

SGER: Computer-Assisted Interpretation of Citizen Input in Rebuilding Lower Manhattan

Daniel O. Kutz, Javed Mostafa
School of Library and Information
Science
Indiana University
Bloomington, Indiana
{dokutz, jm}@indiana.edu

Monique Girard, David Stark
Center on Organizational Innovation
Columbia University
New York, New York

{mds54, dcs36}@columbia.edu

ABSTRACT

Computer-mediated communication, specifically facilitated by the Internet, offers new channels for citizens to express their views to elected officials, government agencies, and civic associations. Individuals have embraced email, mailing lists, and web-based forms as a method of registering their opinions, frequently resulting in tens of thousands of comments addressed to a given entity. The resulting deluge poses a technical and political challenge: how can official recipients, who have invited citizen input, make sense of commentary received on such a massive scale? How can meaningful patterns be efficiently and effectively identified? We propose the use of computer-assisted interpretation and dynamic visualization complemented by human analysis as one means of addressing these challenges.

Keywords

Digital libraries, information-retrieval, information visualization

1. INTRODUCTION

The Center for Organizational Innovation at Columbia University has put together an extensive archive of materials generated from the deliberation process that took place after 9/11. One of the major ideals promoted in the mobilization effort was that the participatory involvement by the public at large should be an integral part in the rebuilding process. This resulted in the development of several events by organizations that used the Internet as a medium for querying the public and communicating deliberative results.

A breadth of data from this mobilization has been made available to us for analysis. This includes, statements from the 5,000 plus participants of the July, 2002 Listening to the City meeting [1], arguably the largest town hall meeting ever held; some 19,000 statements gathered throughout New York City during the Imagine New York Envisioning Workshops [2]; and numerous on-line chat groups, press coverage, reports, meeting minutes, hearings, as well as petitions, letters, and comment cards.

2. DATA

In particular we examine the Imagine NY project, which was run by the Municipal Art Society, which worked with existing groups in an attempt to garner feedback and information from the public with regard to rebuilding the World Trade Center and the World Trade Center Memorial.

A typical Imagine NY event would take 1-4 hours, would work with preexisting groups, and would involve around ten participants. Larger groups could participate, but would be required to coordinate the sessions via smaller subgroups of participants. 130 venues held a total of 230 Imagine NY workshops. During the workshop ideas and thoughts generated by the participants were summarized and transcribed. Once the workshops had finished (a three-month process) Imagine NY took these generated ideas and identified and categorized various themes and visions that emerged from the discussion. These statements and resultant categories were then made available online for viewing on the Imagine NY website [2].

3. BROAD GOALS

The data we have has a strong historical value, and provides insight into the thoughts, ideas, and concerns of New Yorkers, and the public at large. However, instead of generating a static digital archive for use solely by scholars, ideally this material should be accessible to everyone, in a dynamic nature, so that users are not only limited to browsing this corpus of data, but can also explore, manipulate, and contribute themselves.

To help achieve this our initial goal is to use techniques for analyzing known concepts (for example, concepts generated as priori hypotheses from social theories) by applying constraints on the supervised mining process and displaying and visualizing the results in a way that will allow the user to get a deeper understanding of the data, and facilitate interpretation by scholars.

Specifically the system we are developing should:

- 1) Given corpora from different sources, apply supervised, unsupervised, and hybrid means to automatically generate concepts;
- 2) Rank concepts according to their prevalence and importance (weighted using different techniques);
- 3) Allow a means to show context of the concepts, i.e., link back to original corpus/corpora and show key information related to the origin of the concepts. For example, genre (chat, email, website, etc.), speaker/writer, event, date/time, or sentences in proximity;

4) Provide a general means to discover association among concepts; here interactive visualizations show association among discovered and seed concepts;

5) The concept and association discovery should be time-mapped, i.e., if corpus has date of origin/publication then we should be able to trace the evolution of concepts over time; this may permit conducting analysis at a finer grain level; we could for example track convergences around key concepts (growing acceptance and popularity) or divergences away from concepts (disfavor and loss in popularity) as dynamic visualizations or "animations" of concept spaces;

7) Engage sociologists and citizens to validate the outcome of the analytic tools.

