Self-Es: The Role of Emails-to-Self in Personal Information Management

Horațiu Bota[*]
School of Computing Science
University of Glasgow
h.bota.1@research.gla.ac.uk

Paul N. Bennett,
Ahmed Hassan Awadallah,
Susan T. Dumais
Microsoft Research
Redmond, WA
{pauben, hassanam, sdumais}@microsoft.com

ABSTRACT

Email has been central to online communication for the past two decades. Through constant use, new information flows are being defined around users' interactions with emails. Alongside traditional messages, the email inbox is an always-available repository of to-do lists, reminders, files and notes. In this paper, we investigate the use of self-addressed emails (*self-Es*) as an information management tool, by analysing both: (i) responses to a survey about email use; and (ii) a collection of user donated self-addressed emails. Our results show that sending self-Es is a frequent behaviour among the users we questioned. In addition, we find that to-dos and reminders are the most popular type of information contained in emails-to-self. Our findings have direct implications for the development of systems that support novel interactions with information flows centred around email.

Keywords

Personal information management; email search.

1. INTRODUCTION

While email has taken a central role in communication, research has primarily focused on the role of email as a *person-to-person* communication tool. However, another phenomenon has occasionally been noted in research [18]: many users often send emails to themselves, where the user sending the mail is the single and only recipient of the mail. For brevity we term such an email a *self-E*[1] (e.g. see Figure 1). For example Bellotti et al.[1] describes such emails as being used as a repository of "to-do" and "to read" items. Shokouhi *et al.* [6] mentions the use for long-term archiving of information; they also note that users have reported difficulty when searching for such mails since common search filters also return results where the user has included themselves in the to/cc with other additional recipients. In this paper, we focus solely on email as a person-to-*self* communication tool via analysis of self-Es.

From: Alex A. Author <a.a.author@acm.org>
To: Alex A. Author <a.a.author@acm.org>
Cc: –
Bcc: –
Subject: after vacation
Sent: Fri 9/2/2016 2:40 PM

\+ Set meeting with Sara
\+ Check whether John resolved the dependency
\+ Ask Pat to pushback timeline

Figure 1: A (hypothetical) example of a self-E. Self-Es are emails where the *only* recipient is the sender. We separate this from the case where the sender may cc himself/herself in addition to others.

Given that a user who sends a self-E is the only recipient of the email, such emails raise interesting questions in *personal* information management. In order to assist the user in taking action on such emails, we must first understand the types of information managed in such a way and then identify likely actions. For example, if a self-E is a "to-do" item such as "write the report", the system can offer to set a reminder. If the item contains driving directions, it may be ranked preferentially on a mobile device. If the item contains meeting notes, then it might be ranked highly based on similarity with a user's calendar to surface at an appropriate time. If the message is sent immediately before a weekend or vacation, the system may surface it to the top when the user returns. Likewise, other user assistance or filtering might be appropriate depending on whether the email contains receipts, images, attachments or keywords such as "test", etc.

However, before such user aids can be designed we must first understand the spectrum of intents users have when sending a self-E. In this paper, we provide the first such characterization using three data sources: (1) we conduct a survey of self-E behaviour and intents; (2) we analyze self-E behaviour in a publicly available email corpus; (3)

[*]Work performed while at Microsoft Research.
[1]Not to be confused with a selfie photograph!

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

CHIIR '17, March 07-11, 2017, Oslo, Norway
© 2017 ACM. ISBN 978-1-4503-4677-1/17/03…$15.00
DOI: http://dx.doi.org/10.1145/3020165.3020189

we analyze self-Es that were contributed and labeled by the original senders. In addition, we characterize the frequency of such intents and build classifiers to predict a common intent in self-Es: reminder or to-do intent. The research questions we want to address through our analysis are:

(RQ1): **How many email users send self-Es? How often?** In particular, we want to find out the proportion of users that engages in self-E behaviour and how frequently they send emails to themselves.

(RQ2): **Why do users email themselves?** We want to discover what user tasks drive self-E behaviour, and the types of information contained by emails-to-self.

(RQ3): **Can reminder intent be detected in self-Es?** Specifically, can we use email classification techniques to detect self-Es that serve as reminders or to-do items?

2. RELATED WORK

Email has been central to on-line communication for the past two decades. Because of its central role in communication, extensive effort has been allocated to understanding user interactions with email [5, 18, 19]. Extending these efforts, our work brings together two central areas of research on email: (i) information and task management in email and (ii) email classification.

2.1 Information management in email

Email is a critical tool for communication and collaboration. Whittaker and Sidner [19] studied how users manage their email. They used the term *email overload* to refer to the phenomenon whereby users are overwhelmed by email. They show that even though email was designed for communication, it is actually being used for other tasks, like task management (e.g., preserving task context and monitoring task progress) and organizing longer term information (personal archiving). More recent work has shown that users receive and retain more email [8] reinforcing email's role as a digital archive.

Prior work investigated the use of email for task management [18]. The authors discuss how people not only use emails to communicate about tasks but also send themselves email to put messages in the inbox as reminders and links to useful information. An in-depth look at email as a task management tool is provided in Bellotti et al.[1], in which the authors report on a multi-phase study of email users.

Our work is similar to this line of research in that we focus on email usage for task management. Our work extends this line of work by focusing on characterizing and understanding the different intents of self-addressed emails. The characterization is done using multiple data sources including surveys, emails in a publicly available dataset and emails labelled and contributed by the original sender.

2.2 Email classification

Carvalho and Cohen [5] considered classification of email messages into a set of "speech acts" in email, such as a request or a commitment. They showed that using text features and correlation among email messages in the same thread can improve email-act classification. Sappelli et al. [15] proposed a taxonomy of tasks that are expressed through email messages. They manually annotated two email datasets and evaluated the validity of the dimensions in the taxonomy in addition to investigating the potential for automatic e-mail classification.

Corston-Oliver et al. [17] developed a system for automatically identifying action items in email messages. They used the system to produce a task-focused summary of a message that consists of a list of action items extracted from the message. Bennett and Carbonell [2] also studied the task of action item detection in email. They report that using enriched feature sets, such as n-grams, and contextual cues improves performance by up to 10% over bag-of-word features. Shen et al. [16] predicted tasks associated with an incoming email by leveraging email sender, recipients, and distinct subject words. They found the body words to not provide additional prediction value.

