Characterizing Email Search using Large-scale Behavioral Logs and Surveys

Qingyao Ai[1], Susan T. Dumais[2], Nick Craswell[2], Dan Liebling[2]
[1]University of Massachusetts Amherst, MA, USA
aiqy@cs.umass.edu
[2]Microsoft Research, Redmond, WA, USA
{sdumais, nickcr, danl}@microsoft.com

ABSTRACT

As the number of email users and messages continues to grow, search is becoming more important for finding information in personal archives. In spite of its importance, email search is much less studied than web search, particularly using large-scale behavioral log analysis. In this paper we report the results of a large-scale log analysis of email search and complement this with a survey to better understand email search intent and success. We characterize email search behaviors and highlight differences from web search. When searching for email, people know many attributes about what they are looking for; they often look for specific known items; their queries are shorter and they click on fewer items than in web search. Although repeat queries are common in both email and web search, repeat visits to the same search result are much less common in email search suggesting that the same query is used for different search intents over time. We consider search intent from multiple angles. In email search logs, we find that people use email search not just to find information but also to perform tasks such as cleanup or organization, and that the distribution of actions they perform depends on the type of query. In our survey, people reported that they looked for specific information in both email search and web search, but they were much less likely to search for general information on a topic in email. The differences in overall behavior, re-finding patterns and search intents we observed between email and web search have important implications for the design of email search algorithms and interfaces.

Keywords

Email Search; Re-finding; User Evaluation; Query Log Analysis

1. INTRODUCTION

Email has been a key communication and productivity tool for decades. Despite the proliferation of alternative communication methods such as instant messaging and social networks, 78% of American office workers maintain that email is "very important" for their jobs [24]. However, workers suffer from "email overload" [15] as email volumes continue to grow. Email clients, from legacy UNIX tools like Elm to web sites like Gmail or outlook.com, offer affordances for organizing email. Folders allow users to impose some structure on their email. Labels, colors, and flags allow visual semantic categorization without the drawback of a strict folder structure. All of these measures require the user to "stick with" their organizational plan. However, evidence from Yahoo! shows that most Yahoo! mail users have no folders and of those that do, less than 10% actively use them [14]. Although email search in many forms has existed for a long time, Google's introduction of Gmail in 2004 provided a popular interface where search was the primary organizational method. At the time of its introduction, Gmail intentionally did not provide a mechanism to delete mail. Most modern web mail clients offer search features, yet email search is understudied compared to web search.

Just as web search differs from traditional library search, email search differs from web search. Web search engines cover billions of documents, whereas mail clients top out on the order of 10^5 messages. Structurally, web search engines use hyperlink structure between documents and behavioral click data to rank results. Email corpora are personal, with few explicit links and no aggregate clicks across users. In addition, email messages (unlike web pages) provide structured metadata like sender address, subject, and timestamps which provide important cues to organize search results.

Most prior research on email search uses small-scale qualitative studies or lab-based experiments. These studies provide insight into the organizational and information finding practices, but typically lack the breadth that large-scale studies from log data provide. Notable exceptions include *Stuff I've Seen* [10], a document and email search client, deployed to hundreds of users, and recent work by Carmel et al. [3] who studied relevance ranking of email using a few thousand opt-in users. However, neither of them provide a comprehensive overview of both how users express email information needs and the actions they take on found items.

In this paper, we conduct a large-scale log analysis of email search interactions using the Outlook web client. These data enable us to identify similarities and differences from web search in query formulation and in re-finding. To gain initial insights about email search intents, we use log analyses to analyze the actions that people take on the emails they retrieve. We also use a survey to understand how people

WWW 2017, April 3–7, 2017, Perth, Australia.
ACM 978-1-4503-4913-0/17/04.
http://dx.doi.org/10.1145/3038912.3052615

search email and the web and what they know about items they search for. Finally, We discuss how this knowledge can be used to improve the design of email search interfaces.

2. RELATED WORK

Several areas of prior research are relevant to the results presented in this paper, including general research on email practices, email search, re-finding and search intents.

Email and Email Search. Email has been one of the most common communication and personal information management technologies for decades, and it quickly went far beyond its original purpose for asynchronous communications [16]. Much of the early research on email focused on how people organized and managed their email. Whittaker and Sidner [35] proposed the concept of *email overload* to describe the usage of emails beyond communication needs, such as task management and personal archiving. They identified three common strategies for handling email overload: *no filers* who forego using folders by relying instead on browsing or search, *frequent filers* who minimize the number of messages in their inbox by putting emails in folders, and *spring cleaners* who organize their mail only occasionally. A decade later, Fisher et al. [13] found similar results in their study of mailboxes at a large tech company. More recently, email management strategies have been described in terms of *cleaners/keepers* by Gwizdka [16].

With the increase of email accounts and messages over time, email management has become more of a challenge [8, 19, 15, 3]. In the context of web email, studies from Koren et al. [20] and Grbovic et al. [14] show that 70% of users do not define any folders and 90% of the rest do not actually use folders that they create. They argue that search is an increasingly important alternative to human-generated folders and tags. Whittaker et al. [34] revisited the email overload problem by examining the relationship between email management strategies and subsequent retrieval strategies. They noticed that some users expended *preparatory* effort to create complex folder structures to promote re-finding while others were more *opportunistic* and used search and threading to reduce the need to file manually. In a log study of 345 users, they found that people who create complex folder structures did indeed rely on them for retrieval, but that search and threading provided more effective finding.

