.

Categories and Subject Descriptors

General Terms

Keywords
Visualization, blogging, online discussion, local community.

1. INTRODUCTION
Online deliberation is a term associated with an emerging body of practice and research dedicated to fostering purposeful discourse over the Internet. This process of online deliberation includes both knowledge acquisition and knowledge transfer from one participating unit to another. The literature review and fieldwork have revealed some interesting contrasts seen in the usage of discussion forums and blogs for online discussion.

There are numerous politically active individuals and groups discussing civic issues and exchanging information online. Typically, the online systems they use aim to aggregate public deliberation within a centralized site or forum. While these centralized online discussion forums have been successful in stimulating deliberation, they have several limitations. These include the tendency to attract the usual activists, difficulties in scaling up beyond this core group and limited breadth of information exchanged [9, 12].

Unlike discussion forums, weblog networks are decentralized in nature and comprise of scattered yet interlinked individual weblogs. In weblog networks, individual bloggers define the discussion agenda that they are interested in, and users can offer informal observations on various issues without the constraints of rules and formality associated with online forums. This relaxed and individualistic control structure of weblog networks appeal to a number of users who are not activist yet want to express an opinion on a particular theme once in a while. For example, the majority of bloggers are not political activists, but they do tend to be relatively well informed on a variety of topics and issues. Blogging is regarded a social activity [16] and motivations for participation vary greatly [15, 16].

The decentralized nature of blogs makes them easy to use because users can easily offer informal observations on various issues. Participating in centralized forums can dissuade all but the most motivated and determined users. Instead of trying to find a centralized site where conversation is directed, bloggers set up their own sites (blogs), usually at no cost and with relative ease, and just start writing about various topics. Typically, these share personal experiences [15], although political opinions, observations and information are scattered throughout many of these blogs.

When interlinked, these scattered users exhibit a self-organizing social system that allows individuals to interact and share ideas and information among themselves [9, 10]. Ironically, the distributed nature of blogs that makes them attractive to a broad and diverse set of voices from politically less active citizens, is also fundamental to the problems associated with using blogs for citizen-to-citizen deliberation (what we call deliberation in the wild). By disposing of formal structure, blogs lose the strict categorization and aggregation of discussions that more structured approaches such as forums enjoy. The lack of such organization makes discovering content and joining discussions of interest difficult. In addition, the ‘discoverability’ of a blog via web search is in proportion to its search rankings or popularity, which tends to bias the search results towards the opinions of more popular blogs and websites, and bury the results from other less popular but more locally relevant blogs. To address these problems, we have designed a tool to help find and participate in citizen-to-citizen discussion and deliberation that takes place in blogs.
2. BACKGROUND AND RELATED WORK

A blog is a web-based publication technology consisting primarily of periodic articles, most often sorted in reverse chronological order. Blog is the most recent application in the domain of computer-mediated communication technologies that enables a user to publish content online with ease. Blogs can be hosted by dedicated blog hosting services, or they can be run using blog software on regular web hosting services, such as the aggregator site for Southwest Virginia blogs (http://www.swvanews.com).

The blog entry (also called a post or a message) is the basic unit of a blog’s content. The style of entries varies widely, from short passages that simply link to other content, to analysis of quoted content from a news article or other blog, to lengthy segments of original content. The most recent entry, displayed first on the page, is what a visitor is most likely to see at the first glance and to read first. The writer of these blog entries is normally called a blogger. Besides the entry text itself, entries usually have both a header and footer with additional pieces of information [7, 15]. A blog entry usually consists of the following:

- **Title**: main title, or headline, of the post.
- **Body**: main content of the post.
- **Permalink**: the URL of the full, individual post.
- **Post Date**: date and time the post was published.

The social network [20] of blogs (world of bloggers) is usually called the blogosphere. The growth that the blogosphere has been enjoying was detected early on by the Pew Internet & American Life Project [14]. More recently, July 2006 [17], 39% of internet users were reading blogs, a 129% jump from the 17% users who were reading during February 2004 survey [14]. 8% of Internet users, or about 12 million American adults, had their own blogs to publish information and exchange ideas.

