
Do you know what you did last summer? Visualizing personal behavior in Google services

Zhicheng Liu

Georgia Institute of Technology
School of Interactive Computing
Atlanta, GA 30332 USA
zliu6@gatech.edu

Sarita Yardi

Georgia Institute of Technology
School of Interactive Computing
Atlanta, GA 30332 USA
syardi3@gatech.edu

John Stasko

Georgia Institute of Technology
School of Interactive Computing
Atlanta, GA 30332 USA
stasko@cc.gatech.edu

Abstract

People use web-based services extensively, but have few tools at their disposal to view and monitor their past uses. We present a visualization tool that pulls user data from Google services such as Blogger, Calendar, and Documents, and shows users who they have interacted with and what content they have shared. The goal of this research is to understand what people are able to report about their own past behavior and to design a tool to support self-monitoring and reflection. We also consider questions of privacy, data synthesis, and social visualization.

Keywords

Guides, instructions, author's kit, conference publications

ACM Classification Keywords

H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

General Terms

Social visualization, Google services, monitoring personal behavior

Copyright is held by the author/owner(s).

CHI 2010, April 10–15, 2010, Atlanta, Georgia, USA.

ACM 978-1-60558-930-5/10/04.

Introduction

It is human nature that people are inherently interested in understanding their own behavior. Throughout the history of man, technologies such as cave wall paintings, speech, printed books and documents, movies, television, and photographs have played an important role in producing artifacts that archive human behavior. These technologies facilitate personal reflection and storytelling, and shape the identity of individuals and cultures.

As people spend more and more time online, their digital footprints leave traces of their lives all over the Web. In particular, social networks and content networks (e.g. Google Docs) contain a wealth of data about past behavior. In this paper we present our work in progress on visualizing personal information extracted using a subset of Google Data APIs [1]**Error! Reference source not found.** Our discussion focuses on two major design considerations: privacy and data synthesis. We also briefly describe our user study design for measuring users' ability to recall and report their own behavior.

Visualizing Personal Behavior

Interactive visualization has been applied as an effective technique to present snapshots of users' digital life. Visualizations of personal email archives, for example, reveal social networks and temporal rhythms of interactions around an individual. They also suggest that users often *want* to be shown their own behavior. Studies of email interactions found that users were surprised to see emerging patterns and trends in their use, and were also frequently eager to share the visualizations to people involved [3][4].

The web-based nature of many services and user interaction poses a number of limitations and constraints in providing pervasive and universal support of personal data visualization. First, the control of such data is often in the hands of service providers rather than end users. Minimizing the work required from users while maintaining a minimal risk of privacy intrusion is an ongoing challenge. Second, many web applications are closed and do not easily interface with other applications. Such barriers fragment users' data across the web. Visualizations based on these disparate data sources are rarely able to present a coherent and integrated experience to the users.

Data protocols and APIs released by major service providers such as Google and Facebook present an opportunity to investigate ways of resolving these tensions. We first focus on two design considerations: privacy and data synthesis, and then briefly describe our visualization design and user study plans.

Privacy Considerations

For a casual user, it may be desirable that the visualization of her personal data can be created with minimal effort. This, however, implies that she is not in control of the process of data retrieval, analysis and visualizing. Since personal data is often safeguarded and only accessible after logging in, the user has to grant authorization for a third party to access her account and data. A number of privacy concerns arise.

The burden of managing the risk of privacy intrusions often falls on service providers. Secure authorization standards such as OAuth are designed to address this issue. Our tool uses Google's implementation of the OAuth standard. Users are redirected from our

application to Google's login page, where they are invited to login to their Google username and password. If they grant access (Figure 1), Google issues a security token to our application that we use to retrieve the user's personal data. The user can revoke this token at any time.

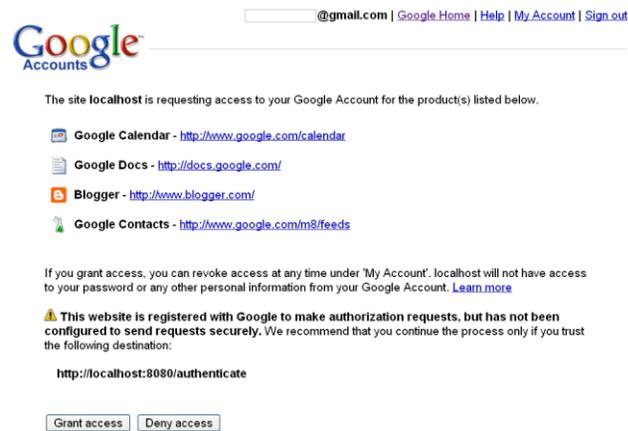


Figure 1. User Authorization to access Personal Data

Data Synthesis

Our tool currently retrieves data from four Google services (Contacts, Documents, Blogger and Calendar) because these services are generally widely used (see **Error! Reference source not found.**). The Contacts API provides information about all the people who have interacted with a given user through any Google service, such as Gmail, Google Talk and Google Documents. Information about contacts includes names, email addresses, gender, group, and time when they last updated their profile. The Documents API provides information about the documents authored or