Several studies have focused on developing effective search systems for email [10, 7, 26]. Dumais et al.'s *Stuff I've Seen* system [10] provided unified access to desktop information including emails, files, web history, etc. In their study of 235 people over several months, they found that email was the most commonly retrieved source of information. Although their search interface provided a rich list view of search results that enabled people to sort by different email attributes or a best-match score based on BM25, people sorted by date most often even when the default was best-match ranking. They reported that people-centric queries were important in email search – approximately 20% of queries contained a person's name or email address. Cutrell et al. [7] reported similar findings from their *Phlat* system.

Some researchers have looked at developing improved ranking models [6, 23, 33, 36, 3]. Using data from the TREC Enterprise track [5], Craswell et al. [6] combined email metadata with email content using BM25F. Ogilvie and Callan [23] took a language modeling approach to combine evidence from the text of the message, the subject, other messages in the thread, and messages that are in reply to the message. Weerkamp et al. [33] explored incorporating thread, mailing list, and community content levels for email ranking. Recently, Carmel et al. [3], developed a two-stage email ranking technique that first matched messages with query keywords and then used a learning-to-rank approach to combine a set of message-specific features. They used an offline evaluation with Yahoo! web email service using 100K queries from 10K opt-in users from enterprise and web emails, but did not report user analysis from online usage.

Re-finding Behavior. Large scale analysis of re-finding behavior has proven beneficial for understanding user intents in web search. Teevan et al. [31] examined queries issued to Yahoo! for 117 people over a year and analyzed re-finding behavior. They found that 39% of queries exhibit re-finding behavior (i.e., the searcher clicked on results that they had previously found). A subset of 24% of queries were repeated queries that result in a click on the same URL, so called *navigational* queries. Further, Tyler and Teevan [32] looked more deeply at re-finding queries and identified several interesting phenomena. For example, re-finding queries tend to converge when a person repeatedly uses a search engine to find the same result, and there are different patterns of re-finding within and across sessions. Harvey and Elsweiler [18] replicated Teevan et al.'s [31] analysis for email rather than web searches. They analyzed log data from 47 people who opted-in to use a browser extension over 4 months [11]. Different from Teevan et al. [31], they found more repeat queries and fewer repeat clicks, but whether their results will hold in more general web email services remains unknown. As far as we know, there is no similar analysis on email search at a much larger scale.

Search Intent. Search intent is another aspect that has been well studied in web search but not email search. A taxonomy of web search intents was proposed by Broder [2] in 2002 and extended by Rose and Levinson [28]. From the perspective of user needs, they classified web queries as navigational (reach a particular site), informational (acquire the information that is assumed to be present on one or more pages), and transactional (perform some web-mediated activities). For email search, there has been some work on how people organize emails [35, 13, 17] and perform email-mediated tasks [4, 12, 21, 9]. More recently, Sappelli et al. [30] categorized email tasks from multiple dimensions through the analysis of message content. Qadir et al. [25] proposed a latent activity model for workplace emails and demonstrated its effectiveness through the task of recipient recommendation. Narang et al. [22] analyzed the relationship between a person's email organizational practices and their search activity using large-scale email enterprise activity logs.

Compared to previous research, we conduct a large-scale analysis of email search behavior with millions of queries and users and complement it with a survey to understand email search intents. We analyze re-finding behavior at scale and examine actions that people perform on search results as additional evidence on why people are searching for email.

3. METHODOLOGY

We studied email search using two complementary techniques: a large-scale log analysis and a survey. The log analysis provides valuable information about general patterns of email search behavior. In our study, we used a sample of

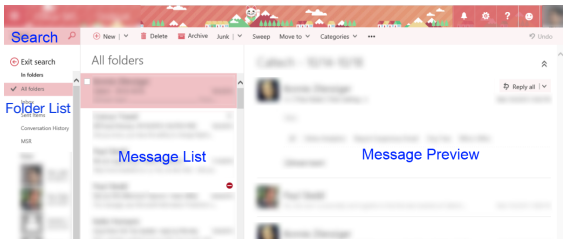

Figure 1: An example search result page for email search in the Outlook web client.

email search logs from the Outlook web client. We analyzed user behavior and compared our results with similar studies of web search interactions to identify similarities and differences between email search and web search. We developed and executed a separate survey to better understand properties of email search and the intents behind different search behavior. The survey covered various aspects of search such as query formulation, the reason for the search, search success, and search experience.

3.1 Log data

We analyzed a sample of email logs from the Outlook web client over a period from July 25, 2015 to July 31, 2015. Although the sample included enterprise and personal accounts in both desktop and mobile settings, the vast majority of the activities were from the desktop in enterprises. The Outlook web client can be used with multiple browsers on desktop and mobile devices, with multiple configurations, but a typical configuration is shown in Figure 1. The search box is located on the top left of the interface, with a folder list in the left, a message list in the middle and the message preview pane on the right. When the user submits a query, results are shown in the message list. Search results are not shown in the message list until the user finishes typing and submits their query. Results are summarized by showing their sender, subject, body text snippet, time and flags such as having an attachment. When a message is selected, it is shown in the preview pane. When selected, the message is shown in the context of its conversation, so the user can view the entire conversation thread without needing to search again or select another message.

Our analysis focused on search-related email activities such as queries, clicks, and sessions. Our definition of a search session is similar to that used in studies of web search. Specifically, we define an email search session as a sequence of activities following a search query such that the gap between successive actions is less than 10 minutes. A gap longer than this, followed by a search, starts a new search session. A session can also be terminated when the user explicitly clicks the exit button near the search query box. In total, our log analysis is based on nearly 1.4 million email sessions covering over 2.0 million queries from 711,000 users. Although we studied user interactions for only one week, people had access to their full email stores and could search for information that had been received long before the week we studied. We followed corporate privacy and anonymity guidelines with regard to handling of log data. The logs we used do not provide the text of the queries, messages, or headers. We were able to view the output of query classifiers run on query text, but not the queries themselves. The logs that we studied contain two types of interactions:

- Interactions on search result lists: query ID, internal message IDs of the results, and clicks within the results.
- Interactions on messages: Open, reply, forward, delete, move, link click, open attachments, etc.