The blogosphere comprises of a few densely connected weblog groups, but the majority of weblog groups are only partially interconnected [8]. Herring and colleagues further argue that the blogosphere is sporadically conversational in nature. Other studies on weblog networks and blogosphere reveal the self-organizing characteristics of weblog networks that results into highly connected groups around particular topics and themes.

Kumar et al. have studied and modeled sudden bursts of connectivity within blogosphere over based on an analysis of the evolving link structure [13]. Their study shows that individual bloggers come together to discuss events. This collective conversation gets bigger to take the form of a spike. Gruhl et al. furthered the work of Kumar et al. and found that discussion in blog communities was generally composed of chatter (ongoing discussion whose subtopic flow is determined by decisions of the authors), although, at times spikes (short-term, high-intensity discussion of real-world events that were relevant to the topic) appeared and a large number of bloggers were exposed to this spread [5].

Kumar et al.’s very large-scale study of interlinking among blogs found that the phenomenon of dynamic conversations across small micro-communities (virtual, not physical) is extremely widespread, and correlated with sustained blogging activity [13]. They also address a possible counter-argument by showing that such interlinking is not simply a function of the size of the blogosphere as a whole.

Ultimately, inter-linking of the sort required to get conversations started has the relatively invisible prerequisite of “inter-reading” [4]. In other words, given the huge scale of the blogosphere, why are any two people likely to be reading each other's blogs, and, ultimately, writing about and sending their readers to each other's blogs? We believe that if we constrain the set of blogs based on location, and if the set of bloggers is sufficiently diverse, then interlinked conversations are likely to be about regional issues. A physical trainer, a semi-retired executive, a lawyer, and a firefighter probably read blogs from people in similar roles from other places, but are more likely to read each other if they are affected by the same civic/political issues, get the same daily newspaper, have kids at the same school, belong to the same church or civic organization, or have other real-world connections that are predicated on living near each other.

Although blogs are sometimes perceived as merely “personal diaries” rather than places for “deliberative practice”, we believe that this distinction is simply not meaningful. Herring et al. [7, 8] places weblogs somewhere between asynchronous discussion forums and standard webpages to emphasize that although most blogs seem to be personal in nature, these blogs share characteristics with other genre of offline discussion, such as newspaper editorials. Deliberative activity, taken to mean something broader than rigid, formal deliberation, is more likely to occur in the context of a more personal online space (where there are fewer rules and constraints), in the same sense that real-world civic and political talk is more likely to take place around the water-cooler, at the dinner table, or in the stands at a little league game than at a town council meeting [11].

Our visualization tool’s goal is to allow the discovery of online discourses that take place in blogs. The value of a tool for discovering this discourse is that it mediates between the needs of the discourse-producers (citizens, who would prefer to choose or create their own "comfortable setting" for discourse) and discourse-consumers (other citizens, government officials, and researchers), who would prefer that bloggers discussed and deliberated in the same setting, to allow discourses to be found with ease.

3. VISUALIZING BLOGS WITH VIZBLOG

Previous research [1, 2, 3, 8, 17] has shown that public discussion forums, message boards, newsgroups and most recently blogs are means for citizens to share information (and misinformation), views and opinions. However, the vast volume of blogs accessible online render exploring content and locating discussions of interest fairly difficult. Our visualization tool, VizBlog (Figure 1), scans a predetermined set of blogs of local origin and creates a network visualization of citizen discussions. Through association and content analysis, blog entries are represented by nodes and linked to each other to form clusters of related local content.
The tool broadly follows the design guideline of Visual Information-Seeking Mantra (Overview first, zoom and filter, then details-on-demand) [19], and provides techniques for Information Visualization tasks of overview, zoom, filtering, details on demand and relations to support the user goals of discovery and insight. The following sections describe the tool and its various features that aim at facilitating its use and information discovery.

3.1 Pre-Processing Blogs for VizBlog

The data used in the visualization is generated by a preprocessor that parses a set of blogs and creates a GraphML1 document. The preprocessor uses the vector space model [18] to determine the similarity coefficient between blog entries. In addition, it creates a list of top keywords from the content of the blog entries. We took the set of blogs for our case study from a regional blog aggregator (http://www.swvanews.com) that gathers a large set of blogs (about 40 and growing) discussing various local (and non-local) topics from Southwest Virginia (see Figure).

