MUSE: Reviving Memories Using Email Archives

Sudheendra Hangal   *Monica S. Lam*   *Jeffrey Heer*
Computer Science Department
Stanford University
{hangal, lam, jheer}@cs.stanford.edu

ABSTRACT

Email archives silently record our actions and thoughts over the years, forming a passively acquired and detailed life-log that contains rich material for reminiscing on our lives. However, exploratory browsing of archives containing thousands of messages is tedious without effective ways to guide the user towards interesting events and messages. We present MUSE (Memories USing Email), a system that combines data mining techniques and an interactive interface to help users browse a long-term email archive. MUSE analyzes the contents of the archive and generates a set of cues that help to spark users' memories: communication activity with inferred social groups, a summary of recurring named entities, occurrence of sentimental words, and image attachments. These cues serve as salient entry points into a browsing interface that enables faceted navigation and rapid skimming of email messages. In our user studies, we found that users generally enjoyed browsing their archives with MUSE, and extracted a range of benefits, from summarizing work progress to renewing friendships and making serendipitous discoveries.

ACM Classification: H.5.2 Information Interfaces and Presentation.

Keywords: Life-logging, Email, Data mining, User interfaces, Visualization.

INTRODUCTION

Email is one of the Internet's most enduring "killer applications" with over 1.8 billion users and 2.9 billion accounts worldwide, as of 2010 [26]. Given the availability of free email services with virtually unlimited storage, we expect that millions of mainstream users will amass large email repositories, perhaps spanning over half a century, across their lifetimes. Already, it is common to find people with email archives spanning several decades.

In much of the wired world, email is used for many daily activities, for everything from setting up meetings to processing business workflow; from making purchases on the Internet to sending emotional messages of love, joy, and condolence; from sending oneself reminders to sharing an interesting web link with friends, and so on. Unlike blogs, diaries and journals, email archives silently capture our experiences *in situ*, as they arise in our communication, with no additional action needed to record them. Indeed, email has become a de facto medium of record; many people consciously deposit important information into email, knowing they can look it up later, and thereby use their email account as an informal backup device. Therefore, email archives contain or reflect memories that are extremely valuable for the purposes of reminiscence. We surmise that for a significant fraction of computer users, more characters are typed into email than into any other application, and these characters, accumulated over a lifetime, can provide a powerful window into history.

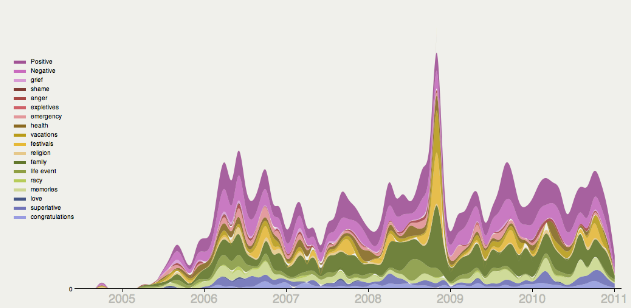


Figure 1: A MUSE visualization of email sentiment. A stacked graph shows the number of email messages reflecting a particular sentiment category over time.

An email archive is one example of a life-logging device that captures and stores different forms of personal data across a long period of time [10]. Ironically, while service providers routinely mine email and other personal data for advertising purposes, there are relatively few tools to help individuals make sense of their own digital assets. Even if all personal data is captured, as Petrelli et al. remark: "*This may be the fate of lifelog data, stored somewhere and ignored, if the owner is not given tools for sorting, clearing and distilling what is of value*" [24].

However, distilling valuable information is a difficult problem in practice, particularly in a domain like email that consists mainly of free-form text along with some metadata and attachments. What kinds of information do users find valuable when detailed, day-by-day records of their communication are available? What techniques enable users to conveniently browse their life-logs and identify valuable information? How do these techniques interact with the user's own knowledge and memory? We believe email is a good platform for studying these questions, as many people already

UIST'11, October 16-19, 2011, Santa Barbara, CA.

have archives spanning relatively long periods of time.

While email archives contain a wealth of valuable information, they are also voluminous; a life-long archive can easily run into tens or hundreds of thousands of messages. Traditional email clients allow users to examine one message at a time, and to filter and to issue search queries. These clients are ill-suited for browsing a large-scale archive, where a user may not know exactly what to look for. Like others before us [35], we see a two part solution to this challenge. First, we can automatically generate cues likely to orient the user towards messages of interest. Second, once the user has acquired a cue, we can encourage exploration of the messages related to that cue and to other related cues. Such exploration mirrors the natural organization of episodic memory of autobiographical events [32].

To explore the efficacy of different types of cues and browsing techniques, we have built a program called MUSE, for *Memories USing Email*[1]. MUSE processes the contents of a user's email archive and generates several types of cues. It also provides a browsing environment that enables the user to follow these cues by rapidly skimming related messages and forking off other trails of exploration. As an illustration, Figure 1 shows how Muse presents a timeline summarizing the sentiments expressed in the archive.

Contributions

In this paper, we make the following research contributions:

1. Based on iterative design and user feedback, we identify four types of cues useful for reminiscence: communication activity with inferred social groups, a summary of recurring named entities over time, occurrence of sentimental words, and image attachments.

2. We propose specific and relatively lightweight mining techniques to identify, organize and present these cues, as well as techniques to enable rapid and exploratory browsing of associated messages and other related cues.

3. We discuss some of the memories evoked by MUSE in users reminiscing with their own email archives, and present insights about the efficacy of different cues. Our studies uncover a range of benefits for exploratory browsing with MUSE, from summarizing work progress to renewing old friendships and identifying milestones.

To date, MUSE has been used to explore email archives containing up to 50,000 messages.

RELATED WORK

While there has been much prior work in various forms of email analysis, there are relatively few usable systems that let end users explore large scale email archives. MUSE draws on several lines of research, summarized below.

Email Analysis

The closest related work to MUSE is Themail [35], which aims to help users reflect on the dyadic relationship with each one of their contacts. Themail visualizes single words that have a high TF-IDF score in emails exchanged with a contact over time. Our starting point for generating important name cues in MUSE was similar to Themail; however, we found that this approach tended to generate noisy words lacking any context and was inefficient when browsing thousands of contacts accumulated in a long-term archive. MUSE solves these problems with its use of named entities and automatic inference of social groups, and introduces additional types of cues; it also provides a more sophisticated exploratory browsing environment.

Other related work includes research on "email rhythms" to provide insights about patterns of communication [19, 23, 33, 34]. SocialFlows [17], GroupGenie [25] and ContactMap [36] help to organize a user's social contacts. These tools are focused on messaging patterns and not the actual message contents.

