Analyzing and Predicting Task Reminders

David Graus[*]
University of Amsterdam
Science Park 904, Amsterdam, The Netherlands
d.p.graus@uva.nl

Paul N. Bennett, Ryen W. White, Eric Horvitz
Microsoft Research
One Microsoft Way, Redmond, WA USA
{pauben, ryenw, horvitz} @microsoft.com

ABSTRACT
Automated personal assistants such as Siri, Cortana, and Google Now provide services to help users accomplish tasks, including tools to set reminders. We study how people specify and use reminders. Our study analyzes a sample of six months of logs of user-specified reminders from Cortana (Microsoft's intelligent personal assistant), the first large-scale analysis of such reminders. We focus our analyses on time-based reminders, the most common type of reminder found in the logs. We perform a data-driven analysis to identify common categories of tasks that give rise to these reminders across a large number of users, and we arrange these tasks into a taxonomy. We identify temporal patterns linked to the type of task, time of creation, and terms in the reminder text. Finally, we show that these patterns generalize by addressing a prediction task. Specifically, we show that a reminder's creation time is a strong feature in predicting the notification time, and that including the reminder text further improves prediction accuracy. The results have implications for the design of systems aimed at helping people to complete tasks and to plan future activities.

Keywords
Reminders, prospective memory, intelligent assistant, log studies.

1. INTRODUCTION
Automated personal assistants such as Siri, Cortana, Google Now, Echo, and M support a range of reactive and proactive scenarios, ranging from question answering to alerting about plane flights and traffic. Several of these personal assistants provide reminder services aimed at helping people to remember future tasks that they may otherwise forget. We perform an exploratory analysis of a large-scale log of user-created reminders within Microsoft Cortana aimed at understanding users' needs and enhancing the system's reminding services.

Table 1 presents an example of the types of reminder dialogs recorded in the dataset. These logs offer insights about the reminder generation process, including the types of tasks for which people formulate reminders, task descriptions, the times that reminders of different types are created, and the periods of time between the creation of reminders and notifications. Beyond analysis of the nature and timing of reminders, we demonstrate how information about patterns of reminder usage and general trends seen across users can be harnessed to assist people with setting reminders. We focus primarily on reminders for tasks planned for a future time. We make the following contributions in this paper:

Table 1. Example interaction sequence for setting a reminder.

Turn	Who	Text
1	User	Remind me to do the laundry.
2	System	When would you like to be reminded?
3	User	Sunday at noon.
4	System	Alright, remind you to do the laundry at 12:00PM on Sunday, is that right?
5	User	Yes.
6	System	Great, I'll remind you! {success chime}

- Study the creation of time-based reminders at scale in natural settings, revealing common reminders specified across users.
- Develop a taxonomy of task types for these common reminders.
- Study temporal patterns in reminder setting and notification, demonstrating noteworthy patterns.
- Build models that predict the desired timing of reminders, demonstrating a direction in harnessing the patterns.

The findings provide insights about the tasks and goals of users in the real world and about the behaviors and needs of people with regards to memory and reminding. They also support efforts on modeling the tasks and goals of users.

2. RELATED WORK
Several areas of research are relevant to our research on tasks and reminders. We focus largely on research on memory and completing planned tasks. We review research in the following areas: (i) reminders, (ii) memory aids, (iii) prospective memory, and (iv) mining and modeling human behavior at scale.

2.1 Reminders
Several systems have been developed to help remind people about future actions [8,9,21,24,30], many of which leverage contextual signals for more accurate reminding. These systems can help generate reminders associated with a range of future actions, including location, events, activities, people, and time. Two of the most commonly supported types of reminders are location- and time-based (and combinations thereof [8,28]). Location-based reminders fire when people are at or near locations of interest [26,36]. Time-based reminders are set and triggered based on time [14,18], including those based on elapsed time-on-task [6]. While time-based reminders can provide value to many users, particular groups may especially benefit from time-based reminders. These include the elderly [29], those with memory impairments [22], and people seeking to comply with prescribed medications [17]. In this paper, we study time-based reminders in the Cortana reminder service. We omit location- and person-based reminders, as they are less common in our data, and more challenging to study across users per their reliance on personal context and relationships between the user and the locations and persons to trigger the reminders.

[*] Work performed during an internship at Microsoft Research.

UMAP '16, July 13-17, 2016, Halifax, NS, Canada
© 2016 ACM. ISBN 978-1-4503-4370-1/16/07…$15.00
DOI: http://dx.doi.org/10.1145/2930238.2930239

2.2 Memory Aids

Memory aids help people to remember past events and information. Studies have shown that people leverage both their own memories via recall strategies and the use of external memory aids to increase the likelihood of recall [19]. Aids can assume different forms, including paper [27] to electronic alternatives [3,15,33]. One example of a computer-based memory aid is the Remembrance Agent [33], which uses context information, e.g., words typed into a text processor to retrieve similar documents. People have been shown to use standard computer facilities to support future reminding (e.g., positioning documents in noticeable places on the computer desktop) [3]. Such uses can be inadequate for a number of reasons, including the lack of alerting [14]. Other work has focused on the use of machine learning to predict forgetting, and the need for reminding about events [21]. Cortana is an example of an interactive and intelligent external memory aid. Studying usage patterns and user behavior enables us to better understand users' needs, develop improved methods for system-user interaction and collaboration, and more generally, enhance our understanding of the types of tasks where memory aids provide value.

2.3 Prospective Memory

Prospective memory (PM) refers to the ability to remember actions to be performed at a future time [5,11]. Beyond simply remembering, successful prospective memory requires recall at the appropriate moment. PM failures have been an area of study [13,35], and studies have shown that failures can be linked to external factors such as interruptions [7,31]. Prospective tasks are usually divided into time-based tasks and event-based tasks [11]. Time-based tasks are tasks targeted for execution at a specific future time, while event-based tasks are performed when a particular situation or event occurs, triggered by external cues, e.g., person, location, or object [12]. Laboratory studies of PM have largely focused on retention and retrieval performance of event-based PM as this is straightforward to operationalize in an experimental setting. Time-based PM is a largely overlooked type in PM studies [10], as this type of self-generated PM is difficult to model in a laboratory setting. The Cortana reminder logs that we study represent a rich resource of real-life time-based PM instances. They provide insights in the type and nature of tasks that users are likely to forget to execute.

