Email Visualizations to Aid Communications

Steven L. Rohall, Daniel Gruen, Paul Moody, and Seymour Kellerman
IBM Research
One Rogers Street, Cambridge, MA 02142
firstname_lastname@lotus.com

Abstract

Electronic mail has become the most widely used business productivity application. However, people increasingly feel frustrated by their email. They are overwhelmed by the volume, lose important items, and feel pressure to respond quickly. Though email usage has changed, our email clients largely have not.

This paper describes our work on using various information visualization techniques (trees, timelines, and low-resolution overviews) in a new email client. These visualizations highlight the relationships among messages and between the people who exchange email.

1. Introduction

Electronic mail has become the primary business productivity application. It has emerged as the most-used communications tool in the US and Canada and, according to an Institute for the Future study, 97% of workers report using email daily or several times each week [1]. In fact, US workers average 49 minutes a day managing email, and 25% spend more than one hour per day on that task [2].

Despite our reliance on electronic mail, our tools for handling email have failed to keep pace [3]. First, workers feel **overwhelmed** by their email. The average user gets approximately 24 messages per day while "high-volume" users can easily get several hundred [1][4]. Ironically, 34% of internal business messages were deemed "unnecessary" [2]. Second, people use their email inboxes to manage their tasks, yet they complain **"things fall through the cracks."** Current organizational structures within email clients, such as folders, prove inadequate, especially for high-volume email users [5]. Finally, people feel pressure to be **more responsive** in replying to email messages, reporting that 27% of messages sent "require" immediate attention [2].

Email clients have changed little since they were first invented. Although today's email clients are more graphical with onscreen buttons, pull-down menus and rich-text display, they are essentially "souped-up" cousins of the email clients from thirty years ago. Most email clients today have the same set of features and organizational structures: multiple folders in which messages can be filed, a textual listing of the messages within a given folder, and the ability to preview a selected message. However, studies have shown that folder systems quickly degrade with the number of messages people receive [5]. Most people end up keeping all of their email in one large folder [6]. The content and use of email has also changed. In addition to traditional letters, email now consists of invitations, receipts, transactions, discussions, conversations, tasks, and newsletters, to name a few variations [3].

The opportunity exists for reinventing the email client, moving it from the current electronic analog of physical mail to a tool that allows users to manage all of their digital communications. New visualizations of the information contained within electronic mail inboxes are a key piece of the solution.

2. Our approach

Email has several numeric attributes (e.g., message size, number of attachments) that are easy, but not very useful, to visualize. Therefore, we have had to look deeper into the structure of electronic communications to determine salient features that should be visualized. We have focused on three such features: message threads, time, and the content of documents.

2.1. Message threads

The idea of message (or conversation) threads is borrowed from discussion databases. At its simplest, a conversation thread in electronic mail represents the series of replies to a message and the replies to those replies. Some email systems have support for message threads. Typically, they display a textual representation of the

thread tree using a Windows Explorer-like tree control.

In discussion databases, threads are complete. Since every message in a discussion database is archived, the entire tree of messages in a thread can be displayed. In fact, several systems have attacked the problem of visualizing the conversation threads that exist within discussion databases: Chat Circles [7], Loom [7], and Conversation Map [8] are examples. Threaded Chat [9] investigated using threads to structure the discourse in a public "chat" room, essentially a synchronous form of discussion database.

Within electronic mail, threads are frequently incomplete. Each person has a separate representation of a thread of messages in an email exchange. In particular, a user may choose to delete some received messages or not save copies of sent messages. As a result, email systems that do *post-hoc* analysis for thread determination (e.g., by using "In Reply To" headers) create threads that are incomplete and have holes. The system we are prototyping does a small amount of bookkeeping whenever a message is sent or received so that we always have complete message thread information.

Once we have complete message thread information, we can turn our attention to the visualization of those trees. Much research has been devoted to the visualization of trees. However, the bulk of that work has concentrated on the display of trees with hundreds or thousands of nodes. Our problem is simpler: message trees, from our observations, are not very deep nor very bushy in general. This allows us to use a simple graphical representation of the message thread and to highlight the interesting relationships among the people involved in the conversation.