4. CURRENT RESULTS

To aid our understanding of this data, we use the framing of a moral world, a theory that asserts that conflict and mitigation are primarily driven by the participants having differing worldviews [2]. We expand on this theory to develop eight moral worlds (civic, inspired, domestic, market, industrial, green, opinion, valor) for identifying issues and concerns within the statements collected by Imagine NY. Using this framework, an online system was developed that allows a human analyst to select a random sampling of statements and apply moral world or multiple moral world codes to the examined statement. The subset coding of statements acts as a seed for which our system can be used to dynamically link all statements. This enables the viewer to better understand the data, provides summative statistics of the worlds identified (figure 1), and displays change in moral worlds over time and location.

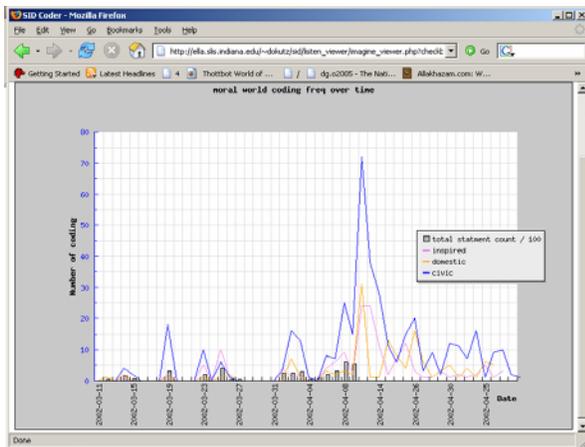


Figure 1. coded results by workshop date.

Our system facilitates the dynamic viewing of generated thoughts and ideas from the Imagine NY events (figure 2). Generated statements from the workshops have been coded for specific moral worlds. A viewer is able to examine the subset of coded

statements, and from there examine similar statements that match the moral world, or other keywords of interest.

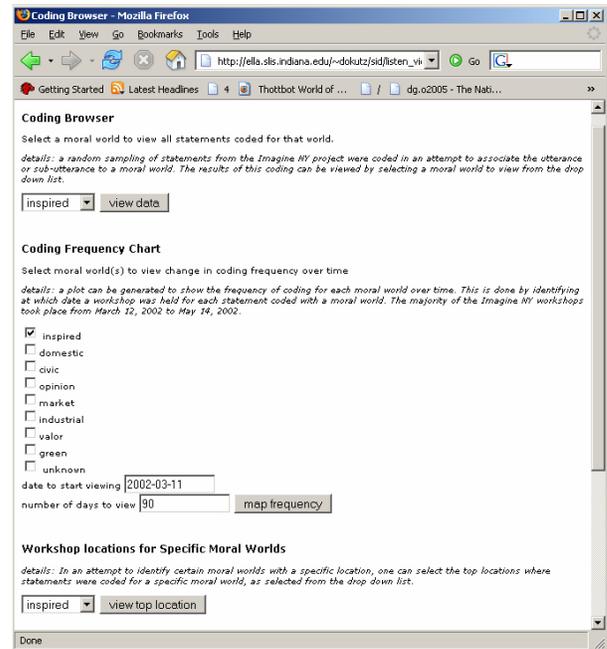


Figure 2. data viewer.

5. CONCLUSION

The advantage of our methodology over the manual coding of all statements, as conducted by Imagine NY, is that it can be implemented by fewer people, and coding from this subset can be extrapolated to the larger dataset. Furthermore, the conceptual or high-level codes represented by the moral world codes are of value for decision makers attempting to develop a quick understanding of civic concerns. This methodology is also valuable from the perspective of digital library or electronic corpus management, as it adds value to the captured data collection by allowing for user-annotation capabilities and by offering the ability to interactively explore data.

6. REFERENCES

- [1] Listening to the City Website
<http://www.listeningtothecity.org/>
- [2] Imagine NY Website
<http://www.imagineyny.com/>
- [3] Boltanski, L. and Thévenot, L. The Sociology of Critical Capacity. *European Journal of Social Theory* (1999), 2(3):359-377.