2.4 Mining and Modeling User Activity

Large-scale user logs from many users have been used for a range of different purposes to improve online services and advance our understanding of how people use systems. Search engine queries and search-result clicks have been used to understand how people seek information online [37], train search engine ranking algorithms to better serve user needs [1,20], and more generally, teach us about how humans behave in the world [34]. Although large-scale log analysis of online behavior has focused largely on search and browsing activity, recent work has targeted the large-scale usage of communication tools such as email [23] and instant messaging [25]. In the case of intelligent agents, analyzing user logs may support inferring users' intents [2] or current activities [32]. Where previous work addressed modeling users' long-term goals [4], Cortana reminder logs can help us to understand short term goals.

2.5 Contributions

We extend previous studies in several ways. We present the first study of the creation of common time-based reminders at scale in natural settings. Second, we develop a taxonomy of types of time-based reminders, facilitated by the data we have about the reminders created by a large populations of users. Third, we characterize important aspects of the reminder generation process, including their nature (e.g., reminding about ongoing versus planned activities), and the relationship between the reminder text and the time of setting the reminder and alerting to remind. Finally, to show that the patterns that we uncover represent general trends, we build predictive models of when reminders should fire.

3. REMINDER TYPES

We first investigate user behavior around reminder creation by studying common tasks linked to setting reminders. We focus on the question: *I a a a a ?* To answer this question, we extract reminders that are observed frequently and across multiple users. Then, we categorize the reminders in a task taxonomy to better understand the task types associated with the reminders.

3.1 Reminder Composition

In the left column of Table 2, we present three examples of common reminders. The examples show a structure that is frequently observed in the logged reminders. Reminders are typically composed as predicate sentences. They contain a phrase related to an action that the user would like to perform (typically a verb phrase) and a referenced object that is the target of the action to be performed.

Table 2. Example reminders as predicates.

Reminder	Predicate
R a a	Take out (me, the trash)
R	Put (me, clothes in dryer)
R a a	Get (me, cash from the bank)

3.2 Data

A session for setting a reminder consists of a dialog, where the user and the intelligent assistant interact in multiple turns. Typically, the user starts by issuing the command for setting a reminder, and dictates the reminder. Optionally, the user specifies the reminder's notification time. Next, the assistant requests to specify the time (if the user has not yet specified it), or provides a summary of the reminder, i.e., the task description and notification time, asking the user to confirm or change the proposed reminder (see Table 1).

We analyze a sample of two months of Cortana reminder logs, spanning all of January and February 2015. We pre-process this set of reminders by including only reminders from the United States market (the only market which had Cortana enabled on mobile devices at that time). To narrow the scope of our analysis, we focus on time-based reminders and remove location (e.g., *X I a a Y*) and person-based reminders (e.g., *X I Y*), which are less common and more challenging to study across users due to their personal nature. Finally, we retain only reminders that are confirmed by the user (turn 6 in Table 1). The resulting sample contains 576,080 reminders from 92,264 users. For each reminder, we extract the reminder task description and notification time from Cortana's summary (turn 4 in Table 1). We also extract the creation time based on the local time of the user's device. Each reminder is represented by:

$\mathbf{r_{task}}$: The reminder's textual *a* ; i.e., the phrase which encodes the future task or action to be taken, as dictated by the user. We extract the text from Cortana's final summary response (" *a* from turn 4 in Table 1).

$\mathbf{r_{CT}}$: The reminder's *a* . This represents the time at which the user the reminder. We represent r_{CT} as a discretized time-value; Section 4.1 defines the discretization process. We extract this timestamp from the client's device.

r_{NT}: The *a* set for the reminder to fire an alert. This data represents the time at which the user wishes to be reminded about a future task or action. We represent r_{NT} in the same discretized manner as r_{CT}. We extract the notification time from Cortana's summary response (turn 4 in Table 1).

r_{AT}: Subtracting the creation time from the notification time yields the *a*, the delay between the reminder's creation and notification time. Intuitively, reminders with smaller time deltas represent short-term or immediate tasks (“ *a* *a* ”), whereas reminders with larger time deltas represent tasks planned further ahead in time (“ *a* *a* *a* ”).

3.3 Identifying Common Tasks

To understand the common needs that underlie the creation of reminders, we first identify common reminders, i.e., reminders that are frequently observed across multiple users. Studying common reminders can aid system designers in understanding broad usage patterns, and steer the design and implementation of features to better support this usage. We employ a mixed methods approach, comprising data-driven and qualitative methodologies, to extract and identify common task types.

Frequent task description extraction. First, we extract common task descriptions, by leveraging the predicate (verb+object) structure described at the start of this section. To ensure that the underlying task descriptions represent broad tasks, we filter to retain only descriptions that start with a verb (or a multi-word phrasal verb) that occurs at least 500 times, across at least ten users, with at least five objects. This yields a set of 52 frequent verbs,[1] which covers 60.9% of the reminders in our sample. The relatively small number of verbs which cover the majority of reminders in our log indicates that there are likely many common task types that give rise to reminder creation. To analyze the underlying tasks, we include the most common objects, by pruning objects observed less than five times with a verb. This yields a set of 2,484 unique task descriptions (i.e., verb+object), covering 21.7% of our sample log.

Manual labeling. Next, we aim to identify common tasks which underlie the frequent task descriptions, and categorize them into a broader task type taxonomy. Specifically, by manual inspection, we identified several key dimensions that separate tasks. In particular, dimensions that commonly separate tasks are: whether the task represents an interruption or continuation of a user's activity, the context in which the task is to be executed (i.e., at home, at work), and the (expected) duration of the task. This enabled us to label the frequent task descriptions as belonging to one of six broad task types with several subclasses.

3.4 Task Type Taxonomy

In this section we describe each of the six task types in turn, and provide examples of the associated verb+object patterns. The example objects are shown in decreasing order of frequency, starting with the most common. Note that verbs are not uniquely associated with a single task type, but the verb+object-pair may determine the task type (compare, e.g., “ *a* *a* ” to “ *a* ”).

1. Go somewhere (33.0%): One third of the frequent tasks refer to the user moving from one place to another. We distinguish between two subtypes: the first subtype is running an errand (83.2%), where the reminder refers to executing a task at some location (e.g.,). Running an errand represents an interruption of the user's activity, but a task of a relatively small scale, i.e., it represents a task that briefly takes up the user's availability. The second subtype is more comprehensive, and represents tasks which are characterized by a switch of context (16.8%), e.g., moving from one context or activity to another (“ ”, “ *a* ”), which has a larger impact on the user's availability.