Figure 1. An email thread

Figure 1 shows an email thread consisting of ten email messages. In this case, the nodes are colored to show the relationship of the sender of each message to the recipient. Messages colored purple are from someone outside of the recipient's company. Yellow messages represent the recipient's contribution to the thread. The red message is from the recipient's manager, green messages are from others within the recipient's work group, and the light blue message is from someone within the company but outside of the recipient's work group.

The relationship among senders and receivers of messages can be determined easily through the use of a corporate directory. In fact, the email system could track my communications and, for example, highlight a message from someone with whose workgroup I communicate frequently, even if I've never exchanged mail with that particular person.

As a means of organization, threads provide a structure that combines messages that would normally be scattered throughout the inbox. Our prototype email client allows a scattered thread to be collapsed into a single visual entity, reducing inbox clutter. Selecting the collapsed thread displays the detailed tree structure.

2.2. Time and timelines

One attribute of electronic mail that is valuable to visualize is the time when a message was received. In fact, several prototype email clients have used time as a key organizing structure. NEC's VisualMail [10] provided a two-dimensional plot of mail categories versus time. The categories were determined by creating rules based upon user behavior. TimeStore [11] took a similar approach and displayed a matrix of email senders versus time. The time scale could be set to view a season, a month, a week, or a day at a time. The volume of email from a sender was shown as a circle in the corresponding cell of the table—a larger diameter representing more email in that period. Frequent senders of email were grouped at the top of the table to aid the user.

Figure 2. Message tree with timeline

Our approach is to combine the message trees described above with a timeline to produce a more useful visualization.

Figure 2 shows a prototype design for displaying a message tree on a timeline. The vertical lines represent day boundaries. The text in the middle band is the subject of the thread. The color-coding of the nodes is the same as in the previous example and represents the relationship of the message senders to the recipient. Note that time is non-linear in this display; days with little or no activity are shown compressed. This avoids the problem of large gaps in the time display experienced in the VisualMail and TimeStore systems.

One of our main areas of focus with this visualization is the combination of the thread information, time, and other information from the user's mailbox. For example, we imagine that the timeline can be broken to show a large passage of time. This might be useful if I receive email from someone infrequently. In that case, the system could show on the timeline the most recent threads of conversation with that person. Also, we plan to incorporate information from people's calendars to aid in search. For example, I might remember that I received a

certain piece of mail just before I left for vacation last summer. By incorporating these "milestones" on the timeline view, we can aid the user in finding the information.

A few other systems have combined trees and timelines. Derthick and Roth describe a system where a tree of contexts and a timeline can be used for exploratory undo/redo [12]. While their timeline and tree can be used for navigation among the various contexts, the design of the interface was not fully explored. The Netscan system for reading Netnews databases also contains a graphical tree superimposed on a timeline [13]. Netscan divided the timeline into days, but the placement of the message nodes within the day was based upon the order in which messages were posted, not the actual time in which they were received. As a result, days with many messages would get quite wide and falsely implied that a message arrived, say, at the end of the day when, in fact, it may have arrived much earlier. Our idea is to place message nodes proportionally within a day even though the width of a day on the timeline may vary.

2.3. Document content

We have experimented with ways to make the content of email documents apparent. In particular, we have explored the ways in which reduced-resolution overviews can be used to help people find the documents they want.

Figure 3 shows a design for the display of search results. In this case, the search might have been "show the last seven days of email from Dan." Dates are listed across the top and overviews of the message bodies are shown in the column below. Dates without any messages are omitted. This interface allows a user to scan the email in the same way that a pile of physical mail would be scanned. It is easy to separate the short messages **(2)** from the longer messages. Similarly, it is easy to pick out the messages that contain images **(3)**. The messages highlighted in purple **(1)** are all part of the same thread. Finally, with some simple coloring of the text based upon