There is much interest in helping users become more effective in email correspondence. Examples include Xobni (xobni.com), Rapportive (rapportive.com), SNARF [21] and Gmail's Priority Inbox and People Widget. However, these systems do not target interactions with email archives.

Life-logging

Many life-logging systems today are geared towards active life-logging, which involves deliberate actions by the user and/or the use of specific hardware, software or services[2]. Pensieve actively solicits input from the user by periodically emailing personal questions and attempts to create a repository of reminiscences [22]. In personal finance, mining spending data [28] is an example of analyzing a passively captured life-log. MUSE shares this focus on passive life-logging, but works in the domain of email communication.

Legal Discovery and Intelligence Analysis

There are some similarities between MUSE and systems used for legal discovery and intelligence analysis. Users of these systems also need to spot cues and peruse large-scale, loosely organized datasets, often including emails. MUSE shares techniques with such tools; for example, extraction of key entities as in Jigsaw [30], and the need for pre-built views of intelligence corpora [3]. However, the use case for discovery tools (trained analysts making sense of unfamiliar corpora in order to spot suspicious activity) is starkly different from that of MUSE (mainstream consumers using their own email for reminiscence). This leads to different design considerations for MUSE.

Text Analysis and Visualization

The TIARA system integrates text analytics and interactive visualization of email [15], and was tested with users performing focused tasks on a short-duration email corpus belonging to another person. The Parallel Tag Clouds visualization [4] highlights the presence as well as absence of significant terms across the parallel text corpora of different circuit courts. MUSE has somewhat similar requirements for identifying key term in email messages across time, though it detects only the presence of terms, not their absence. Themeriver is a system for visualization of text content acquired over time, such as news articles [12]. Work in the topic detec-

[1] MUSE is publicly available at http://mobisocial.stanford.edu/muse.

[2] See http://personalinformatics.org for a list of life-logging systems.

tion and tracking (TDT) area tends to focus on clustering and labeling the textual content of messages (e.g. [31]). Unlike MUSE, all of the above systems do not target the task of reminiscence, and do not take advantage of personal and social context (such as cohesive groups, sentimental messages and images). However, there are possible synergies between the text analysis techniques used by MUSE and these systems.

Interaction Techniques

Many systems use faceted navigation to support exploratory browsing (e.g., Flamenco [38], Phlat [5] and Stuff I've Seen [6]). MUSE employs similar ideas in the domain of email, where the facets are people, months, years and automatically inferred groups and sentiments. Systems like Cyclostar have demonstrated the effectiveness and versatility of elliptical gestures [18]; the MUSE jog dial we describe in later sections is a simplified form of such an interface.

Sentiment Analysis

There is much work on sentiment analysis of tweets and reviews for the purposes of inferring public reaction to a product (e.g., [1]). We Feel Fine is a visualization of sentiments expressed in public blog posts and has been used to generate hypotheses about sentiments at a societal level [13]. MUSE tracks sentimental words in the personal communication of a single user across time and can potentially be used by an individual to examine hypotheses about herself.

Memory and Archiving Practices

There are several studies that take an ethnographic approach to understanding family memories and archiving practices (e.g., [14]). Elsweiler et al's work on understanding human memory in the context of email refinding concludes that "*. . . although people generally remember quite a lot about their emails, there are situations in which people remember less and in these situations it may be more difficult to refind the information required with existing tools*" [7]. Zalinger's dissertation explores the role of Gmail as storyworld, and its intensive participant interviews confirm the richness of narratives that are embedded in email [39]. These studies establish the value of personal and family archives and provide motivation for experimental systems such as MUSE.

USING MUSE

To run MUSE, a user typically downloads it to her own computer and launches it using Java Webstart. This reassures users about confidentiality, which is important given that email archives frequently contain highly sensitive information, from love letters to financial documents. MUSE starts up an embedded web server in the background, and launches a browser window for the user to interact with it. This interface choice means users can use their favorite browser and its familiar features like multiple tabs and windows, navigation buttons, bookmarks and browser plugins.

The user specifies one or more sources of email to MUSE, including online POP/IMAP servers or mbox format files stored on a local file system. The user can select folders to analyze from each source and optionally apply filters by date range, or tell MUSE to only analyze their sent messages. Focusing on sent messages is a useful heuristic for reminiscence because they are handcrafted by the user and reflect the user's thoughts and actions. In contrast, the INBOX tends to be much larger (on average, 2.5 times as many messages involving 4.8 times as many people, according to an anonymized dataset provided to us for research purposes by Xobni). Much incoming email is delivered via mailing lists, which are sometimes scanned casually or even ignored entirely. Since access to online email providers can be slow, MUSE caches messages once fetched, though the cache can always be cleared under user control.

Once MUSE has fetched the user's messages, it processes and indexes their contents and attachments, and builds an address book of contacts. Typical processing speed on a current-generation laptop or desktop computer is about 1,000 messages per minute. MUSE performs entity resolution by unifying names and email addresses in email headers when either the name or email address (as specified in the RFC-822 email header) is equivalent. This is essential since email addresses and even name spellings for a person are likely to change in a long-term archive. Name equivalence is tested by ignoring case differences and equating commonly used variations in naming, e.g., with or without a middle initial, "Firstname Lastname", "Lastname, Firstname" and "Firstname Lastname - Department". Once the address book is available, MUSE infers social groups based on a grouping algorithm, described in the next section. The user can specify the number of groups to be inferred, but 20 is a reasonable default. Users can also refine the inferred groups manually.

Next, MUSE allows the user to browse four different kinds of cues and the messages associated with them. These cues are described in detail in the next two sections. Since users' tend to become highly engaged while exploring their archives containing years or decades of messages, it is not uncommon for them to spend several hours with MUSE. To allow them to split this time across multiple sessions, MUSE lets users save session state and reload it later without having to reprocess the archive.

MEMORY CUES FOR BROWSING EMAIL ARCHIVES

In this section, we discuss a set of memory cues for browsing email archives and techniques for generating them. We discovered these cues by observing users interact with early versions of MUSE on their archives and analyzing why they appeared to be useful. For each type of cue, we present the intuition, some details of its implementation, and the presentation technique used.

Automatically generated cues do not have to be fully precise or complete; rather, they should work hand-in-hand with a user's memory and the actual content of the messages. In practice, we expect that many cues will be ignored by a user because they are obvious, redundant, noisy or overshadowed by other, stronger cues. However, for the system to be engaging it should surface a relatively high fraction of useful cues that lead to valuable memories and avoid flooding the user with misleading or irrelevant cues. We also prefer techniques that are sufficiently lightweight to run on end-users' own machines and do not require server-class hardware.