Prior research has also looked into grouping emails into a set of activities. Kushmerick and Lau [10] formalized e-commerce activities as finite-state automata, where transitions among states represent messages sent between participants. Dredze et al. [7] used user generated activity labels and classified emails into activities using overlapping participants and content similarity. Qadir et al. [14] introduced a latent activity model for workplace emails. They posed the problem as probabilistic inference in graphical models that jointly capture the interplay between latent activities and the email contexts they govern, such as the recipients, subject and body.

Similar to this line of work, we also build email classification models. Our work is different in that we focus on predicting a common intent in self-Es: reminder or to-do intent. This is one of the most commonly expressed intents in self-addressed emails and identifying it can enable interesting user assistance scenarios. Additionally, since emails-to-self need not follow communication conventions as those emails sent to others (e.g., they may be brief and lack textual context cues), it is an interesting question if they can be automatically classified.

3. DATA AND METHODS

To conduct our analysis, we make use of three complementary sources of data. First, we consider the information available via a public email corpus named "Avocado" . This helps ground our research in a corpus that others can easily access for reproducibility and helps us with an initial perspective on self-E behaviour. However, given the emails in the corpus are from approximately 15 years ago, it is possible that email behaviours have changed. Therefore, we also conducted an email survey and acquired donations of recent self-addressed emails to understand self-E behaviour in a more modern setting and collect explicit intent labels from our study participants – labels which can help us probe self-E motivations more deeply. In this section, we describe the basic characteristics of these data sources. In Section 4, we discuss the implications of each for our research questions on self-E behaviour and intents.

3.1 Avocado

Our first source of data to understand self-E behaviour is a publicly available email corpus. The Avocado Research Email Collection [13] consists of emails from 279 accounts of a now defunct technology company, referred to as "Avocado", that was active in the early 2000s. From this collection, we selected all accounts that belonged to individual people; that is, the accounts were neither shared accounts

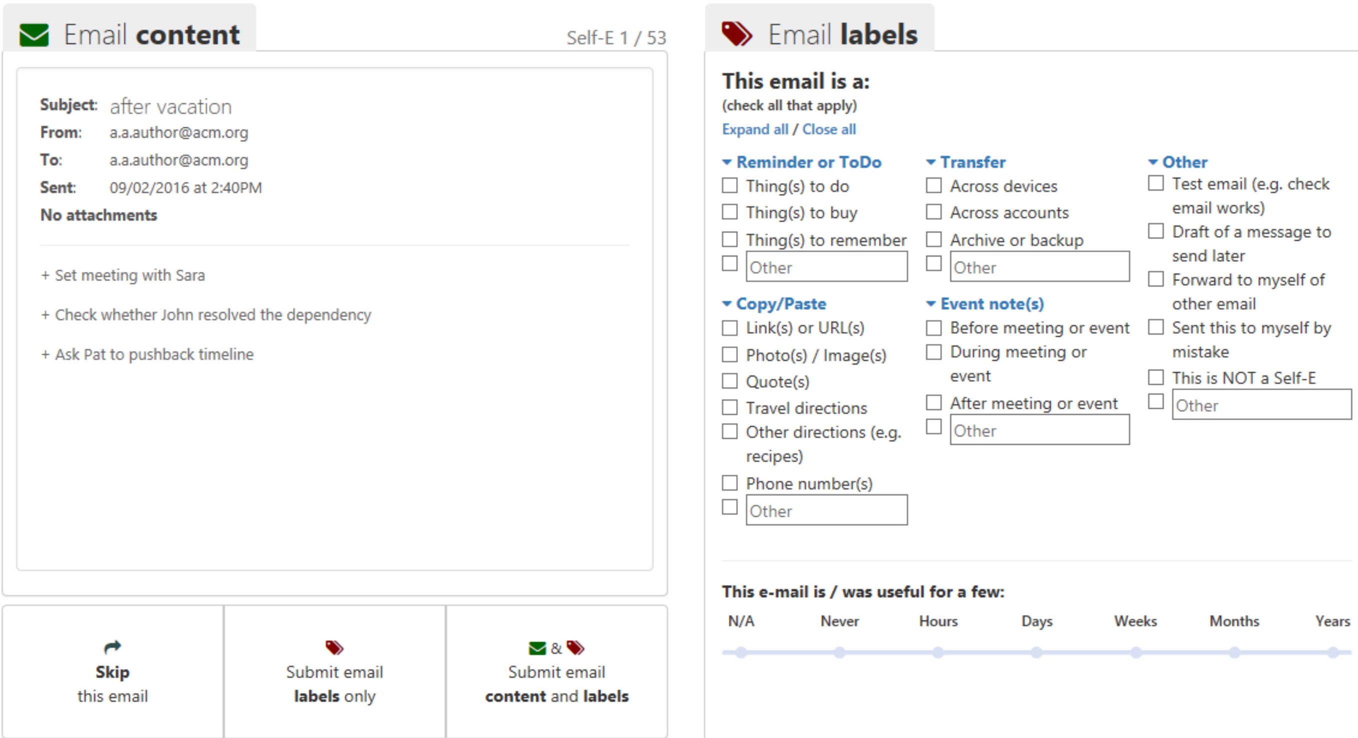


Figure 2: Part of the labelling interface we used to collect self-E donations from users in our organisation. Email content and details are displayed in the panel on the left, email labels are displayed in the panel on the right.

nor system accounts (such as mailing lists or conference room accounts). To ensure a sufficient sample size of email activity, from the accounts that belonged to individual people, we selected all those accounts that were active for at least 50 calendar days, where activity is determined by having at least one email of any type sent or received on a given day. These criteria resulted in 88 email accounts belonging to individual people with sufficient overall email activity. For these accounts, we analysed their entire sent folders for self-E behaviour. Those folders contain 110K (one hundred ten thousand) emails sent to any address(es) by these 88 users.