In addition to behavioral information, the logs also contain query- and message-related information:

- Query operator: the operators used for searching particular fields of messages, such as `from:`, `to:`, `subject:`, etc.
- Message metadata: internal message ID, thread ID, session ID, sender ID, received date.
- Query type: the output of a multi-class classifier that categorized query strings as either Person (people's names), Number, or Company (company names).

3.2 Survey

Our survey was designed to discover why people searched for emails (*intents*), how they formulated their search (*queries*), and what they wanted from the system (*results*). Participants were asked to recall their last email search and to answer questions about their queries and search experience. To facilitate comparison with web search, we asked the same questions about participants' last web searches. Since we were primarily interested in email search, the email-related questions always preceded the web-related questions. Although self-reported data are sometimes unreliable, people are able to recall specific recent events quite accurately [29]. We use a survey rather than a diary study to enable us to consider a broader range of people.

The survey questions focused on the last search people did in email and on the web. We first asked participants to describe in their own words what they were looking for. We then asked what the first query they used to search was, what they remembered about the email/page they were looking for, and why they were searching for the email/page. We asked about search success, the number of queries and the time to complete their search. To better understand re-finding behavior, we asked whether they had issued the query before and whether they had clicked the same result before. Finally, we asked questions about their email and browser software, and optional demographic information (gender and age).

We distributed the same survey both internally within our organization and externally to a panel provided by SurveyGizmo. Internal participants were a random sample of people from our company. External participants, who regularly use web email and were between 20 and 75 years of age, were recruited by SurveyGizmo. External participants were paid individually while the internal participants were not. In total, 324 people responded our survey (63% male, 75% between 25 and 55 years of age) and 138 of them are outside our company. Overall 286 participants recalled their last email search and 272 recalled their last web search. Most participants recalled a search they had conducted within two days before the survey, 183 (64.1%) for email and 207 (76.2%) for web.

4. CHARACTERIZING EMAIL SEARCH

We begin with a high-level characterization of email search compared to web search, starting with the survey data.

4.1 Queries and Domain Knowledge

People's descriptions of what they were looking for were quite rich, although the queries they used were much simpler. For email, the descriptions contained information about the content of the email but also many attributes of the target message (e.g., who sent it, roughly when they had seen it, whether it contained a specific piece of information, etc.). The following examples illustrate many different specific and personally relevant attributes that people know about what they are looking for in their email[1].

- *I was looking for an email from a church member that contained an excel spreadsheet that included job assignments/role and responsibilities for a church event that will take place next month.* [Query: james]
- *A few months back, I used email as a note function to write down the name of a guy I met at the gym. I saw him recently and couldn't remember his name off the top of my head, but I knew I had that note email so I did a search on my phone outlook for gym since I remembered that was in the subject. I found the guys name so that if needed I could address him by name. It's part of working on my Dale Carnegie skills.* [Query: gym]
- *I was trying to find out when my last oil change was. The dealer emails me info once I have an appointment, so I was looking for my last email about it.* [Query: toyota]

In contrast, for web searches, people typically wanted general information and specified little about the desired page.

- *I was looking for cold medicine that would work for post nasal drip and sore throat.* [Query: cold medicine]
- *I was looking up who produced the upcoming X-Men movie.* [Query: x-men apolcolypse]
- *Looking for a certain kind of vehicle for sale within 200 miles of me.* [Query: used cars oklahoma]

The differences seen in these examples are also evident in the distribution of attributes that people said that they knew about what they were looking for in our survey. Figure 2 summarizes the kinds of information that people said they remembered about what they were searching for. In both email and web searches, content was the most common information that people remembered: 62.9% of people recalled something about the content of the emails they were looking for, and 50.7% of people recalled the content of web pages they are looking for. This reflects greater familiarity with email (most of which they have already seen) than web pages (many of which they had not visited before). People also remembered much more metadata about the emails they were looking for than the web pages. In email searches, 52.1% recalled the sender of the email, 27.6% recalled the subject and 20.3% remembered roughly when the email was

[1]Specific names of people and places have been obscured

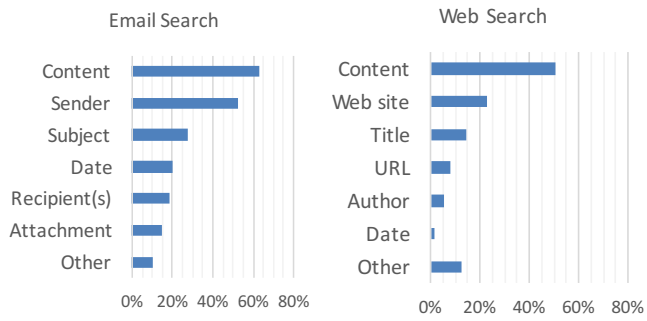

Figure 2: Percent of participants who remembered certain message attributes before doing an email search or a web search.

sent or received. In web search, only 22.8% of people remembered the web site and attributes URL/Author/Date were less than 10%. Most people already had detailed knowledge about the email they wanted before they actually conducted their email search.

Although people remember a lot about what they are looking for in email, it is not always reflected in their query. The examples above show the richness of what people know about what they were looking for in email (e.g., sender, time, subject, links, attachments, etc.). These survey results align with the distribution of attributes that are explicitly specified in email query logs. In the email logs, we find that 18% of queries contain advanced syntax with `from:` occurring in 13.6% and `to:` occurring in 2.7% of queries. Date is seldom specified explicitly (using `date:`), but it is readily available in the email client since results are sorted by date.