Run errand	
grab [something]	laundry, lunch, headphones
get [something]	batteries
pick up [something/someone]	laundry, person, pizza
buy [something]	milk, flowers, coffee, pizza
bring [something]	laptop, lunch, phone charger
drop off [something]	car, dry cleaning, prescription
return [something]	library books
Switch context	
leave (for) [some place]	house, work, airport
come [somewhere]	home, back to work, in
be [somewhere]	be at work, at home
go (to) [somewhere]	gym, work, home, appointment,
stop by [some place]	the bank, at Walmart
have (to) [something]	work, appointment

2. Chores (23.8%): The second most common type of reminders represent daily chores. We distinguish two subtypes: recurring (66.5%) and standalone chores (33.5%). Both types represent smaller-scale tasks which briefly interrupt the user's activity.

Recurring	
take out [something]	trash, bins
feed [something]	dogs, meter, cats, baby
clean [something]	room, house, bathroom
wash [something]	clothes, hair, dishes, car
charge [something]	phone, fitbit, batteries
do [something]	laundry, homework, taxes, yoga
pay [something]	pay rent, bills, phone bill
set [something]	alarm, reminder
Standalone	
write [something]	a check, letter, thank you note
change [something]	laundry, oil, air filter
cancel [something]	amazon prime, netflix
order [something]	pizza, flowers
renew [something]	books, driver's license, passport
book [something]	hotel, flight
mail [something]	letter, package, check
submit [something]	timesheet, timecard, expenses
fill out [something]	application, timesheet, form
print [something]	tickets, paper, boarding pass
pack [something]	lunch, gym clothes, clothes

3. Communicate (21.1%): Next, a common task is to remind to contact (“ *a* ,” “ ,” “ ”) another individual, either a person (e.g., “ ,” “ *a* ,” “ *a* ”), organization/company (“*AT&T*”), or other (“ *a* ,” “ ”). We identify two subtypes: the majority reflects general, unspecified communication (94.7%)

[1] be, book, bring, buy, call, cancel, change, charge, check, clean, come, do, drop off, eat, email, feed, fill out, finish, get, get ready, go (to), grab, have, have to, leave, mail, make, order, pack, pay, pick up, play, print, put, renew, return, schedule, send, set, start, stop, stop by, submit, take, take out, tell, text, turn on, turn off, wash, watch, write.

(e.g., " *a* "), and a smaller part (5.3%) represents coordination or planning tasks (e.g., " *a* *a* "). Both subtypes represent tasks which briefly interrupt the user's activity.

General	
send [something]	email, text, report
email [someone]	dad, mom
text [someone]	mom, dad
call [someone]	mom, dad
tell [someone] [something]	my wife I love her, happy birthday mom
Coordinate	
set [an appointment]	doctors appointment
make [an appointment]	doctors appointment, reservation
schedule [an appointment]	haircut, doctors appointment

4. Manage ongoing external process (12.9%): These reminders represent manipulation of an ongoing, external process, i.e., tasks where the user monitors or interacts with something, e.g., the laundry or oven. These tasks briefly interrupt a user's activity and are less comprehensive than performing a chore.

turn [on/off] [something]	water, oven, stove, heater
check [something]	email, oven, laundry, food
start [something]	dishwasher, laundry
put [something] in [something]	pizza in oven, clothes in dryer
take [something] out	pizza, chicken, laundry

5. Manage ongoing user activity (6.3%): This class of reminders is similar to the previous class. However, as opposed to the user interacting with an external process, they reflect a change in the user's own activity. These tasks incur a higher cost on the user's availability and cognitive load. We distinguish three subtypes: preparing (31.4%), starting (61.4%), and stopping an activity (7.2%).

Activity/Prepare	
get ready [to/for]	work, home
Activity/Start	
start [some activity]	dinner, cooking, studying
make [something]	food, breakfast, grocery list
take [something]	a shower, a break
play [something]	game, xbox, basketball
watch [something]	tv, the walking dead, seahawks game
Activity/Stop	
stop [some activity]	reading, playing
finish [something]	homework, laundry, taxes

6. Eat/consume (2.8%): Another frequent reminder type refers to consuming something, most often food (" *a* ") or medicine (" *a* "). These tasks are small and range from brief interruptions (" *a* ") to longer interruptions (" *a* ").

take [something]	Medicine
eat [something]	lunch, dinner, breakfast, pizza
have [something]	lunch, a snack, breakfast

In summary, we performed a data-driven qualitative analysis and manually labeled frequently occurring task descriptions to identify a set of common underlying tasks. We find that the majority of reminders reflect a plan to travel to a destination, communicate with others, or perform daily chores.

4. REMINDER PATTERNS

Next, we study the temporal patterns of reminders. We seek to understand when people create reminders, when reminders are set to notify users, and the average delay between creation and notification time for different reminders. Such knowledge could prove useful in providing new competencies to reminder services, such as providing likelihoods about when certain tasks tend to happen, suggesting notification times (slot filling), predicting (follow-up) tasks, or proactively blocking out time on people's calendars. In this section we focus on the research question: *Ca a a a , a , a a , a a a a ?*

We study reminders on several levels of granularity. In Section 4.2 we look at global patterns and trends across all reminders. Next, we study temporal patterns per task type in Section 4.3. In Section 4.4, we perform a temporal analysis of task description terms. Finally, we study the relation between reminder creation and notification times in Section 4.5. First, we explain how we represent the reminder's creation and notification time to enable our analyses.

