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. We developed a tool that displays citizen-to-citizen discussion in blogs and reveals some 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 different parts of the graph. The visualization of online discussion can promote participation by highlighting ‘the buzz’ of popular topics and laying out the structure of conversations. In this demo, we will show our tool in use with a digest of regional Southwest Virginia blogs.

Categories and Subject Descriptors
H.5.3 [Group and Organization Interfaces]: Web-based interaction; H.5.4 Hypertext/Hypermedia

General Terms
Design, Human Factors.

Keywords
Visualization, Blogs, Online participation, Digital Government, Human Computer Interaction.

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 weblogs for online deliberation.

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 [1].

Weblogs, however, are quite different from centralized discussion forums in terms of knowledge exchange patterns and the participants they attract and the decentralized structure they employ. 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. In addition, the decentralized nature of blogs make them easy to use because users can offer informal observations on various issues without the constraints of rules and formality associated with online forums. Registering and logging in to centralized forums can dissuade all but the most motivated and determined users. Rules, such as limiting users to two posts per day, may further inhibit the free flow of opinions and other political talk.

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, “my life and experiences” although political opinions,
observations and information are scattered throughout many of the “my life” kinds of blogs). Ironically, these very characteristics that make blogs attractive to a broad and diverse set of voices from politically less active citizens, are also fundamental to the problems associated with using blogs for citizen-to-citizen deliberation (what we call deliberation in the wild). To address the problem caused by an overwhelming number of blogs full of a wide array of topics, we have designed a tool to help find and participate in citizen-to-citizen discussion that takes place in blogs.

2. SCENARIO OF USE
To guide our development, we built several scenarios of use as requirements of how the tool would be used. Here is one such scenario that highlights the parts of the tool described below.

Mark is a member of the Town Council and is involved in town administration including making general town policies. Interacting with local citizens and knowing their concerns about the town helps Mark in his job. In order to increase this interaction, he uses blogs, both as a reader and as a writer. To get relevant information and insights from as many blogs as possible Mark decides to use VizBlog, a blog visualization application. Using VizBlog, he is able to see that the most discussed keywords in the last month among local blogs are “Big Box”. He knows that the topic of discussion is the development plans of a Big Box store behind Margaret Beeks Elementary School. He uses the search facility to filter the visualization to only those entries containing the keywords “Big Box”. He then checks the “Hide unconnected nodes” checkbox and is left with only clusters in the visualization. From the visualization, he is able to see that at least 5 entries directly link to one of his blog entries (where he presented a summary discussion on this topic from the last town meeting). He reads some of the entries by clicking on the appropriate nodes and finds some people against this proposal. Some of the entries link to a common news item of a similar development in a different town. Clicking on the node in the visualization takes him to that news item, and he learns about what development plans worked in that town and what didn’t. Also, he is able to see comments on other entries that link to his blog entry. All this information helps him stay focused and helps him prepare for the next town meeting on this topic.
3. VISUALIZING BLOGS WITH VIZBLOG

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) [2], 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 Visual Representation

VizBlog has been developed using the Prefuse visualization toolkit [3]. 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.

3.1.1 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.

3.1.2 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 2 shows the three types of links and how they are depicted graphically in VizBlog.

![Figure 2: 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 [4]. 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 thicker the line connecting them (see link A and B in Figure 2).

Hyperlinks between blog entries are represented in the visualization via a dashed arrow that is pink in color (see C in Figure 2). The arrow points from the citing node to the cited node.

If two nodes are related by a citation and a similarity coefficient greater than some pre-specified 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 2). The thickness of the arrow is in direct proportion to the value of the similarity coefficient.

3.2 Information Seeking Mantra

3.2.1 Overview

The purpose of the overview in the Visual Information-Seeking Mantra [2] 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 3).

Figure 3: Semantic zooming – varying title length as the zoom value increases from views A to C.

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 4).

Figure 4: A cluster of entries about the retirement of the fire chief of the town of Bristol, TN.

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 5).

Figure 5: Top keywords cloud

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-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 6.

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 7 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.

3.2.3 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 8). Upon left-clicking the node, the relevant blog entry opens up in the default system web browser.

**Figure 7: Identifying clusters is significantly easier after hiding unconnected nodes (B).**

**Figure 8: Mouse-over of a node reveals its title and highlights its neighbours**

**Figure 9: 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 9 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. TECHNICAL REQUIREMENTS
VizBlog has been developed using Java and can be run on any machine with JRE 1.4 or newer installed. It would also require a standard Internet connection (Dial-up, DSL, ISDN, CABLE, or LAN 100MB+) and a web browser such as Firefox, Internet Explorer or Safari in order to read the blog entries.

5. ACKNOWLEDGMENTS
We are grateful to the National Science Foundation Digital Government Program (IIS-0429274) for supporting the research described in this paper.

6. REFERENCES