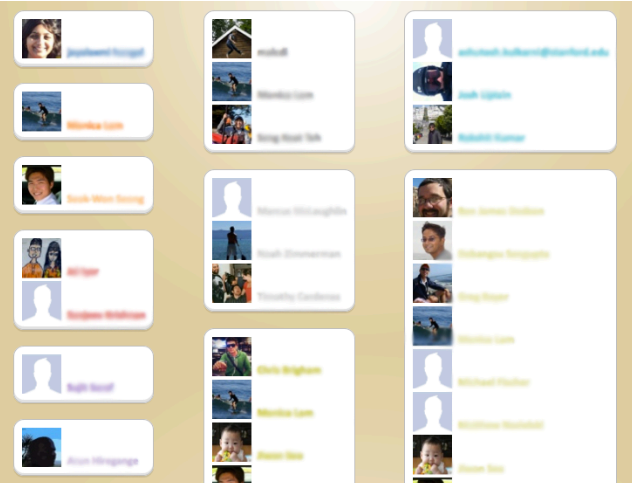


Figure 2: The groups editor showing automatically inferred groups that can be refined by the user. Names are blurred to preserve user privacy.

Group Cues

People routinely interact with thousands of contacts in the course of a few years over email. Since it is difficult to visualize messages, topics, and communication activity with so many individuals, MUSE groups these contacts automatically and lets users explore their communication with the group as a whole. This approach mirrors the way people mentally chunk their contacts into groups like family, colleagues, classmates, neighbors, and so on.

MUSE automatically discovers likely groups by analyzing co-recipiency in messages. It employs a group mining algorithm [25] that satisfies several properties that are important in social contexts. First, a group may consist of people who do not all appear together in any single message (e.g., a user's extended family.) Second, within a group, there may be important subgroups with a significant identity of their own (e.g., siblings as an important subset of extended family). Next, the same person could belong to multiple groups (e.g., a colleague at work may also be part of a hiking group.) And finally, a significant "group", for the purposes of organization, could consist of just one very important person with a high communication volume, such as a spouse or close friend.

Users can optionally refine the inferred groups manually using a drag-and-drop editor. The editor lets users move people between groups and create, clone or delete groups. Users can see the name of each person in a group, but if their browser supports the W3C contacts API [37], they also see the contacts' pictures, if available. For example, Mozilla's Contacts plugins for Firefox can fetch photographs for friends from networks like Facebook, LinkedIn and Gmail. Fig. 2 shows a screenshot of the groups editor with this plugin.

To present the cues associated with groups, MUSE generates a stacked graph visualization with one layer per group as shown in Fig. 3. This visualization lets users spot relative patterns of communication with each group over time, and to correlate them with total communication volume. Most users find that this visualization tells a story of when they started and stopped interacting with various groups, reflecting different phases of their lives. Users can click at any point on the stacked graph to launch into a view containing all the messages exchanged with that group. The view is initialized to the point in time along the X-axis that was clicked. Each group is also assigned a color, which is used to code important terms, as described below.

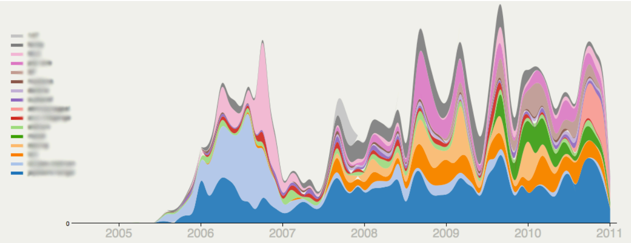


Figure 3: A stacked graph representation of communication with each group over time. Group names are blurred to preserve user privacy.

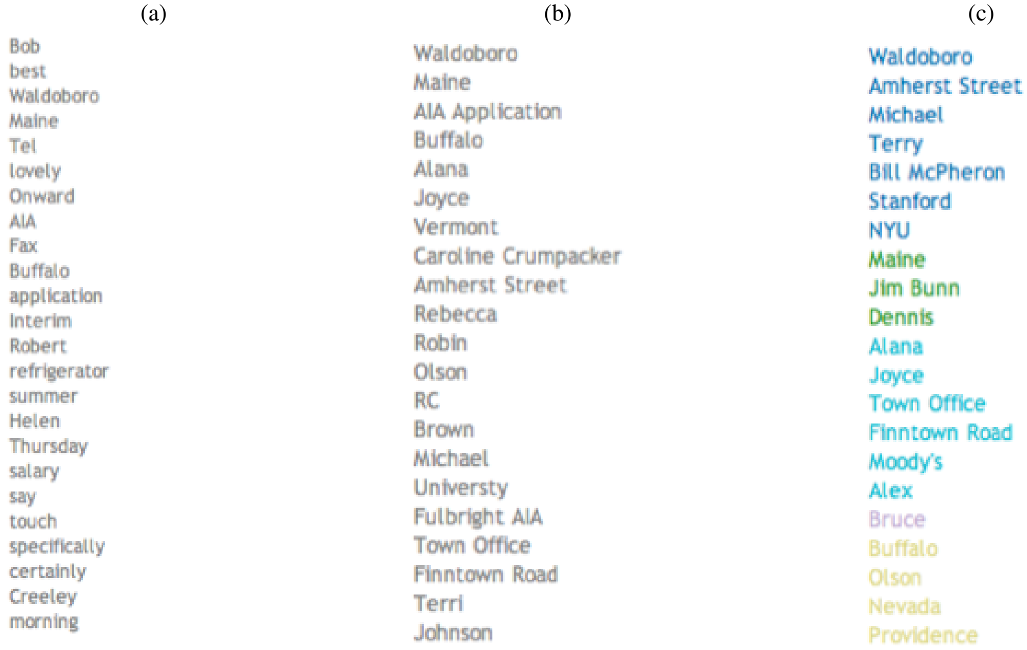
Name Cues

To provide a quick overview of an archive, MUSE creates a summary of important terms on a monthly basis. Terms that make the best cues are often names of various kinds, including people, places, organizations and so on, because names generally tend to carry rich associations in the user's episodic memory [32]. Hence MUSE first extracts named entities from message contents and analyzes these terms. We apply the Named Entity Recognition package from the Stanford NLP toolkit [9], using the default training model. This decision was informed by early experiments using the standard TF-IDF (Term Frequency by Inverse Document Frequency) metric with single words, similar to Themail. The results with this metric were very noisy, despite using appropriate stop-word lists and other heuristics like factoring word commonality into ranking.