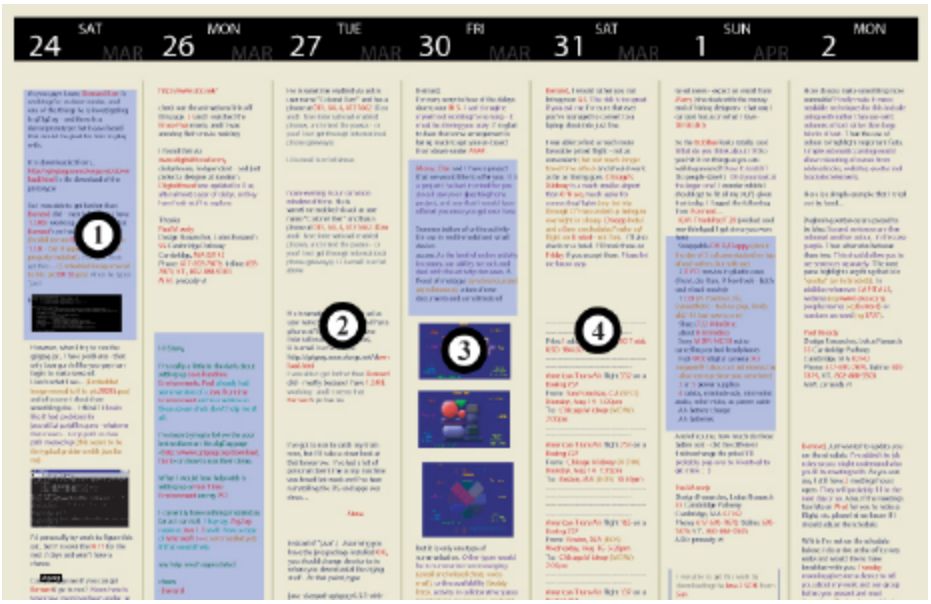

Figure 3. Reduced-resolution document overviews

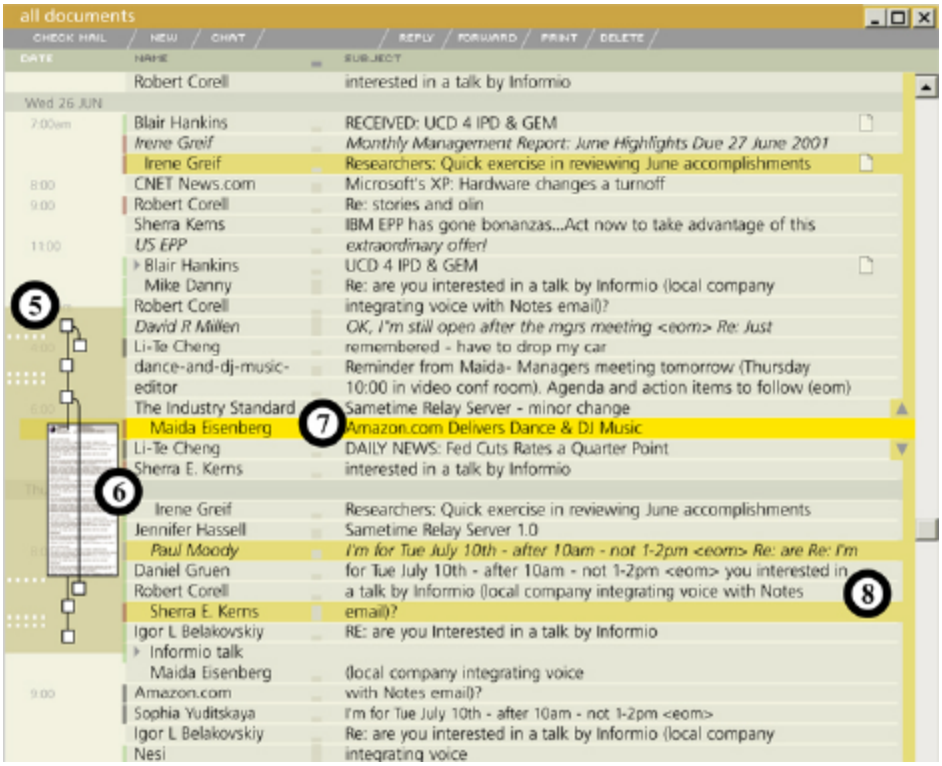

Figure 4. Prototype client interface

extracted features (red to highlight names and dates in this example), message **(4)** stands out from the others. It is, in fact, a travel itinerary.

Although other systems have used reduced-resolution document overviews in their user interface (see, for example [14]), email, in particular, would benefit from such visualization. Email has structure that other client software has failed to exploit. With an overview, people can quickly pick out different types of email (e.g., agendas, online purchase receipts, corporate-wide announcements). Automatic classification of this sort has proven error-prone [3].