The survey suggests that people use shorter queries in email than web search. Similarly, the large-scale logs showed email queries with 1.59 words on average whereas web queries had 3.16 words on average. One reason why people use shorter queries in email is that email results are much easier to quickly scan than web results. Email clients have sorting capabilities and icons support visual scanning to identify the matching attributes. Email search interfaces have more organizational affordances but less query support (query suggestion, spell correction, etc) than web search.

4.2 Sessions and Clicks

In our analysis, we treated each search-result initiated message open as a click. Duplicate clicks (sequential repeat clicks on the same message) were removed in the email and web search logs. As described previously, email and web search sessions are defined in a similar way.

The percentage of queries that had at least one click are similar in email (66%) and web (72%) search. However, the amount of search activity is lower in email than in the web. There are fewer queries per session in email search (1.68) than web search (2.92), and slightly fewer clicks per query in email (2.47) than web search (2.61). In our survey, 83.6% of respondents indicated that they were satisfied with their last email searches and 77.0% of them were satisfied with their last web searches. This suggests that people are able to successfully recognize what they are looking for without much effort in email. This is likely due to a combination of familiarity with email, and the results presentation which

is ordered by an attribute (typically date) and shows other visually salient annotations such as flags and categories.

5. RE-FINDING

Re-finding a web page is a common reason for conducting a web search. People tend to repeat previous searches to re-find the same pages they previously visited [31]. Using our email search logs, we examined properties of re-finding behavior in email search.

5.1 Terminology

In their study of re-finding behavior in web search, Teevan et al. [31] gave formal definitions of different types of queries. We follow their definitions to better compare re-finding in email search and web search.

We categorized queries according to whether they have been issued multiple times by the same user. Teevan et al. [31] explored different query normalization techniques and analyzed their influence on characteristics of re-finding. However, since query strings are not accessible in our email search logs, we can only identify the equivalence between two queries by comparing their hash codes.

In the next sections, we use the following definitions:

- *Equal-Query* queries – Given two queries submitted by a specific user, if they have the exact same text, both of the them are Equal-Query queries.

- *Different-Query* queries – Given the set of queries (Q) submitted by a specific user, if query $q \in Q$ and for any other query $q' \in Q$, q' has different text with q, then q is a Different-Query query.

Click patterns also reveal important properties of re-finding intents. Although the user interface of our web-based email client is not exactly the same as that of search engine result pages, in both cases results have to be selected to see the full text of the email or web page. Thus we follow the same methodology used by Teevan et al. [31] to capture the equivalence and similarity between query click patterns:

- *Equal-Click* queries – Given query q_1 and its click-through set C_1, if there exists query q_2 such that q_2 is issued by the same user and its click-through set $C_2 = C_1$, then both q_1 and q_2 are Equal-click queries (no matter whether they have same query strings or not).

- *Some-Common-Click* queries – Given q_1 and its click-through set C_1, if for any query q_1' ($q_1' \neq q_1$) from the same user, $C_1' \neq C_1$ and there exists a query q_2 (also issued by the same user) such that $C_1 \cap C_2 \neq \emptyset$, then q_1 is a Some-Common-Click query.

- *Overlapping-Click* queries – Equal-Click queries and Some-Common-Click queries are collectively called Overlapping-Click queries

- *No-Common-Click* queries – If query q_1 shares no clicked result with any other queries issued by the same user, then it is a No-Common-Click query.

- *Navigational* queries – Queries where the user issues the same query and clicks on exactly the same results.

5.2 Re-finding Query Distribution

We now compare the re-finding query distribution in email search with previous results in web search. Following the methodology introduced by Teevan et al. [31], we summarize how often the same or different queries are used to find the same or different items. Like Teevan et al. [31], we only show results for queries with at least one click. We note that there are many differences between their web data and our email data that may influence exact percentages (e.g., normalization of queries and URL, time, selection of participants), but not the general patterns.

Table 1 gives the details of the query distribution for web and email search. The left side of the table shows Teevan et al.'s [31] analysis for web search, and the right side shows our analysis for email search. In email search, 38.32% of the queries are Equal-Query queries, that are queries that have been repeated by the same user at least once. This percentage of repeated queries is somewhat higher in email than the repeat percentage in web search (32.59%).

Despite a somewhat higher percentage of Equal-Query queries, the number of queries that have overlapping clicks is lower in email search than web search. In email search logs, 34.12% queries are Overlapping-Click queries compared with 38.84% for web search. More interesting than the overall percentages of same/different queries and same/different clicks is the relationship between the two. In email, only 28% of people who issued the same query clicked on the same results (9.06%/38.32%), whereas 74% of Equal-Query queries in web were also Equal-Click queries (24.02%/32.59%). These so-called navigational queries [31] are much less prevalent in email search than in web search.

In email search, users tend to click different results even with repeated queries. For example, a query like *Amazon* is not intended to go to Amazon.com (as it is in web search), but rather seeks to find an email from or about Amazon, perhaps to find the latest tracking number for an order. Because email is a dynamic collection and the results are typically sorted by date, a simple query like *Amazon* will return the most recent emails first making it easy to keep up to date with the latest correspondence with a company, individual or conversation. Similarly, queries that contain a person's name are common in email search. Even when people issue the same name query, they appear to want information from a different message or thread at different points in time.