Fig. 4 illustrates the difference on an example corpus: a portion of the email archive of noted American poet Robert Creeley, which is hosted at Stanford University Libraries [8]. Using word-based TF-IDF, the top-scoring terms, shown in Fig. 4(a), include generic words like *best*, *lovely*, *say* and *touch* which are unlikely to be useful, and in fact tend to waste the user's time as she tries to understand the context in which they arise[3]. However, we observe from this list that the terms most likely to be interesting are the named entities such as *Waldoboro*, *Helen*, etc. Fig. 4(b) shows the results of first extracting named entities from the message contents, and scoring only these terms. The overall results already appear to be better cues to memory.

We also experimented with 2 other options for identifying key terms: scoring multi-word phrases, and scoring noun phrases extracted with linguistic parsing. We found that users generally react better to named entities since names tend to have deeper associations in the user's memory compared to phrases, which may still lack context.

[3] While the Themail program is not available to us, from the screenshot in the paper, it appears to suffer from the same problem.



(a)

Bob
best
Waldoboro
Maine
Tel
lovely
Onward
AIA
Fax
Buffalo
application
Interim
Robert
refrigerator
summer
Helen
Thursday
salary
say
touch
specifically
certainly
Creeley
morning

(b)

Waldoboro
Maine
AIA Application
Buffalo
Alana
Joyce
Vermont
Caroline Crumpacker
Amherst Street
Rebecca
Robin
Olson
RC
Brown
Michael
Universty
Fulbright AIA
Town Office
Finntown Road
Terri
Johnson

(c)

Waldoboro
Amherst Street
Michael
Terry
Bill McPheron
Stanford
NYU
Maine
Jim Bunn
Dennis
Alana
Joyce
Town Office
Finntown Road
Moody's
Alex
Bruce
Buffalo
Olson
Nevada
Providence

Figure 4: Comparison of top-ranked terms for the same month on a portion of the Robert Creeley archive. (a) single words, (b) named entities, and (c) named entities color coded and clustered by group.

Time-based TF-IDF To score terms, we use a metric based on TF-IDF scoring of the named entities, using the *ntn* variation described by Manning et al [20]. This variation corresponds to natural term frequency, the regular definition of inverse document frequency and no TF normalization. High TF-IDF scores are associated with terms that are specific to a particular document (in our case, the messages for a month) compared to the rest of the corpus.

To further improve term scoring, we introduce a simple, novel variant for *time-based* TF-IDF. The traditional IDF factor penalizes the score of terms that appear across many documents in the entire corpus. In contrast, we compute the IDF factor for a term T in a document D based on the number of documents *preceding* D that contain T.

Intuitively, it makes sense to identify the most significant terms at a particular point in time without knowledge of the future. We have observed that people tend to find the onset of a new term particularly memorable; such scenarios are promoted by the time-based TF-IDF. Examples of terms that benefit from this metric include the name of a newborn family member, or a name like *Obama*, that emerges at some point in time and subsequently becomes commonplace. As time goes by with repeated use of the new term, the IDF score slowly reduces, making the term less prominent.

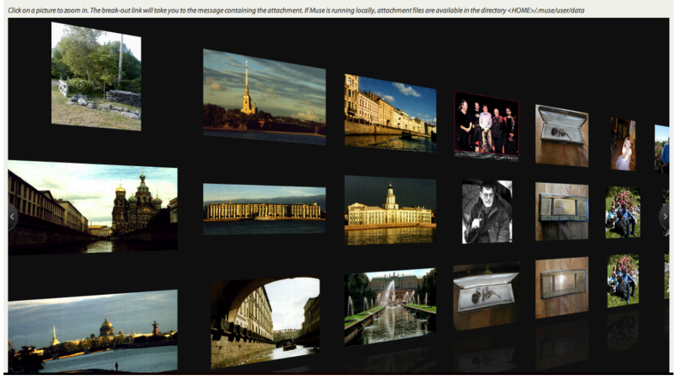
A concrete example from one of the authors' email archives illustrates this point. MUSE highlighted terms related to music in the first two months after he started taking music lessons. Thereafter, the music conversations continued, but these terms disappeared from the monthly summaries because they become relatively common. With the regular version of TF-IDF, these terms do not show up at all because of a relatively low IDF score across the entire corpus.

To present these terms to the user, MUSE lists the top named entities for each month in a month-by-month view, similar to a calendar. Users can ask for more or less terms to be displayed (the default is set to 30 terms for each month). MUSE clusters and color-codes the terms by the inferred group they are most closely associated with. Terms assigned to the same group (and therefore the same color), are displayed together, and are further sorted by descending score. This is useful because terms belonging to a group tend to be related to each other and the user can quickly scan them together; users frequently have varying levels of interest in different groups. An early round of user testing with five users showed that organizing and color coding terms by group was unanimously preferred over just sorting terms by score. Terms not assigned to any group are colored gray and displayed last. The color encoding also makes it easy to scan all terms related to a particular group across different months. Fig. 4(c) shows the names view actually displayed by MUSE for this example.

To avoid the problem of over-representation from a single message, we throttle the number of terms displayed that belong to a single message. After a preset threshold is reached (empirically, four works well), other terms from the message are suppressed, unless they are also present in a different, non-maximally represented message.

Sentiment Cues

During our experiments with early versions of MUSE, we found that many of the messages that engaged users the most during the reminiscence process were those that reflected significant turning points in their life and deep emotions, such as love, joy, grief and anger. There is much evidence that emotional episodes tend to be well-remembered, both for positive and negative emotions [27]. MUSE therefore uses sim-



Click on a picture to zoom in. The break-out link will take you to the message containing the attachment. If Muse is running locally, attachment files are available in the directory <HOME>/.muse/user/data

Figure 5: Attachment wall using the PicLens viewer.

ple sentiment analysis techniques to let users quickly browse messages by the sentiment associated with them.

The most commonly used tools for sentiment analysis such as SentiWordnet [2] and LIWC [16] use word lists to detect sentiment. We have generated our own (English language) lexicon, that consists of 20 categories with terms covering various emotions, family, health, life events, expletives, etc. that may be useful for our domain of reminiscence with email archives. These terms are matched (modulo stemming) with the contents of the message. Significant emotions that can be detected with high certainty, such as congratulations, are assigned their own category. Other kinds of emotions that can be classified with less certainty are grouped into two broad categories for positive (gratitude, pride, joy, humor, etc.) and negative (disappointment, anxiety, worry, etc.) sentiment.