Bloggers gathered on the aggregator site generally live in the area or have some other strong attachment to the region; these include local citizens, students, business-owners, elected officials and candidates, professors, retirees, and many others. There are numerous blog aggregator websites throughout the Internet. For example, greensboro101.com gathers blogs local to Greensboro, North Carolina.

Figure 1: The VizBlog visualization tool (Overview)

Figure 2: http://www.swvanews.com blog aggregator.

\[1\] http://graphml.graphdrawing.org/
3.2 Visual Representation

VizBlog has been developed using the Prefuse visualization toolkit [6]. The visualization is made up of nodes and links, and is animated using a modified version of the force directed layout of Prefuse. The animation can be paused or restarted using the right panel controls. Figure 1 shows the main window for VizBlog. The right panel provides features for navigating, filtering and searching the visualization.

3.2.1 Nodes

Each node in the visualization represents an individual blog entry, and is labeled with the title of the blog entry. The color of all nodes is blue, but it can change dynamically based on whether the node is selected, or is the neighbor of a selected node or if it is a search result. The size of a node is in direct proportion to the number of neighbors, i.e., the number of links it has. Since a link can represent citations or similarity with other nodes, the nodes with these characteristics are enlarged in the visualization. This form of coding stems from the observation that the blog entries that are central to a ‘conversation topic’ receive the most citations, and most of the peripheral blog entries in the conversation are similar to these central entries. The coding is a way to recognize these central nodes.

3.2.2 Links

The links between the nodes represent relationships – Two nodes can be linked if there is a citation from one to another, or if they are similar in content as determined by the preprocessing, or both. This results in three kinds of links in the visualization. Figure 3 shows the three types of links and how they are depicted graphically in VizBlog.

![Figure 3: Visual representation of links. (A) and (B) represent weak and strong links respectively, (C) represents a hyperlink and (D) represents similarity + hyperlink.](image)

Similarity links represent the similarity calculated between the two nodes it connects, as per the vector space model [18]. Similarity is represented visually as a gray-colored undirected line. The thickness of the line is directly proportional to the value of the similarity coefficient between the connected nodes. The more similar two entries are, the thickest the line connecting them (see link A and B in Figure 3).

Hyperlinks between blog entries are represented in the visualization via a dashed arrow that is pink in color (see C in Figure 3). The arrow point from the citing node to the cited node.

If two nodes are related by a citation and a similarity coefficient greater than some prespecified cut-off value, then they are represented by a dark red arrow that points from the citing to the cited node (see arrow D in Figure 3). The thickness of the arrow is in direct proportion to the value of the similarity coefficient.

3.3 Visualization Features

3.3.1 Overview

The purpose of a visualization overview, as defined in the Visual Information-Seeking Mantra [19], is to provide a ‘starting point’ to the users, for them to be able to quickly recognize the major features of the visual space, and start exploration.

Figure 1 shows VizBlog’s initial overview screen. A one month data set for the Southwest Virginia blog aggregator contains about 1000 blog entries. With such high number of nodes, cluttering and occlusion become significant problems at the overview levels. The individual nodes in the layout repel each other slightly, so that the overview is spread out, preventing occlusion. Semantic zooming reduces the visual representation of nodes at overview levels by decreasing the length of the node title, further reducing clutter and occlusion (Figure 4).

![Figure 4: Semantic zooming – varying title length as the zoom value increases from views A to C.](image)

The overview provides easy identification of clusters of inter-related posts. In general, the posts in a cluster revolve around one or two central topics of interest, and the largest nodes in the cluster are the ones closest to the central topic (Figure 5).

![Figure 5: A cluster of entries about the retirement of the fire chief of the town of Bristol, TN.](image)

The tool also displays the top keywords extracted by the preprocessor as a cloud in the right side panel. This cloud provides the viewer a depiction of the most used keywords in the data set (Figure 6).
### 3.3.2 Zoom and Filter

From the overview, the tool provides for zooming into areas of interest to the user via zoom and pan interactions. A user can right-click-and-drag on any point on the display to zoom in or out the display centered at that point. The application also supports scroll-wheel based zooming. In addition, users can zoom into a particular area of interest by pressing the middle mouse button (or scroll-wheel) and drawing a rectangle over the area. Panning is supported via left-click drag on an empty area.