5.3 Metadata

In this section, we further analyze the effect of message metadata on re-finding behaviors in email search. An important characteristic that differentiates email messages from many other text items like web pages is the rich metadata associated with emails. Metadata, such as conversation ID, sender ID, or received date are correlated with the importance of messages [27], and consequentially play an important role in email search. We reconstructed Table 1 to show the effect of using the same conversation ID or sender ID (rather than the exact same message) as the object of re-finding queries. If two messages share the same piece of metadata, we treat them as "same" messages. The results are shown in Table 2. Our analysis shows that people are more likely to click on messages that share same metadata (e.g., same conversation ID) rather than exactly the same message when they issue the same query. When we identify message equivalence using message conversation ID, the per-

Table 1: Distribution of different query types in Teevan et al.[31]'s web search logs and our email search logs.

	Web from Teevan et al. [31]				Email			
	Overlapping-Click URL		No-Common-Click URL	Total	Overlapping-Click Email		No-Common-Click Email	Total
	Equal-Click URL	Some-Common-Click URL			Equal-Click Email	Some-Common-Click Email		
Equal Query	24.02%	4.86%	3.71%	32.59%	9.06%	14.96%	14.30%	38.32%
Different Query	4.91%	5.05%	57.45%	67.41%	2.49%	7.61%	51.58%	61.68%
Total	38.84%		61.16%	100.00%	34.12%		65.88%	100.00%

Table 2: Distribution of different query types in our email search logs when identifying messages with their conversation ID and sender ID.

	Email							
	Overlapping-Click Conversation		No-Common-Click Conversation	Total	Overlapping-Click Sender		No-Common-Click Sender	Total
	Equal-Click Conversation	Some-Common-Click Conversation			Equal-Click Sender	Some-Common-Click Sender		
Equal Query	10.92%	16.08%	11.32%	38.32%	9.18%	15.06%	14.07%	38.33%
Different Query	3.74%	9.52%	48.43%	61.68%	2.53%	7.68%	51.48%	61.67%
Total	40.26%		59.74%	100.00%	34.45%		65.55%	100.00%

centage of Overlapping-Click queries increases from 34.12% to 40.26%. There is an increase of 20.53% in Equal-Query Equal-Click (from 9.06% to 10.92%), 7.49% in Equal-Query SomeCommon-Click (from 14.96% to 16.08%), 50.20% in Different-Query Equal-Click (from 2.49% to 3.74%) and 25.10% in Different-Query SomeCommon-Click (from 7.61% to 9.52%). When re-finding is considered at the level of a conversation thread rather than an individual message, there are more clicks on the same item and fewer on different items. However, there are still far fewer Equal-Query and Equal-Clicks queries than observed in web search.

In contrast to conversation ID, sender ID shows little effect on the query distribution in the re-finding analysis. The percentage of Overlapping-Clicks for sender ID in Table 2 is only slightly higher than that in Table 1. This indicates that sender ID may not be directly correlated with repeated clicks. For example, when repeating a query about a project, users may want the latest email in the same conversation thread but not necessarily from the same sender. Carmel et al.'s study [3] of email ranking also found that features related to the sender address provided little improvement in the performance of email retrieval models.

5.4 Time Decay

To further examine the relation between repeated queries and repeated clicks, we broke the data down by the time between successive repeated queries. We find that the probability of a repeated click for a repeated query is related to the time between the two queries.

As shown in Figure 3, repeated queries issued within the same day have the highest probability (56.62%) of having overlapping clicks. The probability drops gradually as the time gap between two repeated queries becomes longer, falling lower than 30% after three days. This is opposite to what is seen in web search where repeated queries close in time have a lower probability of clicks on the same items compared to queries that were repeated after one or two days [31]. In web search, queries repeated in a short period

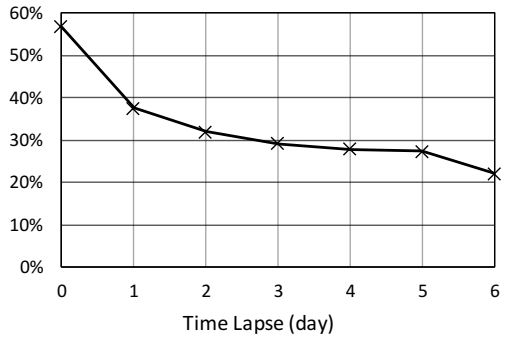

Figure 3: Probability of click overlap for repeated email queries, for various inter-query intervals.

of time are usually intended to find new information on the same topic, and repeated queries that occur over a longer time period are mostly navigational queries which are intended to return to the same page. In contrast, in email search people tend to use the same query to return to the same messages in short periods of time, but use the same queries to find different emails over long time. This pattern is likely influenced by the chronological order used for email search results, which show the most recent matching messages at the top. If new messages that match the query have arrived, it may be easier (and more relevant) to select the new messages than to return to the previously-selected message. Indeed, in our survey, people indicated that they sometimes use search to find the latest information in a conversation.

6. SEARCH INTENTS

In this section, we present qualitative and quantitative analyses to understand people's intents in searching email.

Table 3: Answers and counts for question "what best describes why you were searching?"

Email ($n = 286$)		
Specific need	242	84.6%
Get to a specific email	197	68.9%
Get a specific email, but other emails may serve too	45	15.7%
General need		
No specific email in mind	26	9.1%
Other	18	6.3%
Web ($n = 273$)		
Specific need	135	49.5%
Get to a specific web page	72	26.4%
Get a specific web page, but other pages may serve too	63	23.1%
General need		
No specific web page in mind	117	42.9%
Other	21	7.7%

6.1 Specific or General

In our survey, we asked participants to select the reason that best described why they were searching in email and on the web. Results are listed in Table 3. We aggregated the answers into two high-level categories:

- *Specific* need: The need to get a specific item through search. Users either want a specific email/web page or a group of related email(s)/web page(s).
- *General* need: The need for general information on a topic. Users have a rough idea of what they are want, but do not have a particular email or web page in mind.