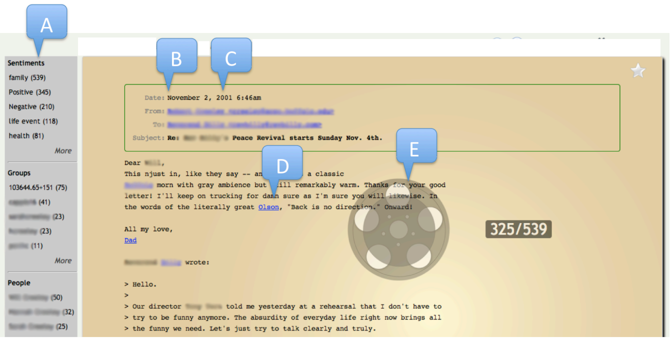
MUSE depicts the frequency of messages reflecting these sentiment categories across time using a stacked graph (see Fig. 1); each layer represents a particular sentiment category. Users can click on a layer to launch into a view containing all the messages reflecting that sentiment; the view is initialized to the point in time along the X-axis that was clicked. Users can also examine how their sentiment graphs vary across different social groups.

We have found that while sentiments are among the noisiest cues provided by MUSE, they are also often the most engaging. Users are curious about interpreting the sentiments graph, especially when exploring their sent messages. Even if a detected sentiment is due to a language artifact, it often gives users a new view into their own use of language.

Picture Cues

Picture attachments in email messages are useful because the vividness of images provides strong cues to memory. Further, pictures are often taken for the explicit purpose of later remembrance and hence may be worth recalling.

MUSE extracts picture attachments from messages (and optionally PDF documents, which are converted to thumbnails) and displays them on the PicLens photo wall (from cooliris.com) which provides a 2.5D zoomable and draggable interface. The images are arranged in reverse chronological order and it is easy for the user to rapidly scan pictures, and pan to different areas of the wall without waiting for the whole wall to render chronologically.



Date: November 2, 2001 6:46am
From: [redacted]
To: [redacted]
Subject: Re: [redacted] Peace Revival starts Sunday Nov. 4th.

Dear [redacted],
This njust in, like they say -- an a classic
[redacted] morn with gray ambience but still remarkably warm. Thanks for your good
letter! I'll keep on trucking for damn sure as I'm sure you will likewise. In
the words of the literally great Olson, "Back is no direction." Onward!

All my love,
Dad

[redacted] wrote:

> Hello.
>
> Our director [redacted] told me yesterday at a rehearsal that I don't have to
> try to be funny anymore. The absurdity of everyday life right now brings all
> the funny we need. Let's just try to talk clearly and truly.

Figure 6: The messages view in MUSE, with 539 messages loaded. (A) Facet panel for sentiments, groups and people. (B) Link to all messages for month and (C) for year. (D) Hyperlink inserted into message contents. (E) Jog dial for rapid skimming. Some details blurred to protect privacy.

We find that users are often pleasantly surprised to rediscover pictures from their email attachments. While many more pictures are shared through formal mechanisms such as photo sharing sites, the memory of pictures sent in email attachments is not refreshed since it is not easy to browse email attachments. Therefore there is a sense of novelty in re-discovering a long forgotten picture.

From the photo wall, users can click on an attachment to go to a view containing the message(s) with that attachment. From that point, the user can continue exploratory browsing using the usual facets such as people, groups, and sentiments. The browsing interface is described in the next section.

EXPLORATORY BROWSING

Generating interesting cues is often merely a step on the way to browsing the actual messages which hold the memorable details. In practice, most users spend more time browsing messages than browsing cues. Therefore it is important to support rapid exploratory browsing of a large set of messages. When a user follows up on a cue, e.g., by clicking on a name in the monthly summaries, or by selecting a point in a stacked graph visualization, MUSE opens a *message view* for the associated messages. The message view displays the actual message header and contents, along with thumbnails of any attachments for the message at the bottom. Multiple views can be simultaneously active in different browser windows or tabs to encourage multiple chains of exploration.

Skimming with an on-screen jog dial

Our original implementation of the message view displayed all the messages in the view, one below the other. This tended to create long pages, and we found that when there were more than about 10-15 messages, users would get bored and stop scanning messages part of the way down the page. While following cues, however, some views (for example, all messages for a group or person) can consist of hundreds of messages and are tedious to click or scroll through.

To alleviate the tedium of scrolling down a long page, we load multiple messages into the browser but display only one message at a time in a fixed message frame. This has the advantage that it fixes the on-screen locations of the message headers and the beginning of the message contents, which are often the most important for sensing the relevance of a message. To enable rapid scanning of messages in the view, we provide a translucent on-screen circular jog dial that is summoned on the spot and dismissed by clicking anywhere in the message frame. The operation of the dial is similar to the physical dial on iPod music players: moving clockwise to the next octant causes the frame to display the next message; moving counter-clockwise displays the previous message. Fig. 6 shows the message view with the jog dial visible. Apart from being somewhat playful, the dial allows fast interactive performance through the use of client-side Javascript – in our experience, users can rapidly rotate the dial to approach a skimming speed of 150 messages a minute while still being able to monitor the content passing by.

The dial affords finer-grained control than keyboard navigation, as users can slow down and speed up as they wish, depending on their interest level in the phase of messages. The jog dial lets them travel relatively long distances (scanning through a few hundred messages is common) without the need for precise cursor positioning, mouse clicks, key presses, or switching gaze from the message view. The circularity of the gesture avoids the need to periodically reposition the cursor, and is a general advantage of elliptical gestures. For fast travel in extremely long views, we also provide the option of tabbing backwards or forwards to move month by month. Of course, the user can always use the keyboard right or left arrows keys to scroll through messages.

It is important to manage browser load for a view with thousands of messages. MUSE maintains a sliding window of pages around the page currently being displayed, and keeps only the pages in that window loaded in the browser. As the user moves along, the Javascript "pages in" new messages to maintain the window around the current page, while retiring pages outside the window. In practice, we find that a default window size of 100 pages (60 pages ahead and 40 pages backward) is adequate for good user experience with no stalls and is easily handled by current browsers.

While we have not formally evaluated the jog dial with respect to alternatives, it is a popular feature with many users of MUSE. We designed the dial to operate primarily with a trackpad, but users have also reported relatively high satisfaction rates with a mouse. We plan to perform detailed standalone evaluation of this interface element in future.