3.2 Survey

As mentioned above, given that most emails in the Avocado collection were sent roughly 15 years ago, it is possible that e-mail behaviour has changed in the interim. As a more current way of identifying the information needs that underlie self-E behaviour, we conducted a survey to look at both self-E and general email practices. Our main objective was to uncover the types of information that users include in their self-addressed emails. The survey was distributed to a random set of employees within Microsoft who were based in the USA, and responses were received from a total of 238 people: 73.7% ($n = 160$) male, 86.6% ($n = 206$) between 25 and 54 years of age, 77.7% ($n = 174$) had a bachelor's degree or higher. Our respondents held various positions in our organisation, with the two largest groups being "Software Developer" (29.4%, $n = 64$) and "Program Manager" (22.9%, $n = 50$) – other occupations including human resources, sales, marketing and design professionals.

Our survey was structured into four separate sections: in the first section, we asked our respondents to report on general aspects of their use of email, such as the number of accounts they use for personal and work email, the number of emails they send daily from their accounts, and their usage of email management tools (e.g. email folders, marking items read or unread, zero inbox). In the second section, we asked our respondents to think back on the last self-addressed email they sent, and describe, in as much detail as possible, the intent behind their most recent self-E, the task that their self-E helped with and the context they were in when they sent it. In the third section, we asked respondents to report on their general self-E behaviour, how often they send themselves emails, and with what intent. Lastly, we asked a series of demographic questions which were optional. Throughout our survey, we collected intent labels by displaying a set of predefined labels to our respondents, as well as allowing free-form text input for user-defined labels. The predefined labels we used were broken down into five high-level categories, with each category containing several subcategories, as shown in Table 1. The intent categories were initially developed by manually inspecting self-Es we identified in Avocado; we defined the final set of intent categories and subcategories after several iterations in which we tested our survey design, using small samples of respondents – results from those pilot iterations are not included in the analysis we present here.

3.3 Labelling tool

Finally, part of our effort to understand self-E behaviour was concerned with creating a current collection of self-addressed emails. Similar to the survey described in Section 3.2, this has the advantage of measuring current practice.

Intent category	Intent subcategory
Reminder or ToDo	Thing(s) to do Thing(s) to buy Thing(s) to remember Other [free-form text input]
Transfer	Across devices Across accounts Archive or backup Other [free-form text input]
Copy/Paste	Link(s) or URL(s) Photo(s) or Image(s) Quote(s) Travel directions Other directions (e.g. recipes) Phone number(s) Other [free-form text input]
Event note(s)	Before meeting or event During meeting or event After meeting or event Other [free-form text input]
Other	Test email (e.g. check email is working) Draft of a message to send later Forward to myself of another email Sent this to myself by mistake This is NOT a Self-E Other [free-form text input]

Table 1: Pre-defined intent labels (categories and subcategories) used for our survey and our labelling tool.

In addition, compared to working with Avocado data, this approach has the advantage that the sender of the email can label their own intent in sending themself a self-E; compared to the survey, the advantage is that this approach does not rely on user memory which might be subject to biases as to the types of self-Es that are recalled. Although privacy concerns prevent us from sharing the self-Es we collected with the external research community, we expect those that reproduce our methodology, in a similar context as ours, will find results consistent with our corpus.

To build this collection, we developed a Web application that allowed users in our organisation to quickly find, label and donate their recent self-Es. The application integrated with the single sign-on mechanism available in our organisation, and after user login, it automatically searched the user's sent item folder of their work email account for self-addressed mail, retrieving their most recent self-Es. The interface, shown in Figure 2, displayed email properties, together with email content and attachments. Users were able to label individual self-Es using the same pre-defined intent categories used in our survey (and discussed in the previous section). In addition to intent labels, they were able to report how long the displayed self-E was useful to them, using a set of pre-defined labels, in order of increasing duration from a *"Few hours"* to a *"Few years"*. In the following sections we refer to these duration labels as self-E lifetime.

Given the privacy concerns that surround sharing email with third parties, in addition to allowing users to donate self-Es and labels, we allowed our users to skip emails they did not want to share or to label self-Es and donate only the labels, without actual self-E content. For all donations, even skipped self-Es, we logged email meta-data, such as email sent date and time, number of characters in subject or body, and the number of attachments. In total, we collected 1274 self-Es, from 101 unique users: 813 with content and labels, 230 with labels only and 231 with just meta-data. The maximum number of donations per user was 88, and the minimum 1; overall, the mean number of contributed self-Es per user was 12.61, with a standard deviation of 13.84.

To have participants use our self-E collection tool, we used the same methodology as for our survey and sent email invitations internally to a random sample of our organisational address book. Please note that there was no overlap between respondents to our survey and users who donated emails using our labelling tool. That is, users who responded to our survey were not invited to use the labelling tool and vice-versa. The data collected using these two methods (survey and labelling tool) represent complementary perspectives on self-E behaviour.

4. RESULTS

4.1 How many users send self-Es?

To answer our first research question (RQ1), we examine all three of our data sources. In particular, we focus on answering this question by characterizing both the percentage of users who email themselves and how often users email themselves.

To estimate the percentage of users who email themselves, we begin by examining the view offered via the set of 88 active users in Avocado. Overall, 81% ($n = 71$) of these users had at least one self-addressed email in their sent folder, with the mean proportion of self-Es in the 110K total sent mails of these users being close to 1%. Thus, we can conclude that the vast majority of users engage in self-E behaviour with an overall prevalence that is small as a percentage but still significant in terms of overall quantity.

Figure 3 shows the cumulative distribution of self-Es in our sample. Examining it, we find that for 25% of the email accounts we analysed, more than 1% of all outgoing email is self-addressed. This can be seen in Figure 3 by noting that the black dashed line crosses the 1% line at 75% (i.e. 25% of users surpass this level of activity). This indicates that there is a reasonable percentage of users (1 in 4) that engage in this behaviour with a higher frequency than average and the mean is not simply dominated by a few outliers.

