You Scratch My Back and I'll Scratch Yours: Combating Email Overload Collaboratively

Manas Tungare
Dept. of Computer Science, Virginia Tech.
Blacksburg, VA, USA.
manas@tungare.name

Manuel A. Pérez-Quiñones
Dept. of Computer Science, Virginia Tech.
Blacksburg, VA, USA.
perez@cs.vt.edu

Abstract
Email is no longer perceived as a communication marvel, but rather as a constant source of information overload. Several studies have shown that accessing, managing, and archiving email threatens to affect users' productivity. While several strategies and tools have been proposed to assuage this burden, none have attempted to empower users to fight the overload collaboratively. We hypothesize that despite differences in email management practices and frequencies of filing among users, there is some degree of similarity in the end-product of the organizational structures reached by those working in close cooperative roles (e.g. members of a research group, employees of an organization). In this paper, we describe a system that enables collaborators to share their filing strategies among themselves. Tags applied by one user are suggested to other recipients of the same email, thereby amortizing the cost of tagging and email management across all stakeholders. We wish to examine if such system support for semi-automated tagging reduces email overload for all users, and whether it leads to overall time savings for an entire enterprise as network effects propagate over time.

CHI 2009, April 4 – 9, 2009, Boston, MA, USA
ACM 978-1-60558-246-7/09/04.

Keywords
Personal Information Management, Email Management, Collaborative Tagging.

ACM Classification Keywords
H.5.2 Information Interfaces and Presentation: User interfaces – Evaluation/ methodology

Introduction & Motivation
Email has slipped from being a killer app for communication to being a constant source of information overload for the modern knowledge worker. Staying in control of one's email inbox is becoming harder and harder, as the sheer volume of incoming email keeps increasing. When faced with burgeoning inboxes and limited time, users must triage their email before they can read and reply.

Apart from volume, another important issue is the emergent usage of email for tasks it was not designed to handle, such as document management, to-do list management, and preserving archives for potential future lookup [11, 16]. Several strands of research have examined email and experimented with various

ways to make it easier to manage. Some of these are based on providing the user with adequate tools to perform the management herself; others strive to automate the management task computationally behind the scenes.

Many tools and techniques that attack the email overload problem tend to view email as personal information (with notable exceptions [6, 5]). In doing so, they miss the intrinsic social nature of email: email is primarily a medium for conversation and transmission of ideas; the role of information archival has been thrust upon it. Based on discussions with several colleagues and from an informal analysis of their email inboxes, we noticed that despite the oft-noted variability in the tendency, techniques and frequency of filing, there is often a lot of similarity among close collaborators in the end-product of the filing process, i.e. the actual organization scheme that is reached as a steady state. Project names and sender/recipient names are often used as folder names. This is in addition to the demonstrated similarity in hierarchies in different collections belonging to the same user (i.e. email, files, bookmarks, etc.) [2]. (In this paper, we do not make a distinction between the use of tags and folders for email organization. In a later section, we discuss the interchangeability of both approaches in our prototype.)

What is evident is that a single strategy alone will not be sufficient to solve the problems with email. The vastly different roles of email require similarly unique strategies for overcoming information overload. Solutions for spam filtering, for example, have little to do with solutions for filing emails. In this paper, we describe our work in enabling collaborators to share their organization schemes among themselves using TagShare, a collaborative email tagging prototype. We wish to determine whether such sharing succeeds in reducing the time spent by each user in managing their email, and whether there is adequate overlap among organization schemes to partially automate the filing task for users based on opt-in criteria. As with any automated organization tools, it is important that users trust the system, and are able to benefit from it. We wish to explore if social recommendations provided by our system satisfy these criteria for users. Finally, we are interested in knowing if such social sharing of email organization schemes motivates some users to adopt strategies they see in other users.

Related Prior Work

Email users have been noted to follow different strategies. Despite the slight difference in the terminology used by various researchers, (filers and pilers [13], prioritizers and archivers [11], no-filers and filers [16], cleaners and keepers [7]), they are generally subdivided into two major camps: those who tend to keep their information organized regularly, and those who tend not to. Whittaker and Sidner identify a third tendency, and refers to such a user as a spring cleaner [16]. Gwizdka further notes [7] that although no-filers and spring cleaners had problems keeping up with task management via email, the advantage that filers gained was at the cost of having to spend time each day regularly in keeping their inboxes trimmed and filed.