co-authored by a given user. Information about the documents includes title, content, authors, creation time and last updated time. The Blogger API provides information on blogs and blog entries created or collaborated by a given user. Information about the blog entries includes title, content, authors, commenters, and creation time. Finally, the Calendar API provides information on calendar events created by or shared to a given user. Information about the events includes name, starting time and ending time.

These four Google service APIs provide a rich set of data with different attributes and characteristics (e.g. temporal, text, social network). However, merging and synthesizing across datasets and presenting a visual representation of the result constitute a major challenge. For example, although a given user can be linked between the Documents API and the Contacts API by their email address, personal entities in Blogger data are often anonymous and identifiable only as a nickname or pseudonym. Thus, while we hope to collect and display user activity across platforms, our tool will aggregate behavior across platforms where possible. Blogger users in our user study will login separately to the Blogger visualization.

Visualization Design

Our service is targeted at casual users [2] so the visual design should not be overly complex, but it needs to maintain expressive power over a complex and multi-faceted data collection. We thus believe that a collection of simple interactive visualizations is appropriate. As a start, we built a node-link graph with radial layout to visualize the clusters of social networks across different types of document collaborations (Figure 2). Our work here is ongoing and we are

creating different visualizations and each may be suited for portraying a particular aspect of the data.

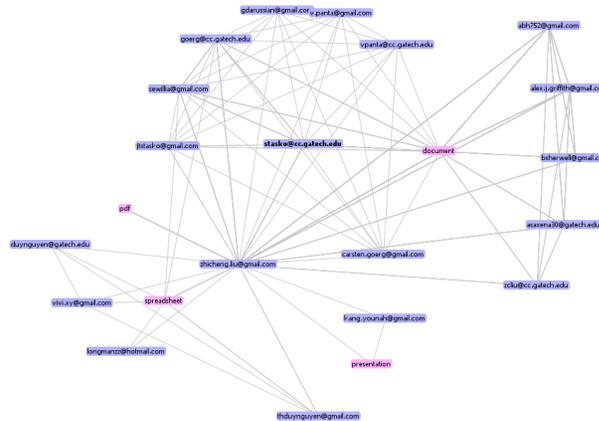


figure 2. Visualization of the social network of contacts (blue) clustered by different types of Google documents (pink).

User Study Challenges

With the visualizations, we plan to conduct a user study that investigates people's ability to recall and report their own activity on Google services. Participants will first complete a pre-survey drawn from an ego-network analysis approach. Ego network analysis is a branch of social network analysis that articulates an "ego" (the person of interest) and her "ties" (relationships with other people). This kind of analysis is useful for measuring an individual's experiences and relationships to others around her. (This is different than social network analysis, which focuses on patterns of interactions and structures among a network of people).

We will evaluate people's ability to report their own behavior by comparing their pre-survey self-reports to measured data revealed in our tool. We anticipate a range of possible predictors for what people are likely to report, such as how recently they used the service, how frequently they use it, uses that were important to them, or uses with people who are meaningful to them.

A number of broader questions arise in this work. People don't always remember people they've seen on the street; we do not have a good baseline for what people *should* be able to remember about their online behavior. How much should people know about what they have done on a site? What information is most important or meaningful? While many fields, like health, are touting the importance of personal information management, it is not clear what behavior people should monitor and how. Through the user study, we hope to gain a better understanding of these issues, which can inform the design of future tools to support visualization and management of personal behavior.

References

- [1] Google Services API directory. <http://code.google.com/apis/gdata/docs/directory.html>
- [2] Pousman, Z., Stasko, J. T. and Mateas, M. Casual Information Visualization: Depictions of Data in Everyday Life. *IEEE Trans. Vis. Comput. Graph.* 13, 6 (2007), 1145-1152.
- [3] Viégas, F.B., Boyd, D., Nguyen, D.H., Potter, J. and Donath, J. Digital artifacts for remembering and storytelling: PostHistory and Social Network Fragments. In *Proc. 37th HICSS '04*, 1-8.
- [4] Viégas, F.B., Golder S. and Donath J. Visualizing Email Content: Portraying Relationships from Conversational Histories. In *Proc. CHI '06*, (2006).