For our project we have implemented a visualization that presents the user his or her email as a personal archive. The metaphor we attempt to capture is that of the scrapbook. Just as one can page through a scrapbook and view pictures and snippets of text, one can page through their email to do roughly the same: find out what he or she was doing at a point in time, with whom they were communicating, and possibly evoke an emotional response. Our implementation is intended to be built on top of an existing email client to give the user a persistent view of the visualization and also allow the user to search their email based on time. In this paper we will discuss our visualization, the results of our implementation, and future directions.

Motivation

The way that people view their email today is roughly the same way people read message boards. Email, however, is far more personal than the typical message board. Individual messages are often sent from one individual to another. Whether the message contains sensitive, personal, or just mundane information, they are not usually meant to be viewed by the public. Additionally, email is both easy to send and easy to store. This makes email a personal space with a large amount of rich data. What our visualization attempts to do is present users with access to some of the rich, personal data that is already being stored.

Audience

This visualization is intended for use on personal inboxes. While it would still work on work or group accounts, the chances of returning results that have emotional significance to the user is much less when most of the emails are work related. Ideally, the user will have used his or her inbox for a large volume of personal transactions over a lengthy period of time. This condition is best facilitated through persistent, web-based email systems. For the purpose of our implementation, our focus is building a visualization for Gmail.

Related Work

The visualization of email as a personal archive has been explored by Donath and others. Of particular interest is her comment that "some of the most important information about a piece of email exists nowhere in the message itself – it is the meaning that the communication has for the recipient (or
sender). Seemingly innocuous notes may be souvenirs of momentous occasions."¹ What Donath suggested is echoed in our motivation; the presentation of data facilitates the user's recall of events, something not readily encoded in an email message.

Implementations of email discussed by Donath include Viegas' PostHistory and Boyd's Social Network Fragments. These visualizations made salient the analyses about email in a social context, particularly revealing tie strength and patterns in networks over time. Our work is differentiated through our approach, in that it is perforce egocentric but can still reveal the same patterns. Whereas existing approaches are aggregations with a few snapshots, ours is the reverse - snapshots of email augmented by aggregations. Both approaches, however, have the end result of evoking memories of past events.

Experience

The visualization is designed to sit at the top of Gmail and can be used passively as well as actively. As such, it is constrained by the limits of hypertext; the interface contains clickable elements that allow the user to further explore his or her email. The time span will be determined by the receive date of the first email in the search page. By only presenting a slice of information from the same time as the current page of emails we hope to facilitate a user experience much like that of looking through a scrapbook.

---

¹ Donath, J. Visualizing Email Archives (draft).
Figure 1: Visualization sitting at the top of the Gmail inbox view. This was our intended vision for the implementation.

As Figure 1 shows, from each time segment 3 images will be randomly selected from attachments and displayed along with the addresses of individuals from whom the user received the most email in that time period. What is shown above is the activity taken once a particular email address is selected. On selection, the most interesting words and phrases will be selected from that person's email using tf-idf. These words and phrases will then appear around their name. The intent is to give not only visual cues in the form of attached images, but also share the ideas and thoughts that were common.
This visualization can be viewed passively by using Gmail normally and allowing the names and images to come and go without interacting with them. Alternately, one could delve into the data already stored in their inbox using the visualization to guide them through their email. The date, images, names, and phrases will all be linked to the query or email that refers to them. For example, if a user clicks on one of the pictures, they will be taken to the email that the picture is attached to. Another use would be to select a name from the visualization in order to create a search for all email from that person.

We intend to analyze both the inbox and sent box separately to see if more memory recall is obtained from mail sent to a user or mail sent from the user.

Results

Google proved to be a very large barrier for us. There was a lot of initial promise in using Google Gears to get email information along with Gmail and from the same information pool. However, information about how Gmail maintains their offline information is sparse and not very useful. Information on Google Gears is more abundant, but we could not get it to access the database that Gmail sets up when you turn on offline access.

We were able to get more useful access to an email archive by downloading the archive to a local computer using IMAP, and converting the archive to XML. This however removed any association with an email client.

Without having an email client to sit on top of, our visualization became the main vehicle for searching through an email archive. In order to simulate an entire email client, we have simplified the email experience to four types of searches (browse, address, email, and search).

In browse view, a user is looking at all the email in their inbox, from newest to oldest. The visualization displays two random images from all the attachments in the inbox. The top nine email addresses by volume of email sent are shown next to the images. Interesting words and word pairs, as returned from the tf-idf algorithm we used, are gathered for all the mail in the inbox and displayed.

Address view only shows information and interesting words for a single sender. Additionally, the email subjects displayed are from the same address. However, we found that more often than not a specific address would not have any images to display. It was decided that the visualization did not have the emotional attachment we were hoping to add to searching an email archive when no pictures were shown, so two random images, much like browse view, from the entire inbox are displayed in this view.
The email view is much like the address view. It shows information about the sender of the email, and related interesting words from all the emails that this sender has sent to the inbox. Also, two random images from the entire inbox are displayed at the top for the same reason as the address view.

Finally, the search view displays information about a specific search. Interesting words found in emails containing the search, as well as the emails that contain the search are shown. If available, images from those emails are displayed.

*Future Directions*

Because email is such an intensely personal set of information, we would like to make the entire process streamlined enough to fit into either a firefox extension or a Greasemonkey script. There are a number of hurdles to get over first. For instance, some way to run tf-idf client side would be necessary. We would also need a much clearer idea of how Gmail queries for email and attachments.

If we had more time we would add an sense of time to the visualization. So that as a user moved through the archive, only images, words, and senders relevant to a specific timeline surrounding a particular email, search term, or address would be shown rather than from the entire inbox. This would allow users to see how their interactions changed over time.