The user can, at any time, zoom out to the overview level by clicking on the “View All” button on the right panel of the main window (see Figure 1).

To aid exploration of the visual space, VizBlog provides several features meant to aid filtering of data items in the visualization. These provide the user mechanisms to perform directed exploration of the visual space for the topics or entries of their interest.

**Filter by similarity.** The similarity slider (shown in the middle of the right panel of Figure 1) lets the users filter off edges based on their strength. This strength is determined by the value of the similarity coefficient for the two nodes connected by the edge. The coefficient values vary between 0 for least similar to 1 for identical. By default, VizBlog displays only those similarity links that have a coefficient of 0.1 or more. By moving the slider, the user can modify this system cut-off. This has an effect of changing the density of the clusters. An example of this change is shown in Figure.

Figure 7: Using similarity slider to filter-off weak links. (A) shows all the links and (B) shows some links filtered out

Further, the tool provides an option to filter off nodes that are not connected to any other node in the current view. The presumption behind this feature is that while exploring the clusters, the user might not be interested in unlinked nodes that are not a part of any cluster. Furthermore, since we are interested in deliberation in the wild with an emphasis on citizen-to-citizen deliberation, blogs that are not connected to other entries are not as relevant to our purposes.

The filtering significantly reduces the number of visible visual items, and increases the visibility of clusters. Figure 8 shows an example of this. By increasing the similarity filter in combination with filtering of unconnected nodes, the user can quickly reduce the number of visual items to view only the most similar entries. These nodes provide the general themes of the clusters they are part of, and for more details, the user can reduce the filter to show the other related nodes.

Figure 8: Identifying clusters is significantly easier after hiding unconnected nodes (B).

### 3.3.3 Details on demand

Eventually, users would want to explore further details of particular entries. In VizBlog there are three ways to obtain more details as needed. First, left clicking on a node opens the particular blog entry in a browser window. Second, mouse over a node displays more information about the node. Finally, searching highlights nodes that match the search term.

A user can mouse over a node to view its full title via a tool-tip. The mouse-over also causes the particular node to change its color to red, and its immediate neighbors are highlighted in orange (Figure 9). Upon left-clicking the node, the relevant blog entry opens up in the default system web browser.

Figure 9: Upon left-clicking the node, the relevant blog entry opens up in the default system web browser.
Figure 9: Mouse-over of a node reveals its title and highlights its neighbours

Figure 10: VizBlog search showing the results of searching for “Virginia”

Search. Often users are interested in particular topics or keywords. The search feature in VizBlog highlights nodes that match a search string (see Figure 10 for an example). A typical problem with search in visualizations is providing context. If, for example, the user is in a zoomed-in view and performs the search, he may not see all the results. The search in VizBlog, along with the visual feedback of color change, provides a textual feedback prompting the user to switch to the overview to view all results.

4. EVALUATION OF VIZBLOG

To evaluate the design of VizBlog, we conducted a usability study using citizens and government officials from the town of Blacksburg, Virginia. The main purpose of the study was to determine whether citizens and government representatives were able to identify topics of discussion with ease. A secondary goal was to evaluate the ease of use of the functionality built into VizBlog.

The evaluation used three different groups of users. The three groups can be characterized as experts, citizens, and town officials. The expert users were Computer Science Graduate students, most of which have experience using advanced computational tools such as VizBlog. The town officials were selected because of their interest in the use of information technology for government purposes. The graduate students and town officials were invited via email to participate in the study. Participants for the citizen group were selected from a pool of participants who had previously participated in a focus group interview in Fall of 2005 and had agreed to be contacted again. We called them and invited them to participate for this study. The participants in the citizen group received $25 for participation in the study.

The graduate students were selected because they are experts in technology use. Findings within this group can help us identify serious failures in the tool. If some functionality cannot be understood by the experts, there is really no hope that the citizens and town officials would find it easy to use.