Gwizdka notes [7] that filing is a cognitively hard activity, and ideally performed as soon as an email is received. Several tools have been designed to assist users in managing their inboxes automatically: the Bifrost Email Organizer [1] filters email according to certain predefined rules that take into account the

number of recipients, references in the body of the message to current calendar events, and whether or not an email was addressed to them directly. While they report positive feedback from their experiments (N=10, from a single organization), it is doubtful whether such fixed filters could be applicable to several thousand users from various organizations.

Fisher et al. describe the lessons learnt from their Social Network and Relationship Finder (SNARF) [5]. They use email history to infer an individual's social graph, and then use that to provide spatial cues for email triage. No metadata is exchanged between users, however, and each user's use of the tool remains isolated. (i.e., no network effects were evident.) Several email providers (e.g. Google Gmail) make use of aggregate reports from their users to identify and target spam. They rely on their users clicking the "Report Spam" button to identify spam, and this knowledge is aggregated and used to mark the same email (or copies of it) as spam for other recipients as well. Garg et al. [6] describe a spam-filtering technique based on the sharing of email filters among collaborating users. Our prototype extends this to sharing not just whether a message is considered spam, but also how it was tagged/filed by one's close social network.

The idea of collaborative management of organizational information has been pursued in several domains. Erickson proposes Group Information Management as a field of inquiry that examines the semi-public sharing of personal information in social circles [4]. He cites the example of calendaring, which has been fairly well-accepted by users, especially corporate users. Users often view and propose events to another user's calendar, thus sharing the burden of scheduling a meeting among all participants (instead of being designated to a single specific individual). SearchTogether [14] is a search environment that allows users to conduct real-time collaborative searches. Bradshaw et al. [3] propose a knowledge artifact that is the basis of information sharing and annotation among research collaborators. In a similar manner, we envision the collaborative tagging of email to be an activity that distributes the task of tagging among several willing participants who also individually stand to gain selfish benefits.

There is significant common ground among users at the same workplace or among frequent collaborators. Tang et al. [15] describe their work in detecting similarities in files across the (publicly-available) home directories of corporate users. They also suggest that such serendipitous discovery of similar files may reveal shared interests. Similarly, we expect that similarities in the filing structures of collaborators' email can help reveal common interests (in addition to assisting in personal email management).

It is important to note that the social approaches to email discussed so far use the edges from a social graph (the relationships) to help manage email. We propose enlisting the assistance of the nodes themselves (one's social contacts) in a symbiotic relationship to simplify email management for both users. Information from the edges is minimally used to infer a collaborative relationship between the sender and recipient.

Jones et al. [8] propose the development of a unified interface for project management that spans discrete personal information collections (files, calendar, etc.). If project names are found to be often used as email

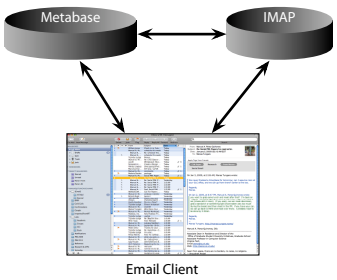

Figure 1: TagShare for Apple Mail

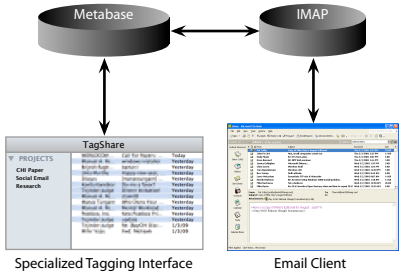

Figure 2: TagShare for other email clients

tags, we can extend project-related attributes from a single user to multiple users. Kaptelinin [9] reported the use of interaction histories to infer project information, which can be used to group related resources and access them in a single project environment. Our prototype infers project names from collaborators' filing habits, and we expect that to assist information management in similar ways.

Scenarios

Here is one of the scenarios that guided the design of our prototype:

Alice, Bob and Charlie are members of a research group and often collaborate in their work. Alice is a frequent filer by nature and prefers to organize her email by tagging it for easy retrieval. Bob files his email irregularly, perhaps every semester. Charlie prefers not to organize his email, and instead relies on keyword searches to locate specific messages. Bob sends a message to Alice and Charlie about a Call-For-Papers and suggests that they should write about their current project, Social Email. Alice, upon receiving the email, applies the tag 'Social Email' to it, and replies saying that she is interested. When Bob and Charlie receive her reply, they see a new option in their email client in addition to the usual user interface. It provides them the option of tagging their email the same way Alice did. Bob agrees to let his mail client tag it as 'Social Email'. In addition, he tags it as 'Papers in Progress'. Alice and Charlie then see their TagShare UI updated to reflect the new tag applied by Bob.