Facets and Hyperlinks

We encourage users to follow top-level cues by using links to related facets and annotating the message contents with the named entities identified as top level name cues. For example, if a name cue is mentioned in a message, we insert a hyperlink for that name to a view containing all the messages with that name. Users can optionally open this new view in a different browser tab without disturbing the current chain of exploration. Similarly, ordered lists of groups, sentiments and people associated with the messages in the current view are shown in a facets panel on the left (see Fig. 6). From any message, the user can click on the month or year in the message header to launch a new view containing all the messages in that time unit. When a message view is generated by querying for a term or sentiment, that term is highlighted in the text of the message.

Throughout the interface, clicking on the name or email address of a person, or the description of a group, can be used to launch a message view containing all the associated messages. Similarly it is possible to launch an attachment wall consisting of all pictures in the messages in the current view. Finally, in addition to browsing facets, users can always type in their own search terms.

USER STUDY

Throughout the development of MUSE, we conducted surveys and formative studies to inform our design. We also conducted two formal studies (with a total of 13 users), as well as other informal testing with early users. The findings led to the current version of MUSE. We now report on a small study with 6 users to test this version.

Methodology

In this study, we recruited 6 participants (P1–6) who had access to relatively long-term email archives. We required them to have at least 5,000 messages (preferably sent by the user), acquired over at least 5 years. We invited three professionals from our university library to participate in this study because they could offer us expert-level feedback based on their experience in dealing with archival material. Two were professional archivists who dealt with special collections, both physical and digital, and the third was a historian and professional curator for the library. The remaining three participants were working professionals. Participants had between 10 and 30 years of work experience, and had all been using email throughout their working lives for both professional and personal purposes. Two of the participants were female, and only one had used an early version of MUSE before this study. Participants were compensated with a $10 gift coupon.

We conducted a pre-study meeting with each user to ascertain the state of their email archives. Without exception, all users had email archives in multiple accounts or sources (online service providers, company account, files on a hard drive, etc.) Some users had archives in older formats (like Eudora) and no longer had the program to read it; in these cases we wrote scripts to help them convert their archives to the mbox format which MUSE can read. There were frequently discrepancies between what users thought they had in which folders, and what material was actually present, pointing to the difficulty of maintaining consistent foldering practices, across time and different accounts, even for professional archivists. In general, it remains a challenge for ordinary users to access their decades old email. Further some service provides like Hotmail support only the POP email protocol, which allows access only to the Inbox folder. One user had a significant number of messages in a Hotmail account; we helped him import the Hotmail messages to Gmail for better access via IMAP. The problem of maintaining access to email in historical formats is well-known to archivists [11], and they pointed us to commercial software that they frequently use to convert from one email format to another.

In the actual study, we gave participants a 5-minute tour of the different types of cues and browsing features of MUSE. We then asked them to spend 30 to 45 minutes examining the cues and following them to generate memories. At the end, we asked them to fill out a detailed survey with 41 questions (5 of the 6 users asked for more time to browse their archives, or if they could return the survey after running MUSE on other parts of their archives, in which case we let them do so). The survey asked them to rate the usefulness of different kinds of cues, and to rate different aspects of the user interface such as the jog dial and faceted navigation. Other questions asked for feedback on sentiment categories that were useful or noisy, and for general comments about the automatically inferred social groups.

Results

Most users thought MUSE provided useful cues to jog their memory and remind them of past incidents, which they had otherwise forgotten. While Musing, people were deeply engrossed in their past. They were reminded of both high-level patterns (P5: "*That year is full of Europe for me, I traveled so much.*") and specific episodes (P1: "*I had to go to the DMV when I moved to a new city.*"). Users were often surprised by the extent of material in their archives (P4: "*Wow – I'm writing a book on Warcraft, and I didn't realize my email had stuff about it back in 1999!*"), and generally enjoyed discovering long-forgotten messages.

Broadly, all four cue types got good ratings from users. On a 5-point scale, the attachments cues were rated the highest with an average of 4.25, closely followed by a tie between monthly terms and sentiments at 4.17 each. The groups cues got an average rating of 3.83. While our sample size is too small to be conclusive, comments from other users of MUSE are consistent with these findings.

Two participants remarked that the monthly terms view was what they would use the most. The picture cues were also popular. One user found valuable pictures of her son's first year lying in her archives. P3: "*I've been looking for these and thought they were lost. Let me save them while I can. . .*"

While we expected that the names and attachments cues would be highly evocative for the user, we were somewhat surprised that users responded well to the simple sentiment cues. The sentiment cues achieved the same average rating as the name cues, which took considerably more time and effort on our part to generate and prioritize. P1 remarked, "*I just keep coming back to the sentiments view, it's so much fun.*" Users also took the sentiment graphs fairly seriously. P5: "*I am relieved to see that 'positive' outweighs 'negative' by a considerable margin, especially in recent years!*"

When we asked users which sentiment categories were accurate, and which ones were not, the responses varied. P5: "*I thought positive and negative were pretty accurate, the negative and angry ones did capture some uncomfortable exchanges with a former roommate (there was money involved)*", and P3: "*The congratulatory messages are pretty useful.*" Users also experienced noisy or incorrect sentiments. P1: "*It thinks there is a lot of religion in my life because I used to work in a theology library. . . well, you know, maybe that's right*" (laughs), or P6: "*I deal a lot with 'Born Digital' documents, so* MUSE *thinks I have a lot of life events.*" For the most part, users ignored misclassifications and focused on the categories that did work well for them. Our observation is that the simplicity of the sentiment categorization has the advantage of being extremely *transparent*, which allows users to easily ignore any parts that do not work well. A surprise to us was that four of our six users voluntarily mentioned that they would like to be able to edit the lexicon and put in their own terms. Perhaps this should not be so surprising seeing the popularity of tools like Google's N-gram viewer that allow people to ask simple questions about the evolution of language.

While the social groups were rated lower than other cues, we noted that the activity graphs with groups straightaway told a story. Most users would look at the graph and say, "*Oh yeah, that makes perfect sense.*" Further, users would often enter group views from the faceted browsing interface (by clicking on a particularly important group or person) and may not have realized that they were following group cues.