4.1 Method

To study common temporal patterns of reminders, we discretize time by dividing each day into the following six four-hour buckets: (i) late night [00:00-04:00), (ii) early morning [04:00-08:00), (iii) morning [08:00-12:00), (iv) afternoon [12:00-16:00), (v) evening [16:00-20:00), and (vi) night [20:00-00:00). By combining this time-of-day division with the days of week we yield a 7 by 6 matrix M, whose columns represent days, and rows times. Each r_{CT} and r_{NT} can be represented as a cell in matrix M, i.e., $M_{i,j}$ where i corresponds to the day of week, and j to the time of day. Furthermore, we distinguish between M^{CT} and M^{NT}, the matrices whose cells contain reminders that are created, or respectively set to notify, at a particular day and time. We represent each reminder as an object $r \in M$, with the attributes described in Section 3.2: the reminder's task description (r_{task}), creation time (r_{CT}), notification time (r_{NT}), and time delta ($r_{\Delta T}$). To study the temporal patterns, we look at the number of reminders that are created, or whose notifications are set, per cell. We compute conditional probabilities over the cells in M^{CT} and M^{NT}, where the reminder's creation or notification time is conditioned on the task type, time, or the terms in the task description.

$$P(r_{CT} = X \mid w) = \frac{|\, r \in R :\; w \in r_{task} \wedge r_{CT} = X \,|}{|\, w \in r_{task}, r \in M^{CT} \,|} \qquad \textbf{Eq. 1}$$

$$P(r_{NT} = X \mid w) = \frac{|\, r \in R :\; w \in r_{task} \wedge r_{NT} = X \,|}{|\, w \in r_{task}, r \in M^{NT} \,|} \qquad \textbf{Eq. 2}$$

To estimate the conditional probability of a notification or creation time, given a term from the task description, we take the set of reminders containing term , that are created or whose notification is set at time X, over the total number of reminders which contain the word (see Eq.1 and Eq.2). By computing this probability for each cell in M^{NT} or M^{CT}, (i.e., $\sum_{i,j \in M} P(r_{NT} = i,j \mid w)$) we generate a probability distribution over matrix M.

$$P(r_{NT} = X \mid r_{CT} = i,j\,) = \sum_{i,j \in M^{CT}} \frac{|\, r \in M^{CT}_{i,j} : r_{NT} = X \,|}{|\, r \in M^{CT}_{i,j} \,|} \qquad \textbf{Eq. 3}$$

To study the common patterns of the periods of time between the creation of reminders and notifications, we estimate a probability distribution for a reminder's notification time given a creation time (see Eq. 3). We compute this probability by taking the reminders in each cell of M^{CT} that have their notifications set to fire at time X, over all the reminders in that cell.

Finally, we study the delays between setting and executing reminders, by collecting counts and plotting histograms of $r_{\Delta T}$ of reminders for a given subset, e.g., $|\, r_{CT} \in R : w \in r \,|$ or $|\, r_{CT} \in R : r_{CT} = X \,|$.

4.2 Global Patterns

We now describe broad patterns of usage, and answer the following questions: “*A a a (. . - a), a a a (. . - a)?*” and *H a a a a a ?* To answer these questions, we examine the temporal patterns in our log data over the aggregate of all reminders in the two-month sample (576,080 reminders).

Figure 1 shows the prior probability of a reminder’s creation time, $P(r_{CT})$, and notification time, $P(r_{NT})$, in each cell in M^{CT} and M^{NT}. Looking at Figure 1, we see that in our sample, planning (reminder creation) most frequently happens later in the day, more so than during office hours (morning and midday). This observation could be explained by the user’s availability; users may have more time to interact with their mobile devices in the evenings. Additionally, the end of the day is a natural period of time to “wrap up the day,” i.e., looking backward and forward to completed and future tasks.

Turning our attention to notification times, the right plot of Figure 1 shows a slightly different pattern: people execute tasks (i.e. notifications trigger) throughout the day, from morning to evening. This shows that users want to be reminded of tasks throughout the day, in different contexts (e.g., both at home and at work). This is reflected in our task-type taxonomy, where tasks are related to both contexts. We also note how slightly more notifications trigger on weekdays than in weekends, and more notifications trigger at the start and end of the workweek. This observation may be attributed to the same phenomenon for reminder creation; users may tend to employ reminders for activities that switch between week and weekend contexts. Finally, comparing the two plots shows the notification times are slightly less uniformly distributed than creation times, e.g., users create reminders late at night, when it is relatively unlikely for notifications to fire.

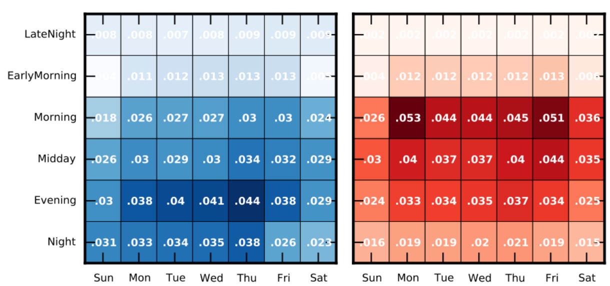

Figure 1. Distribution of reminder creation times (left plot) and reminder notification times (right plot) for all reminders in two-month sample (*n*=576,080).

Next, to determine how far in advance users typically plan, we look at the delays between reminder creation and notification in Figure 2. The top plot shows distinct spikes around five-minute intervals, which are due to reminders with a relative time indication (e.g.: “ *a a 5* ”). These intervals are more likely to come to mind than more obscure time horizons (e.g., *a a a 6.34*). The second and third plots clearly illustrate that the majority of reminders have a short delay: around 25% of the reminders are set to notify within the same hour (second plot), and around 80% of the reminders are set to notify within 24 hours (third plot). Interestingly, there is a small hump around 8-9 hours in the second plot, which may be explained by reminders that span a night, e.g., created at the end of the day, to notify early the next day), or a ‘working day’ (creation in the morning, notification at the end of the day).

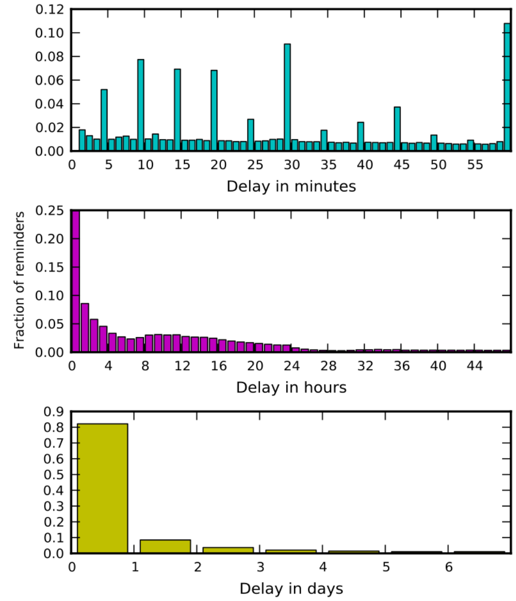

Figure 2. Histograms of delays (in minutes, hours, and days) between reminder creation and notification times.