According to our survey, *Specific* needs are more common in email search. As shown in Table 3, 242 of 286 participants (84.6%) were searching for a specific email or a specific piece of information in an email. Many of them mentioned a particular item of interest (link, message, presentation, etc.) in their descriptions. In comparison, *General* needs are much less common in email search. Only 9.1% of people did not have a specific email in mind before search. An example of a general need in email search was: "*I was looking for background information on a particular bug that I knew someone had emailed about.*", but such needs are not common in our survey responses. The proportion of *General* needs in email search (9.1%) is four times less than in web search (42.9%). In web search, people often use search engines as a means to learn and discover new information. However, in email search, people already have good knowledge of their email corpus and are not as likely to want new information or to learn new things from their email archives.

6.2 Post-Query Actions in Email Search

Selecting a search result does not necessarily represent the final goal of an email search. One way to get additional insights about the tasks that people are trying to accomplish in email search is to look at what actions they performed after a search. Our email search log recorded a variety of actions on email such as forwarding, deleting, clicking a link and opening an attachment. In this section, we analyze users' interactions to understand their tasks in email search.

Table 4: Relative counts (Actions), query coverage (Queries) and average frequency (Freq.) of different actions in email search.

	Actions	Queries	Freq.
Select:			
MessageSelected	76.53%	99.50%	3.6
Respond:			
ReplyAll	1.58%	6.90%	1.1
Forward	1.37%	5.51%	1.2
Reply	1.09%	4.78%	1.1
Organize/delete:			
Delete	12.44%	4.13%	14.0
Move	1.00%	0.45%	10.3
MarkAsRead	0.38%	0.49%	3.6
Flag	0.12%	0.38%	1.4
MarkAsUnread	0.12%	0.38%	1.4
FlagComplete	0.05%	0.15%	1.4
FlagCleared	0.01%	0.04%	1.4
Open attachment/link:			
OpenedAnAttachment	3.68%	9.24%	1.9
LinkClicked	1.63%	5.15%	1.5

In Table 4, we computed three characteristics of actions:

- *Actions*: the proportion of times that each action occurs. (The numbers of all actions sum to 100%.)
- *Queries*: the proportion of queries in which each action occured. Because a query can contain multiple actions, the sum of numbers is greater than 100%.
- *Frequency*: the average frequency of each action in queries where the action occurred.

MessageSelected is the most common action in email search - it occurred in 99.50% of queries and accounted for 76.53% of all actions. After MessageSelected, OpenedAnAttachment has the highest query coverage (9.24%), followed by ReplyAll (6.90%), Forward (5.51%), LinkClicked (5.15%), Reply (4.78%) and Delete (4.13%). The popularity of OpenedAnAttachment and LinkClicked indicate that refinding external content via attachments or links is a very common goal in email search. In addition, the frequent appearance of ReplyAll, Forward and Reply reflects that communication is still an important need in email search. Although it only occurs in 4.13% of queries, Delete is a frequent action accounting for 12.44% of all actions. This phenomenon suggests that people use search as a method for cleaning their mailboxes. In these cases, the search query acts more like a filter for finding emails to delete than a request for information. There is no analog of this phenomenon in web search.

As for the action frequency (*Frequency* column in Table 4), most actions like ReplyAll, Forward and Reply appear once on average, indicating that people usually respond to one message per query. However, three of the organizing/cleaning actions (Delete, Move and MarkAsRead) tend to appear in groups. The frequencies per query for Delete and Move are more than 10 on average, suggesting that people use search to perform bulk cleaning and organization of

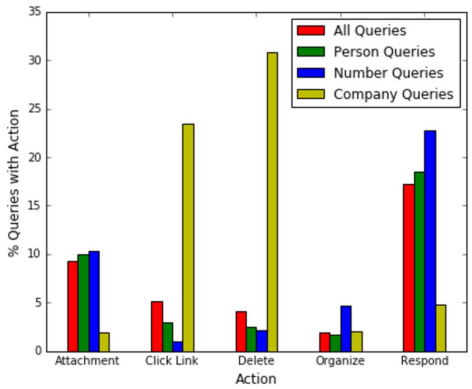

Figure 4: Percentages of queries with actions in different query types. *All Queries* refers to all queries including *Person, Number* and *Company* queries.

their email. The web interface we studied showed 25 results per page, along with an check box to select all messages, and we saw many instances in which 25 messages were deleted at a time.

We now examine the distributions of actions for different query types. We classified email search queries into three classes: Person (people's names, using an existing internal classifier), Number (using regular expressions), and Company (using a list of popular companies). It is difficult to evaluate the accuracy of our classifiers since we cannot look at the text of email queries or the content of email message. Our goal was not to develop a powerful classifier but to use the classification to see whether there are differences in actions for different queries. To simplify the presentation, we grouped Reply/ReplyAll/Forward into *Respond*, and Flag/FlagCleared/FlagComplete, MarkAsRead/MarkAsUnread and Move into *Organize*. The percentages of queries with each type of actions are shown in Figure 4. OpenAnAttachment and Respond were more likely to appear in queries with person names or numbers, but very unlikely to appear in queries with company names. In contrast, LinkClicked and Delete often appeared in queries with company names but rarely appeared in queries with person names or numbers. It seems that users often issued Person/Number queries to interact with emails and issued Company queries to find links or clean up their mailboxes. Recent studies [1, 14] show more than 90% of non-spam emails in consumer email accounts is machine-generated. Many of these messages are order statuses (e.g. from Amazon.com or Alibaba) and updates (e.g. from Craigslist or Tinder). Therefore, company names in queries serve as good filters when users want to select a link from or clean up these automatically generated emails.