3. Status

We have created design mockups and conducted surveys and interviews of email usage (manuscript in progress). We have also begun implementing a new email client, the design of which is shown in Figure 4. The client combines a traditional list of email messages with a time-based message tree **(5)**. The node for the selected message **(7)**, highlighted with yellow, is replaced with a reduced-resolution overview **(6)**. A dimmer, secondary highlight **(8)** also connects the messages within the thread.

Our goal is to create a prototype that is robust enough for daily use so people can test our ideas on their real email. Our infrastructure is designed to be flexible so that we can try out different visualizations. We are using Lotus Domino for our underlying object store [15]. The user interface is being developed using IBM Sash, a development environment based upon dynamic HTML and JavaScript [16].

4. Conclusion

Electronic mail has become indispensable for business. It is used for workflow applications, e-commerce, meeting invitations and more in addition to traditional messaging. Email clients, however, have changed little and people feel overwhelmed by their email. Information visualization is one tool that can change the way we deal with our electronic communication.

In particular, we need to move beyond visualizing the easy features of email and help people better manage their tasks and relationships. If we visualize the relationship among messages in threads and the structure of documents, we can create a better email client.

5. Acknowledgements

We would like to thank Bernard Kerr for providing design insights and summer interns Kayvon Fatahalian, Evan Jones, and Derek Lam for help with the implementation of the prototype.

6. References

[1] "Pitney Bowes Study Reveals Increased Use of Electronic Communications Tools Among North American and European Workers," Pitney Bowes press release, August 7, 2000.

[2] "The Spam Within: Gartner Says One-Third of Business E-Mail is 'Occupational Spam'," Gartner, Inc. press release, April 19, 2001.

[3] Ducheneaut, N. and V. Bellotti, "E-mail as Habitat," *Interactions*, 8(5), Sept.-Oct. 2001, ACM, pp. 30-38.

[4] Levitt, M., "Email Usage Forecast and Analysis, 2000-2005," IDC Report # W23011, September 2000.

[5] Whittaker, S. and C. Sidner, "Email Overload: Exploring Personal Information Management of Email," *Proceedings of CHI'96*, Vancouver, B.C., April 13-18, 1996, pp. 276-283.

[6] Bälter, O., "Electronic Mail in a Working Context," doctoral dissertation, Royal Institute of Technology, Stockholm, Sweden, 1998.

[7] Donath, J., K. Karahalios, and F. Viegas, "Visualizing Conversations," *Proceedings of HICSS-32*, Maui, HI, January 5-8, 1999.

[8] Sack, W., "Conversation Map: A Content-Based Usenet Newsgroup Browser," *Proceedings of IUI'00*, New Orleans, LA, January 9-12, 2000, pp. 233-240.

[9] Smith, M., J. Cadiz, and B. Burkhalter, "Conversation Trees and Threaded Chats," *Proceedings of CSCW'00*, Philadelphia, PA, December 2-6, 2000, pp. 97-106.

[10] Kudo, M., M. Tanaka, and Y. Koseki, "Information Visualization for Electronic Mail Management," *Proceedings of Visual'97*, San Diego, CA, December 15-17, 1997.

[11] Jovicic, S., and R. Baecker, "Time-Based Archiving and Retrieval of Email," Workshop on History-Keeping in Computer Applications, Human-Computer Interaction Laboratory, University of Maryland, December 3, 1999.

[12] Derthick, M. and S. Roth, "Data Exploration across Temporal Contexts," *Proceedings of IUI'00*, New Orleans, LA, January 9-12, 2000, pp. 60-67.

[13] Smith, M. and A. Fiore, "Visualization Components for Persistent Conversations," *Proceedings of CHI'01*, Seattle, WA, March 31-April 5, 2001, pp. 136-143.

[14] Bederson, B. and J. Hollan, "Pad++: A Zooming Graphical Interface for Exploring Alternate Interface Physics," *Proceedings of UIST'94*, Marina del Rey, CA, November 2-4, 1994, pp. 17-26.

[15] Lotus Domino information can be found at http://www.lotus.com/home.nsf/welcome/domino.

[16] IBM Sash information can be found at http://sash.alphaworks.ibm.com/.