Although it was not an explicit goal of this study, we were curious about how users would react to potentially unpleasant memories being brought up. One user volunteered a reaction that was particularly interesting. P1: "*This reminded me of the stress of looking for a job, how much work goes into a cross-country move, and how hard it was to sell my house after I moved. That sounds unpleasant, but being reminded of these things wasn't a bad experience – it made me reflect on how much I've been through over the past six years, and how glad I am that certain experiences are behind me.*"

While browsing, some users switched into the mode of looking for what their conversations were about at specific points in time. P5: "*Let's see, does Obama show up in the monthly terms around November 2008? Oh yes, he does. Cool!*"

On the interface questions, the average rating of the jog dial was 3.6 on a scale of 5. This rating is somewhat skewed by one user giving a rating of 1; she found it hard to use on her new desktop with trackpad that she was unfamiliar with. Surprisingly, 2 users with a mouse gave the jog dial a 5 rating. In hindsight, mouse vs. trackpad is a condition that we should have controlled for in our study, but we let users choose whatever computer they had available. Making the jog dial more robust and easier for first timers to use, and understanding design issues for mice may be areas of future study. The faceted browsing interface was generally liked (average rating 4.00) as was the fact that a regular web browser could be used to interface to MUSE (average rating 4.25).

In the rest of this section, we present further insights obtained by feedback from people who have used MUSE outside the context of this study.

Summarizing work progress. Several of our users remarked after using MUSE that they would find it useful to summarize their year when writing an annual report or performance re-view. A user commented: "*I wish I had this system at project reviews to quickly scan all the project group messages since the last meeting.*"

Extraction and organization. One user suggested that it would be useful to form a group of all personal contacts and use it to take personal email out with her when leaving a job. She said, "*My husband was leaving the newspaper company he worked at, and spent two days printing out all the personal emails in his work account*", as she rolled her eyes.

A user who is a software entrepreneur said that he had saved his email mainly because it had important documents embedded in it, such as "*company ownership spreadsheets, benefits packages, legal agreements – stuff that's nowhere else.*" MUSE can be used to extract these documents and organize them better.

Family groups. A consistent pattern (also noted by Viégas et al. [35]) was that users tended to spend much of their time browsing messages exchanged with their family group(s), perhaps more than any other group. This may have been due to the long-lived nature of such relationships, which makes introspection on them particularly valuable. In addition, it is common for people to tell distant family members about important milestones and events in their lives, such as job promotions and new romances. One user suggested that we allow editing of results so he could clean them up and share his memories with his family.

Picking up forgotten threads. A few users remarked after using MUSE that they were reminded of unfinished work or projects. For example: "*I'd like to remind my friend that we were planning this trip – wonder why it got dropped and we never went.*" Our hypothesis is that users may also find such life-browsing useful as a reminder of high level goals and ambitions they once had. Interestingly, one user reported that she felt a renewed sense of confidence by looking at her past achievements, an observation also made by Kirk and Sellen [14].

Renewing relationships. Multiple users remarked after reviewing old conversations that they felt bad they were no longer in touch with people who had been very close some years ago: "*I had forgotten that we were such close friends, but then I moved, and we stopped talking*" and "*Wow! I had forgotten how nice one friend was in offering me a temporary place to stay (I ended up staying elsewhere) but she has been a little grumpy lately but I can forgive that a bit now that I remember that incident.*"

Serendipitous discovery. One user when browsing messages noticed that her son's name was part of a message and had a hyperlink. Clicking on the link brought up a view with 224 messages with her son's name. As she skimmed through the messages, she remarked: "*Wow, this offers a pretty complete history of my first son's milestones. There is no other record of this. I've been trying to remember his milestones to compare with my other son's.*"

All users said that MUSE revived their memories of topics that they had otherwise forgotten about. Sometimes these were topics in themselves, and sometimes they were satellite topics around other events that they did remember: "*I had forgotten about that lunch we organized right before my thesis defense.*" This suggests that MUSE can add color and detail even to such "flashbulb" memories of significant and well-remembered events. Further, the archives add concrete evidence to the event; it is well known that even flashbulb memories are prone to incorrect recall with high confidence [27].

Summary

What struck us, as we saw our users' reaction to MUSE, was the variety of ways in which users derived utility and benefits from a system that jogs their memory, from summarizing work to renewing friendships and serendipitous discovery. Though our initial goal with MUSE was only to support the task of reminiscence, the stories above include an example of each one of the "5R's" described by Sellen and Whittaker [29]: recollection, reminiscing, retrieving, reflecting, and remembering intentions. Further, it suggests that browsing and remembering the past can affect the future.

The expert users from our library were also very interested in using MUSE to enable browsing of archives of famous people whose papers they help to acquire and organize. P6: "*We have so many donors wanting to donate their documents (including email) to us, we don't have enough people to look through all the materials.*"

CONCLUSIONS

Millions of people have email archives that are rich in sentiment and meaning to them, forming a passively acquired life-log. Users are surprised by the extent of information directly or indirectly reflected in their archives, and broadly enjoy discovering forgotten topics. We have identified and evaluated four types of cues that are useful, yet lightweight to compute: communication with inferred social groups, recurring named entities, sentimental words, and image attachments. We have found that users derive many different types of benefits from reflecting on their email archives. Further, we found that MUSE can potentially be useful to archivists and other curators of digital content. We discovered that users spend most of the time with the actual message content, so it is necessary to integrate cues with effective browsing of messages. MUSE can be publicly accessed along with supplementary materials for this paper at the URL: http://mobisocial.stanford.edu/muse.

FUTURE WORK

We have made MUSE publicly available and intend to scale our current user study by recruiting diverse users via the web. We plan to improve its usability and features based on feedback from users, and conduct detailed user studies of specific user-interface elements. A limitation of our current user studies is that we do not measure recall, i.e., we do not know what important topics and events were not picked up by MUSE. A controlled experiment to identify such topics may suggest possible improvements.

Acknowledgements

We thank Stu Card, Peter Chan, Phuong Ly, Diana MacLean and Jason Zalinger for useful discussions and feedback on this paper, Stanford University Libraries for access to the

Robert Creeley corpus, and Peter Monaco from Xobni for providing us an anonymized email data set. Users of all versions of MUSE have provided us valuable feedback. We also thank our anonymous reviewers for useful feedback. This research was funded in part by NSF POMI (Programmable Open Mobile Internet) 2020 Expedition Grant 0832820 and the Stanford MobiSocial Computing Laboratory.