This is inspired by several social services that allow their users to gain from the wisdom of crowds. E.g. when bookmarking a new link on the social bookmarking service Del.icio.us, it suggests tags applied to that same link by other users. These are often very similar to the tags a user would have applied without such prompting. The advantage gained is that users do not need an extra step of typing the tag, and thus encourages active tagging. The GraceNote CDDB (Compact Disc Data Base) is a web service that provides listings of CD tracks. It is based on a community effort where the first person to listen to a CD submits a list of track names, and later users are able to tag their music automatically. It also allows for corrections by other users.

Prototype Design

Conversations consisting of multiple email messages are especially suitable for automatic tagging. Tags applied to one email of a conversation can be automatically reused for the rest of the emails of the same conversation. Any further email in that thread/ conversation will have that tag already applied. This information is stored as custom email headers, thus annotating existing data as proposed by [12]. However, in case of email, technological limitations prevent us from using this directly for sharing among users, hence tagging metadata will additionally need to be stored in a separate multi-user metabase.

The dichotomy between tags and folders in diverse email systems raises some concerns: we plan to adopt the GMail approach, where tags (or labels) are used as folder names. Hierarchical folder names are flattened into a single tag name, e.g. a folder named Projects with two subfolders, 2008 and 2009 will be flattened into two tags, 'Projects-2008' and 'Projects-2009'.

To explore our research questions, we are building a prototype based on Apple Mail *(http://www.apple.com/macosx/features/mail.html)* (figure 1). This will allow

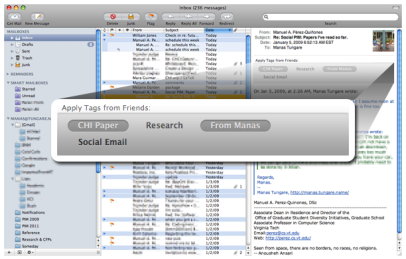

Figure 3: TagShare UI in Apple Mail

users to perform tagging actions as part of their regular workflow, without needing to switch to an external application. The idea of "hooking into" Mail.app was inspired by MailTags *(http://www.indev.ca/MailTags.html)*, a program that allows custom tagging of email inside Mail.app.

For those users who do not use Apple Mail, we plan to create a second prototype, an out-of-process IMAP (Internet Mail Access Protocol) client (figure 2) that will let users filter their email accessed live from any IMAP server. This standalone email management tool directly alters messages stored on an IMAP server. Project information and other metadata applied will then be available to all mail clients that synchronize with that IMAP server. This includes popular email services such as Google Gmail and many corporate email services. While the best integration with a user's workflow can be gained by direct embedding into an email client, we recognize that it is not possible to instrument and support multiple email clients, hence the IMAP-server-based client-independent prototype option.

Metadata about a tagged message will be sent (along with the message-id) to the metabase. This information will then be shared with other recipients of the same message (or the sender herself), but no others. It must be stressed that since neither message subjects nor content is sent to the server, this poses minimal risk of inadvertent disclosure of confidential information to unauthorized parties.

Research Plan & Future Work

We plan to conduct a diary study combined with spot interviews of several participants using the prototypes to manage their email. Since email management practices evolve over a period of time and are intrinsically personal, it is important that any measures be taken after learning effects have been accounted for – at least several months for email, similar to [1]. Kelly [10] describes the methodological challenges in PIM research because artificial experimental tasks lack the personal element, and familiarity with the information influences users' information seeking behavior.

The specific metrics (in addition to content analysis of personal interviews) that we plan to employ are the following: number of messages received, number of tags suggested, number of suggestions accepted, number of messages untagged after automated tagging, frequency of tagging and if it was influenced by the presence of tag suggestions, % of messages left in inbox never tagged, and time required for re-finding tasks with automated tags applied. Pre-questionnaires and interviews will be used to infer a baseline for email habits before the intervention, and timed micro-tasks will be administered during spot checks (e.g. 'please locate the last status update you received from Alice regarding your current project.') [2] found that providing a tool encouraged users to change their strategies. It will be interesting to examine whether such system support for collaborative tagging would lead users to modifying their filing/piling habits.