7. DISCUSSION

In this paper we characterized email search strategies and interaction patterns through large-scale behavioral log analysis and a complementary survey. In our survey, 84.6% of the respondents had a specific email in mind before they conducted their last email search but only 49.5% of respondents had a specific page in mind in their last web search (Table 4). Respondents also indicated that they were successful in 83.6% of their email searches (compared with only 77.0% in their web searches). This highlights the fact that people are familiar with their email corpus and can easily identify relevant results returned by email search engines.

The survey also showed that people remember much more about what they are looking for in email search than in web search (Figure 2), but this knowledge is not always expressed explicitly in their search queries. The log analysis showed that people explicitly specify metadata in only 18% of their queries. Most of these involved the sender or recipient of the email. Although people often recalled roughly when an email of interest was sent or received, they seldom specified this explicitly. However, date information is readily available in the interface since email results are usually ordered by date. Similarly, other properties such as HasAnAttachment or Flag are visible in many email clients including the one that we studied. Although there has been research aimed developing web-style ranking functions for email [3], we believe that an equally fruitful direction is to develop a search engine that can help people articulate what they remember about what they are looking for, or to identify additional metadata to improve the precision of search results, for example by suggesting metadata that is useful in reducing the search space. Email search engines could also provide better search syntax to support deeper semantic or structural queries that allows people to look for a specific table, link, phone number, address or fact within a message.

Re-finding is a common activity in web search. We also found that repeat queries are common in email search (38.2% of queries). However, in contrast to web search, in email search people usually click on different messages when they issue the same query at different times. This appears to be due to the changing nature of email. People want to find new messages in an existing conversation as well as different conversations associated with the same person or topics.

Email search is often task-driven, and our survey and logs provide new insight about this. Our survey showed that 33.5% of participants wanted to transact with the emails they found ("find a link" or "do something on an email"), but for their web searches this was only 7.0%. The log analysis showed that finding specific content (LinkClicked and OpenedAnAttachment) and responding (Reply, ReplyAll, Forward) all occurred after more than 5% of email search queries. Also, we were surprised to see how often people used search not to find information or to respond to people, but to organize and clean their mailboxes. Table 4 showed that more than 4% of email search queries involved Deleting emails and other organizational actions (Move, MarkAsRead, Flag, MarkAsUnread, FlagComplete and FlagCleared) occured with lower frequency. To support these types of organizational needs, search results could provide better capabilities for interacting with sets of items.

Finally, we showed that the distribution of actions depends on the type of query. When queries contained company names, searchers were much more likely to click on a link or delete emails, When queries contained a person name or number, searchers were more likely to open an attachment or respond. Search interfaces could provide more task-based affordances such as enriching snippets to contain links or names of attachments to support likely actions.

8. CONCLUSION AND FUTURE WORK

Email search is an important activity, yet it is not well represented in the existing literature. We presented a large-

scale analysis of search behavior in a web-based email client, complemented by a survey to obtain more details about people's email and web search intents. Our work shows that email search differs from web search in several key ways. Repeat queries occur in both, but repeated results are more common in web search. In email search different messages are selected for the same query, in some cases because they are different messages from the same conversation ID or same sender ID. While the web query "amazon" often leads to the same page, the email query "amazon" often leads to a recent message from that sender.

Search intents are highly contextual, but people rarely express the full context in their query, instead giving a short query and relying on faceting and sorting features in the interface. People use email search to perform a variety of actions, not just to read or respond, but to organize and find links and attachments. We observed search queries, such as company names, being used to move and delete mail in bulk. This can also inform ranking: for such organizational actions, the *group* of messages is relevant together and an ordered ranking is less meaningful. Users also seek information associated with messages such as URLs or attachments. Search interfaces can be improved by surfacing these directly in the search results.

A deeper understanding of email search at scale will drive improvements in user interfaces and ranking. More than 15% of email searches reported in our survey were unsuccessful. Future work will examine the nature of actions with respect to classes of queries, devices and sender-recipient characteristics. Improving the relevance and presentation of email results in a privacy-preserving way will benefit users and continue to support email as a valuable medium.

9. REFERENCES

[1] N. Ailon, Z. S. Karnin, E. Liberty, and Y. Maarek. Threading machine generated email. In *Proceedings of the 6th ACM International Conference on Web Search and Data Mining*, WSDM '13, pages 405–414, Rome, Italy, 2013. ACM.

[2] A. Broder. A taxonomy of web search. *SIGIR Forum*, 36(2):3–10, 2002.

[3] D. Carmel, G. Halawi, L. Lewin-Eytan, Y. Maarek, and A. Raviv. Rank by time or by relevance?: Revisiting email search. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*, CIKM '15, pages 283–292, Melbourne, Australia, 2015. ACM.

[4] 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, Salvador, Brazil, 2005. ACM.

[5] N. Craswell, A. P. de Vries, and I. Soboroff. Overview of the TREC 2005 enterprise track. In *TREC*, 2005.

[6] N. Craswell, H. Zaragoza, and S. Robertson. Microsoft Cambridge at TREC 14: Enterprise track. In *TREC*, 2005.

[7] E. Cutrell, D. Robbins, S. Dumais, and R. Sarin. Fast, flexible filtering with Phlat. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '06, pages 261–270, Montréal, Québec, Canada, 2006. ACM.

[8] L. A. Dabbish and R. E. Kraut. Email overload at work: An analysis of factors associated with email strain. In *Proceedings of the 2006 20th Anniversary Conference on Computer Supported Cooperative Work*, CSCW '06, pages 431–440, Banff, Alberta, Canada, 2006. ACM.