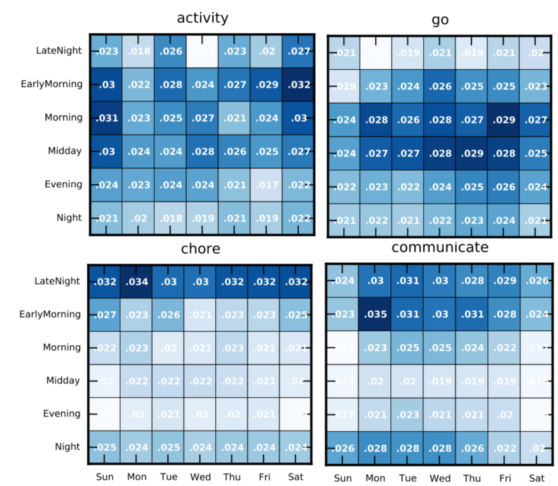

Figure 3. Creation times for different task types.

In summary, we find that on average people tend to set plans in the evening, and execute them throughout the day. Furthermore, most tasks that drive reminder setting are for short-term tasks to be executed in the next 24 hours.

4.3 Task Type

We now explore whether different task are characterized by distinct temporal patterns that differ from the global patterns seen in the previous section. To do so, we use the set of 2,484 frequent reminders to label the two-month sample of reminders. This yields a subset of 125,376 reminders with task type-labels that we use for analysis. We aim to answer the same questions raised in the previous section, but on the level of task type, as opposed to a characterization of the global aggregate.

Creation and Notification Times. First we look at the probability distribution of reminder creation times per task type, i.e., $P(r_{CT} \mid \text{tasktype})$. Looking at the distributions for each task type, we discover two broader groups: per task type, reminders are either created mostly in morning and midday blocks (roughly corresponding to office hours), or outside these blocks. Figure 3 shows examples of both types: “*A* ” and “*G* ” reminders are mostly created during office hours, while e.g., “*C a* ” and “*C* ” reminders are more prone to be created in evenings. Another interesting observation is that activity-related reminders are comparatively frequent on weekends.

Next, we study reminder notification times per task type, i.e., P(r_{NT} | tasktype). Here, a similar pattern emerges. Broadly speaking there are two types of tasks: those set to notify during office hours, and those that trigger outside these hours. See Figure 4 for examples. “*C a* ” and “*G* ” fall under the former type, whereas “*C* ” and “*Ma a* ” fall under the latter. The nature of the tasks explains this distinction: the former relate to work-related tasks (communication, work-related errands), whilst the majority of the latter represent activities that are more common in a home setting (cooking, cleaning).

Taking a closer look at the *C a* task subclasses in Figure 5, we show how “*C a /G a*” and “*C a /C - a* ” differ: the former is more uniformly distributed, whilst the latter is denser around office hours. The general subtask too has comparatively more reminders trigger in weekends, whereas coordinate is more strongly centered on weekdays. The distinct patterns suggest these subclasses indeed represent different types of tasks.

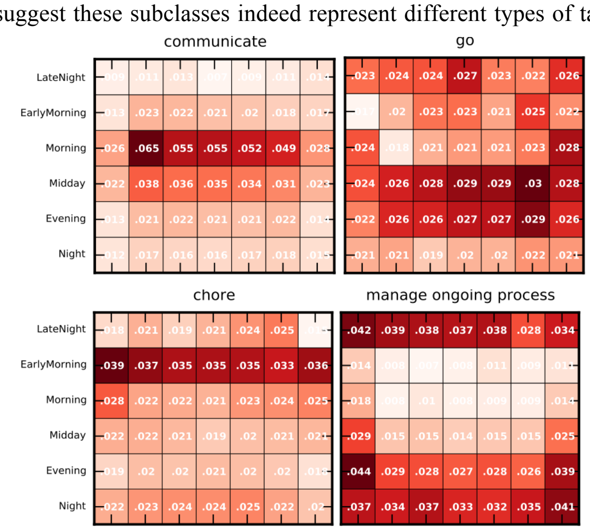

Figure 4. Reminder notification time probability distributions over time, for different task types.

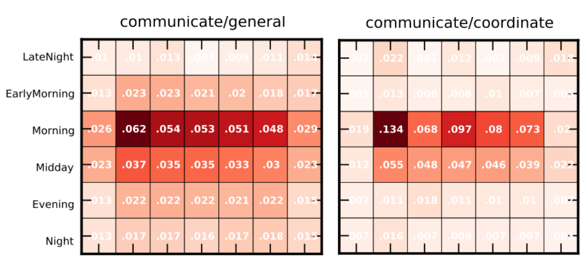

Figure 5. Reminder notification time probability distributions for the *Communicate* subclasses.

Reminder Creation and Notification Delay. To better understand differences in the lead times between reminder creation and notification, we present an overview of the distribution of reminder delays per task type in Figure 6. In general, the lower the boxplot lies on the -axis, the lower the lead time, i.e., the shorter the delay between creating the reminder and executing the task. It is worth comparing, e.g., the plot of “*Ma a* ,” to both “*G* ” or “*C a* ” task types: execution of managing ongoing processes tasks seem to be planned with a much shorter lead time than the other types of task. Considering the nature of the tasks, where ongoing processes often represent the close monitoring or checking of a process (e.g., cooking or cleaning tasks), it is understandable that the delays are on the order of a few minutes, rather than hours. “*C a /C a* ” has the largest delay on average, i.e., it is the task type people plan furthest in advance.

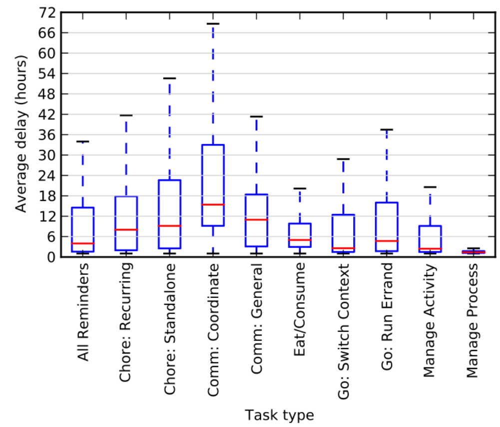

Figure 6. Boxplots showing delay between reminder creation and notification times (*n* = 125,376).

A more detailed examination of the differences between the “Communicate” subtasks, illustrated in Figure 7, we see that “*C - a /G a* ” subtasks are more likely to be executed with lower lead time, as noted by the peak at hour 0 in the top plot. The “*C - a /C a* ” subtask is about as likely to be executed the next day, as seen by the high peak around the 12 hour mark in the bottom plot. Much like the observations made in the previous section, the difference in the patterns between both *C a* subtasks suggests that the distinction between the subtypes is meaningful. Differences are not only found on a semantic level through our qualitative analysis, but also in temporal patterns.