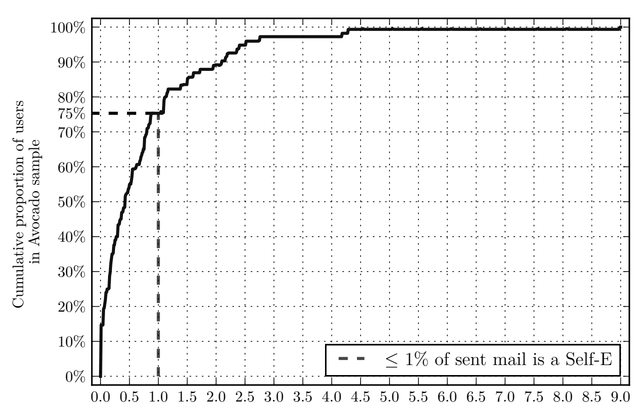


Figure 3: Cumulative distribution of proportion of self-Es per user in our sample of the Avocado email corpus

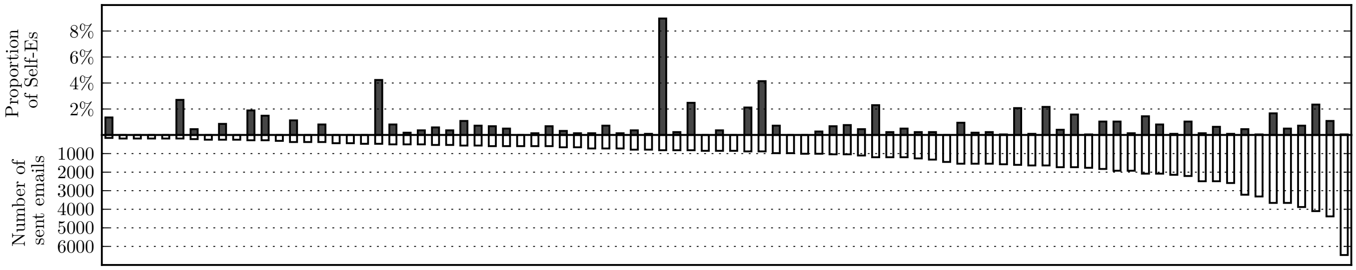


Figure 4: Per user proportion of self-Es in sent mail and sent mail volume in the Avocado corpus. Each pair of bars represents one user in the sample we analysed. The figure is sorted by sent mail volume from the least (*left*) to the most (*right*).

We visualize per user self-E frequency in greater detail in Figure 4. This figure plots both self-E proportion of sent mail and sent mail volume for each of the 88 users. Observe the variation in the height of the dark bars; there is clearly high variance across users with regard to the proportion of sent emails that are self-Es. Furthermore, some users engage in the behaviour more heavily. For example, note in the figure that for one account (who had a total sent mail volume of roughly 1000 mails), more than 8% of all outgoing emails were self-Es! Thus, some small number of users demonstrate this behaviour at a much higher prevalence.

One of the limitations in using the Avocado corpus for self-E analysis is that email habits and practices may have changed since the time period represented by emails in the corpus – approximately 15 years ago. As discussed above, we therefore supplement our understanding from this publicly available corpus by both conducting a survey and collecting email donations. The remainder of the paper focuses on these sources of data.

With regard to the survey, respondents reported whether they sent emails to themselves, how often, and with what intent. Figure 6a shows that almost all respondents to our survey (92%, $n = 219$) have sent at least one email to themselves. In addition, we asked whether they save information that is useful only to themselves in their email accounts, with 81.1% ($n = 193$) reporting that they habitually save useful information in their email. Figure 5 shows that the majority preferred method of saving useful information in email accounts is sending self-Es, namely 71.4% of respondents ($n = 170$) reporting that they regularly save information in their email accounts by sending self-Es.

Overall, our findings show that self-E behaviour is frequent among users of email, ranging from 81% of users in the Avocado email collection to 92% of respondents to our survey having sent at least one self-E. Even more, for those who use email to manage useful information other than traditional email messages, self-addressed emails are preferred over other methods of storing information in email accounts, which suggests that for a large proportion of email users, self-E behaviour has become integrated into general workflows.

How often do users send self-Es?

In addition to measuring the proportion of users with self-E behaviour and proportion relative to all sent mail, another of our aims was to understand how often users send themselves emails. To this end, we asked our survey participants who reported having sent at least one self-E ($n = 219$) to describe how long ago they sent this self-E (figure 6b) and how often they send self-Es in general (figure 6c). The majority of answers reported recent self-Es having been sent: either on the same day as the survey (*"Today"* 13.7%, $n = 30$) or a few days before the survey (*"Few days ago"* 48.9%, $n = 107$); less than 12% of respondents reported their most recent self-E as having been sent a *"Few months ago"* or earlier. With regard to general self-E behaviour, our survey answers suggest that the majority of respondents that reported having sent at least one self-E, send themselves emails *"Several times per month"* (37.4%, $n = 82$) or *"Several times per week"* (24.7%, $n = 54$).

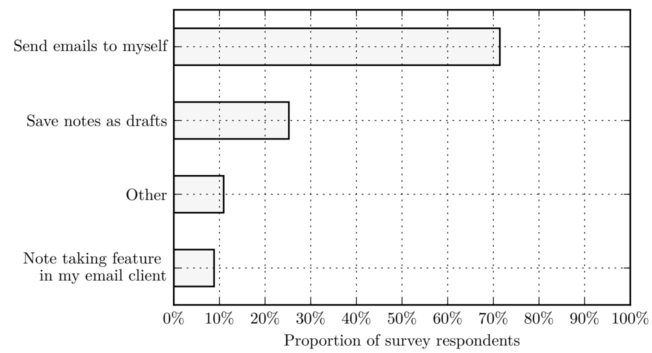
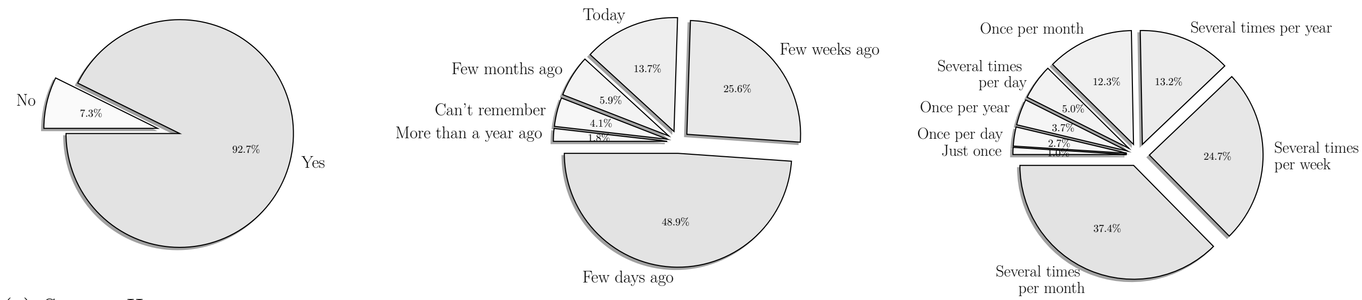


Figure 5: Survey: How do you save information that is useful only to you in your email account?