VizBlog helps interested citizens discover discussion on topics of interest. VizBlog also should help government officials and staff to ‘see’ what a portion of the public is saying about local issues on the blogosphere. Without VizBlog, there are just too many individual blog entries to be able to make much sense of the overarching topics of discussion taking place among them.

For this study, we used a one-month snapshot from the Southwest Virginia blog aggregator. The month used spans from February 21st till March 21st 2007. This includes just under 1000 blog entries.

4.1 Study Protocol

Each participant was allotted 1 hour. We greeted them on arrival and informed them about the purpose of the evaluation and also about the tool in general. We then explained the procedure of the study. We gave each participant two copies of an Informed Consent form, which if they agreed, they signed and were instructed to keep one copy for their records.

We gave a short tutorial to each participant. The tutorial had explicit instructions on how to perform a series of tasks with VizBlog. The goal was that the participants would learn how to use the system by focusing on how to use the different features of VizBlog. One of the tutorial tasks was how to use the Search box. The steps were:

1. Type “Global Warming” and press Enter.
2. Open the blog in a Web browser by left clicking on the entry.
3. Close the browser to return to the visualization.

In addition, we provided the participants with a reference sheet and asked them to read it before continuing. This reference sheet had a glossary of terms used in the software and also included screen shots with explanations about its functionality.

Once the participants completed the tutorial we asked them to fill out an online pre-evaluation survey. This questionnaire included a demographic survey and also a survey about the participants’ experience using the Internet and reading/writing blogs.

Following the pre-evaluation questionnaire we handed the participants a sheet that contained a set of pre-designed benchmark tasks. These were a set of 6 tasks that covered most of the functionality of the tool. The tasks used were:

1. Pick the largest cluster and identify the topic of discussion.
2. Identify the top 3 keywords in the visualization.
3. Can you find a cluster that talks about the “Marsh Fork Arrests”??
4. Find a blog that talks about "Fire Chief Vinson's retirement" and that discusses "house fires".
5. Identify a blog entry that cites more than two other blog entries.
6. Identify a blog entry that you find interesting.

We did not time the users on any of these tasks. We informed them that there was no right or wrong way of performing the tasks and that this evaluation was to test the tool and not them. During the evaluation we encouraged the participants to “think aloud” while performing the tasks. Following these tasks, we asked the participants to fill out an online post-evaluation questionnaire.

4.2 Usability Lab Setup
The usability lab was setup in a closed lab in the Corporate Research Center at Virginia Tech. This room was not accessible to the general public when the study was in session. The evaluator and the participant were both present in the room at the same time. Once the tutorial was completed, communication between the evaluator and the participant occurred only in an event when the participant had questions or comments. In general, there was no communication between the evaluator and the participant, during the time of the actual tasks.

The entire study was audio recorded and the on-screen activity was captured. Camtasia was used as the screen capture and voice recording software to record all data for later reference. An external microphone was used for audio recording.

4.3 Pre-evaluation Questionnaire
We used an online pre-evaluation questionnaire to gather demographic information about participants (e.g. job, age) and to assess their use of the Internet. This section describes the results from this questionnaire. Figure shows the Internet and blog use for all three groups.

The age for all participants ranged from 20 to 71 yrs. We had 10 participants in the citizen group, 10 in the students group, and 3 town officials. The data shows that about 65% (15 out of 23) of the participants read news online using a web browser. Only about 17% (4 out of 23) of the participants use an RSS reader, the remaining users do not regularly read news online.

4.3.1 Citizen Group
There were a total of 10 participants in the citizen group. Their ages ranged from 40 to 71 yrs. From the survey, we learned that participants from the citizen group overall used the Internet regularly. They use it more for personal reasons than for work. Most of the participants read political blogs a few times a month and local blogs less than once a month. These participants have almost never written blogs; only 2 of them do it at all. One of them writes a blog once a week and the other one several times a month.