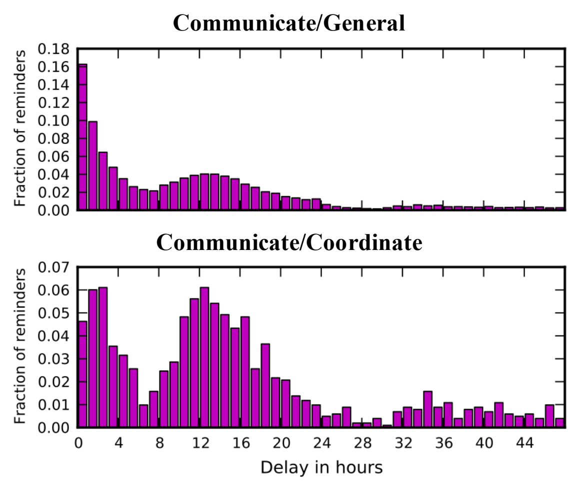

Figure 7. Delay (lead time) between reminder creation and notification for “Communicate” subtasks.

In summary, we have clearly shown how task type-specific temporal patterns differ from the aggregate patterns in Section 4.2.

4.4 Terms

One can hypothesize that the terms in task descriptions show distinct temporal patterns, i.e., reminders that contain the term “ *a*” are likely to trigger around dinner time. Presence of these temporal patterns may be leveraged for reminder creation or notification time prediction. To study this, we manually inspected the temporal distribution of task descriptions' terms of the 500 most frequent terms. More specifically, we compute conditional probabilities for a cell in M^{CT} or M^{NT} given a term (see Eq. 1 and Eq. 2). We found several intuitive patterns, which we illustrate below with examples. These patterns provide intuition behind the terms we use as features in predictive modeling, discussed in Section 5.

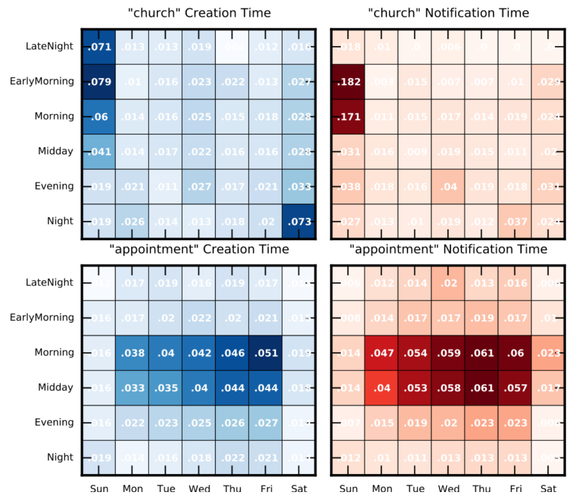

Figure 8. Creation & Notification times for reminders with the terms “church” (top) & “appointment” (bottom).

Figure 8 shows creation and notification times of task descriptions which contain the terms “church” or “appointment.” The “appointment” plot shows a strong pattern around the morning and midday blocks, representing office hours. Reminders that contain “church” show a clear pattern too; they are largely created from Saturday night through Sunday morning, and are set to notify on Sunday early morning and mornings. When we examine the delays between reminder creation and notification, clear patterns emerge.

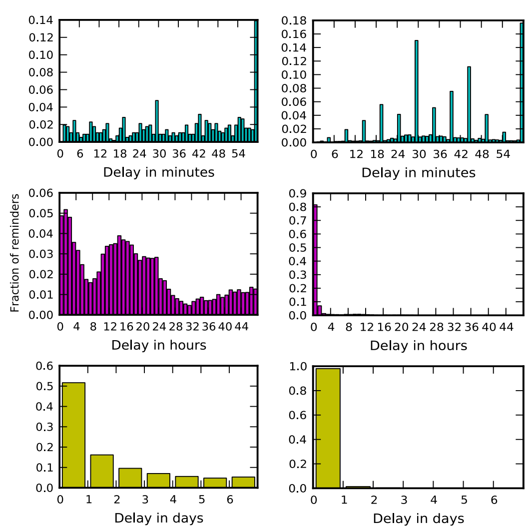

Figure 9. Delays between reminder creation and notification for reminder task descriptions containing the terms “appointment” (left) and “laundry” (right).

In Figure 9, we compare the average delays of reminders containing the term “appointment” to “laundry.” Clearly, on average, “appointment” reminders have longer delays, reflected by the nature of the task (which may involve other individuals and hence require more planning), whereas “laundry” reminders are more likely to reflect short-term tasks (which may be performed individually). In summary, we see distinct temporal patterns in task descriptions' terms. In Section 5, we study the generalizability of these patterns.

4.5 Time

Finally, we look at correlations between reminders' creation and notification times. Motivated by the observations that most reminders are set to notify shortly after they are created, we study the probability of a reminder's notification time given its creation time, $P(r_{NT} \mid r_{CT})$. See Figure 10 for examples. Looking at the plots in detail, we see how reminders across different creation times appear similar: they are most likely to have their notification fire within the same cell or the next, confirming earlier observations that the majority of reminders are short-term (i.e., same cell). However, upon closer inspection, we see that as the reminder's creation time moves towards later during the day, reminders are more likely to be set to notify the next day. Furthermore, in the third plot from the left, we see how reminders created on Friday evenings have a small but substantial probability of having their notification fire on Monday morning (i.e., the reminder spans the weekend). These patterns show how delay between reminder creation and notification time is low on average, but the length of delay is not independent from the creation time.

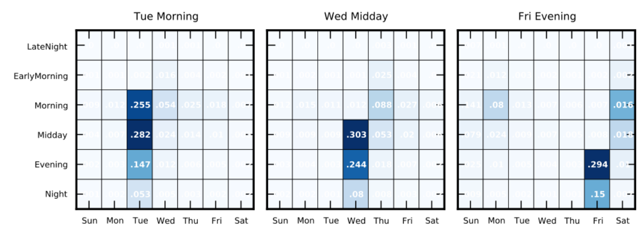

Figure 10. $P(r_{NT} \mid r_{CT})$ for three different r_{CT}.

In summary, we have shown distinct temporal patterns of reminders of different task types, and of the terms in task descriptions. Finally, we have shown that a reminder's notification time is most likely set shortly after creation time, but the later in the day a reminder is created, the more likely the notification time is further in the future.