[9] D. Di Castro, Z. Karnin, L. Lewin-Eytan, and Y. Maarek. You've got mail, and here is what you could do with it!: Analyzing and predicting actions on email messages. In *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining*, WSDM '16, pages 307–316, San Francisco, California, USA, 2016. ACM.

[10] S. T. 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 the 26th Annual International ACM SIGIR Conference on Research and Development in Informaion Retrieval*, SIGIR '03, pages 72–79, Toronto, Canada, 2003. ACM.

[11] D. Elsweiler, M. Harvey, and M. Hacker. Understanding re-finding behavior in naturalistic email interaction logs. In *Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '11, pages 35–44, Beijing, China, 2011. ACM.

[12] D. Elsweiler and I. Ruthven. Towards task-based personal information management evaluations. In *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '07, pages 23–30, Amsterdam, The Netherlands, 2007. ACM.

[13] D. Fisher, A. J. Brush, E. Gleave, and M. A. Smith. Revisiting Whittaker & Sidner's "email overload" ten years later. In *Proceedings of the 2006 20th Anniversary Conference on Computer Supported Cooperative Work*, CSCW '06, pages 309–312, Banff, Alberta, Canada, 2006. ACM.

[14] M. Grbovic, G. Halawi, Z. Karnin, and Y. Maarek. How many folders do you really need?: Classifying email into a handful of categories. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, CIKM '14, pages 869–878, Shanghai, China, 2014. ACM.

[15] C. Grevet, D. Choi, D. Kumar, and E. Gilbert. Overload is overloaded: Email in the age of Gmail. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, pages 793–802, Toronto, Ontario, Canada, 2014. ACM.

[16] J. Gwizdka. Reinventing the inbox: Supporting the management of pending tasks in email. In *CHI '02 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '02, pages 550–551, Minneapolis, Minnesota, USA, 2002. ACM.

[17] J. Gwizdka. Email task management styles: The cleaners and the keepers. In *CHI '04 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '04, pages 1235–1238, Vienna, Austria, 2004. ACM.

[18] M. Harvey and D. Elsweiler. Exploring query patterns in email search. In *Advances in Information Retrieval: 34th European Conference on IR Research, ECIR*

2012, volume 7224 of *Lecture Notes in Computer Science*, pages 25–36. Springer, Berlin, Heidelberg, 2012.

[19] B. Hogan and D. Fisher. A scale for measuring email overload. *Microsoft Research*, MSR-TR-2006-65, 2006.

[20] Y. Koren, E. Liberty, Y. Maarek, and R. Sandler. Automatically tagging email by leveraging other users' folders. In *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '11, pages 913–921, San Diego, California, USA, 2011. ACM.

[21] S. Michel and I. Weber. Rethinking email message and people search. In *Proceedings of the 18th International Conference on World Wide Web*, WWW '09, pages 1107–1108, Madrid, Spain, 2009. ACM.

[22] K. Narang, S. T. Dumais, N. Craswell, D. Liebling, and Q. Ai. Large-scale analysis of email search and organizational strategies. In *Proceedings of the 2017 ACM on Conference on Human Information Interaction and Retrieval*, CHIIR '17, Oslo, Norway, 2017. ACM.

[23] P. Ogilvie and J. Callan. Experiments with language models for known-item finding of e-mail messages. In *TREC*, 2005.

[24] K. Purcell and L. Rainie. Email and the internet are the dominant technological tools in American workplaces| Pew Research Center. *Pew Research Center*, 2014.

[25] 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, 2016. Association for Computational Linguistics.

[26] P. Ramarao, S. Iyengar, P. Chitnis, R. Udupa, and B. Ashok. Inlook: Revisiting email search experience. In *Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '16, pages 1117–1120, Pisa, Italy, 2016. ACM.

[27] K. Rector and J. Hailpern. MinEMail: SMS alert system for managing critical emails. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, pages 783–792, Toronto, Ontario, Canada, 2014. ACM.

[28] D. E. Rose and D. Levinson. Understanding user goals in web search. In *Proceedings of the 13th International Conference on World Wide Web*, WWW '04, pages 13–19, New York, NY, USA, 2004. ACM.

[29] D. M. Russell and E. H. Chi. Looking back: Retrospective study methods for HCI. In *Ways of Knowing in HCI*, pages 373–393. Springer, New York, NY, 2014.

[30] M. Sappelli, G. Pasi, S. Verberne, M. de Boer, and W. Kraaij. Assessing e-mail intent and tasks in e-mail messages. *Information Sciences*, 358–359:1–17, 2016.

[31] J. Teevan, E. Adar, R. Jones, and M. A. S. Potts. Information re-retrieval: Repeat queries in Yahoo's logs. In *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '07, pages 151–158, Amsterdam, The Netherlands, 2007. ACM.

[32] S. K. Tyler and J. Teevan. Large scale query log analysis of re-finding. In *Proceedings of the Third ACM International Conference on Web Search and Data Mining*, WSDM '10, pages 191–200, New York, NY, USA, 2010. ACM.

[33] W. Weerkamp, K. Balog, and M. de Rijke. Using contextual information to improve search in email archives. In *Advances in Information Retrieval: 31th European Conference on IR Research, ECIR 2009*, volume 5478 of *Lecture Notes in Computer Science*, pages 400–411. Springer, Berlin, Heidelberg, 2009.

[34] S. Whittaker, T. Matthews, J. Cerruti, H. Badenes, and J. Tang. Am I wasting my time organizing email?: A study of email refinding. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pages 3449–3458, Vancouver, BC, Canada, 2011. ACM.

[35] 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, Vancouver, British Columbia, Canada, 1996. ACM.

[36] T. Xu and D. W. Oard. Exploring example-based person search in email. In *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '12, pages 1067–1068, Portland, Oregon, USA, 2012. ACM.