4.3.2 Town Officials
A total of 3 town officials participated in this evaluation. Their ages ranged between 38 to 58 yrs. The data shows that town officials use the Internet on a regular basis for work and personal reasons equally. They read blogs on an average more than once a week and read more political blogs than local blogs. They also visit the Southwest Virginia aggregator once a month. Town officials wrote blogs less than once a month.

The only characteristic that differentiates the citizen group from the town official group is in the use of the Internet at work. Both use the Internet equally for personal reasons. The other attributes are closely matched. Since we were able to attract only 3 participants from the town, and because their profiles do not differ much, we lump them together for the rest of the analysis done in the paper.

4.3.3 Computer Science Graduate Students
A total of 10 graduate students participated in this evaluation. These graduate students are considered experts in this evaluation. Their ages ranged from 22 to 27 yrs. Most of the Computer Science students used the Internet, both for work and personal reasons on a daily basis. They read blogs more than once a week but not so much of Political or even less local blogs. These participants wrote blogs a little more than once a month, with two of them writing several times a week.

4.4 Task Completion
A major goal of our usability evaluation is to identify if participants can complete the tasks enumerated in section 4.1. Overall and across all users and tasks, 85% (117 out of 138) of tasks were completed successfully. The completion rate among different tasks and groups varied a little, but in general users were able to use VizBlog to do the work requested.

![Average Internet and Blog Use](image)

Figure 11: Average Internet and blog use. Choices were: 6 = Several times a day, 5 = At least once a day, 4 = Several times a week, 3 = Once a week, 2 = Several times a month, 1 = Once a month, and 0 = Never

![Percentage of Participants that Completed Each Task](image)

Figure 12: Percentage of participants that completed each task

The graph in Figure 12 shows that on an average 93% of all the participants were able to identify a particular cluster. 76% of the
participants were able to identify the discussion topic in a cluster. About 86% of the participants were able to identify the top three keywords in the visualization. 87% of the participants were able to find a cluster that discussed a given topic. 73% were able to identify a particular blog entry (e.g., talked about “house fires”) that was part of a larger cluster (e.g., “chief vinson retirement”). All participants were able to find two blogs that cited each other.

However, when we break up the completion rates by groups, we found that the student group performed strong in all tasks (at least 80% completion in all tasks) but the citizen/town group did not fare well in a couple of the tasks. Figure 13 shows each group’s performance for each task. Task 2 and 4, in particular, were difficult for the citizen/town group. Task 5, in our opinion, was the most difficult task in the study. The next section provides some insights as to why some users had trouble completing some tasks using VizBlog.

Easy to identify blogs quoting or citing other blogs: Ninety percent of students strongly agreed/agreed that it was easy to identify blogs citing other blogs. The citizen/town group agreement was lower, at 77%. A participant from the graduate student group said “I liked the citation relationships; it made it very easy for me to see how various blogs are connected, which is good considering that if you read a lot of blogs you may be going back and forth.”

Easy to identify the top keywords: The student group agreed (80%) that it was easy to identify the top keywords in the visualization using the keyword cloud. One of the participants said “I liked the keyword cloud the most. It showed the hot topics clearly.” 77% of the citizen/town found it easy to use too.

Easy to use search: Using the search is where the difference between the students and the citizen/town was the most pronounced. 100% of the students found the search easy to use. However, only 62% of the citizen/town found it easy to use. Four participants were neutral and one disagreed that it was easy to use. It seems that participants in the citizen/town group did not understand how the search worked, and thus had trouble figuring out what happened when they searched for keywords. Our current design for the search might be more appropriate for experienced computer users. We have to rethink how to better present the search and the search results for novices.

Easy to find specific blog entries using search: Although some participants had problems with the search, all participants were able to find specific blog entries using the search feature. Nevertheless, only 80% of the student group and 77% of the citizen/town group agreed that it was easy to do.

4.5 Post-evaluation Questionnaire

The post-evaluation questionnaire helped us gain insight about the usability of the tool and the participants experience using the tool.

Ease of use: Citizens and town officials were mostly neutral about the ease of use of VizBlog. But most of the graduate students agreed that this tool was easy to use. One of the student participants said “this tool provided a global view of all the blog entries… VizBlog really made it easy for me to look for information I was interested in.” Another participant mentioned, “It was so easy to navigate through the information, though the number of blog entries was large.” A participant from the Citizen group said “VizBlog brings many voices together and provides a way of managing them and navigating among them.”