5. PREDICTING NOTIFICATION TIME

In the previous section we have shown temporal patterns in reminder creation and notification time of four types: aggregate patterns, task type-related, term-based, and time-based. To study whether these patterns can be harnessed for aiding users in reminder setting, we address a prediction task. Specifically, we turn to the task of predicting the day of the week in which a task is most likely to happen (i.e., predicting r_{NT}). Motivated by our observation that the majority of the reminders are set to trigger soon after being set (Section 4.2 and 4.5), and by the patterns we observed of the terms from task descriptions (Section 4.4), we aim to answer the following research questions: *I a - a a ?* and *D - a a a a ?* Rather than seeking to harness a predictive model about timing, the primary aim of the experiments is to study whether the patterns discussed earlier generalize, as demonstrated by their contribution to predictive performance.

We cast the task of predicting the day of week a reminder is set to notify as multiclass classification, where each day of the week corresponds to a class. The input to our predictive model is the reminder's task description (r_{task}), creation time (r_{CT}), and the target class is the notification time (r_{NT}) day of week. We measure the predictive power of the patterns identified in the previous sections via term-based and (creation) time-based features. Specifically, for term-based features, we extract unigram bag of word features, and our time-based features correspond to R_{CT}'s time of day (row) and day of week (column), and the minutes since the start of week.

5.1 Experimental Setup

We use Gradient Boosted Decision Trees for classification. This method has proven to be robust and efficient in large-scale learning problems [16]. The ability of this method to deal with non-linearity in the feature space and heterogeneous features make it a natural choice. To address the multiclass nature of our problem, we employ a one vs. all classification strategy, where we train seven binary classifiers and output the prediction with the highest confidence as final prediction. We compare the accuracy to randomly picking a day of the week (with accuracy of $1/7 \approx 0.1429$) and to a more competitive baseline which predicts the notification will be for the same day that the notification was created (BL-SameDay).

For the experiments, we sample six months of data (January through June 2015). All data were filtered according to the process described in Section 3.2, resulting in a total of 1,509,340 reminders. We split this data sequentially: the first 70% (approx. January 1 to May 7) forms the training set and the last 30% (approx. May 8 to June 30) forms the test set. We use the first two months of the training set for the analysis described in Sections 3 and 4 as well as for parameter tuning before we retrained on the entire training set.

In the next section, we report predictive performance on the held-out test set. Specifically, we report macro and micro-averaged accuracy over the classes (Macro and Micro, respectively). We compare three systems: one that leverages time features based on the reminder's creation time (Time only), one with term features (Terms only), and finally a model that leverages both types of features (Full model). We test for statistical significance using t-tests, comparing our predictive models against BL-SameDay. The symbols ▲ and ▼ denote statistically significant differences (greater than the baseline and worse than the baseline, respectively) at $\alpha = 0.01$.

6. RESULTS

Table 4 shows the results of our prediction task. First we note that the baseline of predicting the notification time to be the same day as the creation time performs much better than random (at 0.1429). This indicates users mostly set reminders to plan for events in the short-term. Next, we see that the *T* model significantly improves over the baseline, indicating that the reminder creation time helps further improve prediction accuracy. As noted earlier, tasks planned late at night are more likely to be executed on a different day, and the use of creation time helps leverage this and more general patterns. Finally, the model that uses only features based on the task description (*T*) performs better than random, but does not outperform the baseline. However, when combined with the time model (*F*) we see an increase of 8.2% relative to the time only model. We conclude that the creation time provides the most information for predicting when a task will be performed, the task description provides significant additional information, primarily when the description is used in combination with the reminder's creation time.

7. DISCUSSION AND IMPLICATIONS

Through log analyses we have shown common reminders and made an attempt at identifying and categorizing the types of tasks that underlie them. We identified that the majority of reminders in our sample refer to either daily household chores, running errands, or switching contexts. We have shown there are distinct temporal patterns across reminders and reminder types. Finally, we demonstrated that we can leverage these patterns to predict the day of the week that a reminder is most likely to trigger, i.e., the day the task is most likely to be executed. The findings have implications for designing systems to help with task completion, and more generally for developing technology to reduce prospective memory failures.

Table 4. Average accuracy of the day-of-week prediction task.

Run	Micro	Macro	Error reduction
Full model	**0.6788**▲	**0.6761**▲	**+0.3381**
Time only	0.6279▲	0.6258▲	+0.2333
Terms only	0.1777▼	0.1772▼	−0.6944
BL-SameDay	0.5147	0.5165	

There are several limitations in our log analyses. First, we performed this analysis on a specific subset of reminders: reminders from one geographic locale and for a single type of reminder: time-based. There are opportunities to understand cultural and linguistic factors in reminder creation by considering reminders from multiple regions. We additionally seek to investigate other types of reminders, such as those involving people, places, and events. Second, it is difficult to quantify the comprehensiveness of the task type taxonomy, which covers common reminders. The ontology may therefore not cover more intricate, personal, specific, or complex reminders, the nature of which needs to be better understood. Finally, our approach and analysis is entirely log based. The taxonomy's categories were manually labeled, and we make inferences and assumptions about the tasks that people are engaged in. User studies are needed to better understand the reminder process, including the generation and value of reminders, including how people behave when they are notified.

We explored a potential use of predictions about the day of the week that a reminder will trigger. Understanding when people tend to perform tasks is useful more generally, e.g., for effective resource scheduling or tailored advertising purposes. Understanding task durations could be useful in developing systems to automatically terminate ongoing tasks or allocate time for task completion.

8. CONCLUSIONS AND FUTURE WORK

We performed an analysis of a large corpus of reminder data collected in the wild. We identified common task types in frequent reminders seen across multiple users. We found that users largely remind themselves to go somewhere, communicate, or perform daily chores. Furthermore, we show how reminders display different temporal patterns depending on the task type that they represent, the reminder's creation time, and the terms in the task description. Finally, we show that we can use these identified patterns to predict when a reminder is set to trigger. Specifically, we confirm that the reminder's creation time is a strong indicator of notification time, but that including the task description further improves accuracy over the strongest baseline, with a 33% reduction in error. Future work includes developing more sophisticated models (e.g., considering personalized signals) to improve prediction performance. In the long-term, we believe that insights and predictions about tasks and the use of reminders can help with building and fielding systems with the ability to proactively reserve time, manage conflicts, remind people about tasks they might forget, and, more generally, to help people achieve their goals.