Summary

We describe a system to enable email users to share their organizational strategies with immediate collaborators and an experiment to examine whether automated collaborative tagging can assist users in email management. We hypothesize that despite habitual differences in email management techniques among users, there is some degree of similarity in the end-product of the organizational structures reached by users working in close cooperative roles (e.g. members

of a research group, employees of an organization). Via collaborative tagging of email for groups of users, we plan to study if such system support for semi-automatic social information management can assist users in overcoming the email overload problems they face today.

References

[1] O. Bälter and C. L. Sidner. Bifrost inbox organizer: giving users control over the inbox. In NordiCHI '02: Proceedings of the second Nordic conference on Human-computer interaction, pages 111–118, New York, NY, USA, 2002. ACM Press.

[2] R. Boardman and M. A. Sasse. "Stuff goes into the computer and doesn't come out": a cross-tool study of personal information management. In CHI '04: Proceedings of the SIGCHI conference on Human factors in computing systems, pages 583–590, New York, NY, USA, 2004. ACM Press.

[3] S. Bradshaw, M. Light, and D. Eichmann. (Bee)Dancing on the Boundary between PIM and GIM. In Proceedings of the 2nd Invitational Workshop on Personal Information Management at SIGIR 2006., 2006.

[4] T. Erickson. From PIM to GIM: Personal Information Management in Group Contexts. Commun. ACM, 49(1):74–75, 2006.

[5] D. Fisher, A. Brush, B. Hogan, M. Smith, and A. Jacobs. Using social metadata in email triage: Lessons from the field. Human Interface and the Management of Information. Interacting in Information Environments, pages 13–22, 2007.

[6] A. Garg, R. Battiti, and R. G. Cascella. "may i borrow your filter? " exchanging filters to combat spam in a community. In AINA '06: Proceedings of the 20th International Conference on Advanced Information Networking and Applications - Volume 2 (AINA'06), pages 489–493, Washington, DC, USA, 2006. IEEE Computer Society.

[7] J. Gwizdka. Email task management styles: the cleaners and the keepers. In CHI '04: CHI '04 extended abstracts on Human factors in computing systems, pages 1235–1238, New York, NY, USA, 2004. ACM Press.

[8] W. Jones, P. Klasnja, A. Civan, and M. Adcock. The personal project planner: Planning to organize personal information. In CHI 2008 Proceedings of the ACM Conference on Human Factors in Computing Systems, page 681, 2008.

[9] V. Kaptelinin. Umea: translating interaction histories into project contexts. In CHI '03: Proceedings of the SIGCHI conference on Human factors in computing systems, pages 353–360, New York, NY, USA, 2003. ACM Press.

[10] D. Kelly. Evaluating personal information management behaviors and tools. Commun. ACM, 49(1):84–86, 2006.

[11] W. E. Mackay. More than just a communication system: diversity in the use of electronic mail. In CSCW '88: Proceedings of the 1988 ACM conference on Computer-supported cooperative work, pages 344–353, New York, NY, USA, 1988. ACM Press.

[12] D. Maier, A. Halevy, M. Bates, H. Bruce, B. Bederson, and M. Knox. Enhancements of personal information. In Breakout Group Summary, PIM Workshop 2005, 2005.

[13] T. W. Malone. How do people organize their desks? : Implications for the design of office information systems. ACM Trans. Inf. Syst., 1(1):99–112, 1983.

[14] M. R. Morris and E. Horvitz. Searchtogether: an interface for collaborative web search. In UIST '07: Proceedings of the 20th annual ACM symposium on User interface software and technology, pages 3–12, New York, NY, USA, 2007. ACM.

[15] J. Tang, T. Lau, J. Lin, and C. Drews. Consolidarity: Exploring patterns of social commonality among consolidated file storage. In HCI Consortium Meeting, February 2006. HCI Consortium, 2006.

[16] S. Whittaker and C. Sidner. Email overload: exploring personal information management of email. In CHI '96: Proceedings of the SIGCHI conference on Human factors in computing systems, pages 276–283, New York, NY, USA, 1996. ACM Press.