Features were intuitive: Just over half (7 of 13, 54%) of the citizen/town participants agreed or strongly agreed that VizBlog was intuitive. One town official said “Sense of Overview was great. I felt I zoomed in mentally on a topic even as I zoomed in visually, sense of knowing where I was and being oriented to an organized set of information as opposed to wandering around randomly on the Internet in the dark.” The graduate students, on the other hand, agreed unanimously that the features of VizBlog were intuitive.

Easy to identify blogs with similar content: All students (100%) and 92% of the citizen/town participants strongly agreed/agreed that it was easy to identify blogs with similar contents. One of the graduate students said “This tool gave me a view of the entire blog space, and highly important components by forming clusters”. A participant from the citizen/town group said “The graphic representation of the relationship between articles could be useful for someone trying to locate threads and linking to articles of their interest or posting their on entries at these points.”

Easy to find specific blog entries using search: Although some participants had problems with the search, all participants were able to find specific blog entries using the search feature. Nevertheless, only 80% of the student group and 77% of the citizen/town group agreed that it was easy to do.

5. DISCUSSION OF RESULTS

Overall the participants found the VizBlog interface intuitive and easy to use, as discussed in the previous section. We asked in the post-questionnaire three general questions: What did you like most and least about VizBlog, and what changes would you
suggest to make it easier to use. This section summarizes these results organized around the visualization features of VizBlog.

It is worth noting that the large amount of information managed by VizBlog can be overwhelming. Organizing 1000 nodes in a way that users can make sense of its structure is a challenging task. We feel that our design has done a good job at handling the amount of information. Nevertheless, there were a couple of people who mentioned that the large amount of information was overwhelming and they were stressed out when performing the tasks. Some users were disappointed that they could not maximize the visualization to fit the screen. For VizBlog to be truly effective, we need to do a better job addressing the handling of the large volume of data in future iterations of the tool.

5.1 Overview
The participants liked how the visualization provided a central place to find blogs of similar content. They also mentioned how the graphical representation of blog entries was useful in locating threads of discussion and helped them link to articles of interest or post their comments on entries of interest. Most of the participants pointed out that this tool was easy to navigate and that it saved time and reduced the frustration involved in looking for specific topics of interest. Majority of the participants from all the groups liked how VizBlog provided a global view of all the blogs as it gave them the ability to see discussions at a glance. They also mentioned that the similarity links and hyperlinks between entries made it easier to identify relationships between entries. Also the color-coding made easier to distinguish blog entries that were connected to other blog entries, search results and the top keywords.

Overall the users thought that the various clusters of related topics of discussion offered valuable insight in identifying the most discussed topics and other topics of interest. The similarity links and hyperlinks between blog entries made it easier to identify similar topics. Zooming out to the overview level in the visualization, helped users distinguish the denser clusters more clearly from the others. Also the similarity slider and the filter to view clusters only helped filter out blog entries that were not related for that particular value of similarity.

Cloud. Most of the participants said the keyword cloud would help users of this tool identify the most talked about topics of discussion. Others added that the cloud makes it easy to search for topics of interest. Some said the cloud helped as a cross-reference, where they could verify if the most talked about topic was in fact the largest cluster. One participant suggested that when a keyword is clicked, the visualization should zoom in to the heaviest cluster of that contains that keyword. Most participants liked it because it displayed the top 20 keywords from the visualization and in one glance they were able to identify the most discussed topics.

The very term “cloud” was confusing to some users, particularly those that were not “computer savvy”. They were actually looking for something in the shape of a cloud. These participants suggested that there should be a popup in the keyword cloud that points out to them that these are the top 20 keywords in the visualization.

5.1.2 Zoom and Filter
Panning the interface, zooming in and out and also moving the nodes around seemed intuitive to most users. Filtering, however, raised a few questions. They suggested that a filter to turn off clusters that are not of interest would be useful in reducing screen clutter. Also checkboxes could be used to filter out blog entries based on author and/or subject. One participant suggested having a filter that filters blog entries based on the date of publication of the entry so they can easily focus on recent entries.