To examine how often users send self-Es using the donated self-Es described in Section 3.3, for each user we estimate the average rate of self-Es by dividing the number of self-Es they donated by the duration in weeks between their most recent self-E donated to us, and their oldest self-E, as determined by email sent date. Figure 7 shows the distribution of the average number of self-Es per user, with approximately 65% of users sending one or more self-Es per week. This is in general agreement with the survey data where 62.6% reported their last self-E was within a week and 32.4% reported a frequency of at least two self-Es per week. Thus multiple data sources indicate that although the proportion of sent mails may be small, sending self-Es is a regular weekly habit for the majority of users – and can even be a daily habit for some users. This is supported both by 7.7% ($n = 17$)



(a) Survey: Have you ever sent an email to yourself, and only yourself, containing information that is useful to you?

(b) Survey: Approximately how long ago did you send yourself your most recent Self-E?

(c) Survey: How often do you send yourself self-Es?

Figure 6: Proportion and recency of sending self-Es in a survey of 238 people at a large technology corporation.

of survey respondents who reported sending self-Es once or several times per day, as well as the information in figure 7 which shows that roughly 8% of email donors have an average weekly self-E rate of 8 or more self-Es. This highlights, from different perspectives, a substantial population of users that is active daily in their interaction with self-Es.

4.2 Why do users email themselves?

Next, we turn to our second research question (RQ2), why do users email themselves? In particular, we wanted to understand the tasks and information needs that drive self-E behaviour. In both our survey and our self-E labelling tool, we asked participants to describe the intents behind their self-Es. In our survey, we asked respondents to describe the intent behind (i) their most recent self-E and (ii) the intents behind their self-Es in general. In our labelling tool, participants were able to directly attach intent labels to their own self-Es. In both the survey and labeling tool, we allowed users to select multiple intents from a set of predefined labels (e.g., "Things to do", "Things to buy") but also allowed them to define their own intent labels as well. Since multiple labels were allowed for the same self-E, when we normalize by the number of total labels (greater than or equal to the number of mails), we refer to it as a percentage of labels (sums to 1 across intent labels). When we normalize by the number of mails we refer to it as a percentage of mails (sums to more than 1 across intent labels).

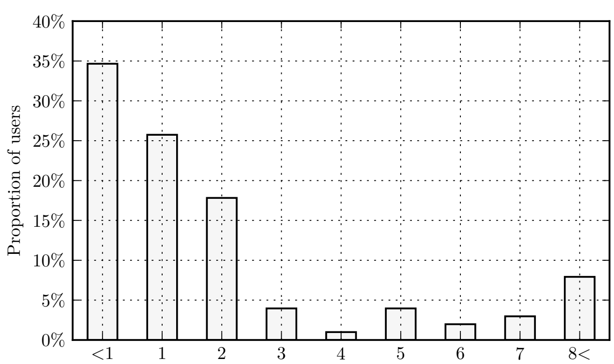


Figure 7: Average number of self-Es per week

Figure 8 shows the percentage of intent labels from the labelling tool (on the *left*) as well as the absolute differences from this distribution, observed in our two survey questions regarding self-E intent (on the *right*). When the survey differences are above zero, it means that survey respondents report a higher percentage of those intent labels than observed when donations are collected. Likewise below zero, it means that survey users report a lower percentage of those intent labels when compared to donations. These differences may be explained by a tendency to have a bias toward certain intents in memory when asked to recall a specific instance – while this aspect bears further investigation, we note that the relative ranking of the intent categories is consistent for both survey and donations: figure 9 shows the proportion of intent labels in both collected self-Es and survey answers, ranked by frequency in the donated self-E collection.

In particular, the most popular intents, for both labelled self-Es and survey answers, are *"Things to do / remember"*, *"File transfer across devices"* and *"Link(s) or URL(s)"*, with *"Things to do / remember"* being the most popular overall. Within the donations, approximately 37% of the labels are *"Reminder or ToDo"* category, and as a percentage of mails, 53% of all donated self-Es had some form of reminder or to-do intent label attached. Among the *"Other"* categories of intent labels, which were free-form text, users reported sending themselves test emails, passwords, photos of receipts, code snippets or mistakenly sending themselves emails. Overall, these distributions emphasize that while "Reminders or To-Dos" form an important large subclass of self-Es they are still only slightly more than half of self-Es and that the remainder captures a rich set of other information management practices where the appropriate automated user assistance may differ by intent.

What types of tasks drive self-Es?

To further understand both the breadth and complexity of self-E intents, one of the questions in our survey asked respondents to describe their most recent self-E in as much detail as possible, including information about why they sent the email to themselves, as well as the context they were in. In this section, we reproduce some of the answers to that question, in order to better illustrate the tasks and intents that appear to drive self-E behaviour, as described by the respondents to our survey. As discussed in the previous section, task management, in particular reminders and to-dos,