Acknowledgements The first author is supported by the Netherlands Organisation for Scientific Research (NWO) under project nr 727.011.005.

REFERENCES

1. Agichtein, E., Brill, E., and Dumais, S. (2006). Improving web search ranking by incorporating user behavior information. *P . SIGIR* (pp. 19–26).
2. Armentano, M.G. and Amandi, A.A. (2009). Recognition of User Intentions for Interface Agents with Variable Order Markov Models. *P . UMAP '09* (pp. 173-184).
3. Barreau, D. and Nardi, B.A. (1995). Finding and reminding: File organization from the desktop. *ACM SIGCHI B* , 27(3), 39–43.
4. Barua, D., Kay, J., Kummerfeld, B., Paris, C. (2014). Modelling Long Term Goals. *P . UMAP 2014*. pp 1-12.
5. Brandimonte, M.A., Einstein, G.O., and McDaniel, M.A. (2014). *P M : T a a a* . Psychology Press.
6. Chaurasia, P., McClean, S.D., Nugent, C., and Scotney, B. (2014). A duration-based online reminder system. *I . J. P a C a C a* , 10(3), 337–366.
7. Czerwinski, M., Horvitz, E., and Wilhite, S. (2004). A diary study of task switching and interruptions. *P . SIGCHI* (pp. 175–182).
8. DeVaul, R., Clarkson, B., and Pentland, A. (2000). The Memory Glasses: Towards a wearable context aware, situation-appropriate reminder system. *P . SIGCHI W S a I a U C* .
9. Dey, A.K. and Abowd, G.D. (2000). Cybreminder: A context-aware system for supporting reminders. In *Ha a U C* (pp. 172-186).
10. Dismukes, R. K. (2012). Prospective Memory in Workplace and Everyday Situations. In *C D P - a S* , 21 (4): 215--220.
11. Einstein, G.O. and McDaniel, M.A. (1990). Normal aging and prospective memory. *J. E a P : L a - , M , a C* , 16(4), 717.
12. Ellis, J. (1996). Prospective memory or the realization of delayed intentions: A conceptual framework for research. *P - : T a a a* , 1–22.
13. Ellis, J. and Kvavilashvili, L. (2000). Prospective memory in 2000: Past, present and future directions. *A C P* , 14, 1–9.
14. Fertig, Scott, Freeman, E., and Gelernter, D. (1996). "Finding and reminding" reconsidered. *SIGCHI B* , 28(1), 66–69.
15. Fertig, S., Freeman, E., and Gelernter, D. (1996). Lifestreams: An alternative to the desktop metaphor. *P . SIGCHI C - C a* (pp. 410–411).
16. Friedman, J.H. (2002). Stochastic gradient boosting. *C . S a . Da a A a* . 38(4), 367–378.
17. Hanauer, D.A., Wentzell, K., Laffel, N., and Laffel, L.M. (2009). Computerized Automated Reminder Diabetes System (CARDS): E-mail and SMS cell phone text messaging reminders to support diabetes management. *D a T - a T a* , 11(2), 99–106.
18. Hicks, J.L., Marsh, R.L., and Cook, G.I. (2005). Task interference in time-based, event-based, and dual intention prospective memory conditions. *J. M a La a* , 53(3), 430–444.
19. Intons-Peterson, M.J. and Fournier, J. (1986). External and internal memory aids: When and how often do we use them? *J a E a P : G a* 115(3), 267.
20. Joachims, T. (2002). Optimizing search engines using clickthrough data. *P . SIGKDD* (pp. 133–142).
21. Kamar, E. and Horvitz, E. (2011). Jogger: Models for context-sensitive reminding. *P . AAMS* (pp. 1089–1090).
22. Kapur, N., Glisky, E.L., and Wilson, B.A. (2004). Technological memory aids for people with memory deficits. *N - a R a a* , 14(1-2), 41–60.
23. Koren, Y., Liberty, E., Maarek, Y., and Sandler, R. (2011). Automatically tagging email by leveraging other users' folders. *P . SIGKDD* (pp. 913–921).
24. Lamming, M. and Flynn, M. (1994). Forget-me-not: Intimate computing in support of human memory. *P . FRIEND21* (pp. 2–4).
25. Leskovec, J. and Horvitz, E. (2008). Planetary-scale views on a large instant-messaging network. *P . WWW* (pp. 915–924).
26. Ludford, P.J., Frankowski, D., Reily, K., Wilms, K., and Terveen, L. (2006). Because i carry my cell phone anyway: Functional location-based reminder applications. *P . SIGCHI* (pp. 889–898).
27. Malone, T.W. (1983). How do people organize their desks? Implications for the design of office information systems. *ACM TOIS*, 1(1), 99–112.
28. Marmasse, N. and Schmandt, C. (2000). Location-aware information delivery with commotion. *P . Ha a U - C* (pp. 157–171).
29. McGee-Lennon, M.R., Wolters, M.K., and Brewster, S. (2011). User-centred multimodal reminders for assistive living. *P . SIGCHI* (pp. 2105–2114).
30. McGee-Lennon, M., Wolters, M., McLachlan, R., Brewster, S., and Hall, C. (2011). Name that tune: Musicons as reminders in the home. *P . SIGCHI* (pp. 2803–2806).
31. O'Connail, B. and Frohlich, D. (1995). Timespace in the workplace: Dealing with interruptions. *P . SIGCHI C - C a* (pp. 262–263).
32. Partridge, K. and Price, R (2009). Enhancing mobile recommender systems with activity prediction. *P . UMAP 09*. 307-318.
33. Rhodes, B. and Starner, T. (1996). Remembrance Agent: A continuously running automated information retrieval system. *P . P a a A a I A a M A T* (pp. 487–495).
34. Richardson, M. (2008). Learning about the world through long-term query logs. *ACM TWEB*, 2(4), 21.
35. Sellen, A.J., Louie, G, Harris, J.E., and Wilkins, A.J. (1996). What brings intentions to mind? An in situ study of prospective memory. *M* , 5(4), 483–507.
36. Tu, Y., Chen, L., Lv, M., Ye, Y., Huang, W., and Chen, G. (2013). ireminder: An intuitive location-based reminder that knows where you are going. *IJHCI*, 29(12), 838–850.
37. White, R.W. and Drucker, S.M. (2007). Investigating behavioral variability in web search. *P . WWW* (pp. 21–30).