The similarity slider was possibly the most problematic feature in the tool. Users found the similarity slider counter-intuitive. It seemed to them like a similarity value of zero would display no links as opposed to all links being present. They were unsure of how this slider worked. The value associated with the similarity projected an inverse relationship to the users. Users expected that a value of 0 should show no relationships and a value of 1 would show all relationships. The reality was that a value of 0 set the threshold to show all relationships greater than or equal to 0, thus showing all similar links. We plan to remove the value from the interface and change the wording and possible the directionality of the slider to address this problem.

Search. Several of the users did not like the way the search results were highlighted. They suggested that the search results should be highlighted in a different color and also be enlarged in size. They were not able to distinguish between the “magenta” color used to highlight the search results and the purple color used for the nodes in general. Some users pointed out that they would like only the search results to be visible for a particular search, and all the other unrelated nodes should be made invisible. Others also suggested having a separate list containing the search results. Some pointed out that the search box should have a white background as opposed to grey. Some suggested that a search history would be helpful to frequent users of this tool. One of the suggestions from a participant was having a “radar view” to aid in easier navigation of the search results.

5.1.3 Details on demand
Right clicking on a node to open a blog entry in the web browser was a little confusing at first to some users. Users sometimes inadvertently pressed the right mouse button down, opening a node in a browser when they did not intend to do so.

Some participants did not like how the nodes displayed only 10 characters of the entire title and they had to zoom in to see the rest. Though the tooltip displays the entire title when you mouse over the node, but the tooltip does not last more than 3 seconds and users found that annoying. The participants suggested displaying the entire title of the blog entries and to provide some mechanism where in the tooltip would stay visible longer. Some even suggested having the name of the author and timestamp appear in the tooltip in addition to the blog title.

Two of the participants asked whether this tool provided support for color-blind as well as blind people. Since one of our participants was color blind, it was very hard for the user to differentiate between the colors when searching for a topic. They suggested using different bold shapes and enlarging the search results to differentiate the results. This is clearly a problem that we must address in future versions of our tool.

6. CONCLUSIONS
Since weblogs have become a means for online deliberation for citizens and town leaders, a visualization tool like VizBlog would make exploration of this content easier. The usability evaluation of this tool proved that VizBlog did make it easier for users to...
identify topics of interest from the visualization. This visualization could also help interested citizens discover discussion on topics of interest and also help them communicate with other citizens and join in discussion online. VizBlog could help government officials and staff to see what a broader and more diverse public is expressing about local issues and concerns. Without this visualization, there are just too many individual blog entries to be able to make much sense of the overarching topics of discussion taking place among them.

The results of the usability evaluation of VizBlog suggested that the tool was easy to use and did provide insight into ongoing discussions in blogs. This usability evaluation was a spot test to verify whether users were able to identify topics of interest from the visualization. Nevertheless, a number of usability problems were found in this study. For the tool to be used by the population at large, we need to make it easier to use for computer novices. We found that most of the people who had difficulty using the tool were the computer novices, those that use computers for personal reasons and not necessarily for work.

In a future version of VizBlog, we plan to enable citizens to select which groups of blogs to visualize. That way, users can select the blogs of their choice and our tool will gather the entries, and provide the interactive visualization to allow users to explore the online deliberations. We also plan to release VizBlog as an open source project, once it has reach a stable point in its usability and development cycle.

7. ACKNOWLEDGMENTS
We are grateful to the National Science Foundation Digital Government Program (IIS-0429274) for supporting the research described in this paper. We would also like to acknowledge the generous advice of our outside advisors, John M. Carroll, Mary Beth Rosson, and Joseph Schmitz. We are indebted to Spencer Lee, Hyung Nam Kim, Uma Murthy, Alain Fabian, Jaideep Godara and B. Joon Kim for their assistance with this research.

8. REFERENCES
[16] Nardi, B., Schiano, D., & Gumbrecht, M. Blogging as social activity, or, would you let 900 million people read your diary? CSCW’04, Nov. 6-10, 2004, Chicago-IL.