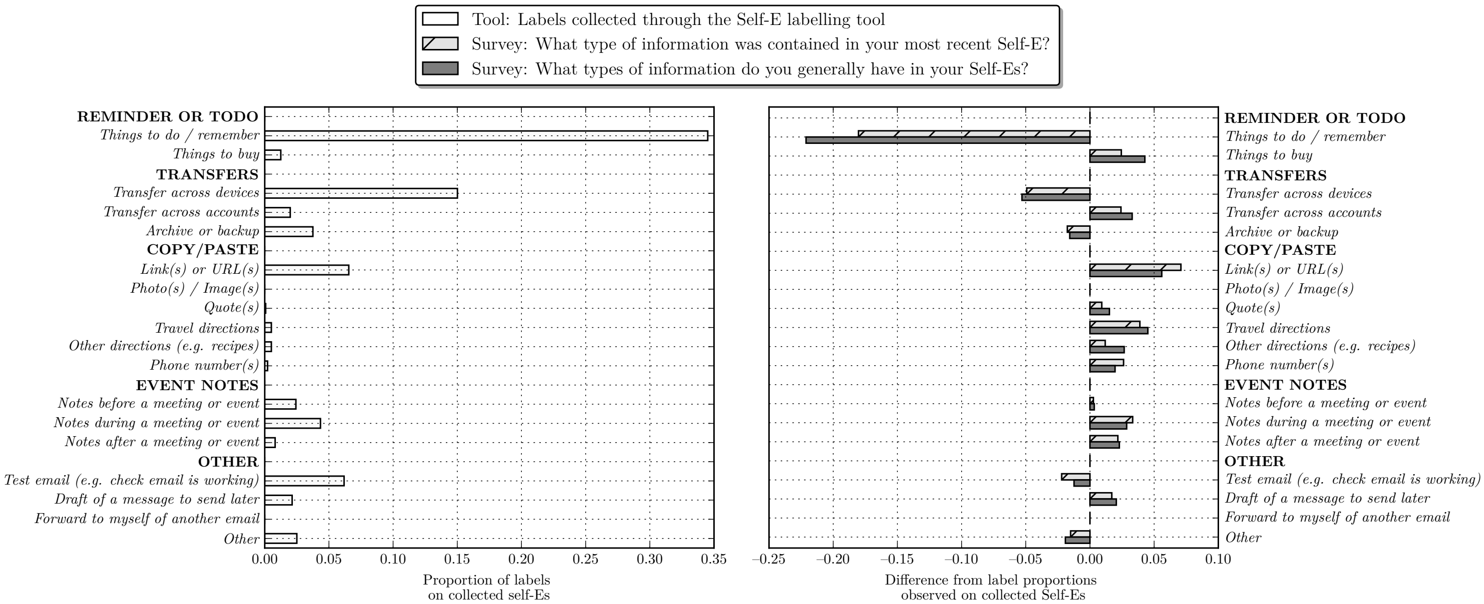


Figure 8: *Self-e* intent labels based on donated emails (left) and survey answers (right)

drives a large proportion of self-E behaviour; in the words of our survey respondents:

1. [My recent self-E contained] *A bullet-pointed list of ideas, tasks and reminders that I sent myself from my phone so that I remembered to break those out into Outlook tasks and later.*
2. *I always send mail to myself when I have some information that I need to save for later use and I am not sure where will I be when I will need to use this information. This mostly happens for personal scenarios and rarely for work scenarios.*
3. *I copied an image from a web site to remind myself of a To Do in a week's time. The To Do is to purchase something, which has a validity date on it and doesn't become valid for another week. I had just arrived back home from vacation and was checking my email when I logged this as a To Do.*
4. *I leverage flagged email option to maintain my TODO list. If I need to add a task for myself by myself, I send an email to myself. I like one place to track every thing whether it is email, follow up or tasks. In my role and most of the management role, email or sometime calendar activities is the main task tracker.*

It is interesting to note that users have integrated some self-E intents with existing information management tools (e.g. *"Outlook tasks"*, *"flagged email"*) and extend their usefulness. In addition, the answers we received describe self-Es as a tool for cross-boundary self-communication: whether across devices (e.g. *"sent myself from my phone"* [to Outlook]), across spaces (e.g. *"not sure where I will be"*) or across time (e.g. *"To Do in a week's time"*). This suggests the ubiquity of email, combined with its familiarity of use, could be further leveraged to develop novel email client features which further support self-E intents.

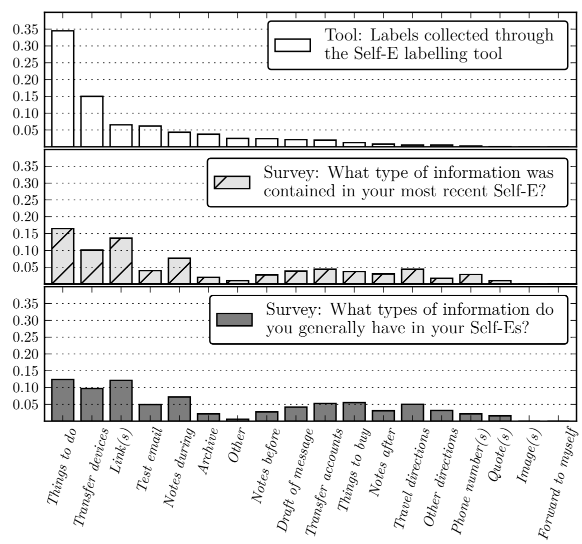


Figure 9: Absolute proportion of *Self-e* intent labels in donated emails (top) and survey answers (middle and bottom). Ordered by label frequency in donated emails (top). Label names are truncated.

How long are self-Es useful after being sent?

In addition to intent labels, we asked our participants to attach lifetime labels to the self-Es they donated. Specifically, users were asked to answer the following question: "How long is or was this self-e useful for you?", using a set of predefined answers as lifetime labels for the self-Es they donated. Self-E lifetime is interesting to explore because it has implications for search or the proactive display of self-Es. For example, knowing how long a self-E is useful can be used to keep that self-E displayed at the top of email stacks, or in task management interfaces, only for the duration of their usefulness. Figure 10 shows the distribution of duration labels over intent labels (each row sums to 100%).

It is interesting to note that both *"Reminders and ToDos"* and *"Copy/Paste"* self-E lifetimes have heavier distribution tails, with these types of self-Es being useful up to a few weeks or months. This suggests they are partially used as archiving devices for various pieces of information. If an email client enhanced these types of self-E with appropriate meta-data – e.g. indicating a significant portion of a self-E's text originated from a copy/paste operation – later rank operations might be able to promote such self-Es in response to user searches, since their longer lifetimes imply they are more likely to be searched for in the future.

Overall, the majority of self-Es have relatively short lifetimes, from a few hours to a few days, which suggests that most self-Es, across intents, are used for immediate or short-running tasks, rather than longer-running ones. In contrast to the other categories of intents, the "Other" self-Es that were labelled as never useful were mostly *"Test emails"* to check email delivery or correct email formatting; even though these emails achieve their purpose, they are generally perceived as never useful. As mentioned above, this has implications for email client design, suggesting that features centred around emails-to-self should be aware of self-E pro-actively rank or display self-Es based on their age and intent.

