VizBlog: A Discovery Tool for the Blogosphere

Manuel A. Pérez-Quiñones
Andrea Kavanaugh

Uma Murthy
Philip Isenhour

Jaideep Godara
Spencer Lee
Alain Fabian

Center for Human Computer Interaction
Virginia Tech
Blacksburg VA, USA 24061-0106
(540) 231-6000
{perez,kavan,umurthy,isenhour,jgodara,zamfir,alfabian}@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. We are developing a tool to display citizen-to-citizen discussion in blogs and to reveal 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
Blogs, Online Participation, Digital Government, Visualization, 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.

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 [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 of 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 “South Main Street
Business Revival”. He knows that topic of discussion is the revival of South Main Street with new shopping and entertainment businesses. He uses the search facility to filter the visualization to only those entries containing the keywords “south main street”. From the visualization, he is able to see that at least 3 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 in a different town. Clicking on the website node takes him to that news item, and he learns about what development items 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. VizBlog

We created VizBlog using the Prefuse visualization tool kit [2]. Each blog entry is represented by a node, labeled with the blog title. The links between nodes represent either direct links or similarity in content. For direct links, VizBlog shows a solid line of single pixel width. This edge connects two blogs where one directly links to the other blog. This occurs when a blogger writes a comment about another blogger’s posting.

The other type of association between blog entries is one of similarity. Two blogs might be similar if their content contains several words in common. In particular, we calculate a similarity index between two blog entries by using a vector space model [3] comparing the most frequent keywords of the blogs. This computation is done for all pairs of blog entries and the similarity is shown using a variable width edge. The thicker the connecting edge, the more similar the two blog entries are. Figure 1 below shows an example visualization showing similarity links between nodes.

![Figure 1. Visualization of Blog Entries – line thickness indicates how similar two entries are](image)

Having calculated the frequency of words in all the blog entries being displayed, the tool provides a “cloud” view (see Figure 2) of all the top keywords. The font size of the word is proportional to how popular that term is in the current set of blog entries being displayed. In Figure 2, for example, the word ‘mountain’ is very popular, thus larger than most of the other words. Clicking on this word produces a search with the word “mountain.”

![Figure 2. Cloud of keywords](image)

VizBlog provides a basic search capability that allows users to search for particular words among the top keywords. The result of the search is highlighting those nodes in the graph that contain the word. Clicking on a node allows the user to open that blog entry on his/her web browser. Figure 3 shows VizBlog after doing a search for Appalachian.

![Figure 3. Search in VizBlog with search bar and results enlarged for illustration purposes](image)

4. ACKNOWLEDGMENTS

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

5. REFERENCES