REFERENCES

1. S. Asur and B. A. Huberman. Predicting the future with social media. *CoRR*, abs/1003.5699, 2010.
2. S. Baccianella, A. Esuli, and F. Sebastiani. SentiWordNet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In *Proceedings of Language Resources and Evaluation (LREC'10)*. European Language Resources Association, May 2010.
3. G. Chin, Jr., O. A. Kuchar, and K. E. Wolf. Exploring the analytical processes of intelligence analysts. In *Proceedings of CHI-2009*. ACM, 2009.
4. C. Collins, M. Watternberg, and F. Viegas. Parallel tag clouds to explore and analyze faceted text corpora. In *Proceedings of VAST '09*. IEEE, 2009.
5. E. Cutrell, D. Robbins, S. Dumais, and R. Sarin. Fast, flexible filtering with Phlat. In *Proceedings of CHI '06*. ACM, 2006.
6. S. Dumais, E. Cutrell, J. Cadiz, G. Jancke, R. Sarin, and D. C. Robbins. Stuff I've Seen: a system for personal information retrieval and re-use. In *Proceedings of SIGIR '03*. ACM, 2003.
7. D. Elsweiler, M. Baillie, and I. Ruthven. Exploring memory in email refinding. *ACM Trans. Inf. Syst.*, 26(4):1–36, 2008.
8. Email archives. Robert Creeley Papers, M0662. Dept. of Special Collections, Stanford University Libraries, Stanford, CA.
9. J. R. Finkel, T. Grenager, and C. Manning. Incorporating non-local information into information extraction systems by gibbs sampling. In *Proceedings of ACL '05*. Association for Computational Linguistics, 2005.
10. J. Gemmell, G. Bell, and R. Lueder. MyLifeBits: a personal database for everything. *CACM*, 49(1), 2006.
11. A. Goethals and W. Gogel. Reshaping the repository: The challenge of email archiving. In *Proceedings of the 7th International Conference on Preservation of Digital Objects*, 2010.
12. S. Havre, E. Hetzler, P. Whitney, and L. Nowell. Themeriver: visualizing thematic changes in large document collections. *IEEE Transactions on Visualization and Computer Graphics*, 8(1):9–20, 2002.
13. S. D. Kamvar and J. Harris. We Feel Fine and searching the emotional web. In *Proceedings of WSDM-2011*. ACM, 2011.
14. D. S. Kirk and A. Sellen. On human remains: Values and practice in the home archiving of cherished objects. *ACM TOCHI*, 17(3):10:1–10:43, 2010.
15. S. Liu, M. X. Zhou, S. Pan, W. Qian, W. Cai, and X. Lian. Interactive, topic-based visual text summarization and analysis. In *Proceedings of CIKM '09*, pages 543–552, 2009.
16. LLWC Inc. Linguistic Inquiry and Word Count. `http://www.liwc.net`.
17. D. MacLean, S. Hangal, S. K. Teh, M. S. Lam, and J. Heer. Groups without tears: mining social topologies from email. In *Proceedings of IUI-2011*. ACM, 2011.
18. S. Malacria, E. Lecolinet, and Y. Guiard. Clutch-free panning and integrated pan-zoom control on touch-sensitive surfaces: the cyclostar approach. In *Proceedings of CHI '10*. ACM, 2010.
19. M. Mandic and A. Kerne. Using intimacy, chronology and zooming to visualize rhythms in email experience. In *Proceedings of CHI '05 (extended abstracts)*. ACM, 2005.
20. C. D. Manning, P. Raghavan, and H. Schutze. *Introduction to Information Retrieval*, page 127. Cambridge University Press, 2008.
21. C. Neustaedter, A. Brush, M. A. Smith, and D. Fisher. The social network and relationship finder: Social sorting for email triage. In *Proceedings of CEAS '05*, 2005.
22. S. T. Peesapati, V. Schwanda, J. Schultz, M. Lepage, S. Jeong, and D. Cosley. Pensieve: supporting everyday reminiscence. In *Proceedings of CHI '10*. ACM, 2010.
23. A. Perer, B. Shneiderman, and D. W. Oard. Using rhythms of relationships to understand e-mail archives. *J. Am. Soc. Inf. Sci. Technol.*, 57(14):1936–1948, 2006.
24. D. Petrelli, S. Whittaker, and J. Brockmeier. Autotypography: what can physical mementos tell us about digital memories? In *Proceedings of CHI '08*. ACM, 2008.
25. T. J. Purtell, D. MacLean, S. K. Teh, S. Hangal, M. S. Lam, and J. Heer. An algorithm and analysis of social topologies from email and photo tags. In *Proceedings of the Fifth ACM Workshop on Social Network Mining and Analysis (SNAKDD)*. ACM, 2011.
26. Radicati Group Inc. Email statistics report, 2010-2014. `http://www.radicati.com/?p=5282`.
27. D. Reisberg and P. Hertel. *Memory and Emotion*. Oxford University Press, 2004.
28. J. Schwarz, J. Mankoff, and H. S. Matthews. Reflections of everyday activities in spending data. In *Proceedings of CHI '09*. ACM, 2009.
29. A. J. Sellen and S. Whittaker. Beyond total capture: a constructive critique of lifelogging. *CACM*, 53:70–77, May 2010.
30. J. Stasko, C. Gorg, and Z. Liu. Jigsaw: supporting investigative analysis through interactive visualization. *Information Visualization*, 7:118–132, 2008.
31. A. C. Surendran, J. C. Platt, and E. Renshaw. Automatic discovery of personal topics to organize email. In *Proceedings of CEAS '05*, 2005.
32. E. Tulving. *Elements of Episodic Memory*. Oxford University Press, 1983.
33. J. R. Tyler and J. C. Tang. When can I expect an email response? A study of rhythms in email usage. In *Proceedings of ECSCW'03*. Kluwer Academic Publishers, 2003.
34. F. B. Viégas, D. Boyd, D. H. Nguyen, J. Potter, and J. Donath. Digital artifacts for remembering and storytelling: Posthistory and social network fragments. In *Proceedings of HICSS '04*. IEEE Computer Society, 2004.
35. F. B. Viégas, S. Golder, and J. Donath. Visualizing email content: portraying relationships from conversational histories. In *Proceedings of CHI '06*. ACM, 2006.
36. S. Whittaker, Q. Jones, B. A. Nardi, M. Creech, L. Terveen, E. Isaacs, and J. Hainsworth. ContactMap: Organizing communication in a social desktop. *ACM TOCHI*, 11(4):445–471, 2004.
37. World Wide Web Consortium. W3c contacts api working draft. `http://www.w3.org/TR/contacts-api`.
38. K.-P. Yee, K. Swearingen, K. Li, and M. Hearst. Faceted metadata for image search and browsing. In *Proceedings of CHI '03*. ACM, 2003.
39. J. Zalinger. *Gmail as storyworld: How technology shapes your life narrative*. Unpublished dissertation, Rensselaer Polytechnic Institute, 2011.