4.3 Can reminder intent be detected in self-Es?

Given how frequent the *"Reminders or ToDo"* intent is among the emails we collected, we set out to build a classifier that detects reminder self-Es. In particular, we wanted to classify self-Es under the broad category of *"Reminder or ToDo"* items, and not its subclasses (*"Things to do"*, *"Things to remember"* or *"Things to buy"*). From a systems perspective, being able to detect a user's need to be reminded of a to-do self-E can be leveraged for proactive notifications or better email search results. From a research point of view, identifying intent within less formally composed emails is an interesting and challenging problem. Even though others have looked at predicting "to-do" intent in email [16], prior work has focused on detecting "to-do" intent in general *person-to-person* email, which tends to follow

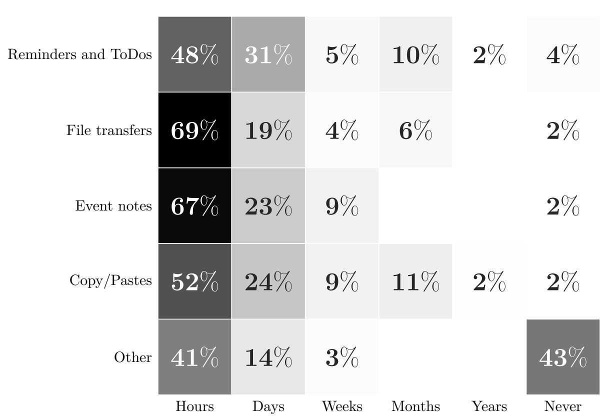


Figure 10: Labelling tool: How long was this self-E useful?

Feature set	**Mean accuracy**	
	Flat	Stacked
All features	0.651 ($\pm$0.046)	0.786 ($\pm$0.041)
Content features only	0.606 ($\pm$0.067)	0.747 ($\pm$0.048)
Metadata features only	0.619 ($\pm$0.074)	0.641 ($\pm$0.057)

Table 2: Mean classification accuracy and standard deviation over 5 fold cross-validation. For the stacked approach, "All features" significantly improves classification accuracy over both "Content features only" ($t(8) = 2.709, p < 0.05$) and "Metadata features only" ($t(8) = 9.006, p < 0.01$) – determined using an unpaired, one-tail t-test, corrected for multiple comparisons. For the flat approach, there are no significant differences between feature sets.

more formal conversational conventions (e.g. *"Will send you the PDF tonight"*) as opposed to self-Es (e.g. *"PDF Alex"*).

Because we wanted to use text-based features for our classification approach, we use only the donated self-Es that include both content and labels ($n = 813$ emails). The self-Es donated with content *and* labels included 48% ($n = 388$) *"Reminder or ToDo"* self-Es and a roughly equal proportion of other categories (52%, $n = 425$). Thus, our data set is nearly naturally balanced. We report accuracy and area under the ROC curve. We note that random performance would achieve 52% accuracy and 0.5 area under the ROC curve.

We experimented with different feature sets: (i) bag-of-words representation of self-E subject and body, which we refer to as content features, (ii) features extracted from email metadata and header properties, such as the number and type of attachments, sent date and time (e.g. sent in the morning, sent in the weekend), which we refer to as metadata features and (iii) a combination of the previous, which we refer to as mixed features. For our text features, we used simple term frequency.

We could have simply learned a binary classifier by treating all labels that are not "Reminder or ToDo" as negatives. This approach yielded an accuracy better than random using all features (65.1% $\pm$ 4.6%) but we desired to do better possibly by leveraging the other labels beyond simply treating them as negatives. To do so, we could have used a structured prediction approach [9, 11] to jointly predict all labels even though our only true target was the "Reminder or ToDo" class. Instead, we leveraged information from the other labels by initially predicting each of the other high-level self-E intent categories (*"Transfers"*, *"Copy/Paste"*, *"Event Notes"*, and *"Other"*) and using those predictions as features to the final classifier on *"Reminder or ToDo"* prediction. By using nested cross-validation on the training set, we ensure the training data is kept free of bias and we never train on the test data.

This is similar to the refined expert approach of Bennett & Nguyen [3] where a stacking approach [4, 20] is extended with related predictions and draws from the multi-task learning literature [12] where learning a set of related tasks may simplify learning the final target. Table 2 shows the mean accuracy (and standard deviation) of our classifier over

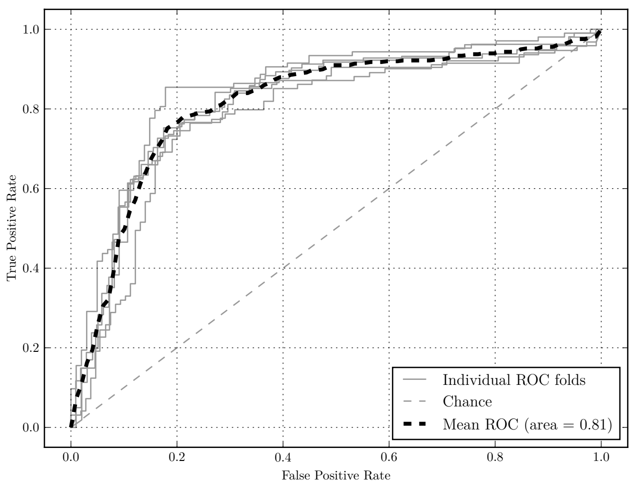


Figure 11: "Reminder or ToDo" intent classifier, using the stacked approach: ROC over 5 fold cross-validation using "All features".

stratified[2] five-fold cross-validation using both the simple (flat) approach and the stacked approach with predictions for each of the related tasks. Note that predicting the related targets using "All features" set yields a significant gain relative to simple binary prediction (78.6% accuracy versus 65.1%; statistical significance determined using an unpaired, one-tail t-test on classifier accuracy scores over 5 random folds of the data: $t(8) = 5.396$, $p < 0.01$). This indicates that there is utility in collecting the other labels even if "Reminder or ToDo" is the final target. Furthermore, comparing the feature sets reported in the table, we see that the content features provide the most value. However, using both content and meta-data features improves beyond content alone.

Finally, figure 11 displays the receiver operating characteristic (ROC) curves for each of the folds using all features, as well as the mean ROC curve, and the area under the curve (see legend). Overall, our classifier performs well, with a mean accuracy of 78.6% self-Es being correctly identified as *"Reminder or ToDo"* items or not. Best performance, in terms of accuracy, is achieved when using both text and email features. Given the small amount of data we had access to, this is an encouraging first step in classifying reminder-oriented self-Es, and we speculate that more data would significantly improve this classification approach.

5. CONCLUSIONS

In order to assist users in taking action on self-addressed emails, we must first understand the types of information managed through self-Es, and identify likely actions users take on this type of mail. Understanding the intent behind self-Es can allow systems to pro-actively support user interaction with self-addressed emails. In this paper we present the first complete characterization of email as a person-to-self communication device. Especially in the context of task management, other authors [1, 6, 18] have noted the use of emails to oneself (and only oneself) before. However, the literature was lacking a characterization of the proportion of users that engage in this self-E behaviour, the frequency with which they do it, and the full spectrum of the types of intents captured by such self-Es. Our analysis from several data sources suggests that the vast majority of users, 81% to 92%, engage in this behaviour at some time and that 32.4% to 40% of users send two or more self-Es per week. Furthermore, sending self-Es is a part of regular behavior for a substantial percentage of users, with approximately 8% of users sending self-Es on a daily basis, on average. Even though in terms of sent mail, self-Es form just 1% of all sent mails, they are part of a regular weekly pattern for the majority of users and are used at least monthly by around 3 out of 4 users.

Furthermore, our analysis illustrated that while the *"Reminder or ToDo"* intent is most common, self-Es actually have a broad spectrum of intents. Given this variety, providing the appropriate support may be different depending on self-E intent. For example, pro-actively surfacing *"Reminder or ToDo"* self-Es in a few days after they are sent to ensure they are not lost in a sudden inbox surge of mails from others may be the best support. In addition, *"Copy/Paste"* intents which constitute verbatim information such as URLs, images, quotes, travel directions, other directions, and phone numbers tend to have a longer lifespan than most other types of self-Es, which implies they are also more likely than other self-Es to be searched for in the future. Augmenting self-Es with longer lifespan with metadata – for instance, in the case of *"Copy/Paste"* self-Es, explicitly flagging copy/paste content – could enable search to distinguish these intents from other self-Es and promote them in searches.

Finally, given past work on extracting "ToDos", reminders or action items from mails to others, it is an interesting question of whether "Reminder or ToDo" intent self-Es can be distinguished from other types of intent. This may be an especially challenging problem since self-Es may be briefer and lack formal grammar and conventions that help distinguish such speech acts in mails to others. We demonstrate that using features of the email body, meta-data and predictions of belonging to other self-E intents that we can distinguish this special class of self-Es well above random. This bodes well for future technology that aims to support "Reminder or ToDo" intents specifically.

6. REFERENCES

[1] V. Bellotti, N. Ducheneaut, M. Howard, and I. Smith. Taking email to task: The design and evaluation of a task management centered email tool. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '03, pages 345–352, New York, NY, USA, 2003. ACM.

[2] P. N. Bennett and J. Carbonell. Detecting action-items in e-mail. In *Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '05, pages 585–586, New York, NY, USA, 2005. ACM.

[3] P. N. Bennett and N. Nguyen. Refined experts: Improving classification in large taxonomies. In *Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '09, pages 11–18, New York, NY, USA, 2009. ACM.

[2]Each fold preserved the proportion of samples for each class.

[4] L. Breiman. Stacked regressions. *Machine Learning*, 24(1):49–64, 1996.

[5] V. R. Carvalho and W. W. Cohen. On the collective classification of email "speech acts". In *Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '05, pages 345–352, New York, NY, USA, 2005. ACM.

[6] M. E. Cecchinato, A. Sellen, M. Shokouhi, and G. Smyth. Finding email in a multi-account, multi-device world. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, pages 1200–1210, New York, NY, USA, 2016. ACM.

[7] M. Dredze, H. M. Wallach, D. Puller, and F. Pereira. Generating summary keywords for emails using topics. In *Proceedings of the 13th International Conference on Intelligent User Interfaces*, IUI '08, pages 199–206, New York, NY, USA, 2008. ACM.

[8] B. V. Hanrahan, M. A. Pérez-Quiñones, and D. Martin. Attending to email. *Interacting with Computers*, 2014.

[9] T. Joachims, T. Hofmann, Y. Yue, and C.-N. Yu. Predicting structured objects with support vector machines. *Communications of the ACM, Research Highlight*, 52(11):97–104, November 2009.

[10] N. Kushmerick and T. Lau. Automated email activity management: An unsupervised learning approach. In *Proceedings of the 10th International Conference on Intelligent User Interfaces*, IUI '05, pages 67–74, New York, NY, USA, 2005. ACM.

[11] J. D. Lafferty, A. McCallum, and F. C. N. Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proceedings of the 18th International Conference on Machine Learning*, ICML '01, pages 282–289, San Francisco, CA, USA, 2001. Morgan Kaufmann Publishers Inc.

[12] R. Michalski and G. Tecuci, editors. *Machine Learning: A Multistrategy Approach*, volume IV. Morgan Kaufmann Publishers, Inc., 1994.

[13] D. Oard. Avocado research email collection. *Philadelphia: Linguistic Data Consortium*, 2015.

[14] A. Qadir, M. Gamon, P. Pantel, and A. H. Awadallah. Activity modeling in email. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1452–1462, San Diego, California, June 2016. Association for Computational Linguistics.

[15] M. Sappelli, G. Pasi, S. Verberne, M. de Boer, and W. Kraaij. Assessing e-mail intent and tasks in e-mail messages. *Inf. Sci.*, 358(C):1–17, Sept. 2016.

[16] J. Shen, L. Li, T. G. Dietterich, and J. L. Herlocker. A hybrid learning system for recognizing user tasks from desktop activities and email messages. In *Proceedings of the 11th International Conference on Intelligent User Interfaces*, IUI '06, pages 86–92, New York, NY, USA, 2006. ACM.

[17] M. G. R. C. Simon H. Corston-Oliver, Eric Ringger. Task-focused summarization of email. Association for Computational Linguistics, July 2004.

[18] S. Whittaker, V. Bellotti, and J. Gwizdka. Email in personal information management. *Commun. ACM*, 49(1):68–73, Jan. 2006.

[19] S. Whittaker and C. Sidner. Email overload: Exploring personal information management of email. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '96, pages 276–283, New York, NY, USA, 1996. ACM.

[20] D. H. Wolpert. Stacked generalization. *Neural networks*, 5(2):241–259, 1992.